3001
GSFC
NASA
1964
vol. 1

TICATIONS OF
HARD SPACE FLIGHT CENT



PRICE(S) $

Hard copy (HC)

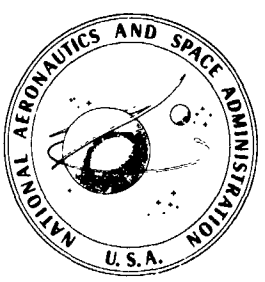
Microfiche (MF)

ff 653 July 65

Volume I.
SPACE SCIENCES

N66 37946 N66 37973

PUBLICATIONS OF GODDARD SPACE FLIGHT CENTER 1964



Volume I
SPACE SCIENCES

FOREWORD

This publication is a collection of articles, papers, talks, and reports generated by the scientific and engineering staff of Goddard Space Flight Center in the year 1964. Many of these articles were originally published in scientific or engineering Journals or as official NASA technical publications, while others are documents of a more informal nature. All are reprinted here as nearly verbatim as typography and format will permit.

These articles are grouped into broad subject categories, but no detailed subdivision has been made. Within each category, the articles are arranged alphabetically by author. An overall author index is given in the back of the volume.

The years 1963, 1964, and 1965 are being published as whole-year issues, and the resulting size dictates the use of two volumes; the first volume is titled Space Sciences, and the second Space Technology. It is anticipated, however, that future issues will be quarterly single volumes.

PRECEDING PAGE BLANK NOT FILMED.

CONTENTS

FOREWORD iii

I. ASTRONOMY AND ASTROPHYSICS

"B STAR COLORS BETWEEN 2000 AND 3000 ANGSTROMS," *by* A. Boggess, III 1
"INTERSTELLAR EXTINCTION IN THE MIDDLE ULTRAVIOLET," *by* A. Boggess, III, *and* J. Borgman 4
"ORIGIN OF THE SOLAR SYSTEM," *by* A. G. W. Cameron 7
"GRAVITATIONAL COLLAPSE," *by* H.-Y. Chiu 13
"SUPERNOVAE, NEUTRINOS, AND NEUTRON STARS," *by* H.-Y. Chiu 27
"SURFACE X-RAY EMISSION FROM NEUTRON STARS," *by* H.-Y. Chiu *and* E. E. Salpeter 55
"ESCAPE OF STARS DURING THE COLLISION OF TWO GALAXIES," *by* G. Contopoulos *and* G. Bozis 58
"DYNAMICS AND EVOLUTION OF CLOSE BINARY SYSTEMS," *by* S.-S. Huang 67
"EFFECT OF NON-SYNCHRONOUS ROTATION ON CLOSE BINARY STARS," *by* S.-S. Huang 108
"AN INTERPRETATION OF ϵ AURIGAE" *by* S.-S. Huang 113
"THE MOTION OF GASEOUS STREAMS IN THE BINARY SYSTEM," *by* S.-S. Huang 120
"ROTATIONAL BEHAVIOR OF THE MAIN-SEQUENCE STARS AND ITS PLAUSIBLE CONSEQUENCES CONCERNING FORMATION OF PLANETARY SYSTEMS," *by* S.-S. Huang 129
"SEQUENCE OF EVENTS IN THE EARLY PHASE OF THE SOLAR SYSTEM," *by* S.-S. Huang 137
"PLASMA NEUTRINO EMISSION FROM A HOT, DENSE ELECTRON GAS," *by* C. L. Inman *and* M. A. Ruderman 142
"MAGNETOGRAVITATIONAL INSTABILITY OF ANISOTROPIC PLASMA WITH HALL EFFECT," *by* G. L. Kalra *and* S. P. Talwar 155
"THE DETECTION OF SOLAR NEUTRINOS," *by* H. Reeves 160
"ACCRETION OF INTERSTELLAR MATTER BY MASSIVE OBJECTS," *by* E. E. Salpeter 165
"C_2 SWAN BANDS IN COMETS." *by* R. E. Stockhausen *and* D. E. Osterbrock 169
"EVOLUTION OF O STARS. II. HYDROGEN EXHAUSTION AND GRAVITATIONAL CONTRACTION," *by* R. Stothers 174
"THE RV TAURI POPULATION," *by* R. Stothers 184

II. CELESTIAL MECHANICS AND GEODESY

"LONG-PERIOD CONTRIBUTIONS TO THE DISTURBING FUNCTIONS OF THE EARTH FROM THE SEVENTH, NINTH, AND ELEVENTH ZONAL HARMONICS," *by* T. L. Felsentreger *and* W. J. Wickless 191
"NOTES ON VON ZEIPEL'S METHOD," *by* G. E. O. Giacaglia 196
"A SERIES SOLUTION FOR SOME PERIODIC ORBITS IN THE RESTRICTED THREE-BODY PROBLEM ACCORDING TO THE PERTURBATION METHOD," *by* S.-S. Huang 206
"ON THE APPLICATION OF PFAFF'S METHOD IN THE THEORY OF VARIATION OF ASTRONOMICAL CONSTANTS," *by* P. Musen 215
"ON THE SIMULTANEOUS COMPUTATION OF THE SECULAR AND RESONANCE EFFECTS IN THE MOTION OF CELESTIAL BODIES," *by* P. Musen 233
"DETERMINATION OF THE TRIAXIALITY OF THE EARTH FROM OBSERVATIONS ON THE DRIFT OF THE SYNCOM II SATELLITE," *by* C. A. Wagner 243
"INTRODUCTION TO THE APPLICATION OF VON ZEIPEL'S METHOD," *by* W. J. Wickless 282

PRECEDING PAGE BLANK NOT FILMED.

III. SOLAR PHYSICS

"COMMENTS ON HIGH ENERGY X-RAYS BURSTS OBSERVED BY OSO-1," *by* K. J. FROST 291
"TEMPERATURE OF A GRAY BODY MOST CLOSELY FITTING THE SOLAR EXTRATERRESTRIAL SPECTRUM," *by* W. B. FUSSELL *and* J. B. SCHUTT 294
"DISTRIBUTION IN HELIOGRAPHIC LONGITUDE OF FLARES WHICH PRODUCE ENERGETIC SOLAR PARTICLES," *by* D. E. GUSS 299
"THE EFFECT OF A CORONAL EXPANSION ON THE COMPUTATION OF AN ELECTRON TEMPERATURE FROM EMISSION-LINE INTENSITIES," *by* W. M. NEUPERT 301
"TIME VARIATIONS IN THE SOLAR EXTREME ULTRAVIOLET SPECTRUM OBSERVED BY OSO-1," *by* W. M. NEUPERT 306
"THE UV LINES OF FE XIV IN THE SOLAR CORONA," *by* R. E. STOCKHAUSEN 315
"SOLAR X-RAYS: A COMPARISON WITH MICROWAVE RADIATION," *by* W. A. WHITE 319

IV. IONOSPHERE AND RADIO PHYSICS

"D REGION FORMATION," *by* A. C. AIKIN 325
"SOME RESULTS OF ROCKET EXPERIMENTS IN THE QUIET *D* REGION", *by* A. C. AIKIN, J. A. KANE, *and* J. TROIM 335
"HYDROGEN IONS," *by* S. J. BAUER 343
"MEAN DIURNAL VARIATION OF THE TOPSIDE IONOSPHERE AT MID-LATITUDES," *by* S. J. BAUER *and* L. J. BLUMLE 348
"ELECTRON DISTRIBUTION IN THE IONOSPHERE," *by* R. E. BOURDEAU 355
"RESEARCH WITHIN THE IONOSPHERE," *by* R. E. BOURDEAU 363
"THE TEMPERATURE OF CHARGED PARTICLES IN THE UPPER ATMOSPHERE," *by* R. E. BOURDEAU 376
"POSITIVE ION TEMPERATURES ABOVE THE F-LAYER MAXIMUM," *by* R. L. F. BOYD *and* W. J. RAITT 384
"GEOMAGNETIC CONTROL OF DIFFUSION IN THE UPPER ATMOSPHERE," *by* S. CHANDRA *and* R. A. GOLDBERG 387
"THE EFFECT OF A VARIABLE ELECTRON TEMPERATURE ON THE EQUATORIAL ELECTRON DENSITY DISTRIBUTION IN THE UPPER IONOSPHERE," *by* R. A. GOLDBERG 397
"A SYNOPTIC VIEW OF IONIC CONSTITUTION ABOVE THE F-LAYER MAXIMUM," *by* W. J. RAITT, S. LAFLIN, *and* R. L. F. BOYD 408
"MAGNETIC SHELL ENHANCEMENTS DURING MAGNETIC DISTURBANCES," *by* A. P. WILLMORE *and* C. L. HENDERSON 412

V. FIELDS AND PARTICLES

"MEASUREMENT OF PRIMARY COSMIC RAY CHARGE AND ENERGY SPECTRA DURING 1963," *by* V. K. BALASUBRAHMANYAN *and* F. B. MCDONALD 419
"MAGNETOHYDRODYNAMIC SHOCK STRUCTURE IN A MOVING MAGNETIC FIELD," *by* D. B. BEARD 423
"SOLAR CYCLE CHANGES IN INNER ZONE PROTONS," *by* R. C. BLANCHARD *and* W. N. HESS 431
"A NEUTRAL-POINT EXPANSION OF THE IDEAL MAGNETOSPHERE," *by* R. BLUM 441
"POSSIBLE SOURCE MECHANISM FOR LOW-ENERGY GALACTIC ELECTRONS," *by* K. A. BRUNSTEIN 445
"STUDIES OF SOLAR PROTONS WITH EXPLORERS XII AND XIV," *by* D. A. BRYANT, T. L. CLINE, U. D. DESAI, *and* F. B. MCDONALD 449
"VELOCITY DEPENDENCE AND SOURCE SPECTRA OF SOLAR PROTON EVENTS," *by* D. A. BRYANT, T. L. CLINE, U. D. DESAI, *and* F. B. MCDONALD 466
"COMMENTS ON THE VANGUARD 3 MAGNETIC FIELD DATA AND ANALYSIS," *by* J. C. CAIN *and* S. HENDRICKS 471
"PARTICLE ACCELERATION IN CISLUNAR SPACE," *by* A. G. W. CAMERON 474
"DETECTION OF INTERPLANETARY 3- to 12-MEV ELECTRONS," *by* T. L. CLINE, G. H. LUDWIG, *and* F. B. MCDONALD 476

"LETTER TO THE EDITOR TELEVISION CINEMAPHOTOGRAPHY OF AURORAS AND PRELIMINARY MEASUREMENTS OF AURORAL VELOCITIES," *by* T. N. DAVIS *and* G. T. HICKS 480
"AN EXPERIMENTAL EXAMINATION OF LOW-ENERGY COSMIC RAY HEAVY NUCLEI", *by* C. E. FICHTEL, D. E. GUSS, *and* K. A. NEELAKANTAN 482
"MODULATION OF LOW-ENERGY GALACTIC COSMIC RAYS", *by* C. E. FICHTEL, D. E. GUSS, D. A. KNIFFEN, *and* K. A. NEELAKANTAN 493
"MODULATION OF LOW-ENERGY GALACTIC COSMIC-RAY HYDROGEN AND HELIUM", *by* C. E. FICHTEL, D. E. GUSS, D. A. KNIFFEN, *and* K. A. NEELAKANTAN 496
"MAGNETIC EFFECTS OF THE QUIET TIME PROTON BELT," *by* R. A. HOFFMAN *and* P. A. BRACKEN 498
"AURORA", *by* B. HULTQVIST 513
"AURORA AND THE LOWER IONOSPHERE IN RELATION TO SATELLITE OBSERVATIONS OF PARTICLE PRECIPITATION", *by* B. HULTQVIST 588
"ON THE HEIGHT OF AURORAL ABSORPTION, II", *by* B. HULTQVIST 604
"ON THE STRUCTURE OF A COLLISION FREE WAVE IN A PLASMA TRANSVERSE TO A MAGNETIC FIELD", *by* R. K. JAGGI *and* A. P. STOKES 612
"EVIDENCE FOR RAPID MOTION OF THE OUTER BOUNDARY OF THE MAGNETOSPHERE", *by* A. KONRADI *and* R. L. KAUFMANN 618
"MEASUREMENT OF LOW-ENERGY PRIMARY COSMIC RAY PROTONS ON IMP-1 SATELLITE", *by* F. B. MCDONALD *and* G. H. LUDWIG 628
"SECULAR AND NON-SECULAR BEHAVIOR FOR THE COLD PLASMA EQUATIONS", *by* D. MONTGOMERY *and* D. A. TIDMAN 632
"INITIAL RESULTS OF THE IMP-1 MAGNETIC FIELD EXPERIMENT", *by* N. F. NESS, C. S. SCEARCE, *and* J. B. SEEK 642
"THE MAGNETOHYDRODYNAMIC WAKE OF THE MOON", *by* N. F. NESS 675
"THE MAGNETOSPHERE AND ITS BOUNDARY LAYER", *by* N. F. NESS 689
"ON THE SOLAR ORIGIN OF THE INTERPLANETARY MAGNETIC FIELD", *by* N. F. NESS *and* J. M. WILCOX 697
"SPINOR FORMULATION OF MAGNETOGAS DYNAMICS", *by* L. A. SCHMID 701
"RESULTS FROM THE IMP-I RETARDING POTENTIAL ANALYZER", *by* G. P. SERBU 750
"EFFECT OF COMPRESSIBILITY, ROTATION, AND MAGNETIC FIELD ON THE DRAG ON A SPHERE OSCILLATING IN A CONDUCTING, VISCOUS MEDIUM", *by* M. P. SINGH 757
"ON OSCILLATING SPHERE IN A ROTATING VISCOUS FLUID", *by* M. P. SINGH 768
"VARIATIONS OF GEOMAGNETIC ACTIVITY WITH LUNAR PHASE", *by* H. L. STOLOV *and* A. G. W. CAMERON 770
"PROPAGATION OF HYDROMAGNETIC WAVES IN THE MAGNETOSPHERE", *by* M. SUGIURA 777
"HYDROMAGNETIC STABILITY OF THE MAGNETOSPHERE BOUNDARY", *by* S. P. TALWAR 796
"CERENKOV SCINTILLATION COUNTER MEASUREMENTS OF THE INTENSITY AND MODULATION OF LOW RIGIDITY COSMIC RAYS AND FEATURES OF THE GEOMAGNETIC CUTOFF RIGIDITY", *by* W. R. WEBBER *and* F. B. MCDONALD 802

VI. PLANETOLOGY

"ON THE EARLY HISTORY OF THE EARTH," *by* W. L. DONN, B. D. DONN, *and* W. G. VALENTINE 819

VII. PLANETARY ATMOSPHERES

"LOW INTENSITY DECAMETER EMISSIONS FROM JUPITER", *by* J. K. ALEXANDER *and* R. G. STONE 841
"AN ANALYSIS OF TIROS II RADIATION DATA RECORDED OVER NEW ZEALAND AT NIGHT", *by* L. J. ALLISON 846
"LATITUDINAL DISTRIBUTION OF CLOUD COVER FROM TIROS III PHOTOGRAPHS", *by* A. ARKING 856

"A RADIATION CLIMATOLOGY IN THE VISIBLE AND INFRARED FROM THE TIROS METEOROLOGICAL SATELLITES", *by* W. R. BANDEEN, M. HALEV, *and* I. STRANGE 861
"TIME CORRELATION OF EXTREME ULTRAVIOLET RADIATION AND THERMOSPHERIC TEMPERATURE", *by* R. E. BOURDEAU, S. CHANDRA, *and* W. M. NEUPERT 879
"DETAILED BEHAVIOR OF THE MIDLATITUDE IONOSPHERE FROM THE EXPLORER XVII SATELLITE", *by* L. H. BRACE, N. W. SPENCER, *and* A. DALGARNO 884
"FIRST ELECTROSTATIC PROBE RESULTS FROM EXPLORER 17", *by* L. H. BRACE *and* N. W. SPENCER 902
"ORIGIN AND EVOLUTION OF ATMOSPHERES AND OCEANS", *by* A. G. W. CAMERON, P. J. BRANCAZIO, *and* N. W. PANAGAKOS 907
"THE RED LINE OF ATOMIC OXYGEN IN THE DAY AIRGLOW", *by* A. DALGARNO *and* J. C. G. WALKER 912
"A LATITUDE SURVEY OF THE NIGHT AIRGLOW", *by* T. N. DAVIS *and* L. L. SMITH 926
"THE UPPER ATMOSPHERE OF JUPITER", *by* S. H. GROSS *and* S. I. RASOOL 936
"THE INFRARED INTERFEROMETER SPECTROMETER EXPERIMENT (IRIS)", *by* R. A. HANEL *and* L. CHANEY 946
"RADIATIVE EQUILIBRIUM IN PLANETARY ATMOSPHERES, I. APPLICATION OF THE STRONG LINE ABSORPTION LAW TO THE ATMOSPHERE OF VENUS," *by* R. A. HANEL *and* F. BARTKO 969
"MIDDLE ULTRAVIOLET DAY RADIANCE OF THE ATMOSPHERE", *by* J. P. HENNES, W. B. FOWLER, *and* L. DUNKLEMAN 987
"LYMAN-α RESONANT SCATTERING USING RANDOM STROKES VECTORS," *by* S. O. KASTNER 993
"INFRASONIC WAVES FROM THE AURORAL ZONE", *by* K. MAEDA *and* T. WATANBE 997
"ATMOSPHERIC DENSITIES FROM EXPLORER 17 DENSITY GAUGES AND A COMPARISON WITH SATELLITE DRAG DATA," *by* G. P. NEWTON, R. HOROWITZ, *and* W. PRIESTER 1024
"STRATOSPHERIC TEMPERATURE PATTERNS BASED ON RADIOMETRIC MEASUREMENTS FROM THE TIROS VII SATELLITE", *by* W. NORDBERG, W. R. BANDEEN, G. WARNECKE, *and* V. KUNDE 1027
"CLOUD HEIGHTS AND NIGHT-TIME CLOUD COVER FROM TIROS RADIATION DATA", *by* S. I. RASOOL 1049
"GLOBAL DISTRIBUTION OF THE NET ENERGY BALANCE OF THE ATMOSPHERE FROM TIROS RADIATION DATA", *by* S. I. RASOOL 1055
"INTENSITIES OF 9.4μ AND 10.4μ CO_2 BANDS", *by* S. I. RASOOL 1059
"DATA FROM EXPLORER 17 ON COMPOSITION OF THE UPPER ATMOSPHERE," *by* C. A. REBER 1060
"A NIGHT-TIME MEASUREMENT OF OZONE ABOVE 40 KM," *by* E. I. REED *and* R. SCOLNIK 1066
"SOME ASPECTS OF WIND SHEAR IN THE UPPER ATMOSPHERE", *by* R. G. ROPER 1073
"TEMPERATURE, PRESSURE, DENSITY, AND WIND MEASUREMENTS WITH THE ROCKET GRENADE EXPERIMENT, 1960–1963", *by* W. SMITH, L. KATCHEN, P. SACHER, P. SWARTZ, *and* J. THEON 1106
"NEW KNOWLEDGE OF THE EARTH'S ATMOSPHERE FROM THE AERONOMY SATELLITE (EXPLORER XVII)", *by* N. W. SPENCER, G. P. NEWTON, C. A. REBER, L. H. BRACE, *and* R. HOROWITZ 1148
"THE REFLECTIVITY OF JUPITER IN THE ULTRAVIOLET," *by* T. P. STECHER 1158
"THE UPPER ATMOSPHERE", *by* J. C. G. WALKER 1162
"INFERENCES OF STRATOSPHERIC AND MESOSPHERIC CIRCULATION SYSTEMS FROM ROCKET EXPERIMENTS", *by* G. WARNECKE *and* W. NORDBERG 1172

VIII. GENERAL

"EXAMPLES OF CERTAIN DATA REDUCTION AND MAPPING PROCEDURES UTILIZING TIROS III FIVE-CHANNEL RADIOMETER DATA", *by* L. J. ALLISON *and* G. WARNECKE 1183

"SYMMETRIC EULER ANGLE DECOMPOSITION OF THE TWO-ELECTRON FIXED-NUCLEUS PROBLEM", *by* A. K. BHATIA *and* A. TEMKIN 1198
"SPACE: HIGHLIGHTS OF RECENT RESEARCH", *by* R. JASTROW *and* A. G. W. CAMERON 1216
"g_J OF THE $(5s5p)^3P_1$ LEVEL OF Cd AND THE $(6s6p)^3P_1$ LEVEL OF Hg BY HIGH-FIELD DOUBLE RESONANCE", *by* R. KOHLER *and* P. THADDEUS 1231
"ORIGIN OF DIAMONDS IN THE UREILITES", *by* M. E. LIPSCHUTZ 1239
"THE BACKWARD RECURRENCE METHOD FOR COMPUTING THE REGULAR BESSEL FUNCTION", *by* T. E. MICHELS 1244
"THEORY OF THE 2s AND 2p EXCITATION OF THE HYDROGEN ATOM INDUCED BY ELECTRON IMPACT", *by* K. OMIDVAR 1261
"EFFECTS OF NON-PHOTOCHEMICAL PROCESS ON THE MERIDONAL DISTRIBUTION AND TOTAL AMOUNT OF OZONE IN THE ATMOSPHERE", *by* C. PRABHAKARA 1300
"ELECTRON MICROSCOPY OF METEORITIC AND ARTIFICIALLY SHOCKED GRAPHITE", *by* T. P. SCIACCA *and* M. E. LIPSCHUTZ 1326
"A. C. CONDUCTIVITY OF A PLASMA", *by* C. S. SHEN *and* R. L. W. CHEN 1329
"PARTIAL WAVE THEORY OF DIATOMIC MOLECULES, II", *by* A. TEMKIN *and* A. K. BHATIA 1336
"HYPERFINE STRUCTURE IN THE MICROWAVE SPECTRUM OF NH_2D", *by* P. THADDEUS, L. C. KRISHER, *and* P. CAHILL 1346
AUTHOR INDEX 1353

I. ASTRONOMY AND ASTROPHYSICS

B STAR COLORS BETWEEN 2000 AND 3000 ANGSTROMS*

ALBERT BOGGESS III

Ultraviolet fluxes at 2200 and 2600 Angstroms have been measured for about 180 B and A stars using filter photometers in rockets. Forty-eight of these stars have been observed more than once, and an analysis of these data show the uncertainty of a single observation to be less than 20 percent. A UV-color versus B-V plot of luminosity class III, IV, and V stars with little reddening shows a degree of correlation consistent with the observational uncertainty. The data appears to be of sufficient quality for use in deriving B star temperatures and interstellar scattering curves.

Ultraviolet astronomical observations have suffered due to limitations in detector systems, limitations in photometric standards, the problem of obtaining precision data with miniature instrumentation in the rocket environment, and the inability to: select the stars to be observed. It is these restrictions which have dictated the kinds of observations that have been made. The first program to be undertaken was broadband ultraviolet color and flux measurements using small, simple photometers that scanned the sky with the rocket motion. This data is of value in its own right and also has been particularly useful in the design of more sophisticated experiments. As instrumentation techniques improved, it became possible to fly medium resolution objective grating spectrometers to record ultraviolet flux distributions. This data is discussed in the following paper. Now that rocket pointing systems are becoming available, a relatively high resolution scanning spectrometer is being designed to study selected stars in greater detail.

Perhaps the two greatest difficulties associated with the work done to date are that the rockets could not be pointed and that the equipment is destroyed after each flight. Because of the inability to point the rockets, an essentially random selection of stars is obtained on any given flight, and it is necessary to combine the data from a number of flights to produce a useful set of observations. But here the second difficulty is encountered. Since the instruments are destroyed at the end of flight, the internal consistency of the results is totally dependent upon the maintenance of laboratory photometric standards which can be used to calibrate each set of rocket instrumentation and to transfer the calibration from one flight to the next. Because of the emphasis that has been placed on these standards, and because they can be maintained only by repeated absolute calibration, the uncertainties in the absolute calibration of the data are generally smaller than the random uncertainties within the observations themselves. Only by paying careful attention to these calibrations and by demonstrating their consistency from one flight to the next can it be argued that the data is worth interpreting. My principal aim here is to state what kind of accuracy has been achieved.

In obtaining the filter photometry data, two wavelengths have been used: 2600 and 2200 Angstroms; and the photometers have been flown in pairs: one with a 2600 Angstrom filter and one with a 2200 Angstrom filter. The photometers in a pair have parallel optical axes so that simultaneous flux measurements are made at both wavelengths. Originally 60 mm diameter quartz or calcium fluoride lenses were used as objectives; more recently 150 mm diameter Cassegrain telescopes have been employed as collectors. In either case, the filter is placed behind the collector, a field stop and field lens are placed at the focal plane, and a quartz windowed photomultiplier is

*Published in *Annales d' Astrophysique*, 27(6): 805-808, 1964.

used as a detector. The filter characteristics are shown in Figures 1 and 2.

The first filter has an effective wavelength of 2200 Angstroms with a 210 Angstrom effective width. It is a first order interference filter and exhibits a long wavelength tail which contributes

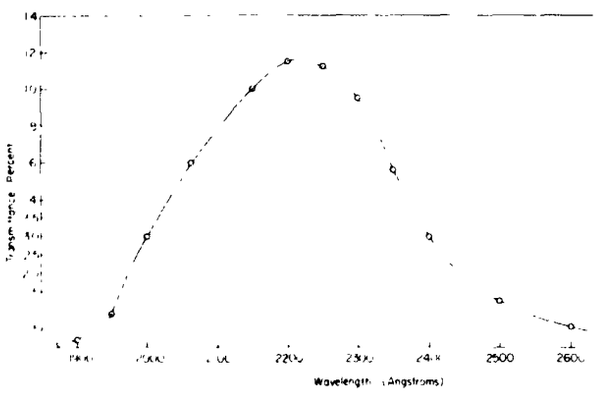


FIGURE 1.— Response curve for 2200 Angstrom photometer.

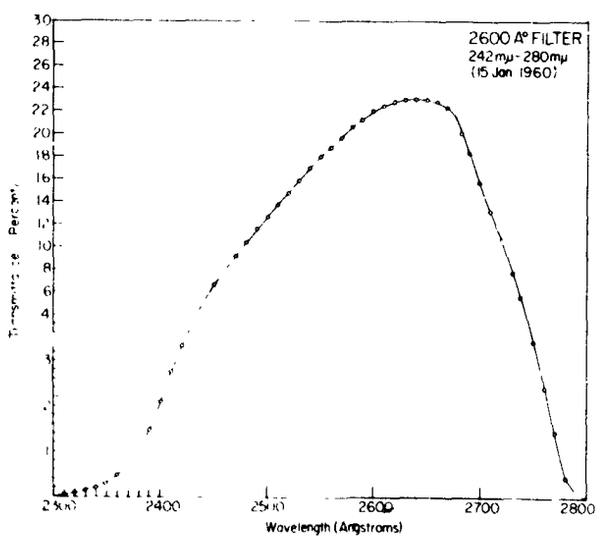


FIGURE 2. —Response curve for 2600 Angstrom photometer.

significantly to the transmitted energy. On the short wavelength side the tail is chopped by the transmission of the optical system. Since the long wavelength tail extends into the bandpass of the 2600 Angstrom filter, its effect on the 2200 Angstrom flux can be allowed for by subtracting a fixed percentage of the flux seen by the 2600 Angstrom filter. This procedure is accurate to within a percent or so for a wide range of spectral gradients. The filter shown in Figure 2 is a sandwich of lead doped potassium bromide-potassium chloride and nickel sulphate hexahydrate crystals together with various high silica glasses. It has an effective wavelength of 2600 Angstroms with a 250 Angstrom bandpass. The calibration of the photometers is fairly straightforward. The transmission of the optical train is measured with a spectrophotometer, and the relative spectral efficiency of the photomultiplier is compared against a known phosphor. Since both filters pass the 2537 Angstrom line of mercury, this line can be used conveniently as an absolute standard in an inverse square law calibration. The mercury standard is maintained by periodically using a thermocouple to compare it to the total radiation from a tungsten lamp calibrated by the National Bureau of Standards.

Three pairs of photometers are flown in a rocket, each pair looking in a different direction in order to cover as much sky as possible. As the rocket spins and processes, the photometers scan the sky, typically seeing an individual star for about two-tenths of a second. In four flights some 300 two color observations on about 180 different stars have been recorded. Forty-eight of these stars have been observed with more than one pair of photometers. Therefore, there are approximately 120 observations, covering forty-eight stars, from which the internal reliability of the data can be judged. Of these forty-eight stars, all observations clustered within 20 percent for forty of them, within 10 percent for seventeen. The uncertainties in this data should be representative of the accuracy of all the observations, so that even for those stars observed only once the data should have an uncertainty of less than 20 percent

It is interesting to see the colors derived for the 48 well observed stars, and to compare them with conventional B-V colors. Here again, one is subjected to a severe selection process. When the selection is restricted to those whose B-V is available, to luminosity classes III, IV, and V, and to those with color excess of .02 magnitude or less, only thirteen of the original forty-eight stars are left, and they are shown in Table 1 and Figure 3. β Per and β Cep are variable, but the variability in (B-V) is small compared to the uncertainty in the ultraviolet colors. ϵ UMa is a peculiar star showing chromium and europium lines; however, with such a small group of stars to work with,

TABLE 1.—*Ultraviolet Colors of Stars With Little Color Excess*

		Sp	B-V	$m_{2200}-m_{2600}$
ζ	UMa	A2V	+.02	−.06
γ	UMa	AOV	.00	.00
α	CrB	AOV	−.02	−.14
ϵ	UMa	AOp	−.03	−.08
β	Per	B8	−.07	−.32
β	Lib	B8V	−.11	−.35
ζ	Dra	B6III	−.12	−.45
τ	Her	B5IV	−.15	-.66
ι	Her	B3V	−.18	−.53
η	UMa	B3V	−.20	−.37
ϵ 1, 2	Sco	B3, 2	−.20	−.36
β	Cep	B2III	−.22	−.49
η	Lup	B2V	−.23	−.55

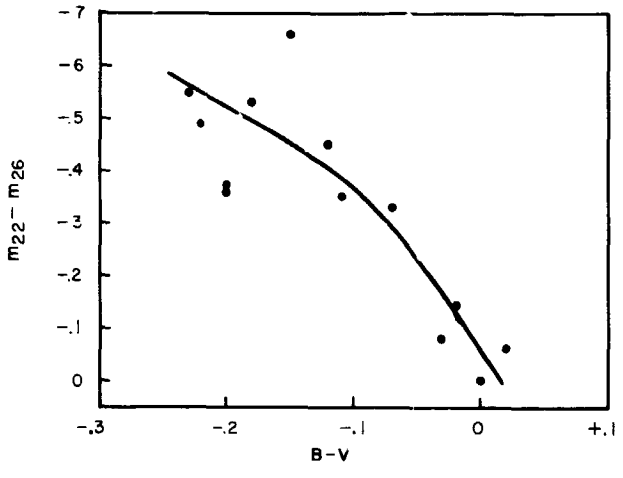


FIGURE 3.—Color-Color diagram for well observed stars of luminosity classes III, IV, and V having $E \leq .02$.

beggars can't be choosers. Finally, η_1 Sco and η_2Sco appear in the photometer field together and this observation must be considered the average of the two stars. Since they are similar in spectral type and color, the observation should still have weight.

The ultraviolet colors listed are absolute. That is, they represent true ratios of the 2200 Angstrom and 2600 Angstrom fluxes. It is fortuitous that early A stars have roughly equal fluxes at these two wavelengths, making the zero point of the ultraviolet color scale nearly correct. The curve drawn in Figure 3 is arbitrary. However, all but three of the stars are within one-tenth magnitude of that curve, and those three are within two-tenths. This degree of correlation appears to be comparable to the accuracy of the data. Such a correlation is very encouraging. It means, for instance, that it should be possible to derive information on interstellar scattering in the ultraviolet, and, of course, to derive a temperature scale for the B stars based on the observed ultraviolet spectral gradient. This phase of the work is just starting, but Stecher will describe some of the first attempts in the following paper in connection with the spectrophotometric data. There again, the matter of calibration is all important. Fortunately there are seven stars in common between the broadband and the spectrophotometric data, so that spectrophotometric fluxes at 2200 and 2600 Angstroms can be compared with fluxes from the filter photometry. The seven stars are ϵ and η UMa, λ Cas, β Cep, τ; ζ and π Sco. The two sets of data, which are obtained by quite different techniques, agree within 20 percent for all seven of these stars, and for four out of the seven there is agreement within 10 percent. In other words, we are now at the stage where the observations are sufficiently reliable to be used with confidence in deriving information on stellar atmospheres and interstellar scattering.

INTERSTELLAR EXTINCTION IN THE MIDDLE ULTRAVIOLET*

ALBERT BOGGESS III AND JAN BORGMAN†

Recently, Boggess (1964) obtained photoelectric photometry of a number of stars in two wavelength bands centered at 2600A and 2200A. A detailed description of the instrumental system will be found in his paper. The accuracy of this rocket-based photometry, as checked by the repeatability of the measurements, is very satisfactory. The observations include six stars in the narrow spectral range 09.5–B1, which appear suitable for deriving a first reliable estimate of the interstellar extinction in the middle ultraviolet. These stars are γ Sco, π Sco, δ Sco, β Sco, ζ Oph, and α Cam, in order of reddening.

The first seven lines of Table 1 give the UBVRIJK photometry of these stars, as far as available at the moment. The data have been taken from Johnson and Borgman (1963) and from more recent work that is in the process of publication. The color excesses E_{B-V} in line 9 have been derived from the MK classification in line 8, using Johnson's (1963) calibration. These color excesses are rather small and an independent deter-

TABLE 1.—*Observations and Inferred Quantities*

		τ Sco	π Sco	δ Sco	β Sco	ζ Oph	α Cam
1	V	2.83	2.88	2.31	2.52	2.56	4.28
2	U−V	−0.26	−0.19	−0.11	−0.08	+0.02	+0.08
3	U−V	−1.27	−1.09	−1.01	−0.89	−0.84	−0.76
4	V−R	−0.18	−0.16:	−0.02	−0.12	+0.08	−0.02
5	V−I	−0.42	−0.32:	−0.16	−0.15	+0.05	+0.06
6	V−J	—	—	—	—	−0.01	—
7	V−K	—	—	—	—	−0.07	—
8	MK	B0 V	B1 V	B0 V	B0.5 V	09.5 V	09.5 Ia
9	E_{B-V}	0.04	0.07	0.19	0.20	0.32	0.38
10	$(N-M)_o$	−0.016	−0.005	−0.012	−0.007	−0.029	−0.021:
11	N−M	+0.005	+0.033	+0.048	+0.057	+0.110	+0.102
12	2.66 E_{N-M}	0.06	0.10	0.16	0.17	0.37	0.35:
13	E_{B-V}	0.05	0.08	0.18	0.18	0.34	0.36
14	F_{2600}	2.49 2/2	2.35 4/3	3.07 2/2	2.74 3/8	1.41 3/2	0.363 1/2
15	$m_{26}-V$	−2.52	−2.51	−2.23	−2.31	−1.63	−1.88
16	$\Delta(m_{26}-V)$	−0.04	−0.11	−0.06	−0.10	.00	−0.02
17	$(m_{26}-V)_{corr.}$	−2.56	−2.62	−2.29	−2.41	−1.63	−1.90
18	F_{2200}	3.70 2/2	2.51 5/3	2.59 3/2	2.38 4/7	1.33 2/2	0.217 2/3
19	$m_{22}-V$	−2.92	−2.55	−2.01	−2.13	−1.54	−1.29
20	$\Delta(m_{22}-V)$	−0.05	−0.14	−0.08	−0.12	.00	−0.03
21	$(m_{22}-V)_{corr.}$	−2.97	−2.69	−2.09	−2.25	−1.54	−1.32

*Published in *The Astrophysical Journal*, 120(4): 1636–1639, 1964.

†Kapteyn Laboratory, Groningen, The Netherlands.

mination seems desirable. For this purpose the intrinsic colors $(N-M)_o$ given by Borgman and Blaauw (1964) have been combined with the observed colors N–M to obtain the color excesses E_{N-M}. Line 12 gives $2.66E_{N-M}$, which is on the E_{B-V} scale as shown by Borgman and Blaauw. In the following, the color excesses $\overline{E_{B-V}}$ of line 13 are adopted; these are the mean values of the entries on lines 9 and 12.

Lines 14 through 17 of Table 1 contain Boggess' 2600A observations and some inferred quantities. F_{2600} is the observed flux in units of 10^{-9} ergs $sec^{-1}cm^{-2}A^{-1}$. The numbers are followed by a weight designation, of which the first figure is the number of observations and the second figure is a weight estimate based on the average quality of the observations. These weight estimates are intended only as a guide when evaluating the data; the relation with mean errors has not been considered. Line 15 gives the color index $m_{26}-V$, based on lines 1 and 14. Though not necessary for the present purpose, some attempt has been made to set the zero point of the color system at AO V. The range in spectral type among the six stars is quite small (09.5–B1); nevertheless the small corrections $\Delta(m_{26}-V)$ were computed in an attempt to remove the intrinsic color differences. Underhill's (1962) models, which were differentially calibrated in $(N-M)_o$, have been used for this purpose. The corrections on line 16 were obtained by interpolating between the models on the basis of the intrinsic colors of line 10 and adopting a zero correction for the intrinsically bluest star, ζ Oph. Hence, line 17 gives the color indices which would have been observed if each star had been intrinsically similar to ζ Oph. It should be pointed out that the corrections of line 16, being small in comparison to the range of m_{26}–V, have little influence on the final result. However, the intended improvement is probably not unrealistic in view of the accuracy of the present observations as well as the very satisfactory agreement between the observed flux and the theoretical atmospheric radiation field in the 2000–3000 Angstrom region (Strömgren 1964, Stecher 1964).

Lines 18 through 21 show the same steps in the reduction of the 2200A photometry as those described above for the 2600A observations. As in line 16, the corrections in line 20 are small.

Giving the same weight to each observation, the following least squares solutions have been derived:

$$^{(m}26^{-V)}\ \text{corr.} = 2.67\,\overline{E_{B-V}} - 2.80 \qquad (1)$$
$$\pm 0.34\ \text{(p.e.)} \quad \pm 0.08\ \text{(p.e.)}$$

and

$$^{(m}22^{-V)}\ \text{corr.} = 4.91\,\overline{E_{B-V}} - 3.12 \qquad (2)$$
$$\pm 0.25\ \text{(p.e.)} \quad \pm 0.06\ \text{(p.e.)}$$

Rejecting the only observation which is single (the 2600A flux of α Cam), solution (1) changes to

$$^{(m}26^{-V)}\ \text{corr.} = 3.33\,\overline{E_{B-V}} - 2.84 \qquad (3)$$
$$\pm 0.26\ \text{(p.e.)} \quad \pm 0.08\ \text{(p.e.)}$$

The uncertainty of the 2600A observation of α Cam is confirmed by the peculiar position of this star in the $^{(m}26^{-V)}$ corr.[9] $\overline{E_{B-V}}$ diagram, which may also be partly caused by uncertain intrinsic colors of this supergiant. We prefer to adopt solution (3) instead of (1). The last terms in each of the relations (1), (2), and (3) are the intrinsic colors of an 09.5V star. The probable errors do not include any zero point error of the color system.

An investigation of the color excess ratios $E_{V-I}/\overline{E_{B-V}}$ and $E_{V-K}/\overline{E_{B-V}}$, based on lines 5, 7 and 13 and the intrinsic colors given by Johnson and Borgman (1963) leads to the conclusion that the reddening law in the direction of ζ Oph and α Cam is probably the "normal" case, which results in $A_V/E_{B-V}=3.1$. The reddening of the other four stars is too small for an evaluation in this respect. It is not unlikely that some of the other stars are affected by an extinction law which is characterized by $A_V/E_{B-V}=4$ as appears to be the case for the most reddened region in the Scorpio-Centaurus group (Borgman 1964). However, the reddening of these four stars is small and the result of the reduction will largely be determined by the reddening of ζ Oph and α Cam. Hence, it is assumed that the reddening law derived from the six stars in Table 1 is essentially a "normal" one which can be represented in other wavelength regions by the Cygnus reddening law given by Johnson and Borgman (1963) (excluding the U-point where the adopted effective wavelength is subject to reconsideration), by the mean

reddening law of Borgman (1961), and by the statistical determination of $A_V/E_{V-B}=3.1$ (see *e.g.* Schmidt-Kaler 1961). These observational results are listed in Table 2 and are illustrated in Figure 1, together with Van de Hulst's dielectric grain extinction curve No. 15 (Van de Hulst 1949). Everything has been normalized to $A_V=3.1$ mag., corresponding to $E_{B-V}=1$ mag. It is evident that Van de Hulst's extinction curve fits the observations quite well to about 4000A but fails to explain the extinction at shorter wavelengths, particularly the new results at 2600A and 2200A. A better fit can probably be obtained with extinction models based on composite particles. A preliminary computation by Greenberg (1964) shows that this may indeed be the case for the carbon core-ice mantle grains suggested by Wickramasinghe (1963).

TABLE 2.—*Normalized Extinction Observations*

Author*	1	E
J & B	0.47	0.22
J & B	0.86	0.85
J & B	1.19	1.45
J & B	1.56	2.24
B	1.73	2.87
J & B	1.83	3.10
B	1.93	3.33
B	2.26	4.02
J & B	2.30	4.10
B	2.49	4.45
B	2.68	4.75
B	2.78	4.90
B	3.01	5.27
B & B	3.85	6.43
B & B	4.54	8.01

*J & B: Johnson and Borgman 1963.
B: Borgman 1961.
B & B: present paper.

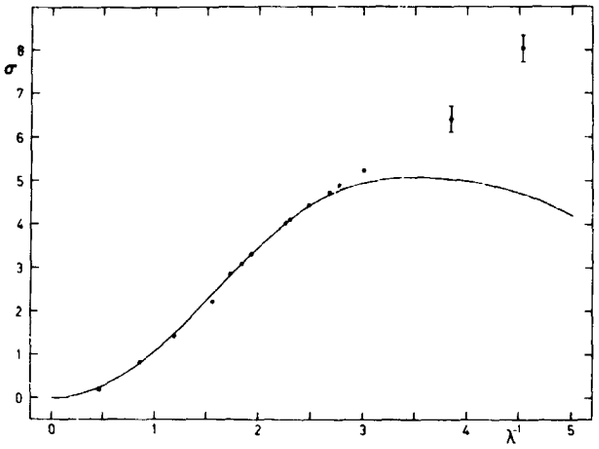


FIGURE 1.—Normalized extinction observations of Table 2. Error bars at the 2600A and 2200A points indicate probable errors. The curve is the theoretical extinction curve No. 15 by Van de Hulst (1949).

REFERENCES

BOGGESS, A. 1964, in preparation.
BORGMAN, J. 1961, *B. A. N.* **16,** 99.
BORGMAN, J. 1964, Report on "Joint Discussion on Local Structure and Motions in the Galaxy", ad. F. J. Kerr, in press.
BORGMAN, J., and BLAAUW, A. 1964, *B.A.N.* **17,** 358.
GREENBERG, J. M. 1964, private communication.
JOHNSON, H. L. 1963, "Basic Astronomical Data", ed. K. Aa. Strand, Univ. of Chicago Press, 1963.
JOHNSON, H. L., and BORGMAN, J. 1963, *B.A.N.* **17,** 115.
SCHMIDT-KALER, Th. 1961, *A.N.* **286,** 113.
STECHER, T. 1964, private communication.
STROMGREN, B. 1964, *Rev. Modern Physics* **36,** 532.
UNDERHILL, A. B. 1962, *Publ. D.A.O.* **11,** 433.
VAN DE HULST, H. C. 1949, *Rech. Obs. Astr. Utrecht* **11,** part 2.
WICKRAMASINGHE, N. C. 1963, *M.N.* **126,** 99.

4 - 2+1-20

ORIGIN OF THE SOLAR SYSTEM*

A. G. W. CAMERON

The question of the origin of the solar system might for convenience be opened with ideas advanced by Descartes a little more than three centuries ago, just about the time the concept of the scientific method as a way of investigating nature was being formulated.

Descartes suggested that the solar system arose from a large vortex containing primitive matter, some going to form the Sun and some, involved in smaller eddies, going to form the planets and their satellites. Considering the state of knowledge of fluid dynamics and science three centuries ago, this was not a bad picture. It is not even grossly different from the picture that many of us have today.

Over the last three centuries, there have been two basic schemes suggested for the formation of the solar system. Each has had ups and downs. Why has this problem been with us so long, and so little progress made with it?

Most of the discussion in the last three centuries has tried to explain a very small number of facts—a few regularities in the solar system: The Sun rotates and the planets are in the same plane around the Sun, and so on. It all amounts to saying merely that the angular momentum vectors in the solar system tend to be aligned. Also, there does seem to be a regular increase in the spacing of planet orbits in the solar system, something that has been called the Titius-Bode law, if it is a law, and no one is sure.

Until recently, that is about all cosmogonists had upon which to base answers to the problem. It is only in the last decade that the solid base of facts needed to make some real progress has been built. This base has come partly from the intensive investigations of meteorites now going on, and from the chemical boundary conditions to the theories of the origin of the solar system that can be set from analysis of the chemical composition and mineralogical properties of meteorites. It has come from new techniques in astronomy, particularly radio astronomy, giving us great insight into the properties of planets and the Moon. And it has come from direct spacecraft explorations, particularly the Mariner flight toward Venus and the discovery that that planet does not have a magnetic field like the Earth's.

But, at the same time, the more discovered about the present composition and properties of the solar system, the more are the facts that need explaining, and the more difficult it is to do this.

Figure 1 (a) will help in reviewing briefly one of the two main ideas about the formation of the solar system. First, there is the idea, about two centuries old, and very popular early in this century, that there was a close passage of two stars that raised huge tides and pulled out filaments of gas. In some way the planets were to have condensed out of the filaments. It is now fairly certain that nothing like this happened, for a variety of reasons. Spitzer, for example, has shown that gas drawn out of a star in this way would be so hot that solids would not condense from it, but the gas would disperse in space instead. Moreover, Lyttleton showed that most material drawn out from a star in this way—more than 99 percent of it—would fall back on the star. Perhaps most important, it is now fairly certain that the Sun cannot contain any appreciable amount of deuterium; it would all have been destroyed by thermonuclear reactions early in its history. But quite large amounts of deuterium, relatively speaking, do occur in the Earth. It cannot have once been in the Sun. Why should other planetary constituents have been?

*Published in *Astronautics & Aerospace Engineering*, August 1964.

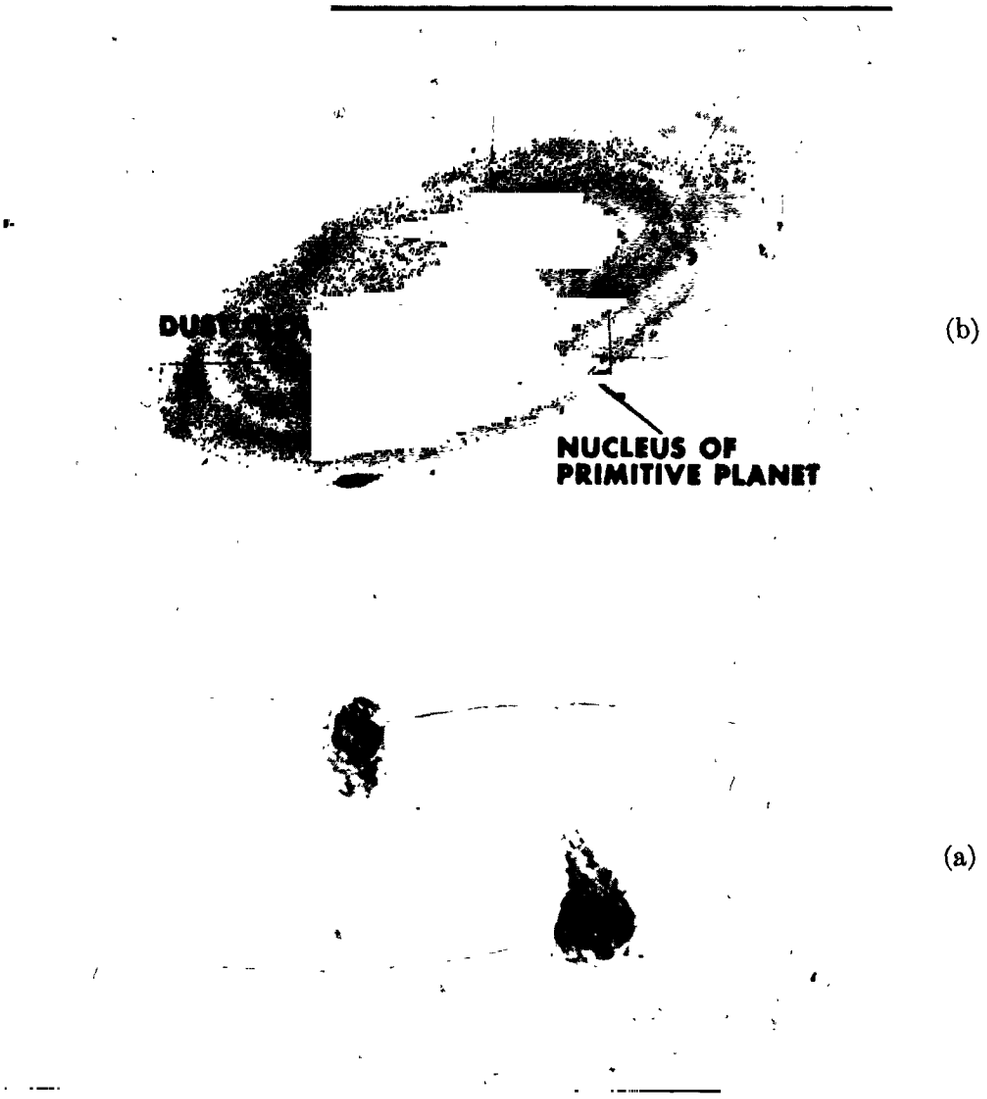


FIGURE 1.—The "fried egg" concept of formation of the solar system (b) is believed generally today, while the once popular supposition that filaments of gas became planets after an encounter of two stars (a) is considered unlikely.

What of the other idea? Figure 1 (b) shows the fried-egg concept which I think most of us tend to believe today, in some form or other, and which is essentially the picture Descartes introduced three centuries ago. There is some sort of a primitive nebula of gas and dust surrounding the Sun, formed as a part of the process which generated the Sun itself.

Figure 2 shows a typical spiral galaxy. Seen in color, the light from the center is reddish, and the light from the distant parts of the disk is bluish. This color difference indicates that the stars in those different parts of the galaxy have quite different characteristics. The reddish color comes from rather old, red giant stars in the central portion of galaxies of this sort, and the blue light from young stars with high surface temperatures.

The outer, bluish stars evolve quite rapidly through their thermonuclear history, and live only a few million years. However, presence of the central red giant stars indicates that the galaxy has been around for a few billion years at least. Thus there is evidence that star formation continues in the outermost parts of the disk in this kind of galaxy, in the region of young stars, where we find the Sun in the Milky Way.

We are beginning to know something about the age of our own galaxy from the study of how stars



FIGURE 2.—Light on the age of our own galaxy is being shed by studies of typical spiral galaxies such as this. When seen in color, light from the center is reddish and that from distant parts of the disc has a blue hue, indicating characteristics of stars in various areas of the galaxy.

have evolved in the star clusters of the Milky Way. Its age seems at least 10-billion years and perhaps as much as 25-billion; say, nominally, 15-billion—much older than the age of the Earth and the meteorites and, presumably, the solar system as a whole, which only goes back some 4½-billion years.

Because the galaxy is much older than the solar system, and because we would expect that star formation should take place much more rapidly in the early stages of the history of our galaxy than at the present time, it is fair to postulate that the conditions in the disk of the Milky Way at the time the Sun was formed were certainly not qualitatively different, and perhaps only slightly different quantitatively, from what we see in space today.

Consider the gas between the stars in the Milky way out of which new stars would have to form. This gas is being continually replenished from the evolution of older stars. There may be some still remaining from the very earliest days of our own galaxy that never formed into stars. But, at any rate, there is certainly a replenishing going on, for instance, through supernova and the less-catastrophic nova explosions. Moreover, some stars throw off continuously a great deal of mass as a very intense stellar "wind." We can look at the regions in our own galaxy where we see very massive hot stars. Because these stars live for such a short time, just a few million years, we know that where we see them must be the position, or very close to the position, where they formed.

Figure 3 shows the Orion Nebula, revealing some details of its structure. In particular, it exhibits patchiness due to dark absorbing regions which are quite inhomogeneously distributed. This is important. The gas, which presumably has to condense to form stars, shows quite large fluctuations in density.



FIGURE 3.—This revealing photo of Orion Nebula shows fluctuation in density of gas, which presumably has to condense to form stars.

There are, then, regions of gas in space having variation in density. The average density overall of the galactic plane is probably about one particle per cubic centimeter. But a typical interstellar cloud may have a density of 10 particles per cc. The upper extreme observed for the particle density in some clouds can be of the order of 1000 particles per cc.

When considering what conditions must be satisfied for gas to become gravitationally unstable and condense, or collapse, to form stars, it is necessary to compare the gravitational potential energy of such a cloud with its internal thermal and magnetic energies. The collapse can take place only if the gravitational potential energy, intrinsically a negative quantity, becomes sufficiently greater than the positive internal energies.

When conditions are attained under which that can take place has been a puzzling question, and has led to different approaches to the subject. These different approaches, in turn, have led to different suggestions as to the primitive solar nebula.

One important approach takes the following tack. Hoyle, Fowler, Greenstein, and others look for cloud complexes large enough to exhibit the necessary gravitational potential energy for collapse. This generally requires some 10^5–10^6 solar masses of material becoming gravitationally unstable and starting to contract as a whole.

Such a region of space, a complex of clouds of gas in space, cannot contract very far before it starts running into difficulties that prevent it from contracting further. One of these may be the magnetic energy contained within it. But a more probable one is the rotation that the cloud possesses.

It is thought that some sort of a general magnetic field exists in space, and that this field's lines of force are reasonably well glued to the material in the cloud. Even in a cloud where the hydrogen and the helium are neutral, starlight keeps such atoms as carbon ionized all the time. These ions will then stay with the lines of force, and the friction of the neutral atoms with these ions will tend to keep the entire mass of the cloud glued to the magnetic lines of force.

So the magnetic field in space can exercise some control over such a large cloud. It probably exerts one main control by requiring the cloud to rotate once during the time it makes one revolution around the center of the galaxy. For the distance of the Sun, that would take about 200-million years.

The rotation may be a bit larger than that, or a bit smaller; but as long as the magnetic field exercises any reasonable degree of control, it cannot be even one order of magnitude different. The conservation of angular momentum will cause the cloud complex to spin more rapidly as it contracts, and the contraction will be halted when the centrifugal force at the equator becomes equal to the gravitational forces.

For this cloud complex to contract further, rotation must twist the lines of force of the general interstellar magnetic field, and in this way transmit angular momentum from the cloud to the surrounding regions. This action will slow the rotation and allow a further contraction of the cloud.

After some period, and the removal of a great amount of angular momentum, the cloud can shrink by a large factor, and small parts of it become unstable toward separate gravitational contraction. Density fluctuations should exist in such a cloud complex and it should fragment into smaller and smaller sub-units, eventually getting down to fragments about the order of one solar mass.

According to the other general point of view, which I advocate, the optimum condition for obtaining the contraction of an initial interstellar cloud does not involve a large cloud complex, but rather a typical cloud, which may have a mass of the order of 1000 solar masses. Assuming this typical cloud, one can calculate that it becomes gravitationally unstable when the density becomes as high as something like 1000 particles per cc—the highest density which is actually observed in space. Such a high density, it should be expected, can only be formed by sliding the material along the lines of force of the interstellar magnetic field. Otherwise, the internal magnetic pressure would be so high that at a density of 1000 particles per cc the cloud would immediately expand to a much smaller density.

When such a cloud becomes gravitationally unstable, difficulties concerning rotational instability tend to disappear, and there is no need to invoke a transmission of angular momentum from

the cloud to the surroundings to bring about a collapse to quite small dimensions and a fragmentation into units of the order of the Sun in mass. Furthermore, collapse takes only about 1 percent of the time it would have taken in the initial state to go through one revolution. So there is in fact no opportunity to twist the lines of force sufficiently to transmit angular momentum to the surroundings, and it is necessary to conserve the initial angular momentum all the way down.

What, then, is the state of the fragment of about one solar mass, the ultimate unit which can result from this general contraction of the cloud in space?

In the picture he put forward, Hoyle assumes sufficient loss of angular momentum to cause his ultimate fragments to contract down as far as about the orbit of Mercury before they start spinning rapidly enough to shed mass in the equatorial plane. Continuing to contract, the Sun would shed more and more mass in the equatorial plane, its initial magnetic field being retained to some extent within the material as it shrank, but linking the gas in the equatorial plane of the Sun to the central body. As the central body shrinks, it spins more and more rapidly, and thus tends to get ahead of the gas in the equatorial plane, which is going around with Keplerian orbital motion. The magnetic linkage acts to spin more rapidly the material shed in the equatorial plane, giving it more angular momentum, and so causing the material to go farther and farther away from the central body itself. In this manner the material in the nebula formed inside the orbit of Mercury eventually passes over the entire region now occupied by the solar-system's planets.

This view of the development of the planets, proposed by Hoyle in conjunction with Fowler and Greenstein, requires that small bodies will chemically condense out of the gas shed in the equatorial plane and accumulate to planetesimals of some tens of meters in dimension. And eventually these "metric planetesimals," as they call them, will accumulate to form the plants.

The inner planets, they explain, form largely out of metals and silicate material—much the sort of thing now observed—because they believe that the first objects to grow and accumulate in space should be the bodies that condense out at the highest temperatures. Then, as the gas goes farther and farther out in space, it drops icy materials—H_2O, NH_3, CH_4—and eventually, by the time the gas has gotten to the outskirts of the solar system, the general region of Uranus and Neptune, the remaining material, mainly hydrogen with some helium, may tend to exaporate into space, or, as Gold has suggested, be blasted out by streams of particles from the Sun.

The other principal feature of the ideas put forward by Hoyle, Fowler, and Greenstein is that during the time that the objects have dimensions of the order of some tens of meters, the Sun exhibits considerable magnetic activity, due to its linkage with the solar nebula, which it is pushing farther and farther away into space. This causes very intense flares, accelerating particles to hundreds of millions of electron volts. These particles intensively bombard the surfaces of the metric planetesimals, thus producing some interesting nuclear effects, such as the lighter elements—the deuterium, lithium, beryllium, and boron—that one finds in the solar system, as well as producing certain isotopic variations observed in meteorites and the Earth. It becomes a very complicated story.

A contrasting view to all this involves my own hypothesis. It starts with a smaller cloud in space, one unable to get rid of any angular momentum. A typical fragment that arises from its contraction will tend to become rotationally unstable at dimensions of the order about 100 AU—quite a bit larger than the dimensions of the solar system. When this object contracts further; it forms a disk of gas only, with no body at the center in hydrostatic equilibrium—that is, no central body to be the Sun.

Given such a primitive disk, how does the solar system form? Here again, we must rely in some way on the magnetic field which spreads through the interstellar medium to begin with, and which has been trapped in the material as the body has shrunk out of space. As there is still a fairly strong density concentration toward the center of the nebula, the material moving about the center of the mass has a shear. Adjacent layers thus tend to go at different speeds, pulling out the lines of force of the magnetic field, so that they lie more and more parallel to the direction of the shear. This squeezes the lines of force together, making the magnetic field stronger, and thus containing more energy.

The energy has to come from somewhere. Since the shearing motion is pumping energy into the magnetic field it has to come at the expense of the shearing motion itself. But the shearing motion in turn depends entirely on the mass distribution in the primitive disk, and the only way in which energy can come from the shear is to redistribute the mass in disk. This would require that most of the matter in the disk flow inward, and some at the outer edge of the disk of course flow outward so that the disk as a whole conserves its angular momentum.

How the Sun would form seems fairly clear. But the only chance to form the planets in this picture comes before there is much redistribution of mass; for the motion of the gas inward to form the Sun—no matter how long it takes, a million years, or what have you—will draw with it all of the little solids, unless they have grown into a large enough body to withstand such motion.

How large must such bodies be to withstand inward movements? It turns out that in the inner part of the solar system they have to be larger than the largest asteroids.

This disk, on the other hand, very clearly has chemically condensed-out material to the extent of at least a hundred times the mass of the present planets. So the planets need not represent a very efficient collection of solid debris in this massive primitive solar nebula. It is only necessary to get 1 percent of the material into the form of planets. So we are interested in a very inefficient method of accumulation. And presumably the great bulk of the material does not grow into bodies at least asteroidal in size.

Accumulating bodies in space presents an extremely difficult problem. If very small bodies come together gently enough, then it may be that they can adhere one to another. The presence of a thick solar nebula can greatly assist such an accumulation process. But take the gas away and let the bodies start colliding with an appreciable relative velocity, and there will tend to be a general shattering and diminishment of bodies rather than an accumulation

There are a number of other points that one might discuss in connection with either of these theories; and they have been discussed in depth in special conferences. Aside from any potential correctness, they have served as grounds for experiments, and in particular have defined some particular things to look for in the scientific investigation of the Moon and the planets.

Earth, of course, presents a disconcerting problem. Its surface record goes back only about 3-billion years, only about two-thirds the age of the Earth. There can be no first-hand investigation of the home planet in its formative stages, only a surmising. And clever theory will not substitute for direct observations in shedding light on the origin of the solar system. We must go to the Moon and to the asteroids and the other bodies in space, particularly the smaller bodies, before we can really try to pin down how it all happened.

BIBLIOGRAPHY

JASTROW, ROBERT and CAMERON, A. G. W., eds., "Origin of the Solar System," Academic Press, New York, 1963

GRAVITATIONAL COLLAPSE*

HONG-YEE CHIU†

With the exception of a few supernova remnants which are in our galaxy, most cosmic radio sources are "radio galaxies." Although flare stars do emit radio waves occasionally, no ordinary stars with strong, steady radio emission have been found. The typical optical power of stars is from 10^{30} ergs/sec (white dwarfs) to 10^{38} ergs/sec (super giants). For comparison, the optical power of the sun is 4×10^{33} ergs/sec. The typical radio power of supernova remnants is around 10^{36} ergs/sec. For a giant galaxy (containing approximately 10^{11}–10^{12} stars with a total mass of around $10^{11}\odot$, where $\odot$ = solar mass = 2×10^{33}g), the optical power is around 10^{44} ergs/sec. Radio emission from normal galaxies is generally weaker, the power ranging from 10^{37}–10^{39} ergs/sec. For certain peculiar galaxies, the so-called "radio galaxies," the radio emission rate ranges from 10^{41}–$10^{44.5}$ ergs/sec.

However, since 1960, another kind of strange object has been found — "quasi-stellar radio sources"; these formed the center of discussion at the symposium. Their radio power is around 10^{44} ergs/sec and their optical power is around 10^{46} ergs/sec, *which is 100 times the total energy output rate of a giant galaxy!* To date nine of these objects have been found. On photographic plates, these peculiar radio sources have star-like appearances, and one of them showed erratic light variations with a time constant of the order of one year and a long-term quasi-period of ten years. A modest estimate of the total energy content of each of these objects (in the form of the kinetic energy of electrons of 1 BeV energy) gives a minimum of 10^{60} ergs, which is the *rest energy* of about 10^{6} suns.

Are "quasi-stellar radio sources" stars? Ordinary stellar-structure theory precludes stars of mass greater than 100$\odot$. In such stars, radiation pressure dominates and because of the thermodynamic properties of radiation, these stars will pulsate with large amlitudes as a result of this pulsation, mass will be ejected, thereby reducing the mass of the star. Are "quasi-stellar radio sources" galaxies? Their angular size is less than 0.5″, while normal galaxies at corresponding distances (determined by their red shift) will have an angular size of the order of 3″. In several cases, the light variation precludes this possibility because galaxies do not have light variations.

So far, the clumsily long name "quasi-stellar radio sources" is used to describe these objects. Because the nature of these objects is entirely unknown, it is hard to prepare a short, appropriate nomenclature for them so that their essential properties are obvious from their name. For convenience, the abbreviated form "*quasar*" will be used throughout this paper.

The idea of having a meeting on gravitational collapse came after the discoveries of the enormous energy output of quasars and the subsequent gravitational-collapse theory of F. Hoyle of Cambridge University and W. A. Fowler of the California Institute of Technology to explain this enormous energy output. When the symposium met, the original hypothesis of Hoyle and Fowler (reviewed at the meeting by E. M. Burbidge of the University of California at La Jolla) was abandoned and was replaced by those to be reported later. Meanwhile, new observational data had been accumulated and were presented at the meeting. It seems safe to say that, at the present time, the interior structure and the physical mechanism of the energy production of quasars are still not understood.

*Published in *Physics Today*, May 1964.

†Goddard Space Flight Center and Columbia University. Mr Chiu is a NRL-NAS Senior Research Associate and also a Research Fellow Academia Sinico, Republic of China.

Fusion of hydrogen into heavier elements will yield an energy of about one percent of the rest energy of matter. If the energy source of the quasars were nuclear in nature, it would be necessary that $10^{5}\odot$ of hydrogen be burned in around 10^{8} years. It is rather unlikely that nuclear energy could be released in unison on such a large scale. Moreover, nuclear reactions usually take place in the interior of a star; the energy released gradually diffuses out, and finally is radiated thermally as black-body radiation, peaked at an energy of around 2 eV. How can the energy of electrons be elevated from a few eV to a few BeV without violating the second law of thermodynamics? Using the most efficient thermodynamic cycle, a nuclear-energy source many orders of magnitude higher is needed[8] Therefore, there is only one energy source left for us to consider: gravitational energy. Incidentally, the gravitational field is the weakest field in nature, but it is the only one that allows us (at least theoretically) to extract up to 8/9 of the rest energy of matter.

After this brief introduction to the central theme of the discussions, I shall now report the individual papers that were presented at the symposium. In order to have continuity in this report, I shall not report the talks chronologically, but in a logical manner.

Reviewing the common properties of radio sources, T. Matthews of Caltech pointed out that radio emission from normal galaxies is rather weak compared with their optical emission. Their radio power is around 10^{37-39} ergs/sec while their optical power is around 10^{44} ergs/sec. Some rare, peculiar spiral galaxies may have a radio emissive power of around 10^{40} ergs/sec. Strong radio galaxies with a radio power of 10^{41-44} ergs/sec are most frequently found among the so-called D systems. These are galaxies with excessively bright nuclei (Seyfert galaxies). Some strong sources are also identified with elliptical galaxies. The optical emission rate of quasars is around 10^{46} ergs/sec and their radio emission rate is around 10^{44} ergs/sec. (See Fig. 1 for a plot of radio and photographic magnitudes of galaxies and quasars.)

To date, a total number of nine quasars have been identified from the contents of the revised 3C catalogue (3C stands for the *Third Cambridge Catalogue of Radio Sources*) which contain about 470 radio sources, many of them still unidentified optically. This catalogue was compiled by a group working under M. Ryle of the University of Cambridge. The two quasars which have been extensively studied and which were extensively discussed at the meeting have catalogue numbers 3C273 and 3C48 respectively.

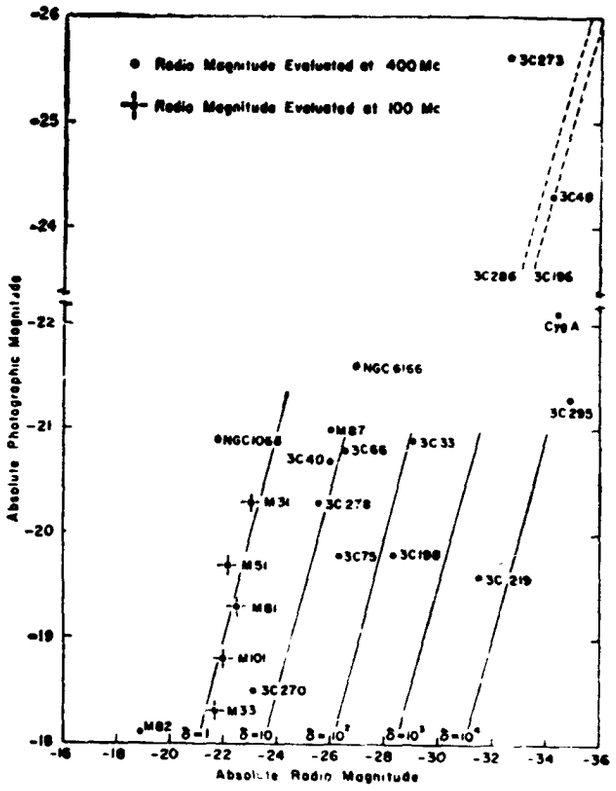


FIGURE 1.—Above, a plot of absolute radio magnitudes and absolute photographic magnitudes of galaxies and quasars. The absolute photographic magnitude, a measure of the optical power, is defined in the footnote on p. 22. The absolute radio magnitude M_r is defined to be $M_r = -54.1 - 2.5 \log_{10} S$ where S is the radio power in watts per second per cycle bandwidth at 400 Mc/sec, with the exception of a few normal galaxies where S is evaluated at 100 Mc/sec; δ is the ratio of radio power to optical power normalized to unity for normal galaxies. Other radio galaxies have a value of δ up to 10^4. Quasars 3 C 273 and 3 C 48 are found in a region in the graph where no galaxies are found. The distance of two other quasars (3 C 286 and 3 C 196) have not been determined and their possible position in this graph are indicated by the dotted lines. The sources of data used here are references 1 and 2.

Radio interferometers with a resolution of the order of 1″ of arc were first used to study the shape of some of the stronger quasars. This method also enables one to locate accurate positions for these sources. This was done chiefly by the

Manchester group (R. P. Allen, B. Anderson, R. G. Conway, H. P. Palmer, V. C. Reddish, and B. Rowson) and by the Caltech group (P. Maltby, T. A. Matthews, and A. T. Moffet). In the radio position of 3C48, an object resembling a 16^m star was found on the photographic plates.* This "star" was surrounded by a nebula measuring $5'' \times 12''$ which does not resemble any galaxy. (See Fig. 2.)



FIGURE 2.—Below, an actual photograph of 3 C 48 (shown in enlarged image at upper right). Objects D and B are reference stars from which the magnitude of 3 C 48 is determined. A faint nebulosity can vaguely be seen. (Photo by T. A. Matthews and A. R. Sandage taken with Hale Telescope, Mt. Palomar, Ref. 3).

At the same time, lunar occultation observations of 3C273 were carried out by the Australian group (C. Hazard, M. B. Mackey, and A. J. Shimmins). This other method of observation provided an optical identification for 3C273. Because of gravitational perturbations, the orbital plane of the moon is not fixed in space but precesses with a period of around 19 years. Thus, the moon, with an angular size of 30 minutes of arc, can occult stars within a belt of 20° as it moves around the earth. When a radio source is occulted, one can learn from the diffraction pattern about the location, shape, and size of the radio source. Fig. 3 shows some observed diffraction patterns of 3C273. In particular, it is easy to tell if a source is an extended one. It was found that 3C273 consists of two sources, *A* and *B;* their separation is $19.5''$ (see Fig. 4). Component *B* is slightly elliptical and four times weaker than component *A*. Using the $200''$ Hale telescope, A. Sandage of Mt. Wilson and Palomar Observatory obtained photographs which show, at the radio position of *A*, a nebula shaped as a jet, pointing away from a 13^m star-like object which is identified as component *B*. One of Sandage's photographs appears in Fig. 5.

Radio spectra for both components of 3C273 and for 3C48 have been obtained. Briefly, the mechanism for radio emission and the nature of radio spectra can be described in the following way. The optical radiation of stars and nebulae is chiefly emitted by thermally excited ions. The spectra of stars and nebulae are rich in absorption or emission lines; their over-all features can be understood in terms of black-body radiation peaked around certain optical wavelengths. The radio spectra that are normally observed can be broadly classified into two categories: thermal and synchrotron emission. In the thermal spectrum, the intensity per unit frequency increases towards shorter wavelengths for emission from "optically thick" emitters. This kind of spectrum resembles that of the long-wavelength end of a black-body thermal radiation, hence its name. Most likely it is produced by electron bremsstrahlung, which occurs when electrons pass very near ions. In the synchrotron spectrum, the intensity decreases towards the shorter wavelengths, and usually the intensity per unit frequency is proportional to $\nu^{-\alpha}$, where ν is the frequency and α is a positive number usually between 0.3 and 1.2 (for most sources, α has a value of around 0.8).

The paths of electrons in a magnetic field are not straight lines but curves. Hence these electrons suffer acceleration while traveling on a

*Superscript *m* stands for the magnitude; $m = -2.5 \log_{10} F + C$ where F is the observed energy flux and C is a constant. At a distance of 32.5 light years the sun has a magnitude of +4.85. This is also the distance at which the absolute magnitude of an object is defined. A 6^m star is just barely visible to the eye. A difference of 5 magnitudes amounts to a difference in measured energy flux of 100.

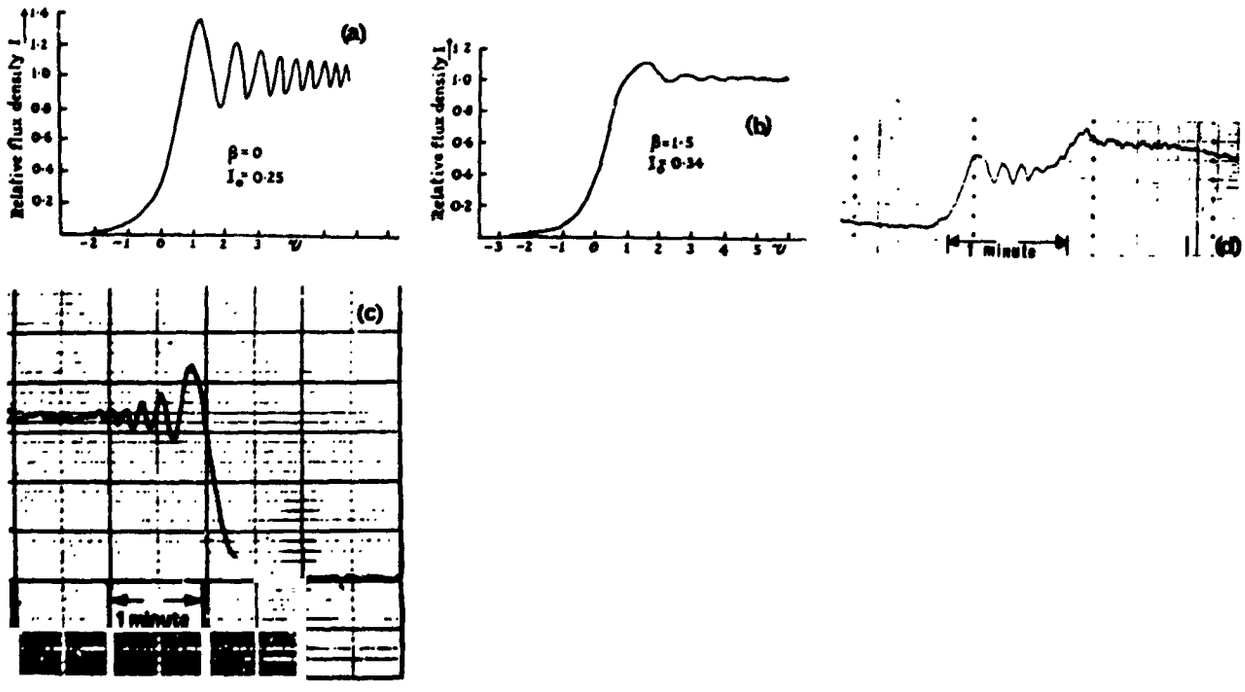


FIGURE 3.—Intensity observed as function of time in lunar occulation: (a) shows diffraction pattern calculated for a point source; (b) shows that of an extended object. Note decrease in amplitude for an extended object. Angular diameter w is related to β by $w=\beta/(C\lambda/2d)^{\frac{1}{2}}$ where λ is the wave length and d the distance to the moon; v is proportional to the time; $v=0$ corresponds to the geometrical shadow. An actual emersion diffraction pattern of 3 C 273 at 410 Mc/sec on August 5, 1962, is shown in (c). Note the similarity to that for a point source (a). Immersion diffraction pattern of October 26, 1962, at 1420 Mc/sec, is shown in (d). Note resolution of 3 C 273 into point source and extended source (Ref. 6, 7).

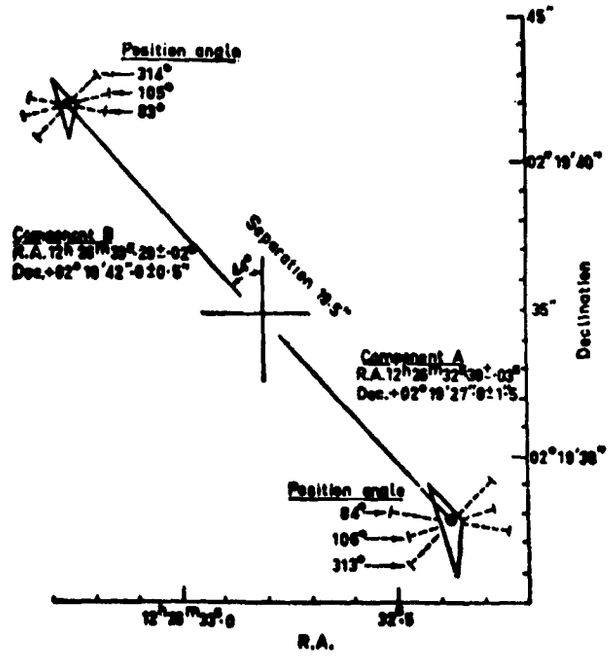


FIGURE 4.—Diagram of the radio source 3 C 273. Sides of full triangles represent widths of equivalent strip source as measured at 410 Mc/sec for each of three position angles indicated (Ref. 6).

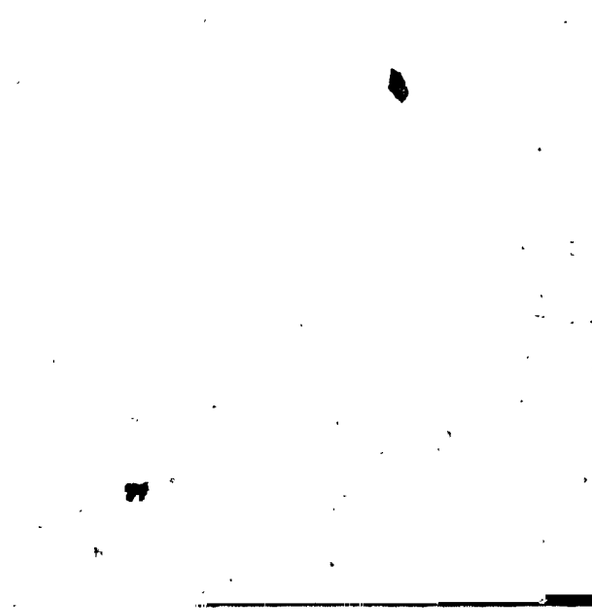


FIGURE 5.—Photograph of 3 C 273. Note the jet which points away from the starlike object. Although its distance is 1.8×10^9 light years, this object is easily seen through an amateur's 6-inch telescope. (Photo by A. R. Sandage with Hale telescope at Mt. Palomar.)

curved path. These "accelerated" electrons then radiate polarized electromagnetic waves, primarily in the radio region. Such radiation from monoenergetic electrons was first observed in electron synchrotrons, hence the name synchrotron radiation. The observed cosmic-ray spectrum is usually

$$N(E)dE = E^{-k}dE$$

where $N(E)dE$ is the flux in the energy band dE and k is a positive number; $k = 1.5$ for high-energy cosmic-ray particles. We may assume that the relativistic electron component of cosmic rays also obeys this law. Then the synchrotron radiation intensity per unit frequency is proportional to $\nu^{-\alpha}$.

The radio spectrum of 3C273 shown in Fig. 6 is a composite of the two sources A and B, and can

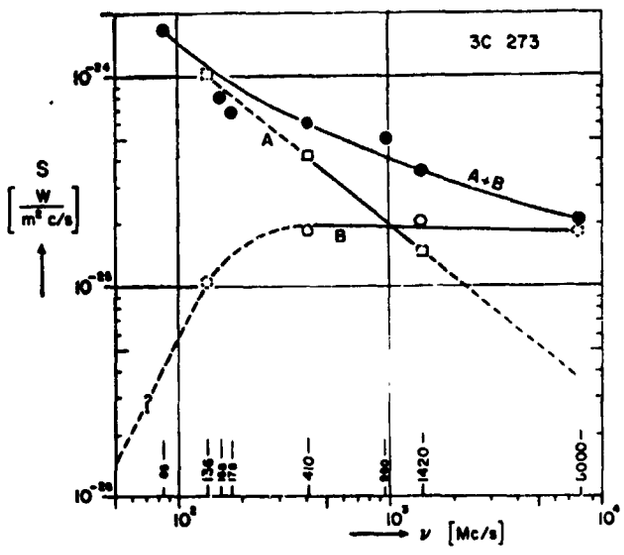


FIGURE 6.—Radio spectrum of 3 C 273. Black dots are measurements of both components. Squares give spectrum of component A; the open circles of component B. The line through the open circles shows how one could fit a thermal spectrum to it. But this interpretation seems highly unrealistic. (Courtesy of W. Priester.)

be resolved as indicated. This resolution was obtained from lunar occultation observations. The spectrum of A seems to belong to the synchrotron type, and that of B to the thermal type. However, this interpretation of the observed spectrum of source B would require a temperature of the order of 10^9 °K. Such a high temperature contradicts the lifetime estimated by present theory.

The optical spectrum of 3C48 has also been available for some time. Several emission lines were found, but no unequivocal identification with lines of known atoms or ions could be made. Moreover, J. Greenstein of Caltech found that the spectra of some fainter quasars, 3C147, 3C196, and 3C286 (also showing somewhat weaker emission lines), looked quite different from that of 3C48.

M. Schmidt studied the spectrum of the bright object, 3C273. He found four emission lines which formed a simple harmonic pattern with separation and intensity decreasing towards the ultraviolet. This series of lines looks like that of any hydrogen-like atoms. Unfortunately, they correspond to no elements found on earth. A unique interpretation can only be made by assuming a red shift ($\Delta\lambda/\lambda$) of 0.16 and that the emitting element is hydrogen. (A far greater red shift has to be assumed if the emitting element consists of other hydrogen-like atoms.) Indeed, with this interpretation, a red-shifted Hα line (shifted from 6563A to 7590Å) was found in the infrared by J. B. Oke of Caltech. Similarly, the line spectra of all other quasars can be interpreted as red-shifted emission lines and forbidden lines. In 3C48 the emission line identified corresponds to that of [Mg II] and the forbidden lines to those of [O II], [Ne III], and [Ne IV]. In 3C48 the red shift is 0.37. The corresponding velocity (if the red shift is interpreted as the Doppler shift) is around 0.15 c and 0.3 c for 3C273 and 3C48, respectively.* Fig. 7 shows the red-shift interpretation of a number of quasars.

Another feature of quasars is that they show a strong ultraviolet excess even at a red shift of $\Delta\lambda/\lambda = 0.8$. Using filter techniques, A. R. Sandage and M. Ryle have successfully identified a few more quasars (3C9, 3C216, 3C245) from a number of objects showing strong ultraviolet excess.

Could quasars be stars in our galaxy with a high receding velocity? This is extremely unlikely. First, the gravitational field of our galaxy cannot retain stars with a velocity $> 10^{93}\,c$. Second, the proper motion (apparent motion of stars with respect to background stars) would appear

*The Doppler shift is $d\nu/\nu = (-\nu/c)\ (1-\nu^2/c^2)^{-1/2}$, where ν is the relative velocity of the source and the observer. For the same velocity most cosmological theories predict a shift $d\nu/\nu$ in between $(-\nu/c)\ (1-\nu^2/c^2)^{-1/2}$ and $-\nu/c$ (linear).

large compared with other stars of much smaller velocity.

W. H. Jefferys of Wesleyan University (Connecticut) studied the proper motion of 3C273. Since 3C273 is fairly bright, its image was found among many sky-patrol plates dating back to 1888. Plates from many observatories using different telescopes, accumulated over the past fifty years, were used to determine the proper motion.

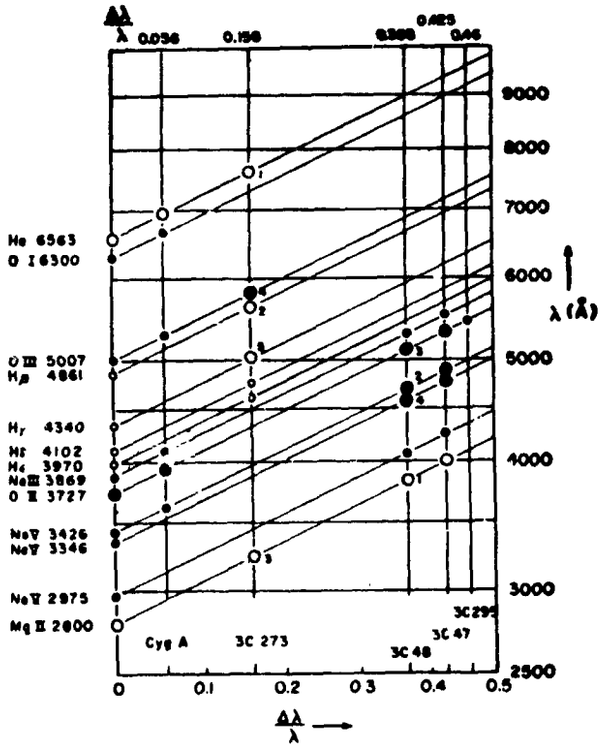


FIGURE 7.—Schematic spectra of 3C 273, 3C 48, and 3C 47 together with Cygnus A and 3C 295. The ordinate is the red shift $\Delta\lambda/\lambda$. Forbidden lines are represented by black dots. The numbers underneath the dots indicate the relative strength of the lines, number 1 being the strongest line in the spectrum. (Courtesy of W. Priester).

The inertial system used is the so-called FK3 system. This refers to a group of more than 900 stars catalogued in the *Dritter (Third) Fundamental-Katalog des Berliner Astronomischen Jahrbuchs* (Kopff, 1934) with the highest precision then available.* The measured proper motion of 3C273 is 0.001 ± 0.0025 seconds of arc per year. The standard deviation is from uncertainties associated with the accuracy of the absolute position of stars in the FK3 system. From the value of proper motion and the velocity derived from the differential rotation of our galaxy, the distance of 3C273 is $>65\,000$ light years, and this puts 3C273 on the border of our galaxy (the size of our galaxy is 10^5 light years). A similar result was also obtained by W. J. Luyten of the University of Minnesota, using more recent plates.

Can the red shift be gravitational in nature? The red shift of light from a self-gravitating body of radius R, mass M, is of the order of $GM/Rc^2 = 10^{96}(M/\odot)(R/R_\odot)^{-1}$. ($R_\odot$ = solar radius $= 7\times10^{10}$ cm). In order to have a red shift as large as 0.1, $(M/\odot)(R/R_\odot)^{-1} \approx 10^5$. For a star of one solar mass and an $R \backsim 10^6$ cm, a density of about 10^{15} g/cm^3 would be required. (E. E. Salpeter of Cornell University reviewed the equilibrium configuration of superdense stars. It is not possible for superdense stars of mass $>>\odot$ to exist, nor can the red shift of light from superdense stars exceed 0.3.) The presence of spectral lines implies a surface temperature of the order of 10^4 °K. From the radius of superdense stars, one can estimate the total radiative power, and thus deduce the distance from the observed flux. If 3C273 were a superdense star, its distance would be 0.3 light years and it would be practically inside our solar system! From the intensity of the forbidden lines J. Greenstein further concluded that the distance would be even smaller. Such stars would have been discovered by Kepler through the failure of Kepler's laws!

One may also assume that quasars have bigger masses and bigger radii than superdense stars. For example, one can assume that 3C273 is a star at a distance of 1000 light years. In order to be compatible with the energy output rate, the radius has to be 10^{13} cm. Then, in order to account for the red shift, the mass has to be $10^7\ \odot$! W. A. Fowler pointed out that such stars cannot even burn hydrogen, because they would fall within their gravitational radius before the temperature would be high enough (10^7 °K) to burn hydrogen!

D. Williams of the University of California at Berkeley reported the observation of interstellar absorption lines in 3C273. When radio waves of distant galaxies reach us, part of the energy at the 21-cm wavelength is absorbed by hydrogen in our

*Recently, the FK 4 Catalogue, containing more than 1,500 stars, has been published.

galaxy. A very weak 21-cm absorption line was found in the radio spectrum of 3C273 with a line shift corresponding to a velocity of 1 km/sec ± 10km/sec. This effect, however, is very marginal.

Since no other rational interpretation can be given to the red shift, one can only attribute it to the cosmological red shift. Using Hubble's law, the distance of 3C273 is placed at 1.8×10^9 light years. This gives an energy output rate for 3C273 of 10^{46} ergs/sec at optical frequency. Some people will question whether Hubble's law is well established. Hubble's law states that the distance of galaxies to us is linearly related to their red shift:

$$\text{Distance} = \frac{\text{Recession velocity}}{H}$$

The proportionality constant $H(\approx 100\text{km-sec}^{-1}\text{-megaparsec}^{91})$* is known as Hubble's constant. $H^{-1}c$ is a measure for the dimension of our universe; at a distance $\approx H^{-1}c$ the red shift is total. One often hears that astronomical observations are not accurate enough to distinguish one type of cosmological theory from another, and many have expressed doubts about applying Hubble's law to obtain the distances of unknown objects from their red shift. The answer is that these two problems are somewhat different in nature. Assuming that the local density of galaxies is the same throughout the universe *at a given time* and that the curvature of the universe is everywhere constant, different cosmological theories will predict different galactic densities as a function of the distance (this galactic density is observed from a single space-time point) and the distance is obtained from the red shift by the above equation. The galactic density observed at a single space-time point would be different for different cosmological theories because of the expansion of the universe, and because the expanding rate as a function of the distance is different in different theories. Unfortunately, the various cosmological theories predict measurable differences in galactic densities only at distances much beyond the reach of the world's most powerful telescope—the 200″ Hale telescope at Mt. Palomar. Hubble's law, however, is established accurately for neighboring galaxies by applying knowledge of the period-luminosity relation of Cepheid variables. (Cepheids are bright variables of luminosity $\backsim 10^5\ L_\odot$, and their period is related to their luminosity by a simple law. Hence, by observing the period and the apparent brightness of Cepheid variables of other galaxies, one can obtain a distance.) Then the absolute brightness of more distant galaxies is related to the red shift through a complicated statistical analysis eliminating the difference in intrinsic luminosities of individual galaxies. The result is a law that can be trusted to within a factor of 2 in distance estimates. Hence we believe that Hubble's law may be applied to quasars to obtain their distance accurately.

H. J. Smith of McDonald Observatory took a bold look at quasars. Is there light variation in these objects? Sky-patrol plates from Harvard University Observatory are available back to 1888. Although a large number of old plates have aged and become unreliable, some of them are still in fairly good condition. In measuring light variations, one only needs to compare the image of 3C273 with neighboring stars. Smith and D. Hoffleit of Yale found a large light variation with an amplitude up to 50 percent and a quasi-cycle of fifteen years (see Fig. 8). A. R. Sandage observed 3C273, 3C48, and 3C196 more recently with a photoelectric device and found some erratic variations up to a few tenths of a magnitude. From statistical theory, if 3C273 were a galaxy with a nominal number of stars ($\backsim 10^{10}$), the expected annual variation would be much less than 10^{-5} of the variation of any star. This limit is obtained by assuming that all stars in the galaxy are variables of the same period. Hence it is not possible that 3C273 and other quasars are compact galaxies. Further, in order that the light from such an object can vary with a time constant $\backsim$ 1 year (the shortest variation observed), the physical dimension of 3C273 cannot be much bigger than the time it takes light signals to go across it—namely, one light year.

Having established that quasars are neither local stars nor galaxies, the next thing to do was to study further the details of their spectra. J. Greenstein reported that these lines have been identified with the forbidden lines O^+, Ne^{++}, and Ne^{++++}. Certain atomic excited states decay through higher-order interactions, and their lifetimes are long ($\sim$1 second). Under ordinary lab-

*1 parsec = 3.25 light years.

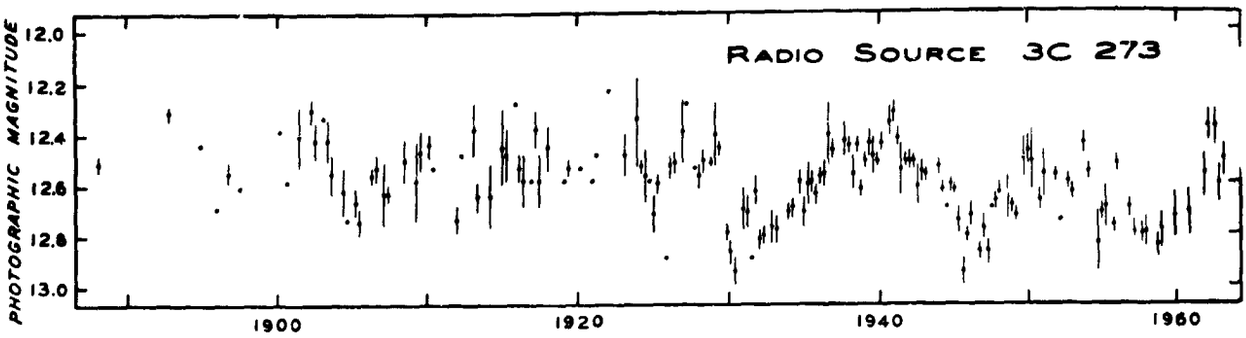


FIGURE 8.—Light variation curve of 3 C 273 as analyzed by H. J. Smith and D. Hoffleit from old sky patrol plates. Data represent statistical average over a few days of observations, and standard deviation is as indicated by vertical line segments. (H. J. Smith pointed out that a big light crash of 3 C 273 occurred towards the end of October 1929. It was a big year for crashes.)

oratory discharge-tube density ($\gtrsim 10^{15}$ particles/cm^3), collisional de-excitation takes place before atoms in these excited states can decay, and the corresponding lines are ordinarily unobservable in the laboratory (forbidden). The presence of forbidden lines indicates low electron density ($< 10^7$/cm^3). As the density increases, forbidden lines gradually disappear.

The observed forbidden lines are [O II], [O III], and [Ne III], but [Mg II] is not a forbidden line. Incidentally, a sizable fraction of energy radiated ($\sim$20%) is contained in these lines. From their observed relative strength the electron number density N_e may be established. Knowing the distance from the red shift, the absolute intensity of the strength of these lines and hence the total number of atoms emitting them may be determined. The table lists Greenstein's results.

Quasar	N_e/cm^3	*Radius (parsec)*	$M/\odot$ *(of emitting parts)*
3C273	10^7	0.6	2×10^5
3C48	10^5	5	2×10^6

If quasars were superdense stars, the differential red shift at different layers would limit the thickness of the emitting layer to that corresponding to the observed width of forbidden lines. This would limit the total number of emitting ions. From this limit and the observed intensity of forbidden lines, Greenstein concluded that quasars would have to be as close as the moon! Such large gravitational perturbations on the solar system could not have escaped the keen eye of Kepler. Hence, from the presence of forbidden lines one can argue indirectly but strongly that the red shift cannot be gravitational.

The lower limit for the masses of 3C273 and 3C48 can be estimated as follows: From the line broadening, we can estimate the kinetic velocity of emitting atoms. It turns out to be 1500 km/sec. In order that these atoms be bound gravitationally at a distance of 0.6 parsec, the center of 3C273 must contain a mass of at least $10^8\odot$. This is the lower limit for the masses of 3C273 and 3C48.

One can also obtain the mass from energy estimates. To know the total energy radiated, one needs to know the age of quasars. The energy output rate is 3×10^{46} ergs/sec, equivalent to the nuclear energy output from the burning of one solar mass of hydrogen in one day. In 3C273, there is a jet, component *A*, 150 000 light years away from the other component, component *B*. It is most likely that the jet was ejected from 3C273. If one accepts this interpretation, then the age of 3C273 must be at least 150 000 years. This gives a value of 10^{59} ergs for the total energy radiated.

But this is not the whole story about the energy content of quasars. The optical emission can be interpreted as resulting from collisional excitation caused by the thermal motion of ions. From the shape of the radio spectrum (Fig. 6), the radio emission is most likely from synchrotron radiation. Thermal radiation with such strong intensity would require a temperature of the order of 10^9 °K and at this temperature its lifetime against neutrino loss is less than a few days. Relativistic electrons moving in a magnetic field can radiate predominantly in the radio region. To date, we do not know of any way of accelerating high-energy electrons with high efficiency, and it seems con-

sistent with known cases (e.g., the Crab Nebula) at about 1 percent of the energy stored in all high-energy particles may be in the form of high-energy electrons. Also, magnetic fields contain energy. A minimum estimate of the total energy consistent with radio data is at least 10^{60} ergs.

The presence of a magnetic field is further supported by the discovery of polarization in the optical region. One does not expect a very large polarization, because the plane of polarization of electromagnetic waves will be rotated while passing through a plasma; the rotation angle is different for different frequencies (Faraday rotation). Hence, for a given band width, the polarization decreases as electromagnetic waves pass through a plasma. At a wavelength of 4500 angstroms, a polarization of 0.45 (± 0.15) percent, based on twelve observations, has been obtained by McDonald. The angle of polarization is defined to $\pm 5°$. This indicates that there is a small component of synchrotron radiation in the optical region as well.

To summarize the discussion on the observed features of quasars, we conclude that (1) quasars are massive objects with a mass $\gtrsim 10^8 \odot$ which are confined to a space of about one parsec; (2) they are not composed of stars but seem to be coherent masses; (3) their lifetime is at least 10^5 years; (4) they radiate radio waves strongly, but radiate most strongly in the optical region; (5) their surface is composed of very tenuous matter of a particle-number density of the order of $10^7/\text{cm}^3$; and (6) their total energy content is larger than the rest energy of $10^6 \odot$.

Now I shall report some theoretical speculations based mostly on the energy requirement of quasars. All of the 10^{60} ergs of energy ($=10^6 \odot c^2$) are in the form of magnetic fields and high-energy particles, and a possible energy source is gravitational energy, which can be released in violent events (e.g., collapse); thus all the energy released may be in the form of high-energy particles. The prevalent theory of gravitation is general relativity. General relativity predicts some singularities, one of which, the Schwarzschild singularity, will be discussed briefly below.

Classically, if one is able to accumulate matter indefinitely and statically, the escape velocity at the surface will eventually exceed the light velocity. As an interesting historical sidelight, I remind the reader that in conjunction with the corpuscular theory of light, Laplace demonstrated in 1795 that a body of the dimension of the orbit of the earth around the sun, with a density of that of the earth, would not allow any of its rays to escape. Hence, he concluded that the most luminous object in the universe may not be visible.

The same conclusion holds in relativity theory. In fact, the mass limit calculated according to Newtonian theory is exactly the same as that predicted in general relativity theory! The light emitted at a distance R from a sphere of mass M is red-shifted by the factor $(1-2GM/Rc^2)^{-\frac{1}{2}}$ as observed by distant observers; for light emitted at the surface, defined by $R^*=2GM/c^2$, the red shift is infinite. R^* is called the gravitational radius or the Schwarzschild radius. The time dilatation factor is also $(1-2GM/Rc^2)^{-\frac{1}{2}}$. To an outside observer, a body falling toward R^* will never reach it, since in approaching it, the time dilatation factor $(1-2GM/Rc^2)^{-\frac{1}{2}}$ diverges and the motion will slow down and eventually stop (time machine!). Signals sent out by the falling body will be more and more red-shifted to very long wavelengths. To a local observer falling towards the center, however, the time of descent is finite. In fact, from the present radius of the sun (7×10^{10} cm), the time of descent towards the center, assuming the sun is within its gravitational radius, is less than one day. For a proton, $R^*=10^{-33}$ cm; for the sun, $R^*=2.6\times 10^5$ cm; and for a massive object $M\sim 10^8 \odot$, $R^*\sim 10^{13}$ cm. For a collapsing star in which the pressure can be neglected, J. R. Oppenheimer and H. Snyder, then of the University of California at Berkeley, found in 1939 that the above conclusion was still valid.

Although no light can cross the gravitational radius, the gravitational field and the electrostatic field of the gravitational singularity (those carried by the so-called longitudinal gravitons and photons) can be felt by external observers. Thus, the Schwarzschild singularity can be detected through its gravitational field, but it cannot be seen. Also, because for a comoving observer the Schwarzschild singularity does not seem to exist, many relativists believe it is just a singularity introduced by the particular coordinate one chooses to describe the gravitational field. Various efforts have been made to find a coordinate system free of the Schwarzschild singularity, including a modification of the topology of the space-time continuum,

but in all cases, singularities always reappear at other places.

In practice, the Schwarzschild singularity sets an ultimate limit on how far a body can collapse gravitationally. This limit is actually approached in the model proposed by F. Hoyle and W. A. Fowler. In a bold attempt, F. Hoyle analyzed the conditions for the formation of large coherent masses. The difficulties of forming large masses are (1) angular momentum and (2) breaking up of large masses to smaller ones prematurely. The period of rotation is proportional to R^{-2} (R is the dimension of the object), but all other time scales (the free-fall time, the time of transit of hydromagnetic waves across a body) are proportional to $R^{-3/2}$ in the nonrelativistic case. Hence, rotation will prevent matter from condensing further when the speed of rotation at the equator equals the escape velocity. As a comparison, rotation becomes important during the formation of the sun when the radius of the protosun is around 10 to 100 times the present value.

In the relativistic region, all time scales have the same dependence on R. Other physical characteristics of the body (density, mass, magnetic field, and so on) then determine which time scale becomes the most important one. In order that a large coherent mass may be formed, the gravitational time scale must be the shortest one. A comparison of the gravitational time scale $[\sim(3\pi/32G\rho)^{\frac{1}{2}}]$ and the hydromagnetic time scale $[\sim(4\pi\rho)^{\frac{1}{2}}R/H]$, where H is the magnetic field, shows that, for a magnetic field of 10^{-6} gauss and a density of 10^{-21} g/cm^3, the gravitational time scale is shorter than that of the hydromagnetic wave when the mass is greater than $10^6\odot$. (As a comparison, the galactic magnetic field is in between 10^{-6} gauss and 10^{-5} gauss, the interstellar matter density is 10^{-24} g/cm^3, and the density of nebulae is 10^{-21} g/cm^3.) From a similar comparison of the rotation time scale ($\sim\omega^{-1}$) with the gravitational time scale, one finds that for $\omega=10^{-15}$ sec^{-1} (the angular velocity of a galaxy), the condition is $\rho>5\times10^{-24}$ g/cm^3. Thus one need not have a spectacular density in order to overcome the limit imposed by angular momentum.

In this simple argument, one does not consider the transfer of angular momentum from one part of the body to the other. If considerable amounts of angular momentum are transferred, the previous argument is not valid. It is not known whether angular momentum may be transmitted in large quantities. Optimistically, one hopes the inner part can fall a great way before transfer of angular momentum becomes important.

Thus, conceivably, a large mass ($\gtrsim10^6\odot$) can condense to a density of around 10^{-16} g/cm^3. At this density, due to rotation, the body becomes appreciably flattened (see Fig. 9). Let the thickness of the disc be W, the diameter be D. Now the time scale for the vertical direction is determined by W and this gravitational time scale is much larger than the hydromagnetic time scale. Fragmentation can now take place in the outer

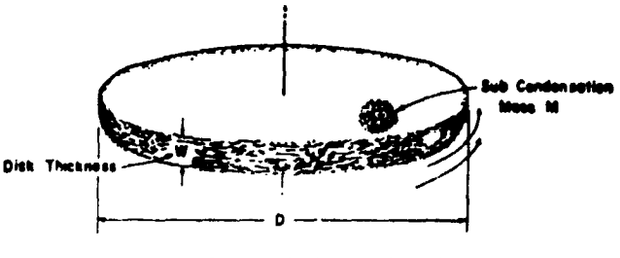


FIGURE 9.—Flattened disc in Hoyle's theory.

part. The mass m in each fragmented blob is given by $m/M\sim(W/D)^2$; for $W/D=10^{-2}$ and $M=10^6\odot$, we have $m=10^2\odot$.

The temperature of the gas can be calculated by applying the virial theorem (which equates the thermal energy of a self-gravitating sphere to its self-gravitational energy). The calculated value for this case is around 10^4 °K, somewhat below the ionization temperature of hydrogen at the corresponding density. Further contraction (now taking place individually in smaller blobs) will raise the temperature, and hence ionize the gas. The ionization of hydrogen requires an energy of about 10^{13} ergs/g and the thermal (and gravitational) energy is about 10^{12} ergs/g at 10^4 °K. The gravitational energy is proportional to R^{-1}. To obtain an energy of 10^{13} ergs/g, a contraction by a factor of ten, or an increase in density by a factor of 10^3 is needed. The density suddenly increases from 10^{-16} g/cm^3 to 10^{-13} g/cm^3.

The gas cloud becomes opaque to radiation only until $\rho\sim10^{-11}$g/cm^3. During the contraction phase, energy has to be supplied to the radiation field as well. Hence the contraction cannot stop until a density of 10^{-11} g/cm^3 is reached. As I

have mentioned earlier, stars of masses $\gtrsim 10^2 \odot$ are not stable against pulsation. A check on the thermodynamic mechanical properties of a $100\odot$ star shows that it is only marginally in stable equilibrium. Now a rapid contraction takes place almost like a collapse, suddenly raising the density from 10^{-16}g/cm^3 to 10^{-11}g/cm^3 in a time comparable to the gravitational free-fall time. Will the collapse stop? In a star of a few solar masses, we are quite sure that the collapse will stop before nuclear energy generation runs wild. The kinetic energy associated with the collapse will be damped out in the subsequent oscillations. In a 10^2 ⊙ star, the collapse will probably overshoot, and the accompanied oscillation will have very large amplitude.

The most efficient damping mechanism is the electron-pair annihilation process of neutrino production. But this is important only when $T > 10^9$ °K; at lower temperature, there is no damping. Since the temperature fluctuation of the star is very large, nuclear reactions (hydrogen burning) will proceed in a somewhat erratic manner. An estimate of the actual time scale for nuclear burning is difficult; but, in order to fit those sporadic light variations in 3C273, the time scale for nuclear burning should be around one year. This rate also fits the observed energy flux (3×10^{46} ergs/sec).

S. A. Colgate of Lawrence Radiation Laboratory and A. G. W. Cameron of Goddard Institute for Space Studies pointed out, however, that F. Hoyle's mechanism does not provide enough high-energy particles for radio emission. Shock waves generated by supernovae, traveling in the very tenuous matter in which these stars are embedded, will accelerate particles to relativistic velocities. A considerable amount of gravitational energy will be released during supernova collapse. Moreover, shock waves will generate hydromagnetic waves throughout the medium, which may explain the long-term light variations (15 years) of 3C273.

Remnants of these massive stars may be in the form of neutron stars or a Schwarzschild singularity, either of which may accelerate particles to relativistic speeds. This may be the mechanism that produced a jet in M87, a galaxy. The idea of acceleration was advanced by Ya. B. Zel'dovich of the USSR.

Another alternative suggested by Hoyle is that the inner region of the big body can collapse gravitationally to extreme density; rotation can either be crushed by gravitational force or dissipated by radiating gravitational waves, assuming a noncylindrical shape was achieved through fission. F. Dyson of the Institute for Advanced Study remarked that the difficulty associated with a gravitational-collapse theory is that it would be over in one day (the local time it takes for a test particle to cross the Schwarzschild singularity of $10^2 \odot$). How can a state of collapse be maintained for over 10^6 years, which would be required in quasars?

T. Gold of Cornell pointed out that in supernova explosions the light output increases rapidly but declines slowly, and this is in contradiction to what is observed in quasars (see Fig. 8).

Hoyle summarized his talk with a diagram which is reproduced here as Fig. 10.

W. A. Fowler then discussed a similar mechanism for massive objects. As condensation takes place, the inner part begins to rotate rapidly, and fission can take place. When the inner part bifurcates into a nonsymmetrical form, gravitational radiation will take place, radiating away an appreciable amount of rotational energy (which is comparable to the total rest energy). This raises the potential energy outside, and part of the matter will have a positive binding energy. Mass will be ejected. The difficulty here, as pointed out by a number of others, is that one cannot avoid a coupling between inside and outside if fast nonsymmetrical rotation takes place.

T. Gold presented the idea that if stars escaping from star clusters can carry away enough angular momentum, star clusters can collapse, and collisions between stars will occur more frequently. Energy released in such collision processes is fairly large, and certain radio phenomena can be explained.

A number of other topics not related to massive objects, but relevant to gravitational collapse or to galactic structure were also discussed.

E. M. Burbidge reviewed several proposed mechanisms for strong radio sources. Originally, a radio source in Cygnus was identified with a pair of galaxies appearing to be in collision. R. Minkowski of the University of California at Berkeley postulated that collision between normal galaxies could generate high-energy particles responsible

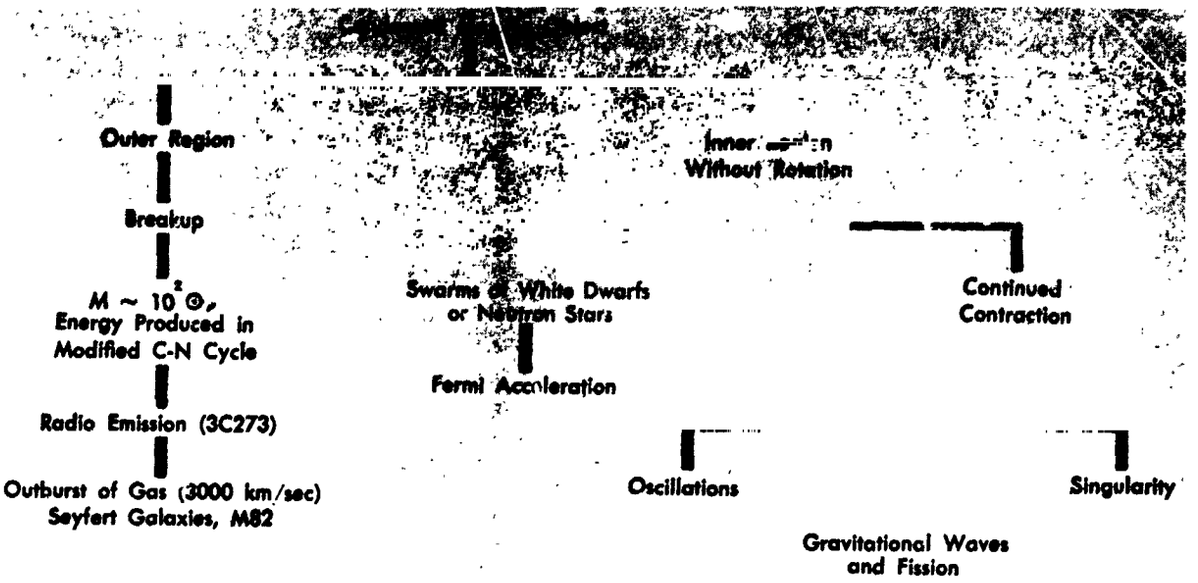


FIGURE 10.—Chart summarizing possible mechanizing of gravitational collapse proposed by F. Hoyle.

for radio emission, but this idea was later rejected because the energy thus released would be far too small. Also, the known number of radio sources is much greater than would be expected from the frequency of galactic collisions for most optically identified sources not associated with clusters of galaxies. G. R. Burbidge of the University of California at La Jolla and F. Hoyle proposed that collisions between galaxies and anti-galaxies (composed of antimatter) might give enough energy. However, there are many inconsistencies in this proposal. For example, why should matter and anti-matter be separated until the right moment to annihilate? Another mechanism is the emission of radio waves through the interaction of a turbulent magnetic field with gases. One can perhaps explain the radio emission from M82 by this mechanism, but in an elliptical galaxy, there is practically no gas! The existence of large scale, strong magnetic fields in elliptical galaxies is also unlikely.

If one accepts supernova remnants as the only radio emitters in a galaxy, one needs a total number of 10^8 supernovae spreading over 10^8 years. I. Shklowsky of the USSR suggested that during the early stages of galactic evolution, rapid star formation can take place, and supernova activities may be stronger than at present. A. G. W. Cameron modified Shklowsky's idea, proposing that in regions rich in gas and low in magnetic field, massive stars can condense simultaneously. After going through their normal evolutionary sequence, they become supernovae—all in around 10^6 years. Remnants of these supernovae can account for the radio emission. The difficulty is the same as in the proposal of G. R. Burbidge and F. Hoyle: in elliptical radio galaxies there appears to be very little gas.

G. R. Burbidge postulated that in the center of a galaxy, if the separation between presupernova stars is around a few light days, the supernova explosion of one of them may trigger the rest of them to become supernovae. In order that his mechanism may work, a stellar density of 10^7 stars/cubic parsec is needed. Such a high density may exist and escape detection. However, one does not know of any trigger mechanism.

H.-Y. Chiu of Goddard Institute for Space Studies reported on a study of the evolution of presupernova stars in which it was found that a very dence core will develop before a star undergoes callapse. Almost all the energy released during collapse will be dissipated in the form of neutrinos, and the dense core will become neutron matter.

E. E. Salpeter reviewed the equilibrium configuration of neutron stars. At a density greater than 10^{12} g/cm^3, matter is composed predominantly of neutrons, and at an even higher density ($> 10^{14}$

g/cm³) hyperons may exist. These neutrons cannot decay because all electron states are filled up. There exists a mass limit for neutron stars, and no static equilibrium configuration is possible if the mass exceeds this limit. This limit for a perfect Fermi gas is 0.76⊙ (Oppenheimer-Volkoff mass limit). If real gases are considered, the limit varies somewhat, but in no case does it exceed 3⊙.

H. Bondi of King's College, London (reported by E. E. Salpeter), derived a rigorous upper limit ($\Delta\lambda/\lambda < 0.615$) for the red shift of light emitted from the surface of neutron stars. A more realistic limit is found to be around 0.3, which incidentally also excludes the possibility that the large red shift (>0.3) of some quasars is a gravitational red shift. In principle, this is the upper limit of the red shift for light emitted from any self-gravitating body.

J. A. Wheeler of Princeton discussed another kind of gravitational singularity. He assumed that one can accumulate matter *statically* until at the center, the time metric g_{00} vanishes (this corresponds to a classical situation in which the gravitational potential *at the center* of the body is equal to its rest energy). Any matter dropped through hole to the center will give away energy equivalent to its rest energy and thus the total gravitational mass (as observed by external observers through its gravitating effect) cannot increase. Such a singularity acts as a gravitational machine, converting all matter into energy. Since the number of nucleons is conserved, Wheeler postulated the existence of a class of new massless bosons, the δ ray, which carries the nucleon number. One of the chairmen remarked that δ rays, if they exist, may be used to transport ponderable matter at the speed of light.

However, it is not known if one can really accumulate matter statically until this kind of gravitational singularity comes into existence. All neutron-star models studied so far indicate that a dynamical collapse will take place when the mass exceeds the critical mass limit (the mass limit was reviewed by Salpeter).

A spirited discussion then took place, reemphasizing some of the points discussed during the meeting. E. L. Schucking of the Southwest Center for Advanced Studies pointed out that other discussions relevant to massive objects, cosmic rays, and cosmic x and γ rays had been omitted from the discussion. P. Bergmann of Yeshiva University remarked that in the past general relativistic effects had been observed only in the weak field limit. Now new developments of astrophysics have made relativity a more physical theory. He expressed hope that a quantized version of relativity theory may play an important part in understanding massive objects. R. Minkowski expressed satisfaction that within the past ten years the technique of optically identifying radio sources had made great progress. On the other hand, a theoretical understanding of the structure of large masses is lacking. In one context, J. R. Oppenheimer, who pioneered the study of neutron stars and gravitational collapse, remarked that the present situation in interpreting quasars resembled that of quantum electrodynamics eighteen years ago, when all one had was confusion and lots of data. If the meeting taught us anything, it has taught us that those two dark clouds in the otherwise clear sky of physics (Kelvin, at the turn of the last century, was referring to natural radioactivity and X rays) are still there. But they are now to be found at the very edge of our universe.

ACKNOWLEDGMENT

The author wishes to thank T. A. Matthews, A. R. Sandage, C. Hazard, M. B. Mackey, A. J. Shimmins, H. J. Smith, D. Hoffleit, W. Priester, and Miss Jane Woodbridge for their help in preparing this report. Thanks are also due to the editors of the *Astrophysical Journal*, *Sky and Telescope*, and *Nature* for permission to use a number of previously published illustrations.

REFERENCES

The following is a list of references with which the author is acquainted. He does not claim that this list is representative or complete.

Radio Spectra

1. R. G. Conway, K. I. Kellermann, and R. J. Long, "The Radio Frequency Spectra of Discrete Radio Sources," Monthly Notices of the Royal Astronomical Society **124**, 343 (1962).
2. R. C. Roeder and G. C. McVittie, "The Luminosity Function of Extragalactic Radio Sources," *Astrophys. J.* **138**, 299 (1963).

3. T. A. MATTHEWS and A. R. SANDAGE, "Optical Identification of 3C48, 3C196, and 3C286 with Stellar Objects," *Astrophys. J.* **138,** 30 (1963).

4. M. RYLE and A. R. SANDAGE, "The Optical Identification of Three New Radio Interferometer Observations-Objects of the 3C48 Class," Astrophys. J. **139,** 419 (1964).

5. L. R. ALLEN, R. H. BROWN, and H. P. PALMER, "Analysis of the Angular Sizes of Radio Sources," Monthly Notices of the Royal Astronomical Society **125,** 57 (1962).

Lunar Occultation Observations

6. C. HAZARD, M. B. MACKEY, and A. J. SHIMMINS, "Investigation of the Radio Source 3C273 by the Method of Lunar Occultation," Nature **197,** 1037 (1963).

7. C. HAZARD, "The Method of Lunar Occultations and Its Application to a Survey of the Radio Source 3C212," Monthly Notices of the Royal Astronomical Society **124,** 27 (1962).

Light Variation of 3C273

8. H. J. SMITH and D. HOFFLEIT, "Light Variations in the Superluminous Radio Galaxy 3C273," Nature **197,** 650 (1963).

9. A. R. SANDAGE "Intensity Variations of 3C48, 3C196, and 3C273 in Optical Wavelengths," Astrophys. J. **139,** 416 (1964).

Red Shift Interpretations

10. M. SCHMIDT, "A Star-like Object with Large Red Shift," Nature **197,** 1040 (1963).

11. J. B. OKE, "Absolute Energy Distribution in the Optical Spectrum of 3C273," Nature **197,** 1040 (1963). (Hα line was found in the infrared spectrum.)

12. J. L. GREENSTEIN and T. A. MATTHEWS, "Red-Shift of the Unusual Radio Source 3C48," Nature **197,** 1041 (1963).

13. J. L. GREENSTEIN, "Quasi-Stellar Radio Sources," Sci. Amer. **209,** 54 (1963).

Neutron Stars and Collapse

14. J. R. OPPENHEIMER and G. M. VOLKOFF, "On Massive Neutron Cores," Physical Review **55,** 374 (1939).

15. J. R. OPPENHEIMER and H. SNYDER, "On Continued Gravitational Contraction," Physical Review **56,** 455 (1939).

Neutrino and Neutron Stars

16. H.-Y. Chiu, "Supernova, Neutron Stars and Neutrinos," Annals of Physics, **26,** 3, 364, (1964).

SUPERNOVAE, NEUTRINOS, AND NEUTRON STARS*

HONG-YEE CHIU†

The final evolution of a pre-supernova star is reviewed. It is now believed that the supernova explosion of a star is first triggered by the collapse of its core, followed by the sudden release of gravitational energy or the sudden ignition of nuclear fuel in the envelope. Among the subjects discussed herein are the causes for stellar collapse, the static structures of white dwarfs and neutron stars, the dynamic collapse of oversized neutron stars, angular momentum and stellar collapse, and the observable features of neutron stars.

1. INTRODUCTION‡

It is now accepted with reasonable assurance that the supernova explosion of a star is first triggered by the collapse of its core,** followed by the sudden release of gravitational energy or the sudden ignition of nuclear fuel in the envelope (References 2–6). A number of suggestions have been made as to the physical processes which may cause the core to collapse suddenly. Among the most promising mechanisms are the disintegration of iron (the then most abundant element) into helium at a temperature of 7×10^{9}°K (Reference 3), and the neutrino process which dissipates stellar energy at a drastic rate of 10^{23} ergs/cm^3-sec at the same temperature (this rate is to be compared with the stellar energy density which is 10^{24} ergs/cm^3 at the same temperature and a density of 10^{6} gm/cm^3) (Reference 5).

As we shall see later, although these two physical processes give rise to different kinds of instabilities, nevertheless they are capable individually of triggering the sudden release of either gravitational or nuclear energy of a star. A detailed computation is underway to determine the range of masses for which either of these two processes comes into play first.***

A sudden release of nuclear or gravitational energy in the envelope will cause it to expand quickly (References 3 and 4). The expanding envelope is directly observable. Owing to its complexity and its dynamical nature, only qualitative conclusions have been presented. Even so, interesting results regarding the origin of cosmic rays (Reference 7) and the origin of elements have been obtained, and qualitative agreements with observational results are quite satisfactory (References 3 and 6).

What is left as the core collapses is not immediately clear. In the past, it was usually assumed that ordinary white dwarfs (whose mean density is around 10^{6} gm/cm^3) are remnants of supernova explosions. The large number of white dwarfs discovered (around 10 percent of the stars in the solar neighborhood are white dwarfs), however, cannot be explained by the relative scarcity of supernovae per galaxy (which occur at a rate of around 1/50—300 years) (References 8 and 9). The other alternative that neutron stars may be the remnants of supernovae has so far been accepted only with skepticism§ (References 10 and

*Published as *NASA Technical Note* D-2112, June 1964. This work was also published in *Annals of Physics*, 26(3):364–410, February 1964.

†Goddard Space Flight Center and Columbia University. Mr. Chiu is a NRC-NAS Senior Research Associate and also a Research Fellow, Academia Sinica, Republic of China.

‡A more complete description of stellar evolution along the present trend of thought may be found in Reference 1.

**The concept of stellar collapse was first introduced by G. Gamow and M. Schonberg (Reference 2).

***H. Y. Chiu and E. E. Salpeter, work in progress.

§However, W. Baade and F. Zwicky believed that supernova explosions are resulted from the formation of a neutron star (Reference 12). This belief is no longer supported by other evidences; see Reference 3.

11). Moreover, there is *no astronomical evidence* yet that such stars even exist.

Theoretically there are also some difficulties. At least one type of pre-supernova stars is believed to have a mass of around $30M_\odot$ ($M_\odot$ = solar mass = 2×10^{33} gm) of which around $20M_\odot$ belongs to the core? Since there exists an upper limit ($\approx0.8M_\odot$) for the mass of a neutron star (and also white dwarfs) beyond which no stable configuration exists, what will be the fate of the massive stellar core? It is unlikely that a large fraction of the $20M^\odot$ which the core possesses will be blown away by any reasonable means. Since the composition of the core is believed to be mainly iron—the end product of thermonuclear evolution, no nuclear energy is available and gravitational energy is incapable of doing so (Section 3.1).

It was also suggested that the collapse would initiate, in addition to an outward shock which ejects matter, also an inward shock wave which compresses the core quickly (Reference 4). When the density reaches 10^{14} gm/cm^3, the inter-nucleon distance will be of the order of the size of the repulsive core ($\approx10^{-14}$cm), and the shock wave will be rebounded by the hard nuclear core. The rebounded shock wave will expand the star to the normal density possessed by a white dwarf, and will eject more mass from the star. This hope, however, was relinquished recently because of neutrino processes. Neutrino processes involving electron neutrinos and μ-neutrinos are extremely efficient in dissipating energy—the relaxation time being around 10^{-5} sec at $T=10^{12}$°K. This would be the temperature of a medium if it is compressed adiabatically to a density of 10^{13} gm/cm^3. This relaxation time (10^{-5} sec) is extremely short as compared with the time the compression takes place, which is calculated to be around 10^{-2} sec. Hence, once compressed to such high densities ($\gtrsim10^{13}$ gm/cm^3), the star will lose all its energy; and it cannot restore itself to a normal density. This is discussed in more detail in Section 3.4.

So far, no stars in any way resembling neutron stars have been discovered. We would like to point out that this is not evidence against the existence of neutron stars. Our atmosphere strongly absorbs electromagnetic radiation whose wavelengths are less than 3000A. The total energy output by a neutron star is proportional to the fourth power of its effective surface temperature T_e. The part of neutron star radiation that will pass through our atmosphere, however, increases only as T, when $T\gg10^4$°K (Section 8.3). For a surface temperature of 10^6°K, the total energy output of a neutron star is around that of the sun, but the part of radiation that can pass through our atmosphere is about 10^{-5} of that of the sun at the same distance. This is too faint to be seen at even a reasonable distance (say, 10 light years).

We shall discuss the nature of various problems related to the formation and detection of neutron stars. Although in the past, several pieces of excellent and wonderful work have been done,† the problem in the large is still unsolved. In this paper we do not attempt to solve this problem. We should like, however, to emphasize the difficulty of the problem and the possible direction of a solution.

Since rotational energy may be quickly dissipated by gravitational radiation when the collapse takes place, we shall ignore rotational effects in our discussions. This is discussed in particular detail in Section 7.

2. CAUSES FOR COLLAPSE OF A STAR

A star is, in the large, a mechanical system in quasi-hydrostatic equilibrium. It has a negative total energy and, in order to be in a stable equilibrium, its total energy must also be a minimum on its energy surface with respect to small perturbations. Moreover, the energetics of a stable star must be such that hydrostatic equilibrium has a useful meaning. When either of these two conditions is not fulfilled, the star will be unstable and, since its total energy is negative to start with, collapse will take place.

When the first condition is not fulfilled, the star is said to have *hydrostatic instability*. When the second condition is not fulfilled, the star is said to have *dynamical instability*. In the following we study them in more detail.

2.1. Hydrostatic Instability

Here we are concerned only with the stability of a star in hydrostatic equilibium under small

†These works are summarised in Section 5.6. References to these works may also be found there.

adiabatic perturbations (Reference 13).* Let E_T be the total energy of a star. E_T is given by

$$E_T = -G\int_0^M \frac{m\,dm}{r} + \int_0^M \epsilon\,dm \quad , \qquad (2.1)$$

where the first term represents the gravitational energy and ϵ is the thermodynamic energy per unit mass, G is the gravitational constant, m is the mass of the star within a sphere of symmetry of radius r, and M is its total mass. In order that the star be in equilibrium, it is necessary that the first variation of E_T be null:

$$\delta E_T = 0 \quad ; \qquad (2.2)$$

and, in order that the equilibrium be a stable one, the sign of the second variation must be positive —that is,

$$\delta^2 E_T > 0 \quad . \qquad (2.3)$$

One can demonstrate that Equation 2.2 gives rise to the equation of hydrostatic equilibrium and Equation 2.3 gives rise to the following inequality:†

$$\int\left(\Gamma - \frac{4}{3}\right) P\,dV > 0 \quad , \qquad (2.4)$$

where P is the hydrostatic pressure and V is the volume of the star; Γ is the adiabatic exponent and is defined as

$$\Gamma = -\frac{v}{P}\left(\frac{dP}{dv}\right)_{ad} \qquad (2.5)$$

The subscript means that this differentiation is carried out under an adiabatic process.

One form of Γ is‡ (see footnote, page 3)

$$\Gamma = \frac{Pv + v\left(\frac{\partial\epsilon}{\partial v}\right)_P}{P\left(\frac{\partial\epsilon}{\partial P}\right)_v} \qquad (2.6)$$

where v is the specific volume (volume per unit mass) and P is the gas pressure. Assuming that ϵ and P_v are all functions of temperature alone, and approximating Δv by v, one obtains the following form for Γ:

$$\Gamma = 1 + \frac{\Delta(Pv)}{\Delta\epsilon} \quad . \qquad (2.7)$$

Normally, for a non-relativistic gas, $Pv = 2/3\epsilon$; hence $\Gamma = 5/3$. For a relativistic gas, $Pv = 1/3\epsilon$; hence $\Gamma = 4/3$.** If, in a particular temperature regime phase transition takes place, the energy of transition has to be included in ϵ., Thus when phase transition occurs (such as ionization dissociation or nuclear disintegration) the value of Γ will drop to below 4/3. For iron-helium transition, the heat of transition ϵ_t is roughly 2.2 Mev/nucleon. The transition occurs at a temperature of 8×10^9°K, at which the thermal energy is around 1 Mev/particle. Since the electrons are quite relativistic, $Pv \approx 1/3\epsilon_e$, where ϵ_e is the thermal energy for the electrons. Using the value $\epsilon_e \approx 1$ Mev, and writing $\epsilon \approx \epsilon_e + \epsilon_t$ (thermal energy of nuclei and radiation energy may be neglected), we find that

$$\Gamma = 1 + \frac{\frac{1}{3}\epsilon_e}{\epsilon_e + \epsilon_t} \approx 1.1 \quad . \qquad (2.8)$$

We remark that the dissociation of iron into helium is not the only mechanism that will cause Γ to drop to below 4/3. Figure 1 shows Γ as a function of temperature for a number of densities for a gas in equilibrium with electron pair creation (Reference 14).*** One finds that the value of Γ may drop to below 4/3 if the density is low enough. For the density-temperature region one usually considers inside a star, the creation of electron-positron pairs will not cause a star to collapse. Only for extremely massive objects, which Flower and Hoyle considered as possible radio sources, will the creation of electron-positron pairs be responsible for the collapse of a star (Reference 15).

The disintegration of iron into helium, however, occurs over an extensive (though narrow) tem-

*Also an unpublished paper by F. J. Dyson: "Hydrostatic Instability of a Star."

†F. J. Dyson has presented a concise derivation of Equation 2.4 (see footnote, page 3), which is briefly described in Reference 5.

‡Derivation of Equation 2.6 may be found in Reference 5.

**Although Equation 2.7 is an approximate expression for Γ, in these two limiting cases it gives the correct value for Γ.

***Also unpublished material by H. Y. Chiu and S. Tsuruta: "Thermodynamic Properties of Hot Matter."

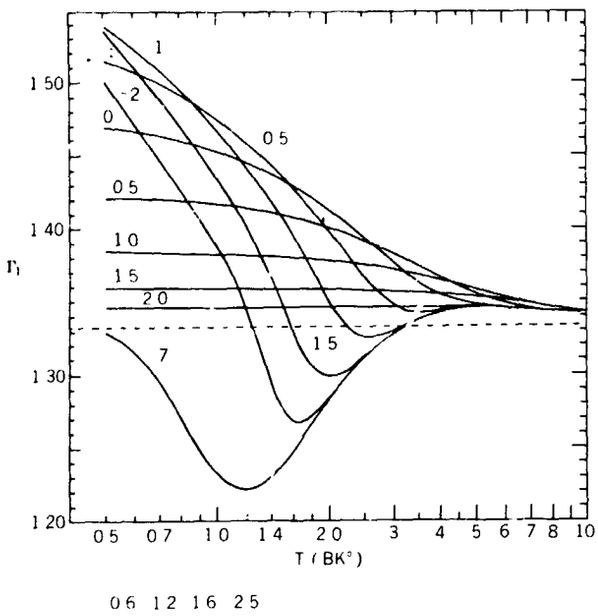


FIGURE 1.—The adiabatic exponent $\Gamma \equiv -\frac{v}{P}\left(\frac{dP}{dv}\right)_{ad}$ for an electron gas in equilibrium with radiation and pair creation is plotted as a function of temperature for a number of densities. The number on the curves refer to $\log_{10} p/p_0$, where $p_0 = 2\times10^6$ gm/cm³.

perature-density region as indicated in Figure 2. Hence, every star which evolves with increasing temperature will come across such a region. In these regions instabilities may occur. Although the sign of the integral in Equation 2.4 has not been proven to be generally negative for a star in

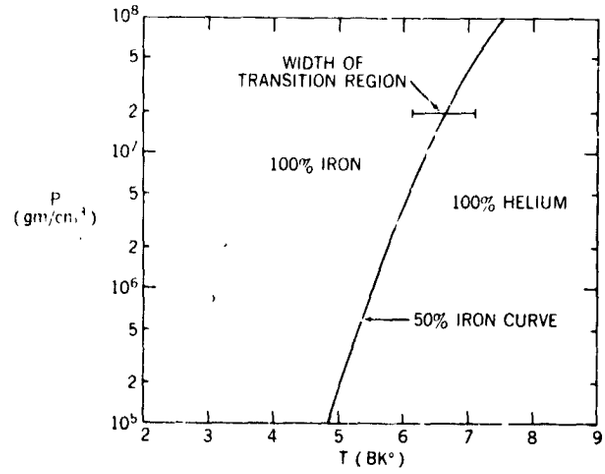


FIGURE 2.—The temperature-density relation, for which the equilibrium ratio Fe^{56} to He^4 is unity, is plotted. The transition region has a width of around $\Delta T \approx 10^9$ °K.

this density-temperature region, in certain simple cases (References 4 and 5) it has negative values. A more extensive work to study this kind of instability is underway.†

2.2. Dynamical Instability of a Star

While a star is stable in a hydrostatic sense, it may still be unstable dynamically, and vice versa. The dynamical stability of a star is related to the energy flow inside a star.

From the condition for hydrostatic equilibrium ($\delta E_T = 0$) one may obtain the equation for hydrostatic equilibrium:

$$\frac{dP}{dr} = -\rho\frac{Gm}{r^2}, \qquad (2.9)$$

where ρ is the density and m was defined previously and is given by

$$m = \int_0^r 4\pi r^2 \rho\, dr \ . \qquad (2.10)$$

A more detailed discussion of Equations 2.9 and 2.10 may be found in standard textbooks (Reference 16). Equation 2.9 is valid only when all matter inside a star does not have macrosopic motion. When a star is contracting rapidly, one must include the acceleration $\dot{v}$ (where v is the velocity and $\dot{v}$ its hydrodynamic derivative). In the absence of relativistic effect, and assuming spherical symmetry, both v and $\dot{v}$ are in the direction of r (notation: v_r). Hence, the equation for hydrostatic equilibrium becomes

$$\frac{dP}{dr} = -\rho\frac{Gm}{r^2} - \rho\dot{v}_r \ ; \qquad (2.11)$$

$\dot{v}_r$ is given by

$$\dot{v}_r = \frac{\partial v_r}{\partial t} + v_r\frac{\partial v_r}{\partial r} \ . \qquad (2.12)$$

It is more convenient to use m as one of the two independent variables since m is independent of t. Then Equation 2.11 becomes (we now drop the subscript r on v and $\dot{v}$).

$$\frac{dP}{dm} = -\frac{G}{4\pi}\frac{m}{r^4} - \frac{\dot{v}}{4\pi r^2} \ . \qquad (2.13)$$

†See footnote *** on, page 1.

Differentiating Equation 2.10, one obtains a relation between r and m:

$$\frac{dr}{dm} = \frac{1}{4\pi r^2 \rho} . \qquad (2.14)$$

Further, we also have the equation of continuity:

$$\dot{\rho} + \rho \operatorname{div} v = \dot{\rho} + \rho \frac{\partial v}{\partial r} = 0 . \qquad (2.15)$$

The contraction of the star is caused by some fast energy dissipation process (like the neutrino process) which depends only on the density ρ and the temperature T. Let the rate of energy dissipation per unit mass be $\epsilon_d(\rho,T)$. The energy balance equation is

$$-T \frac{dS}{dt} = \epsilon_d(\rho,T) , \qquad (2.16)$$

where S is the entropy per unit mass.* Further, thermodynamics supplies us with the following relations:

$$\left.\begin{aligned} S &= S(\rho,T) , \\ P &= P(\rho,T) . \end{aligned}\right\} \qquad (2.17)$$

Hence, the four equations, Equations 2.13 through 2.16, with their four unknown variables (ρ,T,r,v), can now be solved in terms of m and t (which are the independent variables). The boundary conditions are:

$$\left.\begin{aligned} &\text{At } m=0: \\ &\quad r=v=\dot{v}=0, \ \rho \text{ and T are finite,} \\ &\quad \text{and } \frac{dP}{dm} \propto m^{-1/3}; \\ &\text{At some value of r: v, } \dot{v} \text{ are finite;} \\ &\quad \text{and } \rho \text{ and T tend to zero} \\ &\quad \text{simultaneously;} \\ &\text{At } t=0: \\ &\quad \rho=\rho_0(m), \ T=T_0(m), \ v=v(m), \\ &\quad \dot{v}=\dot{v}(m). \end{aligned}\right\} \qquad (2.18)$$

These boundary conditions are physical restrictions on the type of solutions one expects from the set of nonlinear differential equations, Equations 2.13 through 2.16. Their meaning is almost self-evident. Equations 2.13 through 2.16 with the boundary condition Equation 2.18 describe the non-relativistic collapse of a star. If v approaches a certain limit which will be discussed below, it may happen that no solutions could be found which are continuous in both m and t and satisfy the boundary condition (Equation 2.18). In such cases a shock wave will be developed. The behavior of some simple shock wave solutions has been studied by Colgate el al. (Reference 4).

We now study the sign of $\dot{v}$ and the physical limit on v. Let W be the gravitational energy of the star and E its total thermodynamic energy; then the total energy of the star E_T is given by

$$E_T = W + E . \qquad (2.19)$$

The virial theorem gives†

$$-W = (1+\eta) E , \qquad (2.20)$$

where ϵ is a slowly varying constant depending on the gross thermodynamic properties of a star. For all cases, $1>\eta>0$. At $T \approx 6\times10^9$ °K, $\eta \approx 0.1$.†

Approximately one can write

$$W = -G \int \frac{m\,dm}{r} \approx -\frac{GM^2}{R} , \qquad (2.21)$$

where R is the radius of the star, and also

$$E = \int \epsilon \, dm \approx M \cdot R_g T , \qquad (2.22)$$

where ϵ, the thermodynamic energy per unit mass, is approximated by the expression for a per gas:

$$\epsilon \approx R_g T \qquad (2.23)$$

and R_g is the gas constant ($R_g = 8.2\times10^7$ in cgs units). From Equations 2.21 and 2.22, the following relations are obtained:

$$T = \frac{GM}{R_g(1+\eta)} \frac{1}{R} , \qquad (2.24)$$

$$W = -\eta E_T = -\eta M R_g T = -\frac{\eta G}{1+\eta} \frac{M^2}{R} . \qquad (2.25)$$

*This equation is derived on the assumption that the rate of energy transfer by photon is very small compared with the rate of energy dissipation by neutrinos. For the justification of this assumption, see Reference 5.

†See Appendix A.

Since gravitational contraction is the only energy source available at $T \gtrsim 4\times10^9$ °K, above this temperature the energy balance equation becomes

$$-\frac{dW}{dt} = \int \epsilon_d \, dm \approx M\epsilon_d(\rho, T) \, . \quad (2.26)$$

The energy dissipation function ϵ_d for the annihilation process (which is the most important one in this temperature density region) is

$$\epsilon_d = \epsilon_o \frac{T^9}{\rho} \, , \quad (2.27)$$

where $\epsilon_o = 4\times10^{-75}$ ergs/(°K)^9-sec-cm^3.

From Equation 2.24 we can also obtain a dynamical relation between Y and ρ. Since

$$\rho = \frac{4}{3}\pi R^3 \, ,$$

the relation is

$$\rho = \left[\frac{Rg^3}{4G^3M^2}\right] T^3 \equiv AT^3 \, . \quad (2.28)$$

A is a constant defined by the above equation.

From Equations 2.25 through 2.28, we obtain the following relations:

$$-\frac{dR}{dt} = \frac{1+\eta}{\eta GM} \epsilon_o \frac{1}{A}\left[\frac{GM}{R_g(1+\eta)}\right]^6 \frac{1}{R^4} = \frac{C}{R^4} \quad (2.29)$$

and

$$-\frac{d^2R}{dt^2} = -\frac{4C}{R^5}\frac{dR}{dt} = +\frac{4C^2}{R^9} \, . \quad (2.30)$$

C is a constant depending on M and other gas characteristics. We now identify v with dR/dt and $\dot{v}$ with d^2R/dt^2. They both have unchanging negative signs which indicate that the star is collapsing inward, with rapidly increasing speed. Although Equations 2.29 and 2.30 are only approximate solutions, they do indicate strongly that a stellar collapse will eventually occur.

The limit for $\dot{v}$ is set by the gravitational field of the star. Thus,

$$\dot{v} < \frac{MG}{R^2} \, . \quad (2.31)$$

Equation 2.31, together with Equation 2.30, sets a lower limit on R below which no hydrostatic equilibrium is possible:

$$R > \left[\frac{4C^2}{MG}\right]^{1/7} \equiv R_l \, . \quad (2.32)$$

A similar result may be obtained by considering the upper limit for v. The upper limit for v is set by the gravitational energy available at radius R:

$$\frac{1}{2}Mv^2 < \frac{M^2G}{R} \, . \quad (2.33)$$

Equation 2.33 is identical with Equation 2.32 apart from a numerical factor close to unity.

Let $M = 20M_\odot$, $\eta = 0.1$; we find the limiting radius R_l to be

$$R_l \approx 10^9 \text{ cm} \, ,$$

$$C \approx 6.8 \times 10^{31} \text{ (in cgs units)} \, .$$

The corresponding temperature as estimated from Equation 2.28 is roughly 3×10^9 °K. This temperature is below the iron-helium disintegration temperature ($\approx 6\times10^9$ °K.) However, we remark that the above estimate is very crude. The only conclusion we can draw is that both the *annihilation neutrino process* and the *iron-helium disintegration process* are capable of causing a massive star to collapse. To answer this question more quantitatively, one needs to find out the sign of the integral (Equation 2.4) and to solve Equations 2.13 through 2.17 with the boundary condition Equation 2.18 and appropriate input parameters.

The steep dependence of $\dot{v}$ on 1/R ($\propto$ T) as shown in Equation 2.30 indicates that the collapse may take place suddenly.

3. PHYSICAL EVENTS AND CONSEQUENCES OF STELLAR COLLAPSE

We have demonstrated in the last section that, when the gravitational contraction of a star (carried by neutrino emission process) reaches a certain stage, collapse must take place and the concept of hydrostatic equilibrium loses its meaning. We shall find out whether the collapse will

stop and whether the star can eject enough mass so as to become an ordinary white dwarf.

3.1. Impossibility of Ejecting Enough Matter To Become a White Dwarf

In order that a massive core may eject enough mass to become a white dwarf, some physical mechanism must be present to eject almost all of its mass into space.

It is difficult to imagine a practical mechanism to achieve this. However, we shall now demonstrate on energetic grounds that a star cannot eject enough matter into space so that the remaining star becomes a white dwarf. The gravitational energy of a star is GM^2/R. Since no cold star of mass $>1.4M_\odot$ may exist (see Section 4), an energy of the amount

$$\frac{G\left(M^2 - M_\odot{}^2\right)}{R}$$

is needed to dissociate the star from its gravitational binding. This energy must be supplied gravitationally by the remaining core of the star of mass around $M_\odot$, radius r. Hence, to the first approximation

$$\frac{G\left(M^2 - M_\odot{}^2\right)}{R} \approx \frac{GM_\odot{}^2}{r} \quad . \tag{3.1}$$

Consider $M \approx 20M_\odot$, and a radius of 10^9 cm (which corresponds to a temperature of around 1 Mev when instability occurs); we find $r \approx R/400$. The density of the remaining core will be $(400)^3 = 6.4 = 10^7$ higher than that of the star before collapse takes place. From Equation 2.28 the density at the instant of collapse is estimated to be roughly 10^7 gm/cm^3. Hence the remaining core will have a density of at least 10^{13} gm/cm^3, which is the density of a neutron star.

This example is cited just to demonstrate that, when a massive star collapses, there is no possibility that the remaining core will be a white dwarf.

Moreover, most of the energy released during the gravitational collapse of a star will be in the form of neutrinos and cannot be of any practical use. This will be discussed in more detail in Section 3.4.

We now examine the thermodynamic properties of dense matter and the corresponding stars.

3.2. Thermodynamic Properties of Degenerate Matter at Near Zero Temperatures

At zero temperature and reasonably low density ($\rho \lesssim 10^{12}$ gm/cm^3) the energetics of an ionized gas are dominated by the electrons. The available energy states are occupied up to a Fermi momentum p_F corresponding to a Fermi energy E_F. The relation between the electron number density n_e and Fermi momentum p_F and density ρ are as follows:*

$$x = \frac{p_F}{m_e c} \quad , \tag{3.2}$$

$$n_e = \frac{8\pi m_e{}^3 c^3}{3h^3} x^3 = 5.9 \times 10^{29} x^3/\text{cm}^3 \quad , \tag{3.3}$$

$$\rho = 10^6 \frac{A}{Z} x^3 \text{ gm/cm}^3 \quad , \tag{3.4}$$

where A and Z are average values for the mass number and the atomic number for the gas. The electron Fermi energy E_F is given by

$$E_F = m_e c^2 \left[\left(1 + x^2\right)^{1/2} - 1\right] \quad . \tag{3.5}$$

Salpeter (Reference 18) found that in the stellar regime the pressure energy density relation is very well described by that for a perfect Fermi gas All other corrections (Coulomb correction, exchange effect, ion motion, etc.) are very small. These relations for a perfect Fermi gas are (Reference 17):

$$P = \frac{\pi m^4 c^5}{3h^3} f(x) = 6.01 \times 10^{22} f(x) \text{ d/cm}^2 \quad , \tag{3.6}$$

$$E_{kin} = \frac{\pi m^4 c^5}{3h^3} g(x) \quad , \tag{3.7}$$

$$f(x) = x\left(2x^2 - 3\right)\left(x^2 + 1\right)^{1/2} + 3 \sinh^{-1} x \quad , \tag{3.8}$$

$$g(x) = 8x^3\left[\left(x^2 + 1\right)^{1/2} - 1\right] - f(x) \quad , \tag{3.9}$$

where E_{kin} is the kinetic energy density of the electrons. We have the following asymptotic expression:

$$\left.\begin{aligned} f(x) &\to \frac{8}{5} x^5 \\ g(x) &\to \frac{12}{5} x^5 \end{aligned}\right\} \quad x \to 0 \quad , \tag{3.10}$$

*For a description of the theory of a perfect Fermi gas, see Reference 17.

and

$$\left.\begin{aligned} f(x) &\to 2x^4 \\ g(x) &\to 6x^4 \end{aligned}\right\} \quad x \to \infty , \qquad (3.11)$$

Hence

$$P \to \frac{2}{3} E_{kin} \qquad x \to 0 , \qquad (3.12)$$

$$P \to \frac{1}{3} E_{kin} \qquad x \to \infty , \qquad (3.13)$$

and

$$P = \frac{1}{20}\left(\frac{3}{\pi}\right)^{2/3} \frac{h^2}{m} n_e^{5/3} \qquad x \to 0 , \qquad (3.14)$$

$$P = \frac{1}{8}\left(\frac{3}{\pi}\right)^{1/3} hcn_e^{4/3} \qquad x \to \infty . \qquad (3.15)$$

3.3. Composition of Degenerate Matter at Near Zero Temperatures

If the energy of an electron is greater than the ϵ_z, the beta-decay energy for electron emission (counting only the kinetic energy) of the nucleus (Z–1, A), then inverse beta reactions involving (Z, A) will occur:

$$e^- + (Z,A) \to (Z-1,A) + \nu. \qquad (3.16)$$

If $E_F > \epsilon_z$, then the nucleus (Z, A) ceases to be stable and undergoes inverse beta decay to the nucleus (Z–1, A). The nucleus (Z–1, A) cannot decay because all electronic states are occupied up to the Fermi energy $E_F > \epsilon_z$. If the nuclei are in equilibrium, the composition is easily calculated. By minimizing the quantity

$$b^* = \frac{B(Z,A) - ZE_F'}{A} \qquad (3.17)$$

(where E_F' is the Fermi energy of the electron gas minus the neutron-hydrogen atom rest mass energy difference), Salpeter (Reference 18) calculated the equilibrium composition for cold degenerate mater. His result is shown in Table 1.

When $\rho \gtrsim 10^7$, Γ for the electron gas is already very close to 4/3 (Figure 1). The gas sphere is only marginally stable against hydrostatic instability. The inverse beta reactions will cause Γ to be less than 4/3 for the density region $\rho = 10^7$ to $\rho = 10^{12}$ gm/cm^3. Wheeler (Reference 19) has computed the ratio of P to $\rho^{4/3}$; he finds that the ratio actually decreases with ρ in the above density region.* His result is shown in Figure 3.

Table 1

Equilibrium Composition of Cold Matter*

(Z, A)	(26,56)	(28,62)	(28,64)	(28,66)	(28,68)
E_F	0.6	2.5	3.9	6.1	7.0
$\log_{10}\rho$	7.15	8.63	9.15	9.69	9.87
(Z, A)	(30,76)	(30,78)	(30,80)	(32,90)	(38,120)
E_F	8.5	9.5	14.8	20.6	24.0 Free neutrons
$\log_{10}\rho$	10.13	10.28	10.84	11.28	11.53

*ρ is measured in gm/cm^3, E_F in Mev.

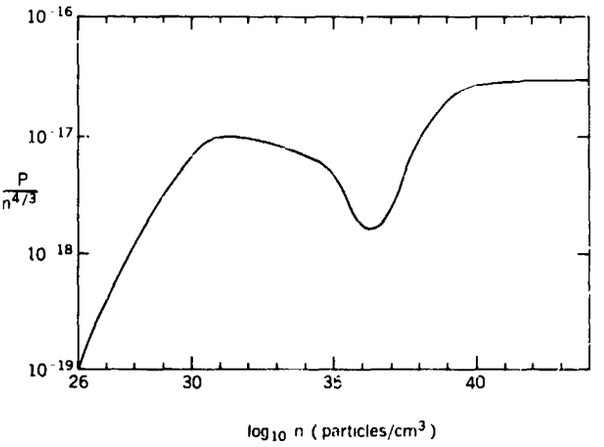


FIGURE 3.—The ratio $P/n^{4/3}$ (where n is the total number density of nucleons) is plotted as a function of n. In between $n = 10^{31}$/cm^3 and $n = 10^{38}$/cm^3 ($\rho = 10^7 \to 1.67 \times 10^{13}$gm/cm^3) $P/n^{4/3}$ decreases as n increases, indicating that the power dependence of P on n is less than 4/3; this result is due to Wheeler.

If all nuclei are not in thermodynamic equilibrium, then the transition may not take place as described by Table 1. Cameron (Reference 20) has studied the rate of nuclear reactions at high density (pycno nuclear reactions). He found that most nuclear reactions leading towards equilibrium state will take place spontaneously even at zero temperature when the density is greater than 10^9 gm/cm^3.

*At zero temperature $\Gamma = (\rho/P)(dP/d\rho)$. If P is expressed as $P = P_0 \rho\Gamma$, where P_0 is a proportionality constant, then Γ is the adiabatic exponent.

We may therefore safely expect that the composition of matter at $\rho > 10^9$ gm/cm^3 will be reasonably given by Table 1 under equilibrium conditions. Wheeler's result (Figure 3) indicates that in-between the density-temperature region $\rho = 10^8$ to $\rho = 10^{12}$ gm/cm^3 there are no equilibrium configurations for a star. Harrison, Wakano, and Wheeler (Reference 19) have also studied the structure of stars at zero temperature; the result is shown in Figure 4, in which the mass as a function to stellar density is plotted. In certain

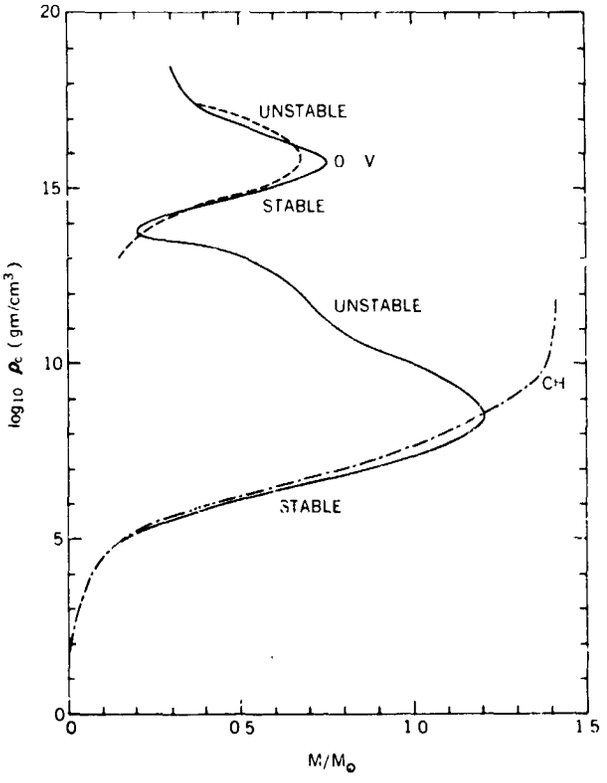


FIGURE 4.--The mass-central density ρ_c relation for zero temperature stars. Given a central density, with the equation of state described in Figure 3 and Section 3.3, the mass may be integrated by using Equations 2.9 and 2.10 or Equations 5.2 and 5.3. Curves marked CH show the nongeneral-relativistic result of Chandrasekhar (Ref. 17), and the curve marked O-V is the general relativistic result of Oppenheimer and Volkoff for an ideal Fermi gas (Ref. 27). This solid curve is due to Wheeler et al. (Refs. 19 and 25.)

density regions where inverse beta reaction occurs, the mass of the star decreases with increasing central density, indicating that such configurations are unstable.

3.4. Neutrino Production at Near Zero Stellar Temperatures

We now turn to the energetics of dense matter at non-zero temperatures. As long as $kT \ll E_F$, the thermodynamic properties of matter are not different from that for the zero temperature case. Hence we shall only consider the case where $kT \approx E_F$. It may be pointed out that, as long as Γ is close to 4/3, the relation $T \propto \rho^{1/3}$ (Equation 2.28) is expected to hold for the temperature and density of a star during its evolution. If $E_F \gg mc^2$, $E_F \propto \rho^{1/3}$. Hence the ratio kT/E_F will be roughly a constant if energy dissipation is not substantial; otherwise it will decrease.

In the following we shall assume that the ratio kT/E_F will be roughly 1. With this assumption, if we find the relaxation time for losing energy to neutrinos is short compared with the time for the star to restore to its original density, then we are back to the zero temperature case.

At $T = 6 \times 10^9$ °K ($kT \approx 0.5$ Mev) the most dominant neutrino process is the annihilation process (References 21, 22, and 23):

$$\gamma \rightleftarrows e^- + e^+ \longrightarrow \nu_e + \bar{\nu}_e \quad . \tag{3.18}$$

At higher temperatures then neutrino processes also must be considered. At low density ($\rho < 10^7$ gm/cm^3, say) the mean free path of neutrinos is longer than the physical dimension of the star. At higher density, the mean free path of neutrinos is no longer compared with the physical dimension of the star. The rate of dissipation of neutrinal energy is then less than the production rate.

At $kT \approx m_\mu c^2$, the creation of π meson pairs and μ meson pairs is not negligible. Because the neutrinos associated with the μ meson do not interact with electrons to the lowest order, the dissipation of stellar energy through μ-neutrino production is more important. We shall also include this in our discussion.

(a) Electron-Positron Pair, π and μ Pair Densities

The number density for the electron pair created, under the condition $E_F/kT \gg 1$, is roughly given by (see, for example, Reference 23)

$$n_e = \frac{8\pi m_e^3 c^3}{h^3} \int_0^\infty \frac{x^2 dx}{\exp\left[(1+x^2)^{1/2}/\theta_e\right] + 1} \tag{3.19}$$

where

$$\theta_e = \frac{kT}{m_e c^2} = \frac{T}{T_e} . \quad (3.20)$$

Numerically

$$T_e = 6 \times 10^9 °K . \quad (3.21)$$

Since $T \gg T_e$, we may use the approximation $(1+x^2)^{1/2} \approx x$. Changing variables from x to $u\mu_m/\theta_e$, we have

$$n_e = \frac{8\pi m_e^3 c^3}{h^3} \theta_e^3 \int_0^\infty \frac{u^2 du}{\exp u + 1} . \quad (3.22)$$

Since

$$\int_0^\infty \frac{u^2 du}{\exp u + 1} = \zeta(3) = 1.2202 , \quad (3.23)$$

where $\zeta(z)$ is the Riemann's zeta function, we find numerically that

$$n_e = 4.4 \times 10^{30} \theta_e^3 \quad (3.24)$$

For the case of π or μ mesons we may define

$$\theta_\mu = \frac{kT}{m_\mu c^2} = \frac{T}{T_\mu} , \quad (3.25)$$

$$\theta_\pi = \frac{kT}{m_\pi c^2} = \frac{T}{T_\pi} , \quad (3.26)$$

$$T_\mu = 1.24 \times 10^{12} °K , \quad (3.27)$$

$$T_\pi = 1.65 \times 10^{12} °K . \quad (3.28)$$

If we neglect the factor unity in Equation 3.19, we have

$$n_\pi = 5.8 \times 10^{37} \theta_\pi K_2\left(\frac{1}{\theta_\pi}\right) , \quad (3.29)$$

$$n_\mu = 1.54 \times 10^{37} \theta_\mu K_2\left(\frac{1}{\theta_\mu}\right) . \quad (3.30)$$

where $K_2(z)$ is the modified Hankel function with the following asymptotic expression:

$$K_2(z) \longrightarrow \frac{2}{z^2} \qquad z \ll 1 , \quad (3.31)$$

$$K_2(z) \longrightarrow \sqrt{\frac{\pi}{2z}}\, e^{-z} \qquad z \gg 1 . \quad (3.32)$$

Table 2 lists the number density of electron pairs and μ meson pairs for $T = 10^{10}°K$ to $10^{12}°K$.

Table 2

Electron, π Meson, and μ Meson Number Densities as a Function of Temperature

(T is measured in °K; n in particles/cm^3).

$\log_{10} T$	$\log_{10} n_e$	$\log_{10} n_\mu$	$\log_{10} n_\pi$
10	31.31	--	--
10.2	31.91	--	--
10.4	32.51	--	--
10.6	33.11	21.7	19.27
10.8	33.71	27.1	26.37
11	34.31	30.5	28.47
11.2	34.91	32.8	31.69
11.4	35.51	34.4	33.69
11.6	36.11	35.6	35.09
11.8	36.71	36.5	36.18
12	37.31	37.2	37.0
12.2	37.91	37.9	--
12.4	38.51	38.5	--
12.6	39.11	39.1	--

(b) Neutrino Cross Sections at High Energy

When the energy difference between neutron and proton can be neglected, the cross section for first order neutrono process is given by

$$\sigma \frac{v}{c} = CG^2 p_\nu^2 , \quad (3.33)$$

where G is the weak interaction constant, v is the relative velocity of the initial system of particles ($\approx c$), and p_ν is the final neutrino momentum. C is a numerical constant listed for a number of reactions in Table 3.

3.5. Relaxation Time for Cooling of a Collapsing Star

At high temperature the energy difference between proton and neutron may be neglected. If

Table 3

Neutrino Reactions

	Reaction	$c \equiv \nu/G^2P_\nu^2$
e - 1	$e^- + e^+ \rightarrow \nu_e + \bar{\nu}_e$	$1/3\pi$
e - 2	$e^- + p \rightarrow n + \nu_e$	$5.5/\pi$
e - 3	$e^+ + n \rightarrow p + \bar{\nu}_e$	$5.5/\pi$
e - 4	$e^- + \gamma \rightarrow e^- + \nu_e + \bar{\nu}_e$	α/π
μ - 1	$\mu^- + \mu^+ \rightarrow \nu_\mu + \bar{\nu}_\mu$	$1/3\pi$
μ - 2	$\mu^- + p \rightarrow n + \nu_\mu$	$5.5/\pi$
μ - 3	$\mu^+ + n \rightarrow p + \bar{\nu}_\mu$	$5.5/\pi$
μ - 4	$\pi \rightarrow \mu + \nu_\mu$	Lifetime $= 2.2 \times 10^{-8}$ sec
μ - 5	$\mu \rightarrow e + \nu_\mu + \nu_e$	Lifetime $= 2.2 \times 10^{-6}$ sec

the relation between the temperature T and the density ρ is given by*

$$\frac{T}{10^{12}\,^\circ K} = \left(\frac{\rho}{10^{13}\ \mathrm{gm/cm^3}}\right)^{1/3}. \qquad (3.34)$$

the neutrino yield from electron-neutrino processes will be roughly the same, except the photo-neutrino process e–4 which is down by a factor of $\alpha/\pi \approx 1/400$. The μ-neutrino processes show a wider variation. All μ capture or annihilation processes have roughly the same cross section as the electron process, as is expected from the theory. However, π and μ mesons can decay into electrons and neutrinos. The rate of decay, to the extent that degeneracy is neglected, is independent of the energy (apart from a time dilatation factor). Up to a temperature of $\approx 10^{12.2}$K, μ-neutrino emission from π-decay is the most important process.

As the density of the star increases, the mean free path of neutrinos becomes smaller. Once the mean free path of the neutrinos becomes smaller than the dimension of the star, the neutrino energy loss rate is no longer the same as the production rate. Let the radius of the star be R, the mean free path for neutrinos be λ, the total stellar neutrino energy content U_ν; then the neutrino luminosity L_ν for a star is

$$L_\nu = \frac{U_\nu}{\frac{R}{c}\cdot\frac{R}{\lambda}}. \qquad (3.35)$$

Equation 3.35 has the following interpretation: When $\lambda = R$, the neutrinos leave the star with the velocity of light and the time constant for doing this is just R/c, the time of transit across the star with a velocity of light. When $\lambda < R$, neutrinos on the average scatter R/λ times before they leak out of the star. Its path length is approximately increased by a factor of R/λ. Hence Equation 3.35 describes the rate of decay of stellar neutrinal energy. The neutrinal energy content of the star may be estimated in the following way: The average energy of each neutrino is around kT. If all neutrinal states are filled up, further neutrino emission processes are inhibited. The total energy content allowable to neutrinos is of the same order as the energy content of the electron pairs, which is of the same order of magnitude as the total energy content.

A recent experiment demonstrated that the neutrino associated with electrons (e-neutrino) and the neutrino associated with μ mesons (μ-neutrino) have different quantum numbers (Reference 24). That is, a μ-neutrino cannot

*Equation 3.34 is consistent with the assumption that $E_F/\tau k \approx 1$.

induce an inverse beta reaction:

$$\left.\begin{aligned} &\bar{\nu}_\mu + p \not\rightarrow n + e^+ , \\ &\nu_\mu + n \not\rightarrow p + e^- , \end{aligned}\right\} \qquad (3.36)$$

nor can an e-neutrino induce inverse μ capture reactions:

$$\left.\begin{aligned} &\bar{\nu}_e + p \not\rightarrow n + \mu^+ , \\ &\nu_e + n \not\rightarrow p + \mu^- . \end{aligned}\right\} \qquad (3.37)$$

Since the average energy of μ-neutrinos coming from π-decay and μ-decay is of the order of mc_μ^2 (or less), inverse μ capture reaction cannot occur. Until $kT \approx m_\mu c^2$ ($T \approx 10^{12}$°K), the number density for μ mesons is small. At the temperature ($T = 1.6\times10^{11}$°K) when the loss of energy due to π decay becomes very important, the mean free path for ν_μ–μ scattering is 10^8 cm, which is around the radius of the star ($\approx 7\times10^7$ cm). Hence μ-neutrinos can dissipate stellar energy rapidly while e-neutrinos cannot. Table 4 summarizes our results. The relaxation time for dissipating stellar energy is shortest at roughly $T = 6.3\times10^{10}$°K, and again at $T \approx 4\times10^{11}$°K. The corresponding relaxation times are roughly 3×10^{-3} and 10^{-4} seconds, respectively. This is roughly the time scale for the star to relax to the zero temperature configuration discussed previously.

Since this time scale is of the same order of magnitude for light to travel across the star, we conclude the core collapses under zero temperature conditions ($E_F/kT \gg 1$). Colgate's proposal (Reference 4) that a shock wave will first compress the star into neutron star density, that it will be rebounded to restore the core to normal density, is somewhat unrealistic.

We have not included in our calculations the red shift of neutrinos due to the gravitational field of the star. The red shift caused by the gravitational field of the star will at most decrease the net loss energy rate by a factor of 2 (Section 8.2, and also Reference 48). If the red shift becomes much larger, then the star approaches a singular position which we shall soon discuss.

4. STATIC STRUCTURE OF WHITE DWARFS AT ZERO TEMPERATURE

Using Equations 3.6 and 3.7 as equations of state, Chandrasekhar (Reference 17) integrated Equations 2.9 and 2.10 to obtain the static structure of white dwarfs. His result is shown in Figure 4 together with Wheeler's, and Oppenheimer and Volkoff's, results. The most interesting feature is that the central density (and the central pressure) become infinite when the mass is still finite. The limiting mass he obtained is $5.75M_\odot/\mu_e^2$ where $\mu_e = \mathbb{A}/\mathbb{Z}$. Taking the most popular value of $\mu_e = 2$, the limiting mass is around $1.44M_\odot$.

Salpeter* worked out an intuitive derivation of the limiting mass, which is reproduced below. The virial theorem tells us that (Equation 2.20)

$$E_G \tilde{=} E .$$

Let E_e be the kinetic energy per electron. Then

$$\left.\begin{aligned} E_G &\approx \frac{GM^2}{R} , \\ E &\approx NE_e \end{aligned}\right\} \qquad (4.1)$$

and

$$M = N\mu_e m_p . \qquad (4.2)$$

Hence

$$E_e = \frac{GM_p^2 \mu_e^2 N}{R} . \qquad (4.3)$$

Let r_e be the average value of the inter-electron spacing and p_e be the average momentum; then the uncertainty principle gives

$$r_e p_e \geq \hbar . \qquad (4.4)$$

Considering only the equal sign case, we find, from the definition of E_e and Equation 4.4, that

$$\begin{aligned} E_e &= \frac{p_e^2}{2m_e} = \frac{m_e c^2}{2}\left(\frac{r_o}{r_e}\right)^2 \qquad p_e \ll m_e c , \\ E_e &= p_e c = m_e c^2\left(\frac{r_o}{r_e}\right) \qquad p_e \gg m_e c , \end{aligned} \qquad (4.5)$$

*Private communication.

Table 4

Neutrino Loss Rates in a Neutron Star.

Temperature (°K)	Density (gm/cm³)	Energy Density (ergs/cm³)	Neutrino Production Rate (ergs/sec-cm³)		Mean Free Path (cm)		Radius of Star (cm)	Relaxation Time (sec)	
			ν_e	ν_μ	ν_e	ν_μ		ν_e	ν_μ
10^{10}	10^7	1.6×10^{26}	10^{26}	--	10^{10}	--	10^9	1	--
	4×10^7	10^{27}	3×10^{27}	--	1.6×10^9	--	6.3×10^8	3×10^{-1}	--
	1.6×10^8	6.3×10^{27}	2×10^{29}	4×10^{12}	2.5×10^8	> Star	4×10^8	3×10^{-2}	--
	6.3×10^8	4×10^{28}	1.3×10^{31}	7×10^{21}	4×10^7	> Star	2.5×10^8	3×10^{-3}	--
	2.5×10^9	2.5×10^{29}	10^{33}	9×10^{28}	6.3×10^6	> Star	1.6×10^8	0.1	3
10^{11}	10^{10}	1.6×10^{30}	Equilibrium	10^{31}	10^6	> Star	10^8	0.3	0.1
	4×10^{10}	10^{31}	Equilibrium	2×10^{34}	1.6×10^5	> Star	6.3×10^7	0.5	5×10^{-4}
	1.6×10^{11}	6.3×10^{31}	Equilibrium	2×10^{36}	2.5×10^4	> Star	4×10^7	1	3×10^{-5}
	6.3×10^{11}	4×10^{32}	Equilibrium	5×10^{37}	4×10^3	> Star	2.5×10^7	--	10^{-5}
	2.5×10^{12}	2.5×10^{33}	Equilibrium	Equilibrium	6×10^2	≈ Star	1.6×10^7	--	10^{-4}
10^{12}	10^{13}	1.6×10^{34}	Equilibrium	Equilibrium	--	≈ Star	10^7	--	10^{-3}

where $r_o = \hbar/m_e c$ is the Compton wavelength for an electron. Combining these two equations, letting $s = r_e/r_o$, one can write

$$E_e = mc^2 \left[\frac{1}{s + 2s^2}\right] . \qquad (4.6)$$

Wheeler (Reference 25) has checked the accuracy of Equation 4.6 against the more complete formula Equations 3.6 and 3.7; he finds the maximum deviation is around 8 percent. Since $R \approx N^{1/3} r_e$, in substituting Equation 4.6 into Equation 4.3, one obtains the following expression:

$$\frac{1}{1 + 2s} = \left[\frac{G m_p^2 \mu_e^2}{r_o mc^2}\right] \cdot N^{2/3} \equiv \mu_e^2 \left(\frac{N}{N_o}\right)^{2/3} , \qquad (4.7)$$

where

$$N_o = \left[\frac{G m_p^2}{\hbar c}\right]^{-3/2} \approx 2N_\odot ; \qquad (4.8)$$

N_o is a characteristic number and $N_\odot$ is the number of particles of the sun. When $N < N_o$, it is possible to find a value of s such that Equation 4.7 is satisfied. When $\mu_e^3 N > N_o$, it is not possible to satisfy Equation 4.7 with a positive value of s Hence, there exists a limiting value of mass beyond which no static equilibrium structure is possible.

Long before the density ρ becomes infinite, inverse beta reactions described in Section 3.3 will occur and Chandrasekhar's description will *no longer* be valid. However, this occurs at a mass very close to his limiting mass; this is shown in Figure 4.

We believe the structure of white dwarfs is well understood. Although the density and the pressure may become infinite at a finite mass, this divergence never occurs in nature as demonstrated by Wheeler (Figure 4). This is, however, not true in the case of neutron stars.

5. THE STATIC STRUCTURE OF NEUTRON STARS

5.1. The Equation of Stellar Structure

Since the numerical value for GM/Rc^2 (the ratio of gravitational energy to rest energy) for a neutron star is close to unity, it is necessary to consider the structure from a general relativistic point of view. Because of neutrino energy dissipation (Section 3.4), the temperature of the

neutron core may always be taken to be zero. The pressure P may be taken to be a function of material density only (P and ρ are all measured in proper coordinates). Further, we consider only the static structure of neutron stars with complete spherical symmetry and no rotation. Under these circumstances, the stress energy tensor becomes

$$T_{\mu\nu} = \begin{pmatrix} -P & 0 & 0 & 0 \\ 0 & -P & 0 & 0 \\ 0 & 0 & -P & 0 \\ 0 & 0 & 0 & \rho c^2 \end{pmatrix} . \qquad (5.1)$$

The equation of stellar structure is a solution of Einstein field equations with a time static and spherically symmetric metric (References 26 and 27). The equation is:

$$\frac{dP}{dr} = \frac{\left(\frac{P}{c^2} + \rho\right) G\left(4\pi r^3 \frac{P}{c^2} + m^*\right)}{r\left(r - \frac{2Gm^*}{c^2}\right)} \qquad (5.2)$$

and

$$\frac{dm^*}{dr} = 4\pi r^2 \rho \, . \qquad (5.3)$$

Neglecting $1/c^2$ terms, Equation 5.2 reduces to that for a not-relativistic gas sphere. At the radius of the star, m^* is actually the gravitational mass observed by a distant observer (by the gravitational field the star produces). In the following we denote the proper mass (the mass before it was assembled into a star) by m, the mass observed by distant observers by an asterisk*.

The analytical properties of Equations 5.2 and 5.3 have been discussed by Oppenheimer and Volkoff (Reference 27). They found that the boundary conditions for an ordinary star apply (e.g., at $r=0$, $m^*=0$, and the central pressure $P_c>0$, etc.).

5.2. The Equation of State

In order that a unique solution for Equations 4.2 and 4.3 be possible for a given set of boundary conditions, P must be a known function of ρ. We shall discuss the equation of state $P(\rho)$ in greater detail. Because the interparticle spacing is of the order of 10^{-13} cm or less, which is roughly the size of elementary particles, it is not possible to exclude the structure of elementary particles from our discussion as in the case of the white dwarfs.

Before we proceed to the detailed structure of the equation of state, we would like to examine the limitations on the equation of state.

5.3. The Unimportance of Quantum-Gravitational Effects

Since the gravitational field inside a neutron star is extremely strong, one might wonder if quantum effects of the gravitational field may interfere with normal interactions among particles. However, quantum effects of gravitational field will, in general, not be important, as we shall show below.

A well-known theorem in Riemannian geometry* states that, in a given space with a non-Euclidean metric, it is always possible to find a coordinate transformation such that locally at any given point the metric $g_{\mu\nu}$ may be reduced to that for a Minkowskian space; that is,

$$g'_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & +1 \end{pmatrix} . \qquad (5.4)$$

Unless the $g_{\mu\nu}$'s change considerably over the dimension of elementary particles ($\approx \hbar/m_\pi c \approx 10^{-13}$ cm), it is not necessary to consider quantum effects of gravitational fields. Even in the extreme case when the mass of a neutron star is close to its critical mass, when the time metric g_{00} vanishes at the center, quantum effects may still not enter. According to Oppenheimer and

*This is easily proven. The line element ds may be written as:

$$ds^2 = \sum_{i,j} g_{ij} dx^i dx^j \, ,$$

where $\{g_{ij}\}$ is the metric tensor and $\{x^i\}$ is a general set of coordinates. Since $\{g_{ij}\}$ is a symmetric tensor, at a particular point $\{g_{ij}\}$ can be diagonalized by "rotating" the $\{x^i\}$ coordinate system at this point. An additional scale change in the rotated $\{x^i\}$ system reduces the line element to

$$ds^2 = \sum_{i,j} \pm \delta_{ij} dx^i dx^j \, ,$$

where δ_{ij} is the Kronecker delta and the sign ($\pm$) is chosen as to be consistent with the signature of Equation 13.1. Since g_{ij} varies with position, obviously this transformation is valid only at the particular point chosen.

Volkoff one can write generally, in the center of a neutron star:

$$g_{oo} = \left(1 - \frac{2Gm^*}{Rc}\right)^{1/2} \exp\left[-\frac{1}{c^2}\int_o^{P_c} \frac{dP}{\frac{P}{c^2} + \rho}\right]; \quad (5.5)$$

when the mass of a neutron star approaches the critical mass, $P_c \to \infty$, the integral in the exponential diverges logarithmically as one approaches the center, and g_{00} vanishes as $1/P$ ($P \to \infty$). Unless $P \to \infty$ abruptly over a dimension of an elementary particle, g_{00} cannot change substantially. This would not occur since the pressure is itself a macroscopic concept. Hence the center of neutron stars is not likely to be the dwelling place of quantum gravitational field theory.

It might be questioned that, since the trajectories of particles with spin are not geodesics (References 28 and 29), this might be an indication that quantum-gravitational effects play an important role. However, the spin-gravity interaction depends on the gradient of the gravitational field (spin-orbit type coupling) (References 28 and 29). Unless g_{00} changes substantially over the dimension of an elementary particle, this interaction may be treated classically. Thus the criteria for replacing a gravitational field by an acceleration is the same for particles with spin or without spin. As long as the pressure retains its macroscopic meaning, locally we can always replace gravitational fields, however strong, by a uniform acceleration. In the following, we therefore neglect the quantum gravitational effects.

5.4. Relativistic Limits on the Equation of State

General relativity does set up certain limits for the equation of state on the basis of positive (or negative, depending on which convention to follow) definiteness of the stress energy tensor, and that signals cannot propagate at a speed faster than light speed.

The speed of sound v_s for a given medium is given as

$$v_s = c\sqrt{\frac{dP}{d\epsilon}} < c, \quad (5.6)$$

where ϵ is the energy density. Integrating Equation 5.6 one obtains

$$P < \epsilon. \quad (5.7)$$

That is, the pressure must not exceed the energy density.

This is more relaxed limit for P.

In the case of a non-interacting ideal gas (including photon gas) the trace of the energy momentum tensor T_{ij} may be demonstrated to have a positive definite value (see, for example, Reference 30); T_{ij} is given by:

$$T_{ij} = \begin{pmatrix} -P & 0 & 0 & 0 \\ 0 & -P & 0 & 0 \\ 0 & 0 & -P & 0 \\ 0 & 0 & 0 & +\epsilon \end{pmatrix}. \quad (5.8)$$

Thus we have a more stringent condition on P:

$$P \leq \frac{1}{3}\epsilon. \quad (5.9)$$

Only in the case of a photon gas one can have $P = \epsilon/3$. This sets an upper limit for the speed of sound to be $(1/\sqrt{3})c$.

Recently, however, Zel'Dovich (Reference 31) constructed an example of a relativistically invariant theory for a classical vector field with a mass, interacting with stationary classical point charges. In this theory the energy density ϵ and the pressure P are

$$\epsilon = nM + \frac{2\pi g^2 n^2}{\mu^2}, \quad (5.10)$$

$$P = \frac{2\pi g^2 n^2}{\mu^2}, \quad (5.11)$$

(units: $\hbar = c = 1$) where M is the mass of the heavy particle, μ the mass of the field quanta, g the charge of the field quanta, and n the number density of heavy particles. In the limit of large n, $\epsilon \propto n^2$ and $\epsilon \approx P$.

One might argue that, since for a particle with mass the stress energy tensor is

$$T_{\mu\nu} = \left(\rho + \frac{1}{c^2}P\right) U^\mu U^\nu - \frac{1}{c^2}\rho\eta^{\mu\nu}, \quad (5.12)$$

(where $\eta^{\mu\nu}$ is the Minkowskian metric, U^μ the f · velocity) and that nothing can be "ilghter" t₁ a photon (for photon the rest mass is zero), $P = 1/3\epsilon$ should be the ultimate limit for all equations of state. Zel'Dovich's example indicates that such intuitive deduction is not correct.

5.5. Detailed Structure of Matter at High Density

The composition of matter at high density was considered in detail by Ambartsumyan and Saakyan (References 32 and 33). When the Fermi energy (including the rest energy) of neutrons exceeds the rest energy of strange particles, strange particles will be created in equilibrium with neutron gas. Although it is necessary to preserve the strangeness conservation law in the production process, in the decay process such a law need not be preserved and the number density for strange particles is governed by their ability to decay. Thus, we do not need to worry about the seemingly strangeness non-conserving character of reactions of the following type:

$$n + n \rightleftarrows \Lambda^\circ + \Lambda^\circ \cdot \qquad (5.13)$$

With this fact in mind, we now write down the reactions we will consider:*

$$\left.\begin{aligned} Y^\circ &\rightleftarrows n \,, \\ e^- + Y^+ &\rightleftarrows n \,, \\ Y^- &\rightleftarrows n + e^- \,, \\ e^- + P &\rightleftarrows n \,, \\ e^- &\rightleftarrows \mu^- \,, \\ e^- &\rightleftarrows \pi^- \,, \end{aligned}\right\} \qquad (5.14)$$

where Y denotes one or more type of strange particles. Let the Fermi energy of particle x be denoted as E_x, its number density by N_x. Since at zero temperature the Fermi energy is the same as the chemical potential, the equilibrium condition for Equation 5.14 is:

$$\left.\begin{aligned} E_{Y^\circ} = E_n = E_e + E_{Y+} = E_e + E_p \,, \\ E_{Y-} = E_n + E_e \,, \\ E_{e-} = E_{\mu-} = m_\pi c^2 \,. \end{aligned}\right\} \qquad (5.15)$$

The reason for setting the chemical potential of π meson to its rest energy is because it is a boson which allows infinite condensation at zero kinetic energy. Further, the charge neutrality condition gives an additional constraint:

$$\Sigma N_{Y+} - \Sigma N_{Y-} - N_e - N_\mu - N_\pi = 0 \,. \qquad (5.16)$$

No π^+, μ^+ are allowed because they will annihilate with π^-, μ^- to become photons (or neutrinos) which subsequently will escape.

From Equations 5.15 and 5.16 the following conclusions may be drawn:

(1) The number density for e and μ are fixed independent of the material density once the Fermi energy of electrons reaches $m_\pi c^2$. At extremely high density only baryons contribute to the pressure. The threshold for π meson creation occurs at a baryon density of 5.86×10^{40} particles/cm³, or a material density of 10^{17} gm/cm³.

(2) π meson is the only boson that can be present. It represents an energy reservoir without contributing to the pressure. Other bosons created (e.g., the intermediate boson) will decay into π mesons directly or indirectly.

The composition of matter becomes very complicated at high energy. Particles become extremely densely packed. At a density of 10^{17} gm/cm³ the interparticle spacing is around 4×10^{-14} cm, which is about the radius of the repulsive core for protons and neutrons. Hence at such high density the interaction among particles is by no means negligible. The interacting energy is around 10 Bev in the cm system (which corresponds to a laboratory energy of 100 Bev). Experimental study of interactions at such a high energy is somewhat beyond the present capability. Our knowledge of the equations of state for closely packed neutron matter is therefore highly uncertain.

*The present approach is due to Ambartsumyan et al. (References 32 and 33).

5.6. Models for Neutron Stars

We have seen that nuclear interactions among elementary particles at extremely high energy play an important role in the macroscopic behavior of dense neutron stars. In the following we shall demonstrate that there may exist a paradox between a very fundamental conservation law (baryon conservation law) and general relativity. From neutron star models with a hypothetical equation we may study the properties of this paradox which, if resolved, should yield us important information on extremely high energy interactions.

We are mostly interested in the singular character of the metric g_{00}. In the following we shall review work done by other physicists on this subject.

(a) Oppenheimer-Volkoff Theory

The very first relativistic model for neutron stars was constructed by Oppenheimer and Volkoff (Reference 27). They assumed an equation of state of an ideal Fermi gas. Their mass-radius relation is demonstrated in Figure 4. The equilibrium mass m* (Equation 5.3) as a function of density shows a maximum of 0.76 $M_\odot$ at a density of around 10^{16} gm/cm^3 and then declines to $0.3M_\odot$ when the pressure P and the density ρ diverge at the center.* Since at the center $m^*=0$, we have, from Equation 5.5,

$$g_{00} = \exp\left[-\frac{1}{c^2}\int_0^{P_c}\frac{dP}{\frac{P}{c^2}+\rho}\right], \qquad (5.17)$$

g_{00} approaches zero when $P_c \to \infty$. Thus a singularity develops at the center when the mass is still finite. This singularity, unlike the Schwarzschild singularity, *cannot* be removed by a mere coordinate transformation (see Section 6). It may be a real singularity.

*(b) Cameron's Model**

Following Oppenheimer-Volkoff's model, several other models have been constructed. Cameron (Reference 10), in particular, used an equation of state computed by Skyrme according to many-body theory, taking into account the unclear interaction at low energy. Salpeter worked out a more complete equation of state computed on the basis of effective range theorgy. It agrees with Skyrme's theory in the low energy limit for which Skyrme's theory applies. Cameron's result is shown in Figure 5. Again, a singularity ($g_{00}=0$) develops at the center of his model star at a finite mass. The value of the critical mass is around $2M_\odot$.

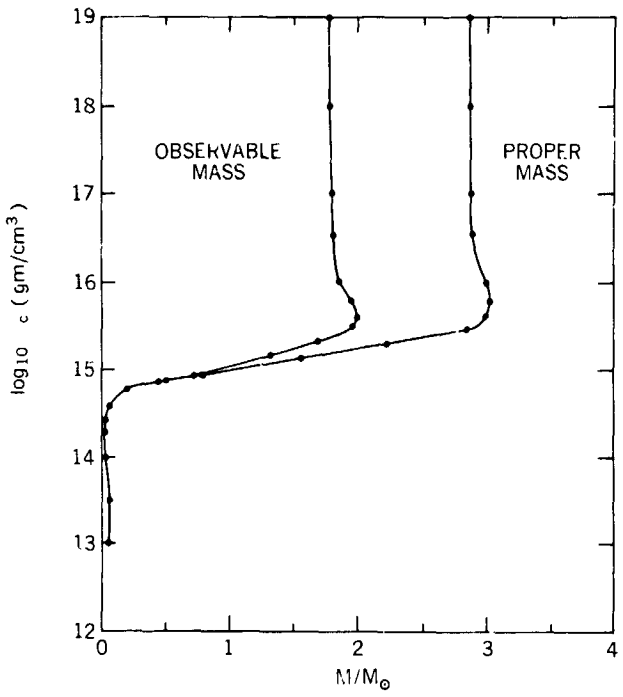


FIGURE 5.—The mass-central density P_c relation in Cameron's model (Ref. 10). Observable mass is m*; the proper mass is m as defined in Section 5.1.

(c) Wheeler et al. Hard Sphere Model

Motivated by the use of repulsive core to interpret low energy nuclear interactions, Wheeler et al. (Reference 19) studied the extreme case: the incompressible fluid for which the equation of state is

$$\left.\begin{aligned} P &= \text{finite} \\ P &= \infty \end{aligned}\right\} \quad \rho < \rho_0 , \qquad (5.18)$$

where ρ_0 is some density. Despite this kind of equation of state if the mass is not too large, in general no singularities will appear; but, for a

*The masses we talk about are masses observed by observers at large distances.

given value of ρ_o, there exists a critical mass for which a singularity ($g_{00}=0$) appears at the center. Moreover, they found that

$$\frac{dM^*}{dM} = \left[1-\left(\frac{M^*}{M_{cr}}\right)^{2/3}\right]^{1/2}, \qquad (5.19)$$

where M^* is the mass of the star observed by distant observer, M is the mass of the material that composes the star before the star is assembled, and M_{cr} is a critical mass depending on the value of P_o. Hence $dM^*/dM=0$ when $M^*=M_{cr}$. Moreover, they found that the ratio M^*/M has a non-zero finite value even when $M^*=M_{cr}$. We shall explore the meaning and implications of this result in the following discussions.

(d) Ambartsumyan-Saakyan Model

Ambartsumyan and Saakyan (References 32 and 33) studied the neutron star models on the basis of their equation of state, taking into account strange particles. Their result is shown in Figure 6. Essentially there is no change in the conclusion that g_{00} vanishes at the center of a finite mass.

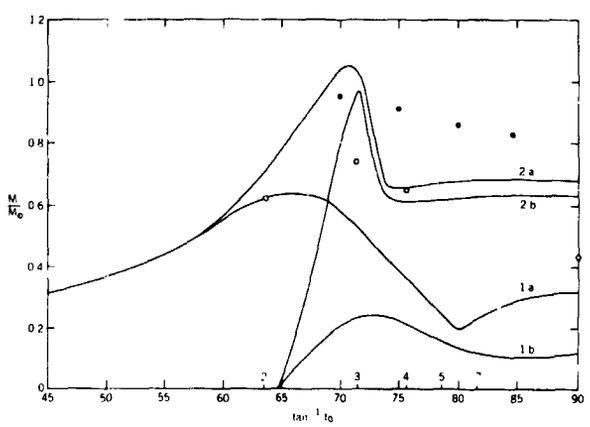


FIGURE 6.—The mass–t relation in Ambartsumyan and Saakyan's model; t_o is the value of t at the center of the star, where

$$t=4\ln\left\{\frac{P_F}{m_n c}+\left[1+\left(\frac{P_F}{m_n c}\right)^2\right]^{\frac{1}{2}}\right\}.$$

Curves 1a and 1b depict the mass of the star and its hyperon core in the case where elementary particles form an ideal gas at any densities. Curves 2a and 2b present the same picture for the case where repulsive forces acting between baryons at high densities (real gas) are taken into account. Black dots indicate the mass of configurations consisting of neutrons, with repulsive forces taken into account. Hollow dots indicate the mass of those configurations consisting of an ideal neutron gas. (Calculations by Oppenheimer and Volkoff, Ref. 27.)

*G. S. Saakyan (Reference 34) recently has repeated the computation by Cameron as described here. He found a mistake in Cameron's paper, and that the value of the critical mass is not as high as Cameron claimed.

6. DYNAMIC COLLAPSE OF OVERSIZED NEUTRON STARS

Although the four treatments of neutron star models have not exhausted the possibility of all types of equation of states, they do indicate that inevitably g_{00} vanishes at the center for a finite neutron star mass. The question arises: Is this singularity unavoidable? Is there no way to save the face of physics by avoiding infinite pressure and density? In this section we would like to examine the problem of the dynamical collapse of a star more closely, and the nature of the singularity with which we are faced.

6.1. Schwarzschild Singularity and the Ultimate Mass Limit

For a point mass, the Schwarzschild solution for the metric has the form (Reference 26):

$$ds^2 = g_{ij}\,dx^{ij} = -\frac{dr^2}{1-\frac{2GM}{rc^2}} - r^2\,d\theta^2 - r^2\sin^2\theta\,d\phi^2 + \left(1-\frac{2GM}{rc^2}\right)dt^2 . \qquad (6.1)$$

For a mass m, a singularity occurs at $r_s=2GM/c^2$. That is, $ds^2=\infty$. Consequently no light signal can cross this boundary defined by $r=r_s$.* For a proton, $r_s\approx 10^{-53}$ cm. Since the radius of a proton is 10^{-13} cm, it does seem that the Schwarzschild singularity may not play as important role in the structure of a proton For massive objects like the sun, the Schwarzschild radius is around 2.6 km, which is about the size of a neutron star of the same mass with a density of around 10^{16} gm.

This singularity has been a subject of discussion among relativists ever since the solution was found. Is it a real singularity? Or is it just a singularity related to a particular coordinate system?

*We might remark that such behavior was first predicted by Laplace on the basis of a corpuscular theory of light. See Reference 35.

By a change of coordinates of $r \to \bar{r}$, $\theta \to \theta$, $\phi \to \phi$, $t \to \bar{t}$, where $\bar{r}$ is related to r by the following relation†:

$$r = \left(1 + \frac{GM}{2\bar{r}c^2}\right)^2 \bar{r} . \qquad (6.2)$$

Equation 6.1 becomes

$$ds^2 = -\left(1 + \frac{m'}{2\bar{r}}\right)^4 \left(d\bar{r}^2 + \bar{r}^2\, d\theta^2 + \bar{r}^2 \sin^2\theta\, d\phi^2\right) + \frac{(1 - m'/2\bar{r})^2}{(1 + m'/2\bar{r})^2}\, dt^2 , \qquad (6.3)$$

where

$$m' = \frac{GM}{c^2}$$

is the geometrized mass. In Equation 6.3 the singularity in r coordinates disappears. Hence it seems that the singularity at $r = r_s$ is a coordinate singularity.

Further, Robertson (Reference 36) has demonstrated that, by a coordinate transformation, one can eliminate this singularity in the following sense: An observer in the flat space observes the free fall of a test particle towards the singularity. The time it takes the test particle to cross the boundary (which is at $r_s = 2GM/c^2$) is infinite; this is because $g_{00} = 0$ at the boundary and the test particle suffers infinite time dilatation as it approaches the singularity. However, if the observer moves with the test particle, the time duration that he observes for the test particle (and himself) to cross the boundary is *finite*. Hence, in the opinion of many relativitists, this singularity is a *coordinate* singularity since it can be removed by a proper choice of coordinate systems (transformation to a co-moving frame).

We regard the Schwarzschild singularity as a *physical singularity* in the following sense. Given a matter distribution such that initially nowhere within the matter distribution is $r_s \leq 2GM/c^2$. An observer observes this distribution of matter in a nearly flat space (since almost all points of the universe are locally nearly flat). Since it takes an infinite time for this distribution of matter to cross the Schwarzschild singularity, we may regard the Schwarzschild singularity to be a real, physical singularity. Once matter is found in a non-singular state, within the finite lifetime of the universe it will remain in the non-singular state. Similarly, if a Schwarzschild singularity were found in nature, it was created with the universe.

Since all massive stars undergo supernova explosion, their cores will inevitably collapse to extremely high densities. It might be worthwhile to investigate whether they collapse asymptotically into within Schwarzschild singularity or not. It is of more importance to investigate, first, whether a singularity similar to Schwarzschild singularity exists in neutron stars with all possibilities of equation of state.

As an interesting episode, we quote a paragraph from Eddington's momumental work on "The Internal Constitution of the Star" (Reference 37):

"The great bulk of these giant stars is due to low density rather than great mass . . . It is rather interesting to notice that Einstein's theory of gravitation has something to say on this point. According to it a star of 250 million km radius could not possibly have so high a density as the sun. Firstly, the force of gravitation would be so great that light would be unable to escape from it, the rays falling back to the star like a stone to the earth. Secondly, the red-shift of the spectral lines would be so great that the spectrum would be shifted out of existence. Thirdly, the mass would produce so much curvature of the space-time metric that space would close up round the star, leaving us outside (i.e., nowhere). The second point gives a more delicate indication and shows that the density is less than 0.001; for even at that density there would be a red-shift of the spectrum too great to be concealed by any probable Doppler effect.

Lest this argument should be regarded by our more conservative readers as ultra-modern, we hasten to add that it is to be found in the writings of Laplace*:

'A luminous star, of the same density as the earth, and whose diameter should be two houndred and fifty times larger than that of the sun, would not, in consequence of its attraction, allow any of its rays to arrive at us; it is therefore possible that the largest luminous bodies in the universe may, through this cause, be invisible.' "

3.2. Will a Massive Star Collapse into a Neutron Star Without Being Singular?

We have surveyed the work of many authors (which by no means exhausted all possible equations of state) and we find that it seems inescapable that g_{00} will vanish at the center of a neutron star if its mass approaches a certain limit, which is around $1M_\odot$.

†This coordinate system is known as the isotropic coordinate system.

*Reference 35

There are several properties characteristic to this kind of treatment of a neutron star:

(1) The rest energy of a particle is $\sqrt{g_{00}}\,mc^2$ and at $g_{00}=0$ it takes no energy to create a pair of particles, nor does one gain any energy by annihilating a pair of particles. All photons emitted at points where $g_{00}=0$ will suffer infinite red shift and there will be no energy associated with such photons.

(2) The ratio of M^*/M_f is a well-defined number even when $M^*=M_{cr}$ in all cases. In Wheeler's work (Reference 19) this ratio is 2.8, and in Cameron's work (Reference 10) this ratio is 1.5.

(3) Although only the simple incompressible fluid case can one calculate $dM^*/dm=0$ at $M^*=M_{cr}$, we believe the result $dM^*/dM=0$ at $M^*=M_{cr}$ is universal because, in the case of incompressible fluid, this is due to $g_{00}=0$ at the center. Thus consider a star at the critical mass. If one adds to it one more gram of mass, this mass will not add to the observed mass of the star, but will be converted into energy—into a kind that can escape, namely, radiation.

5.3. Wheeler's Theory

One may be perturbed by the fact that the star now becomes a gravitational machine by which matter is converted into energy, thus violating the baryon conservation law. Gravitational machines converting mass into energy are nothing new. Any gravitating body, in some sense, is a gravitational machine in the above sense. If we drop matter from infinity to the surface of the sun; roughly 10^{-6} of its mass will be converted into energy which escapes, and this piece of matter will be gravitationally bound to the sun. However, if we pump energy into the sun to disintegrate it, we will reobtain nucleons. There is no ambiguity.

In the case of neutron stars this does not seem to be so. At the critical mass limit the ratio of M^*/M is a fixed constant of the order of 1/3. According to Equation 5.19, the stellar mass approaches the critical mass when the original mass is finite. Beyond this point the star becomes a gravitational machine converting mass into energy. Thus the total number of nucleons is not a well-defined number. When we pump energy into the system to dissociate the star from its gravitational binding, we do not get all nucleons back —we only get a definite number ($\approx 3\cdot M_{cr}/m_p$) back. Thus, the baryon number may not be a meaningful concept for a neutron star.

From this example Wheeler (Reference 19 and 25) concluded that the baryon number conservation law must break down in a very dense star. We are open-minded to this kind of suggestion. Although the baryon conservation law has been demonstrated to high precision to be a valid concept in an ordinary nuclei (Reference 38), it has always been postulated only as an empirical law (e.g., Reference 39).

The origin of this law is still unclear. Moreover, we do not know what this law actually means when baryons are closely packed to within their structure ($\leq$ repulsive core radius). It certainly poses an extremely interesting and meanwhile difficult problem. The final solution, as we have seen in Section 5.3, must come from extremely high energy physics.

6.4. Oppenheimer-Snyder Theory

Oppenheimer and Snyder (Reference 40) had approached the problem of collapse in the following manner: In the absence of the pressure, the collapse may be treated analytically. They computed the world trajectories of material bodies falling towards a common center under the gravitational field produced by these material bodies. They obtained the following result: To an external observer these particles fall toward the Schwarzschild singularity asymptotically and the total time of fall is infinite. This is because of the infinite time dilatation effect mentioned earlier. To a local observer, the time is finite; and, for the case of the sun, it is of the order of 1 day. Since the star continues to fall indefinitely, the star is always in some dynamical motion and this is consistent with the fact that no static structure of such a star exists if its mass exceeds a certain limit.

We do not regard this resolution as satisfactory, since the pressure P cannot be neglected in the case of a neutron star; from what others have demonstrated (References 19, 31, 32, and 33), $P\approx\rho c^2$ and therefore cannot be neglected. Their solution, however, may be of cosmological significance. We shall discuss further possible solutions to this problem in Section 9.

7. ANGULAR MOMENTUM AND STELLAR COLLAPSE

We have delayed the discussion of the effect of angular momentum up to now—not because it is not important, but because it leads to very interesting consequences: namely, the possibility of gravitational radiation from a rapidly rotating neutron star, and the possibility of their detection.

The moment of inertia of a spherical body of radius R, mass M, and of uniform density ρ is:

$$I = \frac{2}{5} MR^2 . \qquad (7.1)$$

The angular momentum $I\omega$ is a conserved quantity (where ω is the angular velocity). For a normal star $R \approx 10^{11}$ cm and the period of rotation is around 1 day. Thus,

$$I\omega \propto R^2 \omega \approx 6 \times 10^{17} . \qquad (7.2)$$

The centrifugal force f_c per unit mass at the equator of the star is

$$f_c = \frac{v_{eq}^2}{R} ,$$

where $v_{eq} = R\omega$ is the speed at the equator. Since $f_c < MG/R^2$ (otherwise the star will be torn apart by rotation), we have the upper limit on R:

$$\frac{(R^2\omega)^2}{R^3} < \frac{MG}{R^3} \qquad (7.3)$$

or

$$R > \frac{(R^2\omega)^2}{MG} . \qquad (7.4)$$

This is the minimum radius to which a rotating star may collapse. Using $M = 2 \odot M$, and the value of $R^2\omega$ given by Equation 7.2, we find that

$$R > 10^8 \text{ cm} , \qquad (7.5)$$

and at this limiting value of R the angular velocity ω is around 1 rad/sec, and the period is around 0.1 sec.

Although a rotating body with cylindrical symmetry does not radiate gravitational waves, there is no reason that a rapidly rotating body should preserve its cylindrical symmetry. The classical Jacobi ellipsoid is an example: a homogeneous, uniform density fluid held together by its gravitational field can assume the shape of an ellipsoid with the axis of rotation perpendicular to the long axis of the ellipsoid. As an estimate we apply to our rotating star the formula of gravitational wave radiating power computed for a spinning rod (References 41, 42, and 43). The radiating power is

$$P = \frac{32GI^2\alpha^6}{5c^5} . \qquad (7.6)$$

The rotational energy is $1/2\ (I\omega^2)$. Hence the relaxation time τ for the star to lose its angular momentum is

$$\tau = \frac{5c^5}{64GI\alpha^4} . \qquad (7.7)$$

For the parameters we consider, we find $\tau \approx 50$ seconds. Although this is longer than the time of dynamical collapse as calculated by Colgate (Reference 7), it is quite small compared with the lifetime of a neutron star, which we shall compute.

Weber thinks such gravitational radiation is detectable (References 41 and 44).

8. OBSERVABLE FEATURES OF NEUTRON STARS

8.1. Energy Content of a Neutron Star

We now shall calculate the energy content of neutron stars. The Fermi energy of neutrons (and hyperons) averaged over the star cannot be greater than $m_n c^2$; otherwise the gravitational energy of the star will exceed its rest energy and the star will be within its Schwarzschild radius. Further, if the temperature is too high, neutrino processes will quickly dissipate its thermal energy. Hence the temperature must be below a certain limit. From these criteria the temperature of neutron stars is thus estimated to be 10^9 °K, assuming they were formed earlier at a high temperature. The relaxation time for cooling by neutrino emission at this temperature by plasma neutrino process and pair annihiliation process (References 21–23 and 45) is greater than 10^3 years, which will be the lifetime of neutron stars against optical emission (see next section).

We now apply the theory of a nearly zero temperature ideal Fermi gas to find the thermal energy. The specific heat per unit mass is given as Reference 17.

$$c_v = \left(\frac{\partial \epsilon}{\partial T}\right)_v = \frac{\pi^3 k}{m_n}\left(\frac{kT}{m_n c^2}\right)\frac{x(x^2+1)^{1/2}}{x^3}$$

$$= \frac{T}{T_0}\frac{x(x^2+1)^{1/2}}{x^3} , \qquad (8.1)$$

where $x = p_F^{(n)}/m_n c^2$ and $p_F^{(n)}$ is the Fermi momentum for the neutron gas. The total thermal energy ϵ_{th} is given by

$$\epsilon_{th} = \int_0^T c^\sigma \, dT = \frac{1}{2} \frac{T^2}{T_0} \frac{x(x^2+1)^{1/2}}{x^3} . \quad (8.2)$$

As we have remarked before, the overall average E_F is less than $m_n c^2$. Therefore x is never very much different from unity. Taking $T = 10^9$ °K and $x = 1$, we find that

$$T_0^{-1} = 0.75 \times 10^{-4} \quad (8.3)$$

and

$$\epsilon_{th} \cong 4.6 \times 10^{13} \text{ ergs/gm} \quad (8.4)$$

The total thermal energy of the neutron star is $\approx M \cdot 10^{14} = 10^{47}$ ergs, which is about the thermal energy of the sun.

The neutron core is not composed of neutrons alone. Hence the thermal energy is a sum over all particle energies. In this case, the value of x will be somewhat smaller than 1. The thermal energy of a neutron star may be several times greater than the value estimated here.

8.2. Physical Properties of the Surface of a Neutron Star

The surface of a neutron star must consist of ordinary matter since, when $\rho < 10^{12}$ gm/cm³, the composition of matter is predominantly iron group elements. Thus we can picture a neutron star as a centrally condensed neutron core, surrounded by a layer of nearly iron group elements. If the interior temperature of a neutron star is taken to be 10^9 °K, then when $\rho > 10^6$ gm/cm³ the iron group elements will become degenerate. Since the thermal conductivity of a neutron core and the degenerate layer is very high, we need only consider the nondegenerate ordinary-stellar-matter envelope which contributes mostly to the overall opacity of the star.

The iron group elements have very high values of Z and the ionization temperature in the k-shell electrons is around 0.015 Mev, which corresponds to a temperature of 10^8 °K. The absorption coefficient exhibits a very sharp edge (absorption edge). At the ionization energy the absorption cross section may be as high as 10^{-20} cm². On the average, the mean free path of radiation may be as short as 0.1 gm/cm². When $T > 10^8$ °K, the main source of opacity is due to Compton scattering. The opacity is a constant. The mean free path of photons is roughly 5 gm/cm².

As a rough estimate we now apply the theory of the white dwarf atmosphere to a neutron star (Reference 46). Assuming the main source of opacity to be due to the absorption edge of k-shell electrons, the radial dependence of temperature and density is:

$$\rho \propto \left(\frac{\gamma}{\beta} - 1\right)^{3.25} , \quad (8.5)$$

$$T \propto \left(\frac{R}{\beta} - 1\right) , \quad (8.6)$$

where R is the radius of the neutron star and the proportional constant is determined by the mean density of the star. For a neutron star of mean density $\langle\rho\rangle = 10^{15}$ gm/cm³, R is around 10^6 cm. Thus

$$\rho \approx \langle\rho\rangle \left(\frac{R}{r} - 1\right)^{3.25} = 10^{15} \left(\frac{\Delta r}{R}\right)^{3.25} . \quad (8.7)$$

Taking $\rho = 10^6$ gm, we find $\Delta r = 10^{-2.77}$ R $\approx 10^3$ cm. Material contained in the envelope in gm/cm² is *roughly* $\int_0^{\Delta r} \rho dr \approx 10^8$ gm/cm². The number of scatterings a photon suffers in leaving the envelope is roughly $\int_0^{\Delta r} \rho dr \lambda \approx 10^8$, where λ is taken to be 5 cm/gm-cm².

To estimate the luminosity, we use an intuitive theory due to J. Steinberger.* Consider the following experimental setup: A series of parallel infinite plates of perfect blackbody properties is set up as in Figure 7. Let the temperature of the nth parallel plate be T_n, and let T_1 be fixed. The energy flow is assumed to be completely radiative. When a steady state of energy flow is reached, it is easy to verify the following relation:

$$\begin{aligned} acT_1^4 - acT_2^4 &= acT_2^4 - acT_3^4 = \cdots \\ &= acT_{n-1}^4 - acT_n^4 = \cdots \\ &= acT_{n-1}^4 - acT_n^4 = acT_n^4 . \end{aligned} \quad (8.8)$$

*Private communication.

where a is the Stefan-Boltzmann constant and c the velocity of light. From this equation we obtain a relation between T_1 and T_m:

$$acT_1^4 = \frac{acT_m^4}{m} . \quad (8.9)$$

Photons are continuously absorbed and reemitted by atoms or nuclei. In the atmosphere of a star the atoms are like the parallel plate in Steinberger's model for radiative transfer. The energy flux radiated from a unit area on the surface of a neutron star is ($T_1 \approx 10^9$ °K)

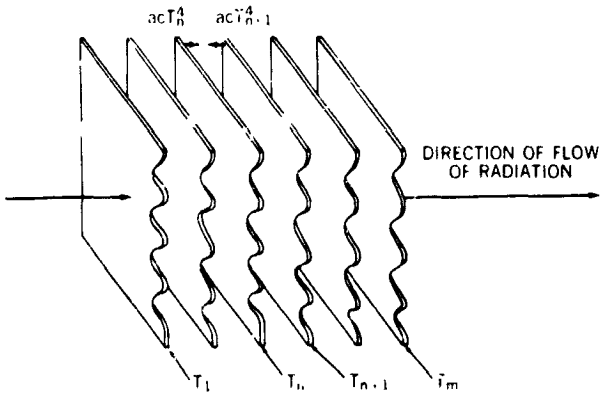


FIGURE 7.—An extremely simplified model of stellar atmospheres. The plates are made from materials with perfect blackbody properties.

$$f = \frac{acT_1^4}{14 \times 10^8} \approx 5 \times 10^{23} \text{ ergs/cm}^2\text{-sec} . \quad (8.10)$$

The overall luminosity L is thus

$$L = 4\pi r^2 f \geq 10^{37} \text{ ergs/sec} . \quad (8.11)$$

The maximum of the spectral frequency corresponds to $T \approx 10^7$ °K. We have computed the energy content of a neutron star to be 10^{47} ergs. A neutron star radiating energy at such a lavish rate will last for at most 1000 years. However, as the temperature is reduced, the rate of energy radiation will be reduced, too; when $T_1 = 10^8$ °K, the luminosity L will be reduced to around 10^{-4} of the value given in Equation 8.11 and the lifetime will be around 10^4 years, with a maximum on the spectral frequency $T \approx 10^6$ °K (0.1 kev).

8.3. Observational Problem

Our atmosphere is transparent only to radiation in the following wavelength region (e.g., Reference 47): 3000A→7000A (visible), 8000A→12,000A (infrared window), 1 cm→10 meters (microwave) 47): 3000A¼7000A (visible), 8000A→12,000A (infrared window), 1 cm→10 meters (microwave), and 200 meters→infinity (low frequency radio wave).

Interstellar matter consists mainly of hydrogen, and it is extremely opaque to radiation of wavelength near the Lyman continuum ($\approx$1000A). Since the absorption coefficient exhibits a sharp edge and has a λ^3 dependence, for frequencies other than that near the Lyman continuum it is relatively transparent.

The surface temperature of neutron stars is around $10^6 \rightarrow 10^7$ °K. Burbidge (Reference 48) has estimated the red shift and he found that, if the neutron stars have a very small envelope (as we have calculated), the light from such stars may suffer a red shift as large as $\lambda/2$. Hence the temperature observed by distant observers will be one-half the value at the surface. The red shift will therefore reduce the observable temperature by at most a factor of 2.

The maximum in Planck's formula for $T = 10^6 \rightarrow 10^7$ °K is between 100A and 10A (in-between UV and x-ray bands). Both lie in a region where interstellar absorption is not large. The mean free path is around 10^5 light years, assuming an interstellar hydrogen density of 1 particle/cm^3. We remark that this wavelength region is very difficult to observe experimentally.

The fraction of this radiation energy R_ν that will pass through our atmospheric window is easily seen to be given by

$$R_\nu = \frac{\int_0^x \frac{x^3\,dx}{e^x - 1}}{\int_0^\infty \frac{x^3\,dx}{e^x - 1}} = \frac{15}{\pi^4} \int_0^x \frac{x^3\,dx}{e^x - 1} , \quad (8.12)$$

where $x = h\nu/kT$. For small values of x, since $e^x \approx 1 + x$, we have

$$R_\nu = \frac{5}{\pi^4} x^3 . \quad (8.13)$$

For $kT = 50$ ev ($T \approx 5 \times 10^5$ °K), $h\nu = 5$ ev (corresponding to $\lambda \approx 3000$A), we find that

$$R_\nu \approx 5 \times 10^{-3} . \quad (8.14)$$

Thus, for a neutron star of internal temperature 10^8 °K it will have a luminosity of $\approx 10^{-6}L_\odot$. For such low luminosity stars, the star has to be, borrowing a comment from Greenstein,* "practically inside the solar system in order to be observable by the world's largest telescope, the 200 inch Mount Palomar telescope." Hence, it is not surprising that no neutron stars have ever been found.

Since in our picture every supernova star will inevitably become a neutron star, we can estimate their numbers in our galaxy. Taking the age of our galaxy to be 10^{10} years, and the rate of a supernova explosion to be roughly 10^{-2}/year (References 8 and 9), we find the total number of neutron stars in our galaxy to be 10^8, of which possibly ten will have a surface temperature of about 10^7 °K, and possibly 10^2 will have a surface temperature of about 10^6 °K.

With the use of proper instruments mounted in an earth satellite now available for scientific research, such neutron stars should be readily detectable. *(See footnote on page 51, however.)*

Moreover, since most stars possess rotation, the rotational energy will be dissipated, during the collapse phase, into gravitational waves. Such waves should be detectable. Instruments designed by Weber and his associates (References 41 and 44), if perfected to their expectation, should be able to detect such waves many galaxies away. Thus their instruments may be used to detect supernova explosions. We remark that their instruments are more sensitive than the neutrino monitor station suggested previously (References 1 and 21).

9. DISCUSSION

The central point of the paradox presented by Wheeler is essentially the following: Given a neutron star, assuming one can add to its mass (hence the nucleon numbers) until it reaches the critical value; then at this critical value a singularity ($g_{00}=0$) develops at the center. Unlike the Schwarzschild singularity this singularity cannot be removed by performing a coordinate transformation. Further addition of mass to the star will not increase its mass nor its nucleon number. All nucleons added to it will be converted into radiation energy. This essentially violates the empirical law that baryons cannot be destroyed in any reaction. Since the existence of the critical mass is inevitable, Wheeler concluded that the baryon conservation law must be violated in super-dense stars.

We have examined this problem and our conclusions are the following:

(a) The absence of quantum-gravitational effects: Inside a neutron star the gravitational field has to interact with the elementary particle field classically. This is because the gradient of the gravitational field is extremely small as compared with the gradients of nucleon, electromagnetic, etc. fields. By the principle of equivalence, one can replace a uniform gravitational field by a uniform acceleration. Thus the equation of state may be computed without regard to any gravitational field present.

(b) Wheeler argued that one cannot get away from this difficulty by proclaiming that the whole star falls through the Schwarzschild singularity. The problem poses, as he calls it, "a problem of principle." No one can forbid an observer from adding one extra gram of matter to a star already at the critical mass.

We would like to remark that it is by no means clear how this singularity is developed. It depends strongly on the equation of state. On the central density-mass curve (Figure 4) a maximum is found. Beyond this maximum, in most cases the gravitational binding energy of the neutron star is *positive*. This means *the critical mass cannot be reached statically.* Moreover, we know that, if the mass decreases with increasing central density, the stellar configuration is not a stable one. Hence, then the problem of the critical mass must be treated dynamically. We also note that the maximum in the mass is reached when the star is still free from singularity.

If the existence of the mass maximum is inevitable, the static problem as posed by Wheeler is irrelevant to the dynamical collapse of a star (the only way a neutron star may be formed).

Neglecting the pressure term in the stress energy tensor, Oppenheimer and Snyder (Reference 40) found that the star collapses toward the Schwarzschild singularity asymptotically. In reality the pressure cannot be neglected—in fact,

*Private communication. The author is indebted to him for an enlightening discussion on the nature of peculiar radio sources which were once considered to be neutron stars. However, J. Greenstein demonstrated that they cannot be neutron stars (Reference 49).

$\approx \rho c^2$. The essential questions are therefore: If we takes into account the pressure, does a real singularity such as $g_{00}=0$ develop at the center first? Or does the star still fall through the Schwarzschild singularity asymptotically? These are certainly relevant questions to ask. They can be answered theoretically be a further study of neutron star models.

(c) Another type of equation of state: S. Weinberg* posed the following argument. An incompressible fluid may be regarded as composed of particles with a repulsive potential between them. Although they are the hardest equations of state one can think of, they also carry an indefinitely large amount of energy. Since the *source* of gravitational field is energy itself, one gets into infinite trouble by letting the energy be infinitely large. He therefore suggested the following: Let there be an attractive force of small range. When matter is compressed hard enough, the attractive force will come into play and local clustering of particles will occur, forming quasi-bound states. The energy of the system will be decreased by the amount of the binding energy. One can thus increase the mass limit for neutron stars by letting local clustering of particles take place.

If one piles all the mass of the universe into a neutron star and considers Wheeler's static problem, then essentially, if one still wants to save the baryon conservation law, the binding energy of these clusters must be of the order of the rest energy of the particles themselves. Such a *super-strong field has not been observed* in present high energy physics experiments.

If the fine structure constant is of order unity, a number of interesting things will happen. Pairs of electrons will be created spontaneously into bound states; yet these bound pairs have zero energy. In order that such thing be possible for baryons, the strength of the coupling $f^2/\hbar c$ must be at least 100 times stronger than that for the strong interaction. In order not to conflict with the present knowledge of elementary particles, the range of this super-strong interaction cannot exceed, say, 10^{-15} cm. Since a binding among all pairs is assumed (attractive force), the interactions must be scalar or tensor in character. The quanta associated with such a field must be around 100 times the π meson mass, or 20 times the nucleon mass.

Of course, what we have suggested is purely speculative. A star which collapses to within its Schwarzschild singularity will gradually become invisible. On the other hand, if a star can get rid of its mass-energy by resorting to baryon nonconservation or the existence of such a hypothetical super-strong interaction, the remaining body should be observable with approximately the surface characteristic discussed in Section 8. They are certainly observable.

10. CONCLUDING REMARKS*

In this paper the final evolution of a pre-supernova star is reviewed. It seems inevitable that the star will end up as a neutron star with a radius of around 10^6 cm and a surface temperature of around 10^6–10^7 °K. A critical mass limit exists for neutron stars. Beyond this mass limit no static structure is possible. Neutron stars can only be detected by extra-terrestrial x-ray telescopes. If detected, they pose interesting questions on our present theory of fundamental particles.

11. ACKNOWLEDGMENTS

The author wishes to thank Professors C. W. Misner and J. A. Wheeler for discussions on the general properties of neutrons stars, and Professor L. Woltjer for a discussion on the character of the atmosphere for a neutron star. He also wishes to thank Dr. Robert Jastrow for his hospitality at the Institute for Space Studies.

*Private communication. The author first learned this idea from Professor C. W. Misner, who believed he was quoting Dr Weinberg. When the author asked Dr. Weinberg, he insisted that he was misquoted by one of us (H. Y. C. or C. W. M.). Nevertheless, he admitted later that this idea might work. His original idea applied to model universes as a whole, but he thinks this idea may work inside a neutron star too. The author feels that, although the present idea originates from misquotations, Dr. Weinberg nevertheless should deserve the credit of being the originator.

*When this article went to press, rocketbound x-ray observations made by Rossi's group (H. Gursky, R. Giacconi, F. R. Paolini, and B. B. Rossi, *Phys. Rev. Letters*, 11: 530 (1963)) and H. Friedman's group (reported in the December meeting of the American Astronomical Society, 1963) reveal the existence of discrete x-ray sources, one of them near the Crab Nebula, the remnant of the supernova of 1054 A.D. The energy flux in the 3A wavelength region is around 10^{-7} ergs/cm²-sec. If the source is indeed neutron stars in the Crab Nebula, the calculated source flux is around 10^{37} ergs/sec. This is consistent with the estimates made in this paper. However, further observations are necessary to clarify if these x-ray sources are indeed neutron stars

A more detailed evaluation of the surface properties of neutron stars has been performed by the author and also by Miss S. Tsuruta and Dr. A. G. W. Cameron. Both works are to be published elsewhere. Results indicated that the crude estimates made in this paper are correct to one order of magnitude.

REFERENCES

1. CHIU, H. Y., "Neutrino Astronomy," Paper presented at the International Congress of Nuclear Physics and Nuclear Medicine, on the occasion of the tricentennial death anniversary of Blaise Pascal on June 28, 1962, at the Clermont University, Clermont-Ferrand, France. (Proceedings to be published.)
2. GAMOW, G., and SCHOENBERG, M., "Neutrino Theory of Stellar Collapse," *Phys. Rev.* 59:539–547, April 1, 1941.
3. BURBIDGE, E. M., BURBIDGE, G. R., FOWLER, W. A., and HOYLE, F., "Synthesis of the Elements in Stars," *Rev. Mod. Phys.* 29(4):547–650, October 1957.
4. COLGATE, S. A., GRASBERGER, W. H., and WHITE, R. H., "The Dynamics of a Supernova Explosion," *J. Phys. Soc. Japan* 17 (Suppl. A-III): 157–160, January 1962.
5. CHIU, H. Y., "Neutrino Emission Processes, Stellar Evolution, and Supernova, II," *Ann. Phys.* 16(3):321–345, December 1961.
6. CAMERON, A. G. W., "A Revised Table of Abundances of the Elements," *Astrophys. J.* 129(3):676–684, May 1959.
7. COLGATE, S. A., and JOHNSON, M. H., "Hydrodynamic Origin of Cosmic Rays," *Phys. Rev. Letters*, 5(6): 235–238, September 15, 1960.
8. ZWICKY, F., "Supernovae" *In:* "Handbuch der Physik," v. 51 (S. Flügge, ed.):766–785, Berlin: Springer-Verlag, 1958.
9. MINKOWSKI, R., "International Cooperative Efforts Directed Toward Optical Identification of Radio Sources," *Proc. Nat. Acad. Sci.* 46(1):13–19, January 1960.
10. CAMERON, A. G. W., "Neutron Star Models," *Astrophys. J.* 130(3):884–894, November 1959.
11. ZWICKY, F., "Collapsed Neutron Stars," *Astrophys. J.* 88:522–525, October 1938.
12. BAADE, W., and ZWICKY, F., "Photographic Light-Curves of the Two Supernovae in IC 4182 and NGC 1003," *Astrophys. J.* 88:411–421, November 1938.
13. LEDOUX, P., "Stellar Stability," *In:* "Handbuch der Physik," v. 51 (S. Flügge, ed.): 605–688, Berlin: Springer-Verlag, 1958.
14. TSURUTA, S., and CHIU, H. Y., "Adiabatic Exponents of a Mixture of Radiation and Electrons," *Astronom. J.* 67, 284 (1962).
15. HOYLE, F., and FOWLER, W. A., "Nature of Strong Radio Sources," *Nature*, 197(4867):533–535, February 9, 1963.
16. SCHWARZSCHILD, M., "Structure and Evolution of the Stars," Princeton, N.J.: Princeton University Press, 1958: Chapter 3.
17. CHANDRASEKHAR, S., "An Introduction to the Study of Stellar Structure," New York: Dover Publications, 1957: Chapter 10.
18. SALPETER, E. E., "Energy and Pressure of a Zero-Temperature Plasma," *Astrophys. J.* 134(3): 682, November 1961.
19. HARRISON, B. K., WAKANO, M., and WHEELER, J. "La Structure et l'Evolution de l'Univers, Brussels: Stoops, 1958.
20. CAMERON, A. G. W., "Pycnonuclear Reactions and Nova Explosions," *Astrophys. J.* 130(3):916–94 November 1959.
21. CHIU, H.-Y., and MORRISON, P. "Neutrino Emissio. from Black-Body Radiation at High Stellar Temperatures," *Phys. Rev. Letters*, 5(12):573–75, December 3, 1960.
22. CHIU, H.-Y., and STABLER, R. C., "Emission of Photoneutrinos and Pair Annihilation Neutrinos from Stars," *Phys. Rev.* 122(4):1317–1322, May 15, 1961.
23. CHIU, H.-Y., "Annihilation Process of Neutrino Production in Stars," Phys. Rev. 123(3):1040–1050, August 1, 1960.
24. DANBY, G., GAILLARD, J-M., GOULIANOS, K., LEDERMAN, L. M., MISTRY, N., SCHWARTZ, M., and STEINBERGER, J., *Phys. Rev. Letters* 9, 460 (1962).
25. WHEELER, J. A., "Superdense Stars," Chapter 10 in *Relativity and Gravitation*, ed. H. Y. Chiu and W. F. Hoffman. New York: Benjamin, 1963.
26. TOLMAN, R. C., "Relativity, Thermodynamics and Cosmology," Oxford, England: Clarendon Press, 1934:239–244.
27. OPPENHEIMER, J. R., and VOLKOFF, G. M., "Massive Neutron Cores," *Phys. Rev.* 55:374–381, February 15, 1939.
28. PAPAPETROU, A., "Spinning Test-Particles in General Relativity, I," *Proc. Roy. Soc.* A 209:248–258, October 23, 1951.
29. CORINALDESI, E., and PAPAPETROU, A., "Spinning Test-Particles in General Relativity, II," *Proc. Roy. Soc.* A 209:259–268, October 23, 1951.
30. LANDAU, L. D., and LIFSHITZ, E., "The Classical Theory of Fields," Tr. from Russian by M. Hamermesh, Cambridge, Massachusetts: Addison-Wesley Press, 1951: 89.
31. ZEL'DOVICH, YA. B., "Equation of State at Ultra-High Densities and its Relativistic Limitations," *Zh. Eksper. i Teor. Fiz. (USSR)*, 41(5)(11)):1609–1615, November 1961; Tr. in *Sov. Phys.*—JETP 14:1143–1962.
32. AMBARTSUMYAN, V. A., and SAAKYAN, G. S., "The Degenerate Superdense Gas of Elementary Particles, *Astron. Zn. (USSR)*, 37(2):193–209, March-April 1960; Tr. in *Sov. Astron. Aj* 4(2):187–201, September–October 1960.
33. AMBARTSUMYAN, V. A., and SAAKYAN, G. S., "The Internal Structure of Hyperon Configurations of Stellar Masses," *Astron. Zh. (USSR)*, 38(6):1016–1024, 1961; Tr. in *Sov. Astron.-Aj*, 5:779, 1962.
34. SAAKYAN, G. S., "Comments on a Paper by A. CAMERON," *Astron. Zh.* 40(1):82–84, January–February 1963; Tr. in *Sov. Astron.-Aj.* 7(1):60–62, July–August 1963.

P
o
6. ROBERTSON, H. P., "Relativistic Cosmology," *Phil. Mag.* 5(suppl.):835–848, May 1928.
37. EDDINGTON, A. S., "The Internal Constitution of the Stars," Cambridge, England: The University Press, 1926; New York: Dover Publications, 1959: Chapter 1.
38. REINES, F., COWAN, C. L., and KRUSE, H. W., "Conservation of the Number of Nucleons," *Phys. Rev.* 109(2):609–610, January 15, 1958.
39. ROMAN, P., "Theory of Elementary Particles," Amsterdam: North-Holland Publishing Co., 1960: 245.
4. OPPENHEIMER, J. R., and SNYDER H., "Continued Gravitational Contraction," *Phys. Rev.* 56:455–459, September 1, 1939.
41. WEBER, J., "General Relativity and Gravitational Waves, New York: Interscience Publishers, 196 Chapters 7 and 8.
42. EINSTEIN, A., "On Gravitational Waves," *Preuss. Acad. Wiss. Berlin* 8:154–167, 1918.
43. EDDINGTON, A. S., "Propagation of Gravitation Waves," *Proc. Roy. Soc.* 102:268–282, December 1, 1922.
44. WEBER, J., "Gravitational Waves," Chapt. 5 in *Relativity and Gravitation*, ed. H. Y. Chiu and W. F. Hoffman, New York: Benjamin, 1963.
45. ADAMS, J. B., RUDERMAN, M. A., and Woo C.-H., "Neutrino Pair Emission by a Stellar Plasma," *Phys. Rev.* 129(3):1383–1390, February 1, 1963.
46. STRÖMGREN, B., *Handb. d. Phys.* 7, 160–161, Berlin: Springer-Verlag, 1936.
47. BYERS, H. R. Chapter 7 in *The Earth as a Planet*, Ed., G. P. KUIPER. The University of Chicago: Chicago Press, 1954.
48. BURBIDGE, G., "The Detection of Stars with Neutron Cores," *Astrophys. J.* 137(3):995–996, April 1, 1963.
49. GREENSTEIN, J. L., and MATTHEWS, T. A., "Red-Shift of the Unusual Radio Source 3C 48," *Nature*, 197(4872):1041–1042, March 16, 1963.

Appendix A
VIRIAL THEOREM

Now we derive the virial theorem. Defining $V \equiv (4\pi/3)r^3$ and multiplying both sides of Equation 2.9 by V, we have

$$V\,dP = -\rho \frac{Gm}{r^2} \cdot \frac{4\pi}{3} r^3\,dr = -\frac{1}{3}\frac{Gm\,dm}{r} \quad . \tag{A1}$$

We now integrate Equation (A1) over the star. Partically integrating $\int V dP$, we have

$$\int V\,dP = PV \Big|_{r=0}^{r=R} - \int P\,dV = -\int P\,dV \tag{A2}$$

since, at $r=0$, $V=0$; and, at $r=R$ (radius of the star), $P=0$. Thus, we have the relation

$$3\int P\,dV = \int \frac{Gm\,dm}{r} = -E_G \quad . \tag{A3}$$

If a relation between P and E (energy density) exists, Equation (A3) relates the gravitational energy of a star to its thermodynamic energy. For a non-relativistic gas $P=2/3E$; and for a relativistic gas, $P=1/3E$. Thus the total energy E_T of a star is

$$E_T = -\eta E_{th} \ , \tag{A4}$$

where $0 \leq \eta \leq 1$.

SURFACE X-RAY EMISSION FROM NEUTRON STARS*

HONG-YEE CHIU† AND E. E. SALPETER‡

Galactic x-ray emission has been observed from rockets[1] and recently[2] two discrete sources of x-rays in the wavelength region of 1.5 to 8 Å have been found. The measured flux over this wavelength region was 10^{-7} erg/cm^2 sec for the stronger source (located near Scorpius) with an angular size less than 5°. The other source is about eight times weaker and seems to coincide with the Crab Nebula. The Crab is known to be the remnant of the supernova in A.D. 1054, three thousand light years away. The stronger Scorpius source may[3] be a (somewhat closer) supernova remnant of similar age.

During a type I supernova explosion, the core of the original star may contract into a hot "neutron star." Such neutron stars were suggested[4] as possible x-ray sources and we give analytic estimates[5] of their observable surface properties as a function of age (for central temperatures of the order of 10^9 °K).

Neutron stars can exist with masses between about $0.2M_\odot$ and $1.5M_\odot$ ($M_\odot = 2\times10^{33}$ g is the solar mass) and have a radius R close to 10 km and interior densities a few times 10^{14} g/cm^3. The neutrons are highly degenerate for central temperatures $T_c \ll 5\times10^{11}$ °K and their average Fermi energy $E_{F,n}$ is about 50 MeV. Because of the high conductivity, the neutron core is essentially isothermal and its total thermal energy E_t is[6] approximately

$$E_t = 7.6\times10^{47}(T_c/10^9\ ^\circ\mathrm{K})^2(M/M_\odot) \times(E_{F,n}/50\ \mathrm{MeV})^{-1}\ \mathrm{ergs}. \quad (1)$$

The neutrons are in equilibrium with a small amount of protons and electrons with $E_{F,n} \approx E_{F,e}$ ($E_{F,p}$ is small). Neutrinos can be emitted by a modified "Urca process" in which two neutrons with suitable momenta (near the Fermi momentum) "scatter" into a final state of $n+p+e^-+\nu$ (the inverse process releases ν). Our present[7] estimate for this Urca energy loss rate, $L_{\nu,\mathrm{Ur}}$ is

$$L_{\nu,\mathrm{Ur}} \sim 2\times10^{36}(T_c/10^9\ ^\circ\mathrm{K})^8(E_{F,n}/50\mathrm{MeV})^{-2.25} \times(M/M_\odot)\ \mathrm{erg/sec}. \quad (2)$$

At densities below 10^{11} g/cm^3, the equilibrium configuration favors "ordinary ionized matter" which constitutes a thin mantle ($\sim$1 km) around the neutron core. Throughout most of the mantle the electrons are realativistically degenerate with a high conductivity, so that the temperature is close to T_c, and one can calculate the "density scale-height" analytically. The plasma neutrino process[8] has a maximum rate at densities near $(T/10^9\ ^\circ\mathrm{K})^3\times4\times10^{10}$ g/cm^3 and we find for the integrated energy loss rate $L_{\nu,\mathrm{pl}}$

$$L_{\nu,\mathrm{pl}} \approx 2\times10^{36}(T_c/10^9\ ^\circ\mathrm{K})^{10}(R/10\ \mathrm{km})^4 \times(M/M_\odot)^{-1}\ \mathrm{erg/sec}. \quad (3)$$

When $\rho < \rho_m \approx (T/10^9\ ^\circ\mathrm{K})^{1.5}\times10^6$ g/cm^3, the electrons become nondegenerate and their conductivity low. The main temperature gradient then occurs in an outer nondegenerate layer (a few meters thick) of "ordinary matter." We assume $T = T_c$ at $\rho = \rho m$, a constant opacity ($\kappa^{-1}\rho^{-1}$ is the photon mean free path), and $A/Z = 2$ in this layer, and obtain its structure analytically. The photon luminosity L_{ph} is then

$$L_{\mathrm{ph}} = 1.6\times10^{37}(0.2/\kappa)(M/M_\odot) \times(T_c/10^9\ ^\circ\mathrm{K})^{1.5}\ \mathrm{erg/sec}, \quad (4)$$

independent of the radius R. For $T_c \geq 5\times10^8$ °K, Thomson scattering of photons by free electrons

*Published in *Physical Review Letters*, 12(15): 413–415, April 13, 1964.

†Goddard Space Flight Center and Columbia University. Mr. Chiu is a NCR–NAS Senior Research Associate, and also a Research Fellow, Academia Sinica, Republic of China.

‡Goddard Space Flight Center and Cornell University, Ithaca, New York.

dominates which gives $\kappa = 0.2$. For $T_c \leq 2\times10^8$ °K the opacity κ depends on temperature (roughly as $\rho^{0.5}T^{-3.5}$), due to photon absorption by a small admixture of bound electrons. We estimate an average value of 0.4 for κ at $T_c = 2\times10^8$ °K.

Table I gives our results for several values of the central temperature for mass $M = 0.5M_\odot$ and radius $R = 10$ km. T_e is the effective blackbody surface temperature corresponding to L_{ph} and the surface area $4\pi R^2$; λ_W is $hc/4.965\kappa T$, the wavelength of the Wien displacement law but with a gravitational red shift of 8 percent included; τ is $E_t(L_\nu + L_{ph})^{-1}$, a measure of the cooling time. A neutron star with $T_c \approx 10^9$ °K in the Crab Nebula (age$\sim10^3$ years) would produce near the earth, according to Table I, an x-ray intensity close to that observed.[2] This agreement is gratifying but in view of the large numerical uncertainties[9] in Table I may still be fortuitous.

Spectral measurements of discrete x-ray sources in the future will be very useful: At short wavelengths ($\lambda \lesssim 1$Å) the rapid fall-off of the neutron star's blackbody spectrum should be easy to distinguish from the flatter spectra expected from other[10] types of x-ray sources. If finer spectral resolution is available, intensity discontinuities may be detectable at wavelengths λ_K corresponding to K-shell absorption edges of medium-size nuclei (e.g., 6.3 Å for Mg and 1.3 Å for Fe, before applying the red-shift correction). Measurement of λ_K (or of absorption lines) would furnish a value for the gravitational red-shift and hence for M/R for the neutron star.

For x-ray wavelengths $\lambda \lesssim 5$ Å, absorption by interstellar gas is small even across the whole galactic disc. If x-ray detection sensitivity can be improved by a factor of about 100 over that in previous flights,[1,2] then any supernova explosion which produced a neutron star during the last one or two thousand years could be detected. There could have been up to about 20 such events. Much older neutron stars with $T_e \lesssim 10^6$ °K would be difficult to detect as x-ray sources because of strong absorption at $\lambda \lesssim 30$ Å by interstellar gas. Because of its small radius the energy radiated directly in visible light by a neutron star ($L_{vis} \gtrsim 4\times10^{27}$ erg/sec $= 10^{-6}L_\odot$) is far too small to be detected in most cases. If the star is situated in a gas cloud (density $\gtrsim 20$ atoms per cm^3) it will ionize a surrounding region (of radius ~ 0.2 light years for $T_e = 10^6$ °K; ~ 0.6 light years for $T_e = 10^7$ °K). Recombination of the ions results in visible light and the region's surface brightness may be (just barely) detectable by terrestrial optical telescopes.

TABLE 1. *Emission characteristics*

Core temp. T_c (°K)	Total thermal energy E_t (ergs)	Neutrino loss rate $L_\nu'_{pl}+L_\nu'_{UR}$ (ergs/sec)	Photon luminosity L_{ph} (ergs/sec)	Surface temp. T_e (°K)	Cooling time τ (years)	Optimum wavelength λW (Å)
2×10^9	1.6×10^{48}	4×10^{39}	2.3×10^{37}	1.3×10^7	10	2.4
1×10^9	4×10^{47}	5×10^{36}	8×10^{36}	1.0×10^7	1×10^3	3.1
5×10^8	1×10^{47}	8×10^{33}	3×10^{36}	8×10^6	1×10^3	3.9
2×10^8	1.6×10^{46}	1×10^{29}	3×10^{35}	4.5×10^6	1.5×10^3	6.8

The chemical composition of the atmosphere of a neutron star is uncertain at the moment, partly because the density scale-heights are so small and diffusion[11] can be much more important than in ordinary stars; H and He from the surface can diffuse in a few years to deeper layers ($T \gtrsim 5\times10^8$ °K) where they are burned up rapidly and these two elements should be completely absent in neutron stars. The equivalent diffusion time for ^{12}C is about 100 years and C is probably strongly (but not completely) depleted in a neutron star about 1000 years old, O and Ne depleted slightly, and any Mg or heavier nuclei (present at the formation of the neutron star) are unburnt. Diffusive separation can lead in a few hundred years to a strongly enhanced abundance in the photosphere of the lightest nuclei (O, Ne, or Mg), unless convection restores mixing, in which case the metals near ^{56}Fe could be most abundant.

We want to thank Dr. A. G. W. Cameron, Dr. G. W. Clark, Dr. F. J. Dyson, and Dr. D. Morton for discussions, and Miss Jane Keston for help in numerical work. One of us (E. E. S.) is indebted to the National Academy of Science for a fellowship.

REFERENCES

1. H. Gursky, R. Giacconi, F. R. Paolini, and B. Rossi, Phys. Rev. Letters **11**, 30 (1963).

2. S. Bowyer, E. T. Byron, T. A. Chubb, and H. Friedman (to be published).

3. Two Arab astronomers, Haly and Albumazar, observed a very bright star, located near this x-ray source, for four months around A.D. 827. They report a brightness "comparable with the quarter moon," which is brighter (and therefore probably closer) than the original Crab explosion. However, no Chinese or Japanese records have been found yet to corroborate this event.

4. H. Y. Chiu, Ann. Phys. **26**, 364 (1964).

5. Such estimates were first made by R. Stabler (Ph.D. thesis, Cornell University, 1960, unpublished), but he used too large a value for the effective opacity.

6. S. Chandrasekhar, *An Introduction to Stellar Structure* (Dover Publications, New York, 1951), Chap. 10.

7. This estimate is essentially the result of a dimensional analysis and could be in error by several orders of magnitude.

8. B. Adams, M. Ruderman, and C. -H. Woo, Phys. Rev. **129**, 1383 (1963). Other neutrino processes give a rate a few orders of magnitude below those discussed in this paper.

9. For fixed T_c, our L_{ph} could be in error by factors of 4 or 5 and T_s by factors of about 1.5. L_ν (and hence the cooling time τ) is uncertain by larger factors. D. Morton (unpublished) has made independent estimates of T_s (but not of τ) which agree well with ours.

10. G. W. Clark, M. Oda, and P. Morrison (unpublished).

11. E. Schatzman, *White Dwarfs* (North-Holland Publishing Company, Amsterdam, 1957).

ESCAPE OF STARS DURING THE COLLISION OF TWO GALAXIES*

GEORGE CONTOPOULOS AND GEORGE BOZIS†

An approximate theory is developed to find out whether a star, moving initially on the plane of symmetry of a galaxy, escapes during the collision with another galaxy of small dimensions. The escaping stars are inside an ellipse on the plane of symmetry of the galaxy. Numerical calculations in a Bottlinger model of a galaxy give results of the same order of magnitude. The total amount of the escaping matter is only a few per cent of the total mass of the galaxy, if the collision velocity is of the order of 2000 km/sec. The behavior of the escaping and non-escaping stars is described. The escaping stars do not follow the colliding galaxy and do not form a bridge or link between the two galaxies.

I. ESCAPING STARS

The purpose of this paper is to consider the effects of the collision of two galaxies on stars moving initially on the plane of symmetry of one galaxy.

If we assume a relative velocity of collision of about 2000 km/sec (Spitzer and Baade 1951), the collision lasts about 10^7 years. During that time the perturbations due to the ellipticity of the first galaxy are insignificant for stellar orbits close to its plane of symmetry (Contopoulos and Bozis 1962; hereafter referred to as "Paper I"). Therefore in order to study approximately the effects of the collision on these orbits we may use a spherical model of a galaxy. The colliding galaxy is also assumed to be spherical but of small dimensions. We will see that, if its diameter is less than 2 kpc, its effects are essentially the same as those of a point mass. In the numerical examples below the mass of both galaxies is assumed to be about 10^{11} $M_\odot$, and the radius of the first galaxy is taken 10 kpc.

In Paper I it has been proved that, if two such galaxies collide with a relative velocity $v=2000$ km/sec, their relative motion may be considered as uniform; in fact the change of velocity is of the order of 30 km/sec and the change of the direction of the velocity is less than 5°.

Further, during the collision, a star moving in a circular orbit on the plane of symmetry of the first galaxy O with radius 8 kpc describes only about 1/20 of a revolution, i.e., about 18°; therefore, we may consider, in a rough approximation, the unperturbed orbit of S to be a straight line, $x'=x_0'$, perpendicular to the axis x', which is directed away from the center O (Fig. 1), and parallel to the axis y'. Under these assumptions the perturbation may be considered as a two-body encounter that lasts as long as the collision itself.

The galaxy O' is assumed to move along the axis z' with uniform velocity $\bar{v}$, reaching the point O_1 on the plane xOy at $t=t_0$. The initial velocity of the star is $\bar{v}_e$; if the perturbing galaxy was not present, S would reach the axis x' at $t=t_1$.

The velocities v and v_e are of the order of 2000 km/sec and 200 km/sec, respectively; therefore, during 10^7 years the galaxy O' moves approximately from $z'=10$ kpc (initial position) to $z'=-10$ kpc (final position), while S moves about 2 kpc along a parallel to the y'-axis. The influence of O' on S when the distance OO' is larger than 10 kpc is very small. In fact, as we will see below, the escaping stars are inside an ellipse whose dimensions are of the order of 4 kpc. Therefore, their minimum distance from O' is about

*Published in *The Astrophysical Journal*, 139(4):1239–1250, May 15, 1964.

†Goddard Space Flight Center and Yale University Observatory.

2 kpc or less. When a point comes from infinity to a distance of 10 kpc the relative velocity changes only by about 1 per cent and the change of its direction is of the order of 2 percent of the angle between the asymptotes. In fact, if v_∞ is the relative velocity before the encounter and v the relative velocity at a distance r from O', then

$$v^2 = \frac{2Gm}{r} + v_\infty^2,$$

hence

$$\Delta v = v - v_\infty \simeq \frac{Gm}{r v_\infty},$$

and for $r = 10$ kpc, $v_\infty = 2000$ km/sec, and $m = 10_{11} M\odot$,

$$\Delta v = 22 \text{ km/sec}.$$

The angle 2ψ between the initial and final velocities is given by $\tan \psi = p/q$, where $x_1^2/p^2 - y_1^2/q^2 = 1$ is the equation of the relative orbit. As q is of the order of 2 kpc, and $p = Gm/v_\infty^2 \simeq 0.1$ kpc, we have $p/q \simeq 0.05$.

The tangent at a point (x_1, y_1) whose distance from the focus is 10 kpc makes an angle ψ' with the axis y_1, where

$$\tan \psi' = \tan \psi \sqrt{\left(1 - \frac{p^2}{x_1^2}\right)},$$

and $[x_1 - \sqrt{(p^2+q^2)}]^2 + y_1^2 = 100$.

Then the deviation $(\psi - \psi')$ is given by

$$\tan(\psi - \psi') = \frac{(p/q)[10 + p - \sqrt{(100+20p-q^2)}]}{10 + p + (p^2/q^2)\sqrt{(100+20p-q^2)}}$$

$$\simeq \frac{p}{q} \cdot \frac{0.05q^2}{10} = 0.005\, pq.$$

Hence

$$\frac{\psi - \psi'}{\psi} \simeq 0.005 q^2,$$

and for $q = 2$, this is 2 per cent. Therefore we may consider the initial and final relative velocities of S with respect to O as being equal to the corresponding velocities in a two-body encounter at infinite distances.

The initial relative velocity of the star S with respect to the galaxy O' is $\bar{v}_0 = \bar{v}_e - \bar{v}$; the corresponding final velocity $\bar{v}_0'$ is equal to $\bar{v}_0$ in measure but not in direction. The final velocity of S with respect to the galaxy O is $\bar{v}_f = \bar{v}_0' + \bar{v}$; if v_f is greater than the velocity of escape v_e, the star escapes from the galaxy O.

In what follows we use a frame of reference where O' is at rest. The galaxy O is moving with respect to O' with velocity $-\bar{v}$ along the axis z' that makes an angle e with the plane xOy (Fig. 1).

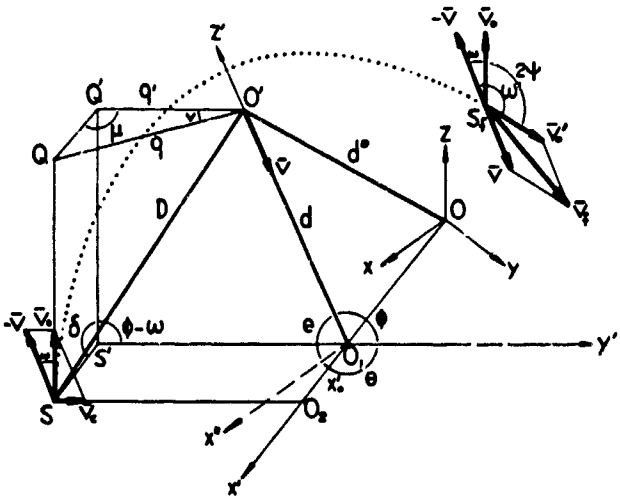


FIGURE 1.—A galaxy O' of mass m is moving downward along the axis z' toward the point O_1 in the plane of symmetry of the galaxy O, of mass M. The axis Ox is parallel to the projection O_1x'' of O_1z' on the plane of symmetry. The distances of O' from O, O_1, and S are d^*, d, and D, respectively.

The projection of z' on the plane xOy makes an angle θ with O_1y', and φ is the angle $z'O_1y'$. Then

$$\cos \varphi = \cos e \cos \theta. \tag{1}$$

At the initial time t we have

$$y' = O_1S' = v_e(t - t_1) \tag{2}$$

and

$$z' = O_1O' = v(t_0 - t). \tag{3}$$

If ω is the angle between $\bar{v}_0$ and $-\bar{v}$, we have

$$\frac{\sin \omega}{v_e} = \frac{\sin \varphi}{v_0} = \frac{\sin(\varphi - \omega)}{v} \tag{4}$$

and

$$v_0^2 = v^2 + v_e^2 + 2\, v\, v_e \cos \varphi. \tag{5}$$

The corresponding angle between $\bar{v}_0'$ and $-\bar{v}$ is ω', and the angle between $\bar{v}_0$ and $\bar{v}_0'$ is 2φ, which

is approximately equal to the angle between the asymptotes of the hyperbolic orbit. Therefore

$$\tan \psi = \frac{Gm}{qv_0^2}, \tag{6}$$

where q is the impact parameter.

The impact parameter is found as follows. Let SS' be parallel to the axis x' and $S'Q'$ parallel to $\bar{v}_0$, $O'Q$ and $O'Q'$ perpendicular to SQ and $S'Q'$, respectively; $S'Q'$ is on the plane $S'O_1O'$ and QQ' is perpendicular to SQ and $S'Q'$. The angle between the planes $QQ'S'$ and $O'Q'S'$ is μ, and the angle between the planes $QO'S$ and $Q'O'S'$ is ν. The angle $Q'S'O_1$ is equal to $\varphi-\omega$, therefore

$$\begin{aligned} q' &= -y' \sin(\varphi-\omega) - z' \sin\omega \\ &= -\left(y' + z'\frac{v_c}{v}\right)\sin(\varphi-\omega) \\ &= v_c(t_1-t_0)\sin(\varphi-\omega) = -y_0' \sin(\varphi-\omega), \end{aligned} \tag{7}$$

where

$$y_0' = v_c(t_0-t_1) \tag{8}$$

is the y' coordinate of S at $t=t_0$. The term q' is positive if $y_0'<0$, and negative if $y_0'>0$. Further

$$S'S = O_1O_2 = x_0', \qquad Q'Q = x_0' \sin\delta, \qquad \text{and } q^2 = q'^2 + x_0'^2 \sin^2\delta - 2q'x_0' \sin\delta\cos\mu. \tag{9}$$

From the trihedral angle $S'Q'SO_1$ we find

$$\cos\delta = \sin(\varphi-\omega)\cos I \tag{10}$$

and

$$\sin\delta\cos\mu = -\cos(\varphi-\omega)\cos I \tag{11}$$

where I is the angle between the planes $y'O_1z'$ and $y'O_1x'$, hence

$$\cos I = \cot\varphi \tan\theta = \frac{\cos e \sin\theta}{\sin\varphi}. \tag{12}$$

Therefore,

$$\begin{aligned} q^2 = x_0'^2[1-\sin^2(\varphi-\omega)\cos^2 I] + y_0'^2\sin^2(\varphi-\omega) \\ -2x_0'y_0' \sin(\varphi-\omega)\cos(\varphi-\omega)\cos I. \end{aligned} \tag{13}$$

A star escapes if

$$v_f^2 = v^2 + v_0^2 - 2vv_0\cos\omega' \geq v_e^2. \tag{14}$$

This is possible only if $v+v_0 \geq v_e$. Then we can set

$$\begin{aligned} -1 \leq \frac{v^2+v_0^2-v_e^2}{2vv_0} = \cos\omega_0' < \frac{v^2+v_0^2-v_c^2}{2vv_0} \\ = \frac{v+v_c\cos\varphi}{v_0} = \cos\omega < 1, \end{aligned} \tag{15}$$

because $v_c<v_e$. Therefore $O<\omega_0'\leq\pi$, and if a star escapes we have

$$\cos\omega' \leq \cos\omega_0'. \tag{16}$$

The angle ω' can be found from the trihedral angle $S_f\bar{v}_0\bar{v}_0'(-\bar{v})$;

$$\cos\omega' = \cos 2\psi\cos\omega + \sin 2\psi \sin\omega\cos(\pi-\nu), \tag{17}$$

where $(\pi-\nu)$ is the angle between the plane $S_f\bar{v}_0\bar{v}_0'$ (i.e., the plane $O'QS$ of the relative orbit) and the plane $S_f\bar{v}_0(-\bar{v})$ (which is parallel to the plane $O'Q'S'$); thus $(\pi-\nu)$ is the complement of the angle $QO'Q'$, and

$$\begin{aligned} \cos\nu = \frac{q'-x_0'\sin\delta\cos\mu}{q} = \tfrac{1}{9}[-y_0'\sin(\varphi-\omega) \\ +x_0'\cos(\varphi-\omega)\cos I]. \end{aligned} \tag{18}$$

If we use relations (6) and (17) in inequality (16) we get

$$\begin{aligned} \{x_0'^2[1-\sin^2(\varphi-\omega)\cos^2 I] + y_0'^2\sin^2(\varphi-\omega) \\ -2x_0'y_0\sin(\varphi-\omega)\cos(\varphi-\omega)\cos I\} \\ (\cos\omega - \cos\omega_0') + \frac{2Gm}{v_0^2}\sin\omega\sin(\varphi-\omega)\,y_0' \\ -\frac{2Gm}{v_0^2}\sin\omega\cos(\varphi-\omega)\cos I\,x_0' \\ \leq \frac{G^2m^2}{v_0^4}(\cos\omega+\cos\omega_0'), \end{aligned} \tag{19}$$

i.e., the escaping stars are inside the ellipse

$$\begin{aligned} x_0'^2(1-\sin^2(\varphi-\omega)\cos^2 I) + y_0'-y_c')^2\sin^2(\varphi-\omega) \\ -2x_0'(y_0'-y_c')\sin(\varphi-\omega)\cos(\varphi-\omega)\cos I \\ = \frac{G^2m^2\sin^2\omega_0'}{v_0^4(\cos\omega-\cos\omega_0')^2}, \end{aligned} \tag{20}$$

whose center is $(x_0'=0,\ y_c')$, where

$$\begin{aligned} y_c' &= -\frac{Gm\sin\omega}{v_0^2\sin(\varphi-\omega)(\cos\omega-\cos\omega_0')} \\ &= -\frac{2Gmv_c}{(v_e^2-v_c^2)v_0}. \end{aligned} \tag{21}$$

The axes of this ellipse form an angle γ with the axes x' and y', where

$$\tan\gamma = -\tan(\varphi-\omega)\cos I = \frac{-\tan(\varphi-\omega)\cos e\sin\theta}{\sin\varphi}. \quad (22)$$

For $0<\theta<\pi$, the angle γ is positive (measured counterclockwise) if $\varphi-\omega>\pi/2$ and negative if $\varphi-\omega<\pi/2$. For $-\pi<\theta<0$, i.e., if the projection of z' on the plane $x'O_1y'$ is on the negative side of the axis x', the angle γ is positive if $\varphi-\omega<\pi/2$ and negative if $\varphi-\omega>\pi/2$. For $\theta=0$ or $\theta=\pi$, the angle $\gamma=0$.

If $\theta=\pi/2$ then $\varphi=\pi/2$, $I=e$, and equations (20)–(22) become

$$x_0'^2(1-\cos^2\omega\cos^2 e)+\cos^2\omega(y_0'-y_c')^2 - 2x_0'(y_0'-y_c)\sin\omega\cos\omega\cos e = \frac{G^2m^2\sin^2\omega_0'}{v_0^4(\cos\omega-\cos\omega_0')^2}, \quad (23)$$

$$y_c' = -\frac{Gm\tan\omega}{v_0^2(\cos\omega-\cos\omega_0')}, \quad (24)$$

and

$$\tan\gamma = -\cot\omega\cos e. \quad (25)$$

From equation (20) we find easily the area of the ellipse

$$E=\pi a_x'a_y' = \frac{\pi G^2m^2\sin^2\omega_0'}{v_0^4(\cos\omega-\cos\omega_0')^2\sin(\varphi-\omega)\sin I}. \quad (26)$$

When $v\to\infty$, then $E\to0$. If $I=\pi/2$, then the ellipse becomes

$$x_0'^2+(y_0'-y_c')^2\sin^2(\varphi-\omega) = \frac{G^2m^2\sin^2\omega_0'}{v_0^4(\cos\omega-\cos\omega_0')^2}. \quad (27)$$

The semi-axes of the ellipse are then

$$a_x = \frac{Gm\sin\omega_0'}{v_0^2(\cos\omega-\cos\omega_0')} = \frac{2Gm\sin\omega_0'\sin(\varphi-\omega)}{(v_e^2-v_c^2)\sin\varphi}, \quad (28)$$

and

$$ay = \frac{Gm\sin\omega_0'}{v_0^2(\cos\omega-\cos\omega_0')\sin(\varphi-\omega)} = \frac{2Gm\sin\omega_0'}{(v_e^2-v_c^2)\sin\varphi}. \quad (29)$$

If $v\to\infty$, then $a_x\to0$ and $a_y\to0$.

The ratio of the axes is

$$a_x/a_y=\sin(\varphi-\omega)=v\sin\varphi/v_0, \quad (30)$$

and the area of the ellipse is

$$E=\pi a_xa_y=\frac{4\pi G^2m^2\sin^2\omega_0'\sin(\varphi-\omega)}{(v_e^2-v_c^2)^2\sin^2\varphi}. \quad (31)$$

Table 1 gives the values of a_x, ay, E, y_c', and ω for the model considered in Section II with $e=\pi/2$, $\varphi=\pi/2$, $v=2000$ km/sec $=2.0454$ kpc/10^6 years, and different minimum distances $r=\sqrt{(\alpha^2+\beta^2)}$ of the colliding galaxy from the center.

TABLE 1

r (kpc)	a_x (kpc)	a_y (kpc)	E (kpc)^2	y_c' (kpc)	ω (degrees)
4	1.41	1.42	6	−0.7	6.22
6	1.94	1.95	12	−1.2	6.55
8	2.41	2.42	18	−1.6	6.19
10	2.80	2.81	25	−1.9	5.74

If the effective radius of the galaxy R_G is about 10 kpc, the above area is between 2 per cent (for $r=4$ kpc) and 8 per cent (for $r=10$ kpc) of the total area.

The proportion of the escaping stars is roughly $E\rho/\pi R_G\bar{\rho}^2$ where $\bar{\rho}$ is the mean density of the galaxy and ρ is the mean density in the region of escape.

Table 2 gives the values of φ, E, y_c', ω for ($e=75°$, $\theta=0°$), ($e=45°$, $\theta=0°$) and ($e=45°$, $\theta=45°$).

The value of γ in the first two cases is zero, and in the last case varies between $-39°.8$ and $-39°.1$.

It is seen that the values of E and y_c' in all three cases are quite near the corresponding values of Table 1. Therefore, the amount of escaping matter from the plane of symmetry of a galaxy does not depend very much on the direction of the collision.

We do not give the values of E, etc., for $r<4$ because near the center the assumption of a binary collision between the star and the approaching galaxy is not even approximately satisfied.

TABLE 2

r (kpc)	$e=75°, \theta=0°, \varphi=75°$			$e=45°, \theta=0°, \varphi=45°$			$e=45°, \theta=45°, \varphi=60°$		
	E (kpc)^2	y_c' (kpc)	ω (degrees)	E (kpc)^2	y_c' (kpc)	ω (degrees)	E (kpc)^2	y_c' (kpc)	ω (degrees)
4	6	−0.7	5.85	7	−0.7	4.10	7	−0.7	5.12
6	11	−1.2	6.15	11	−1.1	4.29	13	−1.2	5.37
8	17	−1.6	5.82	17	−1.5	4.08	20	−1.5	5.09
10	23	−1.9	5 41	23	−1.8	3.80	28	−1.9	4.74

If the diameter of the colliding galaxy is 2 kpc (or less), its cross-section is small, in general, in comparison with the area E of the region of escape; therefore, the amount of escaping matter is essentially the same as in the case of a point mass. If the diameter is larger than 2 kpc, the amount of escaping matter will be smaller.

II. NUMERICAL CALCULATIONS

In order to calculate numerically the orbits of stars during the collision of two galaxies we have chosen a simple model for the first galaxy, in which the forces per unit mass are given by Bottlinger's formula

$$F=\frac{ar}{1+br^3}. \tag{32}$$

This formula has been applied by Lohmann (1953; 1954*a*, *b*) to our Galaxy, as well as to M31 and M33, and by Kerr and de Vancouleurs (1956) to the Large Magellanic Cloud. It is discussed, together with other models, by de Vaucouleurs (1959) and Perek (1962).

The values of a and b were chosen to represent approximately the force function F given by Schmidt (1956) for the plane of symmetry of our Galaxy, namely,

$$a=0.0055(10^6 \text{ years})^{-2}, \quad b=0.012 \text{ kpc}^{-3}.$$

The corresponding potential function is

$$U=-\frac{a}{bk}\left\{\frac{1}{6}\ln\left[\frac{r^2-kr+k^2}{(r+k)^2}\right]-\frac{1}{3}\cot^{-1}\left(\frac{2r-k}{k\sqrt{3}}\right)\right\}, \tag{33}$$

where $k=b^{-1/3}$ and $U_\infty=0$.

In this model the mass of the Galaxy is

$$M=\frac{a}{bG}\simeq 1.02\times 10^{11} M_\odot. \tag{34}$$

The circular velocity is

$$v_c=\sqrt{(Fr)}, \tag{35}$$

and the escape velocity is

$$v_e=\sqrt{(2U)}. \tag{36}$$

Table 3 gives the values of v_c and v_e for different distances r from the center in the above field.

Now we consider the following situation (Fig. 1): A point mass $O'(m)$ goes through a spherical galaxy $O(M)$, represented by the force function (32). This force function is assumed to be stationary; this is justified if the colliding galaxy does not come very close to the center of the galaxy O, because then the amount of escaping

TABLE 3

r (kpc)	v_c (kpc/10^6 years)	v_e (kpc/10^6 years)	r (kpc)	v_c (kpc/10^6 years)	v_e (kpc/10^6 years)
4	0.2231	0.4294	14	0.1783	0.2549
6	.2348	.3750	16	.1676	.2388
8	2220	.3321	18	.1584	.2253
10	.2057	.2997	20	0.1506	0.2138
12	0.1909	0.2748			

matter is rather small. In fact the density corresponding to the force function (32) is

$$\rho=\frac{3a}{4\pi G(1+br^3)^2},$$

and the mean density is

$$\bar{\rho}=\frac{3a}{4\pi GbR_G{}^3},$$

so that if we take a mean radius of the galaxy $R_G=10$ kpc, then the proportion of the escaping stars is

$$\frac{E\rho}{\pi R_G{}^2\bar{\rho}}\simeq\frac{0.04E}{(1+0.012r^3)^2};$$

for $r=10$ kpc this is about 0.6 per cent while for $r=4$ kpc this is 8 percent.

The motion of O' is assumed to be uniform along the line $O'O_1$ where O_1, (α, β) lies on the plane xOy. The plane xOz is taken parallel to the direction $O'O_1$ and e is the angle between this line and the plane xOy. The coordinates of the star S are x, y, z, and d, d^*, D are the distances of O' from O_1, O, and S, respectively.

The unperturbed orbit of S is, in general, a rosette on the plane xOy. In this paper, however, we consider only the case of circular orbits around O.

The perturbation acting on the star is equal to the acceleration of S due to the galaxy O' minus the acceleration of the galaxy O, taken as a whole. As the attraction between O and O' is $mF(d^*)$, the acceleration of O is $(m/M)F(d^*)$. Thus the equations of motion of S are

$$\frac{d^2x}{dt^2}=-F(r^*)\frac{x}{r^*}+\frac{Gm}{D^3}(\alpha+d\cos e-x)-F(d^*)\frac{m}{M}\frac{(\alpha+d\cos e)}{d^*},\quad(37)$$

$$\frac{d^2y}{dt^2}=-F(r^*)\frac{y}{r^*}+\frac{Gm}{D^3}(\beta-y)-F(d^*)\frac{m}{M}\frac{\beta}{d^*},\quad(38)$$

and

$$\frac{d^2z}{dt^2}=-F(r^*)\frac{z}{r^*}=\frac{Gm}{D^3}(d\sin e-z)-F(d^*)\frac{m}{M}\frac{d\sin e}{d^*},\quad(39)$$

where

$$r^*=\sqrt{(x^2+y^2+z^2)},\quad(40)$$

$$d^*=\sqrt{[(\alpha+d\cos e)^2+\beta^2+(d\sin e)^2]},\quad(41)$$

$$D=\sqrt{[(\alpha+d\cos e-x)^2+(\beta-y)^2+(d\sin e-z)^2]},\quad(42)$$

and

$$d=d_0-vt.\quad(43)$$

The numerical values used are the following: $m/M=1$, $MG=a/b$, $v=2000$ km/sec $=2.0454$ kpc/10^6 years, $d_0=100$ kpc, $\alpha=8$ kpc, $\beta=0$, and $e=90°$ (case A), or $e=45°$ (case B). The initial conditions for $t=0$ were taken in such a way that the star, moving in the unperturbed circular orbit, would reach an axis forming an angle g with the positive x-axis (g is positive if measured counterclockwise), when the colliding galaxy reaches the point O_1 (α, β), namely, at about $t=48.89\times10^6$ years.

The greatest part of the calculations was made with the IBM 7090 computer of NASA's Institute for Space Studies, New York. The Runge-Kutta integration method was used with a step 0.1 (10^6 years) and the orbits were calculated for 200×10^6 years. At $t=200\times10^6$ years the colliding galaxy is at a distance of about 300 kpc from the galaxy O and its influence is then negligible. In some cases, when the stars escaped rather fast from the galaxy O, the calculations were stopped earlier.

The initial z coordinates and velocities were all assumed to be zero. Figure 2*A* shows the positions that the stars would have in case A ($e=90°$) at $t=48.89\times10^7$ years (when the colliding galaxy reaches the point O_1), if there was no perturbation. The polar coordinates of these positions are (r, g); therefore the initial polar coordinates, for $t=0$, are

$$r_0=r$$

and

$$g_0\equiv g-48.89\frac{v_c}{r}\times57.^\circ2957795$$

The dots represent stars that are not escaping. The small circles represent escaping stars. The region of escaping stars is approximately a circle of diameter 5 kpc. From Table 1 we see that the ellipse of escaping stars for $r=\sqrt{(\alpha^2+\beta^2)}=8$ kpc is very nearly a circle of diameter 4.8 kpc; the

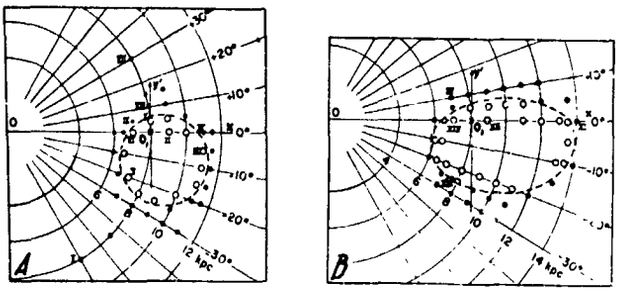


FIGURE 2.—The position (r, g in polor coordinates) of escaping (open circles) and non-escaping (solid circles) stars in their unperturbed orbits when the colliding galaxy is going through the point 0_1 ($\alpha=8$ kpc, $\beta=0$). In case A, $e=90°$; in case B, $e=45°$.

agreement with the calculations is very good in this case.

The center of the circle in Figure 2*A* is not at $x_c'=0$, as it was expected theoretically, but at about $x_c'=1$ kpc, i.e., further out than the point O_1. On the other hand, the y' coordinate of the center is about $y_c'=-1.5$ kpc, i.e., very close to the calculated value. Therefore, the only discrepancy between theory and calculations is that the actual region of escape is further out than expected; this is due, of course, to the fact that the unperturbed orbits are curved and not straight lines as they were assumed in Section I. The only consequence of this effect is that the region of escape actually includes less dense areas, therefore, the amount of escaping matter is smaller.

In case B ($e=45°$) the calculated orbits are marked in Figure 2*B*. The region of escape is approximately an ellipse of dimensions $2a=8$ kpc, $2b=5$ kpc, and the area $E=31$ kpc^2. This is bigger than expected on the basis of the simple theory developed in Section I. However, even in this case the area E is only 10 percent of the effective area in the plane of symmetry of the galaxy O, and as it includes mainly the outer parts of the galaxy the mass of the escaping matter is less than 2 per cent of the total matter near the plane xOy.

The individual orbits of stars during the collision show a number of interesting characteristics. Figures 3 and 4, copied from plots made by the NASA computer, show the main types of orbits encountered, in the case $e=90°$.

A common characteristic of all the orbits is that the perturbations are very small until the colliding galaxy goes through the galaxy O. Until that moment the orbits are practically circles on the xy-plane. After the collision the orbits may be divided into three types: (*a*) orbits of stars that remain far from the region of escape; (*b*) "escaping" orbits; and (*c*) orbits of stars near the region of escape.

If the star is far from the region of escape, its orbit is not changed much by the collision. It remains almost circular, and the z-motion is very small. Such is the case of orbit I ($r=8$ kpc, $g=-60°$).

The orbits of stars inside or near the region of escape can be divided into the following groups (cf. Figs. 2*A* and 3*A*-3*D*): (i) stars with $g\leq 0°$ and $r\geq 8$, or r smaller than 8, but g not near 0°; (ii) stars with $r<8$ and $g\leq 0°$ but g near 0°; and (iii) stars with $g>0°$. We discuss these groups separately.

i) The stars with $g\leq 0°$ that are inside or near the region of escape are accelerated toward the positive y'-axis as the colliding galaxy goes through the galaxy O. If $r>8$ kpc the orbits are deflected to the left. This deflection is bigger whenever the approach to the colliding galaxy is closer. For example, in case II ($r=9$, $g=0°$) the star is ejected abruptly to the left and down, while in case III ($r=11$, $g=-5°$) the star escapes but with smaller velocity. In case IV ($r=11$, $g=0°$) the star does not escape, but its orbit does not differ much from the escaping orbits for $t=200\times 10^6$ years. Only the fact that its velocity is less than the escape velocity makes it sure that it does not escape, but forms a very elongated orbit around the center of the galaxy O.

All orbits near the boundary of the region of escape from ($r=11$, $g=0°$) to ($r=8$, $g=-25°$) are similar to cases III and IV. The y component of the final velocity is negative for large r, while it is positive for r near 8.

The same general behavior is shown by the orbits with $r<8$ but g not close to 0°. For example, case V ($r=7$, $g=-20°$) represents an escaping orbit; the star escapes toward the positive y direction and downward.

ii) The orbits with $r<8$ and g near 0° show an abrupt deviation to the right. This is most marked in case VI ($r=7$, $g=0°$) when the star is ejected abruptly to the right and downward.

iii) The orbits with positive g are usually not escaping. In general, the additional attraction

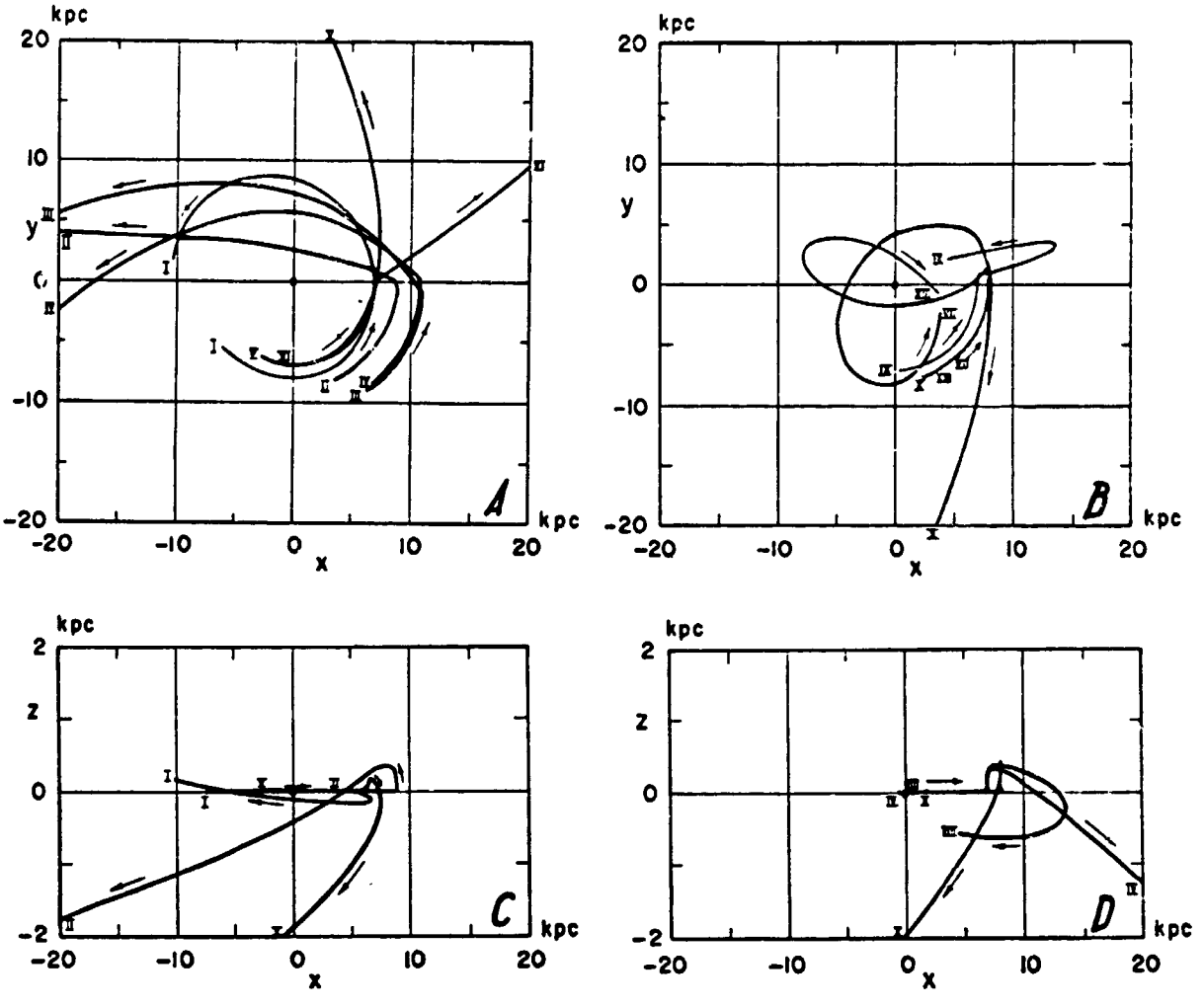


FIGURE 3.—Characteristic cases of orbits during a collision through $O_1(\alpha=8$ kpc, $\beta=0)$ with $e=90°$; I ($r=8$ kpc, $g=-60°$; non-escaping; II ($r=9$ kpc, $g=0°$; escaping); III ($r=11$ kpc, $g=-5°$; escaping); IV ($r=11$ kpc, $g=0°$; non-escaping); V ($r=7$ kpc, $g=-20°$; escaping); VI ($r=7$ kpc, $g=0°$; escaping); VII ($r=8$ kpc, $g=+30°$; non-escaping); VIII ($r=8$ kpc, $g=+10°$; non-escaping); IX ($r=7$ kpc, $g=+5°$; non-escaping); X ($r=8$ kpc, $g=+5°$; escaping). A, B give the projections of the orbits on the xy-plane, while C, D give the projections of some orbits on the xz-plane.

of the colliding galaxy makes the orbits more elongated. Such is the case VII ($r=8$, $g=+30°$), that is rather far from the region of escape. In case VIII ($r=8$, $g=+10°$), however, the deviation is so large, that the orbit reverses the sense of rotation, although the star does not escape. A large deviation, but to the right, is shown in case IX ($r=7$, $g=+5°$). Finally, case X ($r=8$, $g=+5°$) shows an abrupt reversal of the motion of the star and ejection toward the negative y-axis. Such a behavior, however, is quite exceptional.

From the above discussion we conclude that the greatest proportion of the ejected stars escapes toward the left (between the positive y-axis and the negative x-axis) and downward. The stars do not follow the colliding galaxy after the collision and they do not form a bridge or link between the two galaxies.

In case B ($e=45°$) the orbits are similar to those of case A, except as regards the z coordinate (cf. Figs. 2*B* and 4).

i) The stars with $r\geq 8$, $g\leq 0°$ have comparatively large motions upward as the colliding galaxy approaches, which are changed to downward motions later, as in case XI ($r=14$, $g=0°$). This is because in these cases the approaching galaxy comes rather close to these stars before it reaches the plane xOy at O_1.

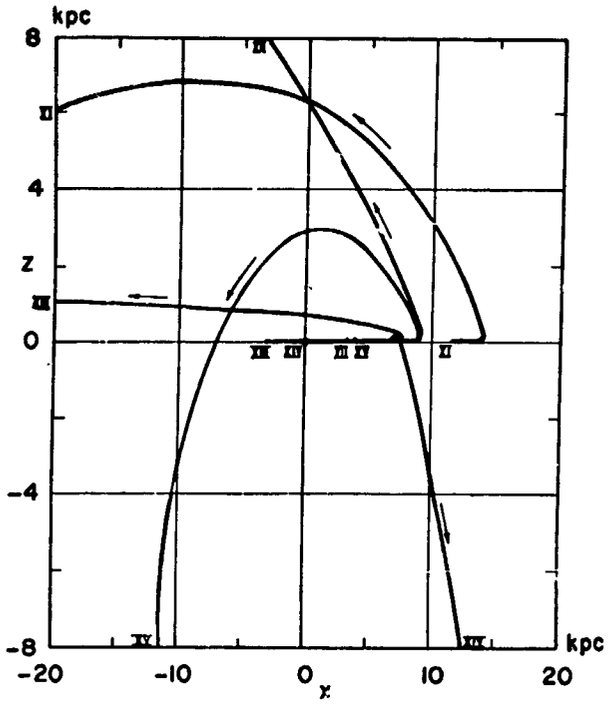


FIGURE 4.—Characteristic cases of orbits during a collision through O_1 ($\alpha=8$ kpc, $\beta=0$) with $e=45°$; XI ($r=14$ kpc, $g=0°$; escaping); XII ($r=9$ kps, $g=0°$; escaping); XIII ($r=8$ kpc, $g=-25°$; escaping); XIV ($r=7$ kpc, $g=0°$; escaping); XV ($r=9$ kpc, $g=+10°$; non-escaping). Only the projections on the xz-plane are given.

If r is near 8 kpc but $r>8$, and $g\simeq 0°$ the upward motions are so fast that the stars eventually escape upward and to the left (case XII, $r=9$, $g=0°$). If r is near 8 kpc but g not near 0° the upward motions are small (case XIII, $r=8$, $g=-25°$), and if $r<8$ kpc the stars go downward, showing only a small upward kink, as in case V. The downward motions in this case can be explained easily, because as the approaching galaxy comes from the right, it does not influence appreciably the stars to the left of O_1 until it goes beyond the point O_1 and comes rather close to them.

ii) The stars with $r<8$ but close to 8 and g near 0° show an abrupt motion downward (case XIV, $r=7$, $g=0°$).

iii) The orbits of stars with $g>0°$ show all the peculiarities mentioned in case A. The z motions are of the form of case XV ($r=9$, $g=+10°$) if $r>8$, and of the form of case XIV, if $r<8$, but with less steep descent.

We conclude that if a galaxy goes through the plane of symmetry of another galaxy at an angle different from 90° it causes the stars to acquire appreciable z velocities. The escaping stars move usually downward, but they do not follow the colliding galaxy in its course.

The final conclusion of this paper is that, during a collision of two galaxies with rather large relative velocity, only a small fraction of matter escapes and this does not form any bridge or link between the galaxies. Therefore the observed bridges between pairs of galaxies probably are not due to collisions with relative velocities of the order of 2000 km/sec. It is possible, however, that much slower collisions, not considered in this paper, may cause much larger changes in the colliding galaxies and eventually explain the formation of bridges between them.

We want to thank Dr. R. Jastrow and the NASA Institute for Space Studies, where most of the numerical work has been done. Some preliminary calculations were made by Dr. C. Goudas in Manchester, and at the Yale Computing Center, by kind permission of Dr. M. Davis; we express our thanks to both of them. The theoretical work has been done at the Astronomical Department of the University of Thessaloniki, Greece, and was supported by the Royal Hellenic Research Foundation.

REFERENCES

CONTOPOULOS, G., and BOZIS, G. 1962, *Contr. Astr. Dept. Univ. Thessaloniki*, No. 7 (Paper I).
KERR, F. J., and VAUCOULEURS, G. DE. 1956, *Australian J. Phys.*, 9, 90.
LOHMANN, W. 1953, *Zs. f. Ap.*, 33, 186.
LOHMANN, W. 1954*a*, *ibid.*, 35, 90.
LOHMANN, W. 1954*b*, *ibid.*, p. 159.
PEREK, L. 1962, *Advances in Astronomy and Astrophysics* (New York and London: Academic Press), 1, 165.
SCHMIDT, M. 1956, *B.A.N.*, 13, 15.
SPITZER, L., and BAADE, W. 1951, *Ap. J.*, 113, 413.
VAUCOULEURS, G. DE. 1959, *Hdb. d. Phys.* (Berlin: Springer-Verlag), 53, 311.

DYNAMICS AND EVOLUTION OF CLOSE BINARY SYSTEMS

SU-SHU HUANG*

N66 37947

1. INTRODUCTION

Since the discovery of the periodicity of the change in light of Algol (β Persei) in 1782 by John Goodricke in England and his correct interpretation for the variability of light of this star as due to the "interpositions of a large body revolving around Algol," the study of light variation together with the accompanying variation in radial velocity, known respectively as the light curve and the (radial) velocity curve, has provided us with the sole means for the empirical study of close binary system. On the other hand, the close binary system can also be understood theoretically from Newtonian mechanics. Thus, our function is to reconcile the observational results in the form and light and velocity curve with the prediction based on the dynamical principles and thereby to derive a consistent picture of the close binary system. Indeed, the history of our study of close binaries shows clearly that, like any field in science where theories and experiments mutually help to make progress, observations of light and velocity curves and calculations based on dynamics act as two facets of a single process of successive approximations that lead finally to the present understanding of the nature of close binaries. Thus, the general form of the light and velocity curve suggests as a first approximation that dynamically the two components behave like two mass points and geometrically look like two spheres. Any analysis of light curves and velocity curves that is based on this assumption which we call the gross, analysis will be discussed in §2 and is included here for those who are not closely associated with observational astronomy.

*Goddard Space Flight Center and Catholic University, Washington, D.C.

Obviously this assumption is far from being true. Indeed, the prediction based on this assumption fails to agree completely with the observed light curve. Hence, the deviation of the observed results from the predicted ones opens up a way to improve the dynamical model of close binaries. Thus, the component stars may not be regarded as point masses and their shape not spherical. Also, other body or bodies than the two components themselves may be present in the system. Such refined treatments are given in the later sections.

Whether these complications can all be determined unambiguously from the photometric and spectroscopic observation with the aide of dynamical principles is not known at present. Is the empirical information—when conditioned by Newtonian mechanics and other laws of physics—comprehensive enough so that a unique determination of the geometrical configuration and physical state of the system can be performed at least in principle? This fundamental question has, to my knowledge, never been asked before, let alone answered. Undeniably this is a question of high mathematical complexity. Whatever the answer may be to this question, we are safe in predicting that such a unique determination is unlikely in practice because it is too difficult to disentangle one effect from many others in a single light and velocity curve, which has a limited accuracy imposed by measuring means. Such a situation which is common to all branches of observational science in general is particularly evident in astrophysics.

If we cannot treat the various deviations from the binary problem in an overall manner, we may nevertheless study them individually and compare

the results of each individual effect with observation. In this way, we can still test the consistency between observation and theory but the uniqueness of interpretation is not warranted by this kind of approach. By necessity it is this kind of approach which astrophysicists heretofore have cultivated in their study of close-binary systems and which we will follow in our treatment.

Next we consider the Roche model and the related problem of mass overflow from the stellar surface. The mass motion resulting from the overflow is then discussed according to different approaches. In the last two sections the origin and evolution of close binary stars are examined in the light of what we have learned from this article as well as from the present understanding of evolution of single stars.

Finally, we would like to add that the main purpose of the present article is to help understand the physical nature of close binary systems. It is intended to serve as an introduction to the basic physics of close binaries. Consequently, theoretical studies of a pure mathematical nature which cannot be matched in the foreseeable future with observation will not be discussed here. Similarly observational results of individual close binaries are mentioned only when they have some bearing either on general principles or on the common nature of the binary star. Because of a limitation of space we shall not be able to review here many physical problems such as models of peculiar systems, physical processes in the gaseous streams, atmospheric eclipses, etc. as well as many technical problems such as the distortions of velocity and light curves due to various causes and the methods of correction, accurate determination of the absolute dimension of eclipsing systems.

For surveys and observations the reader is referred to Kukarkin and Parenago's (1963) article.

While close binaries are not necessarily undergoing eclipse as seen by us, we are interested in only those that do, because the light variation caused by eclipse reveals much information about the nature of the system. It is true that the mere fact of undergoing eclipse does not necessarily warrant a close binary, but its chance of being one is very high. For this reason a theoretical study of close binaries is intimately associated with the observational investigation of eclipsing variables.

For definiteness we shall regard rather arbitrarily all binaries with separations less than 10 times the mean radius of the two component stars as close. Thus, the close binaries can be further classified into three categories: (1) detached systems, (2) semi-contact systems and (3) contact systems as we shall see in §5.

2. OBSERVATIONAL BACKGROUND

In this section we assume that the binary motion may be treated as a two-body problem, i.e. each component acts as a mass point and that the circular disk of each star follows a given law of limb-darkening. Under these simplifying assumptions, the physical state of the two component stars is specified by their masses (M_1, M_2), their radii (R_1, R_2) and their luminosities (L_1, L_2). Altogether there are six independent parameters to be determined for component stars themselves. In accordance with convention, we have denoted here by subscript 1 the quantity that is related to the primary component and by subscript 2 that related to the secondary component. Needless to say, the principal component is the brighter one of the two but is not necessarily the more massive one, although in most cases the brighter one is also the more massive one. From the radii and luminosities, the effective temperatures (T_1, T_2) or equivalently the surface brightness (J_1, J_2) can be readily computed.

The state of a binary can be specified under the assumption just made by the period, P, the eccentricity, e, the semi-major axis, a, of the relative orbit, the inclination i of the orbital plane, and the longitude of the periastron ω as defined by the angular distance of the periastron from the ascending node measured in the direction of orbital motion, and T the time of periastron passage. Altogether there are also six independent parameters to be determined for the binary itself. All these parameters are often called the orbital elements.

In the course of analysis of the velocity curve, it happens that we can also determine the radial velocity—called the γ velocity—of the entire binary system. Although the γ velocity is not connected with the internal properties of a binary system, it is nevertheless regarded also as an orbital element by convention.

In most cases of close binaries the two components move around each other in circular orbits. This gives $e=o$ which is sometimes assumed in the analysis of light curves of eclipsing binaries. Since the periastron loses its meaning when $e=o$, the binary motion is then fixed by the time of the deeper minimum of the light curve instead of the time of periastron passage.

The determination of the orbital elements from the light curve—a process perfected by H. N. Russell, J. E. Merrill, etc.—has been discussed in detail elsewhere in this series (Irwin 1962). While the procedure itself is a very tedious one, the underlying principle and what can be derived from it can be simply stated. First the law of limb-darkening is assumed to be a priori given for each component and is usually written as

$$J=J_c(1-x+x\cos\theta) \qquad (2.1)$$

where J_c is the surface brightness at the center of the disk, x is the coefficient of limb-darkening, and θ is the angle between the line of sight and the stellar radius vector to the point in question. Needless to say, both J_c and x vary with the wavelengths in which the light curve is obtained.

The period of the binary can be obtained very accurately from the long observation of the times of light minimum of the eclipsing binary. Then the shape of the light curve, obtained by observations and all reduced to one single cycle, depends upon the following elements, e, ω, R_1/a, R_2/a, i, $L_1/(L_1+L_2)$, x_1 and x_2. Conversely the observed light curve determines these orbital elements. If the binary satisfies the simplifying assumptions we have made and the light curve is determined in all phases, such a determination is unique in most cases and can be performed in practice, although it is quite complicated when e deviates appreciably from zero and i from $\pi/2$. Usually only the value of $e\cos\omega$ is determined with a high accuracy from the intervals of the minima in such cases. Also, under some circumstances, considerable ranges of various orbital elements lead, within the accuracy of observation, to the identical light curve.

For example, sometimes one cannot even tell from the light curve whether the eclipse is total or annular, consequently we cannot decide which one of the two component stars has a larger radius than the other. The intrinsic difficulty of determination of orbital elements however, arises in the first place from the various effects which are not included in our simplifying assumptions, and in the second place from the imperfection of the observed light curve.

As has been emphasized by Irwin (1962), only in the most favorable cases and with the best observations can we determine x_1 and x_2 from the analysis of the light curve. In general, these two parameters can be assigned beforehand in accordance with the prediction by theory of stellar atmospheres (e.g. Chandrasekhar 1950).

In any case the light curve which gives only the sizes of the stars in terms of their separation does not reveal the complete information about the nature of the binary system. In order to derive the dimension of the system in an absolute measure we depend upon the velocity curve. References to the analysis of velocity curves may be found for example in articles by Struve and Huang (1958), by Petrie (1962) and by others.

For some binaries spectra of both components may be seen on a spectrogram but for others only one is visible, depending upon the relative magnitude between the components and their colors (Hynek 1951). If both components are visible, two velocity curves —one from each component star—give $M_1\sin^3 i$, $M_2\sin^3 i$, $a_1\sin i$, $a_2\sin i$, where a_1 and a_2 are the semi-major axes of the orbit of the two components in the rest frame of reference, i.e.,

$$a=a_1+a_2 \qquad (2.2)$$

The velocity curve also gives $e\cos\omega$, $e\sin\omega$, time of periastron passage, T, and the γ velocity. Since the factor, $\sin i$, is persistently associated with the measured radial velocity, we cannot get rid of it from the spectroscopic results as we have seen. Here i determined from the light curve supplies the missing information. Hence, the spectroscopic results combined with the photometric data determine the properties of the binary system completely, if the two component stars obey our simplified assumption.

If the spectrum of only one component is visible, we can determine $e\cos\omega$, $e\sin\omega$, T, and γ without difficulty just as in the double-lined spectroscopic binary. However, concerning the mass and dimension of the system, we can now obtain

from the velocity curve (of the primary component alone) only the quantity,

$$f_1(M)=\frac{M_2^3 \sin^3 i}{(M_1+M_2)^2}, \tag{2.3}$$

known as the mass function and $a_1 \sin i$. When the results of analysis of both the light curve and the velocity curve are combined, $M_2^3/(M_1+M_2)^2$ and a_1 are obtained. In such a case we cannot determine M_1 and M_2 separately. The missing link is the mass ratio expressed by either

$$\alpha=\frac{M_2}{M_1} \quad \text{or} \quad \mu=\frac{M_2}{M_1+M_2}. \tag{2.4}$$

Thus for an eclipsing binary showing only the spectrum of one component, other means must be sought in order to determine the mass ratio. The means, if available, varies from one case to another and has no general rule. Basically it consists of an estimate from various physical arguments any one of the following parameters, m_1, m_2, α, a, R_1, and R_2 since only one additional parameter is needed to specify completely a single-spectrum eclipsing binary. The fact that a knowledge of either R_1 or R_2 can determine the masses of two components and their mean separation may be seen if we remember that R_1/a and R_2/a can be empirically obtained from the light curve. This gives a from which we can calculate M_1+M_2 from the following well-known relation

$$\left(\frac{P}{2\pi}\right)^2=\frac{a^3}{G(M_1+M_2)}, \tag{2.5}$$

where G is the gravitional constant. Consequently M_1 and M_2 may be separately determined.

In this way the physical and dynamical state of many an eclipsing binary has been painstakingly derived by investigators in the past few decades. Those binaries with apparent photographic magnitudes brighter than 8.5 at maximum have been compiled and studied by Plaut (1950, 1953). Other catalogues of the elements of eclipsing binary systems have been provided recently by Kopal and Shapley (1956) and Wood (1963).

3. DYNAMICAL CONSIDERATIONS OF THE DEVIATION FROM THE TWO-BODY PROBLEM

There are several reasons for which the motion of component stars cannot be regarded as a problem of two bodies. The component stars in a close pair obviously cannot have a spherical symmetry in density distribution. The mass of each star may not be constant. A third body or bodies in the form of stars, planets or resisting medium composed of gases, dust, and larger particles may be present in the binary system. All these effects disturb the orbit of the binary star from what would be expected from Kepler's law of binary motion.

If we regard Kepler's law as the zeroth order approximation to the orbital motion of close binaries, each of those effects mentioned above produces a corresponding perturbation which may be computed in principle to high order of approximation. It is self-evident that starting at the second order approximation, cross terms each arising from two effects will appear in the final result. While it is interesting to investigate the effect of each cause to a high degree of approximation, it is only the result of the first order approximation that the astrophysicist has actually sought in most cases because his main purpose is to identify among various perturbing factors the one that is dominant in any given binary. Needless to say, the result of higher order approximation for a single effect is often useful but we have to be very careful in applying its result quantitatively to the interpretation of observed phenomena because of the presence both of the cross terms just mentioned and of the first-order term due to some effect which has not been considered.

In what follows we shall discuss each of these perturbing factors separately, emphasizing physical significance at the expense of mathematical completeness.

3.1. Departure of Component Stars from the Spherical Symmetry

Because of the rotational and tidal distortions the component star in a close binary system can never be treated as having a spherical symmetry in its density distribution. Thus, the mutual attractive force between the two stars can be resolved into two components: one is the dominant force resulting from the mutual attraction

of two point masses M_1 and M_2 and the other is the perturbing force arising from the deviations from the spherical symmetry. Since the perturbing force is very small compared with the dominant force, the orbital motion should follow closely, with only a slight deviation, to what is predicted in the two-body problem. In other words, the component star may be approximated, in each short time interval, as exercising Keplerian motion. Because of the small perturbing force it is obvious that the orbits defined by two short time-intervals which are themselves separated by a long time-interval, cannot be identical. Thus, we may describe the motion as elliptical but the orbital ellipse changes with time. Obviously the rate of change for each of the orbital elements depends upon the perturbing force and indeed can be expressed in terms of it. The mathematical procedure underlying this physical conception is the method of variation of parameters which is often used to study the motion of a planet under the dominant force of the sun and the perturbing force of other planets. Its application to the study of close binaries was due to Russell (1928) whose calculation was later improved by Sterne (1939 a,b,c). Also, Cowling (1938) studied this problem by an entirely different approach. The presentation adopted in the following discussion follows that given by Sterne (1939 a).

The calculation consists of three steps: (1) to find the disturbing potential field, at the location of the primary due to the undistorted secondary component and due to the primary's own axial rotation, (2) to find both the deformation of the primary due to the disturbing potential obtained in the first step and the potential field at the location of the secondary due to the deformed primary and (3) to find the change of orbital elements because of the potential field obtained in the second step. Since the potential field derived in the first step distorts the primary and the distorted primary in turn produces in the second step a disturbing potential to effect the binary motion, the first two steps may be regarded as a process of successive approximation.

Let S_1 and S_2 be respectively the centers of mass of the primary and the secondary which are at a distance, r, apart. If r_1 and r_2 are respectively the distances of any point from S_1 and S_2, the undistorted secondary gives rise to a potential field (negative values throughout) equal to GM_2/r_2. At point P in the neighborhood of S_1, $1/r_2$ can be expanded in a series. Thus,

$$\frac{GM_2}{r_2}=\frac{GM_2}{r}\sum_{n=0}^{\infty}\left(\frac{r_1}{r}\right)^n P_n(\cos\theta) \qquad (3.1)$$

where P_n is the Legendre coefficient of order n and θ represents the angle PS_1S_2. If the separation between S_1 and S_2 be fixed, the first term is a constant of no consequence while the second term produces an acceleration GM_2/r^2 equal in magnitude for all elements of the primary along S_1S_2. This acceleration is simply the orbital acceleration of the primary as a whole. The remaining terms

$$V_t=\frac{GM_2}{r}\sum_{n=2}^{\infty}\left(\frac{r_1}{r}\right)^n P_n(\cos\theta) \qquad (3.2)$$

constitute the tidal potential and produce tidal distortion in the primary.

In an elliptical orbit, r is actually a function of the time. Thus, a rigorous treatment would be very complicated. The complications have been described by Sterne (1939a). However, as Cowling (1938) pointed out, the time of adjustment of a star to an external gravitational field is of the order of the period of free adiabatic oscillations of the star, which is much smaller than the orbital period of its companion around it. Therefore, it appears probable that the form of the star at any time would approach a state very close to equilibrium form in which the distortion is instantaneously adjusted to the external gravitational field given by equation (3.2). (Thus the longest axis of both component stars is always along the line joining the centers of the two components.)

Now the stars will be assumed to rotate with uniform angular velocities ω_1 and ω_2 about axes normal to the orbital plane. Such an assumption appears to be consistent with observations in general. Rotation of star introduces a disturbing potential which arises from the centrifugal force. Apart from a term which is symmetric with respect to the center of mass of the primary and consequently does not distort the star from a spherical configuration, the disturbing potential, V_r,

$$V_r=-\tfrac{1}{3}r_1^2\omega_1^2 P_2(\cos\theta') \qquad (3.3)$$

where θ' is the co-latitude.

Therefore, the total disturbing potential, V_d imposed on the primary is

$$V_d = V_t + V_r \tag{3.4}$$

and can be expanded in a series of terms of the form

$$c_{n,m} r_1^n P_n^m(\theta,\varphi) \tag{3.5}$$

where (θ,φ) are polar coordinates taken with respect to S_1, P_n^m is a tesseral surface function of (θ, φ) and $c_{n,m}$ is a constant of expansion. This completes the first step of calculation.

The problem of deformation of a sphere which cannot maintain a shearing stress, and consequently of the potential arising from the deformed body, as the result of an imposed disturbing function V_d has been studied by Clairaut, Legendre, Laplace and others. A summary of their works can be found in Tisserand (1891). Here, we shall only outline the general idea of the approach to the problem, omitting detailed calculations.

The equation of the surface of equal density in the distorted primary can be written as

$$r_1 = \xi(1 + \sum_{n,m} Y_n^m), \tag{3.6}$$

where ξ is a parameter (the mean radius) characterizing the surface in question. Thus, the density ρ of the primary is a function of ξ alone—a fact that makes ξ well suited to replace r_1 as the independent variable in the present problem. Thus, the value of the independent variable ξ at the stellar surface is R_1, the mean radius of the primary itself. The Y_n^m's in equation (3.6) are tesseral surface functions with respect to θ and φ and are also functions of ξ. Thus the deformation of the primary will be completely determined if we can find the ξ-dependence of Y_n^m and know the ρ-dependence of ξ.

The total gravitational potential Ψ at any point in the primary is the sum of the gravitational potential, U, due to the deformed primary itself and the imposed disturbing potential, v_d, i.e.,

$$\Psi = U + V_d \tag{3.7}$$

It can be shown that the condition of hydrostatic equilibrium leads to a total differential equation

$$dp = \rho d\Psi \tag{3.8}$$

where p is the pressure. It follows from equation (3.8) that both p and ρ must be functions of Ψ alone. In other words Ψ = constant over any surface of equal density and equal pressure in the primary. Furthermore, since ρ is a function of ξ alone, Ψ must be also a function ξ alone, i.e., independent of θ and φ.

Now U at any point p inside the primary consists of two parts, arising respectively from the mass (1) inside the spherical surface passing through p. (exterior potential) and (2) outside the spherical surface (interior potential). Since the equal-density surfaces are given by equation (3.6), both parts can be expressed in terms of Y_n^m. Actually when we carry out the lengthy calculation, we will find that each part of U can be expressed as a summation of terms involving Y_{nm} and other factors (like ρ, ξ) which do not contain the indices n and m (i.e., independent of θ and φ), if we assume that the distortion given by Y_n^m's is small so that only first order terms of them are retained in the calculation. We have already mentioned that V_d can be expressed as a series whose typical term is given by (3.5). Hence, we can write down Ψ as a sum of three series, two involving Y_n^m and the other involving P_n^m. All terms are indexed by n and m (i.e., indicating their θ and φ dependence). The condition that Ψ is independent of θ and φ makes it necessary that in the expression of Ψ the terms associated with each pair of indices n and m must vanish identically. In this way we derive a differential equation for each Y_n^m. (It is a differential equation because of the transformation from r, to ξ as given by equation (3.6).) If we define

$$\eta_n = \left(\frac{\xi}{Y_n^m}\right)\frac{\partial Y_n^m}{\partial \xi} \tag{3.9}$$

and

$$\rho_m = \frac{3}{\xi^3}\int_0^{\xi} \rho \xi^2 d\xi. \tag{3.10}$$

which is simply the mean density interior to ξ, we can transform the differential equation of Y_n^m into the following form:

$$\xi\frac{d\eta_n}{d\xi} + \eta_n(\eta_n - 1) - n(n+1) + \frac{\partial \rho}{\rho_m}(\eta_n + 1) = 0. \tag{3.11}$$

One reason that we introduce η_n instead of dealing directly with Y_n^m is to eliminate the (θ, φ)

dependence of Y_n^m which is simply p_n^m. Hence equation (3.11) together with p_n^m determines completely the deformation of the primary if we know the values of Y_n^m and η_n at any one point inside the star. This can be easily obtained for the point $\zeta=0$. There is no distortion at the central point. Hence $Y_n^m(0)=0$. In order to derive $\eta_n(0)$ we assume that $\rho(0)$ has a finite value. Setting $\zeta=0$ in equation (3.11) we obtain

$$\eta_n(0)=\eta-2. \tag{3.12}$$

This completes the determination of deformation of the primary.

In the process of determining the deformation we have already derived the potential U that the deformed primary produces at the external points as a series of Y_n^m. Corresponding to the (n, m) component of the distorting potential given by (3.5), the deformed primary produces a component $U_{n,}^m$ to the potential

$$U_{n,m}=c_{n,m}\frac{n+1\eta_n(\mathrm{R}_1)}{n+\eta_n(R_1)}\frac{R_1^{2n+1}}{r_1^{n+1}}P_n^m(\theta,\varphi) \tag{3.13}$$

at external points. Here $\eta_n(R_1)$ is the value of η_n at $\xi=R_1$ (i.e., at the surface). Since $\eta_n(R_1)$ is obtained by integrating equation (3.11) which involves p/ρ_m, the (n,m) component of the exterior potential depends upon the structure of the star.

The result contained in equation (3.13) can now be applied to the present case of tidal and rotational disturbance. The terms of the indices $n=2(m=o)$ in the disturbing potential $v_d-v_r+v_t$ given by equations (3.2) and (3.3) are

$$V_d=\frac{GM_2}{r^3}r_1^2p_2(\cos\theta)-\frac{1}{3}r_1^2\omega_1^2p_2(\cos\theta'). \tag{3.14}$$

Since the result given by (3.13) is independent of the choice of coordinate systems, the primary deformed by Vf of equation (3.14) produces an external gravitational potential field

$$U_2=2\frac{R_1^5}{r_1^3}k_{2,1}\left[\frac{GM_2}{r^3}p_2(\cos\theta)-\frac{\omega_1^2}{3}p_2(\cos\theta')\right], \tag{3.15}$$

where

$$k_{2,\,1}=\frac{3-\eta_2(R_1)}{2[2+\eta_2(R_1)]} \tag{3.16}$$

computed for the primary to which the second subscript is κ is referred to.

The acceleration of the secondary due to U_2 *directed towards* S_1 can be obtained by setting $r_1=r$, $\theta=$ o and $\theta'=\pi/2$ *in* $-\alpha U_2/\delta,r$, and is given by

$$\frac{R_1^5}{r^4}k_{2,\,1}\left(\frac{6GM_2}{r^3}+\omega_1^2\right). \tag{3.17}$$

If we now regard r as the independent variable, the acceleration given by equation (3.17) corresponds to a potential function,

$$\Phi_2=R_1^5k_{2,\,1}\left(\frac{GM_2}{r^6}+\frac{\omega_1^2}{3r^3}\right) \tag{3.18}$$

If the relative orbit is considered, $(1+m_2/m_1))\Phi_2$ will be identical to the disturbing function R of perturbation theory in celestial mechanics. Since the perturbing force is confined to the orbital plane which coincides with the equatorial planes of both components, the change in ω (e.g., Brower and Clemence 1961) is given by

$$\frac{d\omega}{dt}=\left[\frac{1-e^2}{G(M_1+M_2)a}\right]^{\frac{1}{2}}\frac{1}{e}\frac{\partial R}{\partial e} \tag{3.19}$$

In order to derive $\partial R/\partial e$ we have to expand R in terms of e. Since we are interested in the secular motion of the apsides, only non-periodic terms are needed in the evaluation. A detailed calculation by Sterne (1939a) thus derives the secular motion arising from the distortion of order $n=2$ of the primary as

$$k_{2,\,1}\frac{R_1^5}{M_1}\left(\frac{M_1+M_2}{Ga}\right)^{\frac{1}{2}}\left[15\frac{GM_2}{a^4}f_2(e)+\frac{\omega_1^2}{a^3}g_2(e)\right] \tag{3.20}$$

where $f_2(e)$ and $g_2(e)$ are series, convergent for any value of e less than inity,

$$f_2(e)=1+\frac{13}{2}e^2+\ldots, \qquad g_2(e)=1+2e+\ldots$$

Both $f_2(e)$ and $g_2(e)$ can also be expressed in closed expressions (Sterne 1939a) and have been computed by Plavec (1960).

Similarly, the secondary will be distorted by its own rotation and by the attraction of the primary. The perturbation produced by the distorted secondary on the primary can be easily obtained, because of symmetry, by interchanging M_1 and M_2

as well as replacing respectively R_1 and $k_{2,1}$, by R_2 and $k_{2,2}$, in (3.20). Thus the resultant motion of the apsides as compared with the mean orbital motion is

$$\frac{p}{p'}=k_{2,1}\left(\frac{R_1}{a}\right)^5\left[15\frac{M_2}{M_1}f_2(e)+\frac{a^3\omega_1^2}{GM_1}g_2(e)\right]$$
$$+k_{2,2}\left(\frac{R_2}{a}\right)^5\left[15\frac{M_1}{M_2}f_2(e)+\frac{a^3\omega_2^2}{GM_2}g_2(e)\right], \quad (3.21)$$

where p and p' are the orbital and apsidal periods respectively.

If the orbital motion and axial rotation of both components in a close binary are synchronized (Swings 1936), equation (3.21) reduces to

$$\frac{p}{p_1}=k_{2,1}\left(\frac{R_1}{a}\right)^5\left\{\frac{M_2}{M_1}[15f_2(e)+g_2(e)]+g_2(e)\right\}$$
$$+k_{2,2}\left(\frac{R_2}{a}\right)^5\left\{\frac{M_1}{M_2}[15f_2(e)+g_2(e)]+g_2(e)\right\}. \quad (3.22)$$

The contributions arising from the third and fourth order harmonic (tidal) distortion can be similarly obtained. The results can be expressed in terms of $k_{3,i}$ and $k_{4,i}$ $(i=1,2)$ which are functions of $\eta_3(R_i)$ and $\eta_4(R_i)$ respectively in a similar way as $k_{2,i}$ being a function of $\eta_2(R_i)$.

The values of k's can be readily found for any model by integrating (3.11). Thus, values of k_2 have been tabulated for a series of models by Russell (1928). For homogeneous stars, k_2 has a value 3/4; for completely concentrated stars all the k's are zero. In the case of certain polytropic models, the k_n's have been obtained by Chandrasekhar (1933) and later more extensively by Brooker and Olle (1955) for $\eta=2$ to $\eta=7$. The k's for other stellar models have been obtained by Kellar (1948), by Motz (1941, 1950, 1953), by Pike (1955), by Härm and Rogenson (1955), and by Kushwaha (1957).

Table 1 gives the k_2 values for some polytropic gaseous spheres from Brooker and Olle's paper. It illustrates its relation to two other parameters that also depend upon the density distribution in the sphere. The two parameters are the ratio of the central density ρ_c to the mean density $\bar{\rho}$ and the radius of gyration of the sphere. The values $\rho_c/\bar{\rho}$ given in Table 1 are taken from Chandrasekhar's (1939) book. The radius of gyration given in the form of I/MR^2) in the table where I denotes the moment of inertia follows Motz's (1952) calculation. According to the latter. H. N. Russell has suggested that the apsidal motion constant k_2 might depend in a simple way on the radius of gyration of a star. Indeed, in the case of polytropic gaseous spheres and main sequence stellar models, Motz (1952) has found if $\log k_2$ is plotted against $\log[I/(MR^2)]$, all points do fall on a straight line. However, he has also pointed out that the giant models do not follow this linear relationship.

TABLE 1.—*Relation between the Apsidal Motion Constant and other Parameters that Relate to the Density Distribution in the Polytropic Gaseous Sphere*

η	k_2	$\rho c/\bar{P}$	$I/(MR^2)$
0	.75000000	1	.4
1.0	.25990728	3.28987	.26138
1.5	.14327923	5.99071	.20502
2.0	.07393839	11.40254	.15704
2.5	.03485234	23.40646	.11203
3.0	.01444298	54.1825	.07583
3.25	.00869160	88.153	----------
3.5	.00491907	152.884	.04558
4.0	.00119488	622.408	.02358
4.5	.00031609	6189.47	----------
5	0	∞	0

The secular variation of e and a can be similarly investigated. But it is hardly necessary to follow the calculation through. The variation must vanish because the interaction we are considering does not change the total dynamical energy and the total angular momentum of the orbital motion.

The assumption that the axes of rotation of both component stars are perpendicular to the orbital plane is likely true in most close binaries. However, if either or both axes of rotation make a considerable angle with the normal of the orbital plane, the motion of stars in a close binary would be greatly complicated, because we then would have to consider, apart from the motion in the orbital plane, the motion of the orbital plane itself and the motion of the equatorial planes of both components. Brouwer (1946) studied the problem by examining the motion of two rigid spheroids. His treatment was followed by Kopal

(1959) who has included the revolution of the tidal bulge in his considerations.

Thus, if two components have similar distributions of density so as to make corresponding k's for the two components equal, we may express in general the ratio of periods in the following form (Sterne 1939b).

$$\frac{p}{p'}=\alpha_2k_2+\alpha_3k_3+\alpha_4k_4+\ldots \qquad (3.23)$$

where α's can be computed from the orbital elements alone, while k's depend upon the stellar model. For each one-parameter family of stellar models, such as the polytropic family, the observed value of p/p' determines uniquely the parameter, such as the polytropic index. Therefore, it is frequent practice to express the observed motion of apsides in terms of the equivalent index of polytrope for the component stars. If, however, there are many plausible families of stellar models, the apsidal motion does not give any information as to which is the physically correct family. Hence, what the apsidal motion can provide for the astrophysicist is only a consistency check of the computed model.

A large number of papers have been published on the observed motion of apsides of various close binaries. The reader may find the references to these papers in several I.A.U. reports on eclipsing binaries by Kopal (1954, 1957), by O'Connell (1960, 1962) and by Merrill (1964). Also, the catalogues of orbital elements of binaries provided by Kopal and Shapley (1956) and by Wood (1963) give the observed apsidal motion where ever available. The observed results of apsidal motion of some binaries have been analyzed by Luyten, Struve, and Morgan (1939), by Sterne (1939b) and by Schwarzschild (1958).

3.2 The Perturbation Caused by a Third Body

In order to study the perturbation of the orbital elements of a close binary, caused by the presence of a third body, we have to assume that all three bodies behave like point masses. If the three bodies should move at distances of the same order of magnitude from one another, no orbital elements could be defined. Intuitively, such a state of motion is perhaps not stable. Indeed, few such systems have been found observationally What we have actually observed of triple systems usually consists of a close pair accompanied by a third companion which is at a relatively large distance from the close pair. In these cases, the orbit of the close pair and that of the distant companion are well defined and perturbation theories developed for the motion of the moon can be applied directly because the perturbation of the earth-moon system by the sun is, in many respects, similar to the perturbation of the close pair by the distant companion in a triple system. Thus, Slavenas (1927), Lyttleton (1934), Brown (1936, 1937) and Martynov (1948) have respectively applied different lunar theories to the stellar case. In principle, they all derive the change of orbital elements from the disturbing function, but their detailed calculations are so involved that it is impractical to be given in this short article. It will suffice to say that the elements of the close orbits that show secular changes are the line of apsides, the line of nodes and the mean longitude. If the p and p' are respectively the period of close binary and the orbital motion of the third body respectively, the periods of revolution of the apsidal and nodal lines p'' are given by

$$\frac{p''}{p}\sim\left(\frac{p'}{p}\right)^2 \qquad (3.24)$$

in order of magnitude.

Actually, the discovery of triple systems does not depend upon the results of these calculations. Observationally it is the change in the γ velocity obtained from spectroscopic data as well as the change in the period obtained mainly from photometric data that leads to the identification of the third body.

Since the semi-major axis of the relative orbit of the close pair suffers neither secular nor long-periodic perturbations by the third body, its period will remain constant. However, the apparent period as observed between two successive light minima will not be strictly constant because of the motion of apsides and nodes. The variation arising from these causes is small however, since the periods of revolution of the apsidal and nodal lines are very long, as one can see from equation (3.24). The actual observed variation in the apparent period is due to the continuous change in distance of the close pair from the observer. Since the velocity of light is finite, the

change in distance means also a variation in the time-interval between two successive minima.

Consider a rectangular coordinate system with its origin, 0, at the center of mass of the triple system and with its Z-axis coinciding with the line of sight. The distance of the center of mass S^{12} of the eclipsing pair from the XY plane is evidently

$$r \sin i \sin (v+\omega),$$

where r represents the radius vector from 0 to any point on the orbit of S_{12}, i, the angle of inclination of this orbit to the celestial sphere, v the true anomaly of S_{12}, and ω the longitude of the periastron as defined before. If the space motion of the entire triple system produces a γ velocity—denoted by γ_o—in the z-direction, the distance z between the eclipsing system and the observer at any time becomes

$$z=z_o+\gamma_o(t-t_o)+r \sin i \sin(v+\omega), \quad (3.25)$$

where z_o denotes the initial value of z at time t_o of a light minimum of the eclipsing system. Thus, a term $(z-z_o)/c$, where c is the velocity of light, will be introduced into the ephemeris of the light minima.

Any apparent difference in time, arising purely from the light propagation, of some cosmic event is known in astronomy as the time equation or light-time. For example, there is the light-time for converting observed time on the earth to the heliocentric time, because the light carrying the news of the event does not reach the earth and the sun at the same time as the result of a difference in distance. This converting factor can be computed from our knowledge of the earth's position with respect to the sun's. In the present case, the light-time results from the orbital motion of the eclipsing pair around a third body or bodies. Thus, the light-time in this case is the difference (0–C) between the time of observed light minimum and the computed time based on the constancy of the period of the eclipsing pair. A light-time curve can therefore be plotted which, like the velocity curve, yields information as regards the nature of orbit on which the center of mass, S_{12}, of the eclipsing pair move.

The light-time curve—a plot of 0–C against the number of eclipsing cycles—of the Algol system has been extensively studied by many investigators because its records can be traced back nearly two hundred years. Eggen (1948) has shown that the system is composed of four stars, namely the eclipsing pair themselves, a third component causing wriggles in the light-time curve with an orbital period of 1.873 years and a fourth component causing the slow but dominant variations in 0–C with an orbital period of 188.4 years. This interesting system is continuously being studied with the purpose to make a better determination of the periods.

The determination of the (light-time) orbit from the light-time curve, which has been discussed by Woltjer (1922), Martynov (1948), Irwin (1952, 1959) and Kopal (1959) is very similar to one of determining orbital elements from the velocity curve. Indeed, the two problems are basically identical because if we differentiate (3.25) with respect to t, we obtain the radial velocity of S_{12} in its course of orbital revolution around 0. Thus, the light-time curve of an eclipsing pair in a triple system is equivalent in principle to a velocity curve of the motion of the center of mass, S_{12}, of the eclipsing pair around the center of mass, 0, of the entire system.

Actually the velocity curve of S_{12} can be directly determined from observation. Since the motion of S_{12} may be regarded as constant during a few cycles of the orbital motion of the close pair, the γ-velocity determined from the velocity curve during these few cycles of one or both components of the close pair represents the radial motion of S_{12}. Thus, a velocity curve of S_{12} may be obtained by plotting the γ-velocity of the close pair in different epochs. Therefore, we can derive all those orbital elements that are derivable in any single-lined spectroscopic binary. For example, the component, with a period of 1.873 years in the Algol system, mentioned previously, was first detected in this way by McLaughlin (1934).

Both the presence of a third body and the tidal and rotational distortions of component stars produce the effect of apsidal motion. It would be difficult to separate the causes of the observed motion, if there were no other observable criteria. Fortunately, because the presence of the third body in a binary system can be independently detected by the change in γ-velocity or by the light equation, there will be little ambiguity in interpreting the motion of apsides in a close binary. Thus, if the central condensation of the

star is to be studied, we can always choose those close binaries in which no third component is detectable.

3.3 Perturbation by Galactic Objects

A binary system is continually perturbed by the encounters with other objects in the galaxy. As a result the energy is fed into the binary system, thereby increasing the separation of its two components. Ambarzumian (1937) has estimated the effectiveness of the tidal forces, due to the neighboring stars, in modifying the orbital elements of a binary. He has used the two-body approximation of stellar encounters to evaluate the tidal effects of the nearby stars. As Chandrasekhar (1944b) has pointed out, "the essentially characteristic features of the problem are ignored if an attempt is made to evaluate the differential effects of the neighboring stars on the components of a binary along the conventional lines of treating stellar encounters as a series of independent two-body problems."

In view of some recent investigations it becomes apparent that encounters with interstellar clouds play an important role in modifying the peculiar velocities of single stars (Spitzer and Schwarzschild 1951) and in feeding energy into a cluster (Spitzer 1958). On the same ground we may expect that encounters of a binary with interstellar clouds would increase the separation of the binary system. A rigorous treatment of the encounter of a binary with another object belongs to the well-known problem of three bodies in celestial mechanics and a complete solution is still lacking. A brief review of some recent investigations along this line, however, may be found in the book by Leimanis and Minorsky (1958).

In the meantime, a general statistical method has been proposed by Chandrasekhar (1943, 1944a and b) for solving many problems in stellar dynamics. The method first analyzes the nature of the force acting on a star by the rest of the stellar system and then by Markoff's process derives the probability distribution of the force field.

In the case of a binary system it is apparent that the forces exerted by the neighboring stars on the two components differ by a certain amount because of their different positions in space. For any given separation between the components there exists a definite distribution function, $(W(\vec{F}_1, \vec{F}_2)$, which governs the probability that forces of intensities $\vec{F}_1$ and $\vec{F}_2$ respectively, will act simultaneously on the two components of the binary and which can be obtained by Markoff's method (Chandrasekhar 1944a). Thus, the difference,

$$\Delta\vec{F}=\vec{F}_1-\vec{F}_2, \tag{3.26}$$

represents the differential force which tends to accelerate the star "1" relative to the star "2." The differential force may be resolved into two components, a parallel component given by where

$$\Delta F_{\parallel}=(\vec{F}_1-\vec{F}_2)\cdot\vec{k}_1, \tag{3.27}$$

$\vec{k}_1$ is a unit vector parallel to the direction of $\vec{F}_1$ and a perpendicular component. Since the perpendicular component is of a random character, it produces no net effect during a time scale, τ, long compared to the periods of the elementary fluctuations in $\vec{F}$. Therefore, the average net increase in the velocity of the star "1" relative to the star "2" during τ is given by

$$\overline{\Delta v_{1,2}}=\Delta F_{\parallel}\tau. \tag{3.28}$$

If the orbit is circular, it follows from the definition of $W(\vec{F}_1, \vec{F}_2)$ that

$$\Delta v_{1,2}=\tau\int_{-\infty}^{\infty}\int^{\infty}(\vec{F}_1-\vec{F}_2)\cdot\vec{k}_1 W(\vec{F}_1, \vec{F}_2)d\vec{F}_1 d\vec{F}_2. \tag{3.29}$$

The evaluation of this integral is very involved and we can only refer the reader to the original paper by Chandrasekhar (1944a and b). When the separation between the two components, i.e., a, is small compared with the average distance between stars, the result may be approximately written in a simple form as follows:

$$\overline{\Delta v_{1,2}}=4\pi GNMa\tau, \tag{3.30}$$

where M denotes a certain average mass for the field stars and N the number of stars per unit volume.

Instead of the time of dissociation of a wide binary system that Chandrasekhar is interested, we set ourselves to calculate the slow increase in

separation of a close binary due to galactic perturbation.

Since $\overline{\Delta v_{1,2}}$ is the increase in the relative velocity of one component with respect to the other, $\overline{\Delta v_{1,2}}^2/2$ represents the increase in the dynamical energy per unit reduced mass E of the binary, namely

$$\Delta E = \tfrac{1}{2}\overline{\Delta v_{1,2}}^2 \tag{3.31}$$

while E itself is given by

$$E = -\frac{G(M_1+M_2)}{2a}. \tag{3.32}$$

It follows from equations (3.30)--(3.32) that the new separation, denoted by a', after a time τ is given by

$$G(M_1+M_2)\left(\frac{1}{a}-\frac{1}{a'}\right) = (4\pi GNMa\tau)^2. \tag{3.33}$$

This equation reduces to Chandrasekhar's equation for determining the time scale of dissociation, τ_d, when we set $a' \to \infty$.

$$\tau d = \frac{(M_1+M_2)^{1/2}}{4\pi G^{1/2}NMa^{3/2}}. \tag{3.34}$$

Assuming, in the neighborhood of the sun, $N = 0.1$ star/(parsec)^3 and setting $M = 0.5M_\odot$, $M_1+M_2 = 1M_\odot$, Chandrasekhar has obtained from equation (3.34) the following numerical results

$$\tau_d = 2.22\times 10^{15} a^{-3/2} \text{ years} \tag{3.35}$$

if the separation, a, is expressed in astronomical units.

It becomes obvious from equation (3.35) that dissociation of close binaries does not take place in the galactic time scale, say 2×10^{10} years. What we would like to calculate is the small increase in separation during this time. For this purpose let us write

$$a' = a + \Delta a \tag{3.36}$$

Since Δa is small, it can be shown as a first approximation that

$$\frac{\Delta a}{a} = \left(\frac{\tau}{\tau_d}\right)^2 \tag{3.37}$$

which gives numerically

$$\frac{\Delta a}{a} = 2.03\times 10^{-14} \text{ in } 10^{10} \text{ years} \tag{3.38}$$

if a is 1/10 A.U. Therefore, we may conclude that close binaries are not effected by the perturbation due to neighboring stars in any time scale relevant in astronomical disscussions.

Chandrasekhar's investigation of dissolution of binaries forced by the fluctuating field of stars has been followed by Takase (1953) who considers the effect of the interstellar matter instead of stars. He assumes that the interstellar medium may be regarded as a continuous mass distribution with strong density fluctuations in space (Chandrasekhar and Münch 1952). The density fluctuation may be measured by the mean square deviation of density in the medium, denoted by $\overline{(\delta\rho)^2}$. It may be assumed that the scale of fluctuations (i.e., the size of density irregularities) is much larger than the distance between two components of a binary. Then using the relation derived by Osterbrock (1952) between the fluctuating force field and the fluctuating density field and following Chandrasekhar's reasoning, Takase obtains a formula for the time of dissociation of the binary system due to interstellar matter. His formula can be written down by simply replacing $N^2 m^2$ in equation (3.34) by $\overline{(\delta\rho)^2}/\beta$. Since $\overline{(\delta\rho)^2}$ is about the same order of magnitude as $N^2 m^2$, Takase's result does not change our previous conclusion obtained from Chandrasekhar's analysis.

3.4 Resisting Medium

The effect of the resisting medium is to reduce the dynamical energy and angular momentum of a binary system that is embedded in it. Historically, this problem has created interest among astronomers mainly through the study of the origin of our planetary system. As is evident, the matter in the solar nebula in the early phase of the solar system cannot all have condensed into planets, whatever the mechanism of their formation is. A certain amount of gas and dust must have remained in the space around the sun (especially near the fundamental plane). This remnant formed a resisting medium that the new-born planets were submerged. More recently the resisting medium has again been discussed in con-

nection with the problem of accretion of matter by stars (McCrea 1953). But the discussion has been limited only to single stars.

As an illustration of the effect of a resisting medium on the binary motion, let us assume that the medium exerts simply a retarding couple on the binary system. For simplicity, we shall denote by ℓ the retarding couple that acts on unit reduced mass of the system. Hence, the couple acting on the entire system is $M_1M_2\ell/(M_1+M_2)$. We further denote by h and E respectively the angular momentum and dynamical energy per unit reduced mass of the binary system, thus

$$h=r^2\frac{d\theta}{dt} \tag{3.39}$$

while E has already been given in equation (3.32). We obtain

$$\frac{dh}{dt}=\ell \tag{3.40}$$

and

$$\frac{dE}{dt}=\ell\frac{d\theta}{dt}. \tag{3.41}$$

from our assumption. Needless to say, the polar coordinates (r,θ), represent the position of one component star with respect to the other, θ being measured from the periastron.

Since it can be shown from the two-body problem that

$$h^2=G(M_1+M_2)a(1-e^2) \tag{3.42}$$

we find from equations (3.32) and (3.41) that

$$\frac{1}{a^2}\frac{da}{dt}=\frac{2\ell}{G(M_1+M_2)}\frac{d\theta}{dt} \tag{3.43}$$

and from equations (3.39), (3.40) and (3.42) that

$$\frac{de}{dt}=\frac{\ell}{h}[2\cos\theta+e(1+\cos^2\theta)]. \tag{3.44}$$

As the retarding couple acts against the orbital motion, ℓ is negative. Therefore, the semi-major axis, and consequently the period of the orbit decrease in the resisting medium according to equation (3.43) as would be expected. If the resistance to the motion should increase with the speed of stars, their orbital motion would encounter the strongest resistance near the periastron (i.e., $\theta=o$) and the least resistance near the apastron (i.e., $\theta=\pi$). Then de/dt integrated over a complete cycle would be negative, and the eccentricity decreases in the medium. Or if the resistance is independent of the speed, the same result follows because of the second term in the bracket in equation (3.44), the first term in the bracket being zero on the average. However, if the resistance should decrease with increasing speed, the eccentricity could either increase or decrease depending upon the exact law of resistance.

Jeffreys (1918, see also Jeans 1928) who advocated the collisional theory for the formation of the planetary system used this result to explain why planets formed from the matter ejected from the sun by the tidal action of the colliding star should finally settle into nearly circular orbits. According to the current view, the planets were formed in a medium that was already revolving around the sun. Hence, the planet would be rotating with about the same velocity as its surrounding medium from which it had emerged and the medium was no longer a resisting one.

The previous example concerning the motion of new-born planets illustrates very clearly the difficulty of treating the problem of binary motion in a gaseous medium. It shows that the effect of the medium on the binary motion depends critically upon the dynamical state of the medium itself.

Recently, two papers (Fesenkov 1956; Kiang 1963) have appeared which deal with the motion of planets in a medium that is not static (with respect to the sun). Actually, Kiang has considered three plausible cases for the state of the resisting medium: (1) static, (2) rotating freely and (3) rotating uniformly and calculated the effect in a great detail for each case. However, the basic idea in his calculations is identical to what we have described here.

The effect of a resisting medium on the eccentricity of a binary has been recently rediscussed by Varsavsky (1962) who has followed the treatment of Poincaré (1911) and has been able to explain why binaries of shorter and shorter periods have in general, smaller and smaller eccentricities

for their orbits. However it should be noted that Poincaré has assumed that the resistance increases with an increase of relative velocity of two components and with a decrease of separation between them. Obviously this law of resistance has no physical ground. Consequently, Varsavsky's quantitative analysis may not be regarded as realistic although the general idea may be valid.

The most difficult factor in treating the binary motion in a resisting medium is the variation of masses of stars themselves. As we shall see later, the medium may have been created as a result of mass ejected from the component stars. Or the stars may accret mass from the medium in which they are embedded, and which could exercise infinite varieties of laminar or turbulent motion that we can conceive of. All these factors only indicate that a medium surrounding the binary system can give widely divergent results. Indeed if the medium is rotating faster than the binary motion, the orbit will expand instead of shrinking. Therefore, mathematically it may be interesting to study the effect on the binary motion by assuming a particular state of motion for the medium (with or without an accompanying change in stellar masses), the result will represent only one of a large number of possible cases that may or may not actually confront us. What is important astrophysically is to find out the actual circumstance under which a particular binary is found: Is the binary system embedded in a medium of its own creation? Is the medium resisting or accelerating? Does either of the component stars lose or gain mass? While we cannot answer all these questions from the scanty data we have accumulated for any binary system, these are astrophysically interesting questions. It follows that it is not urgent to make any elaborate calculation for a few possible cases. Rather we should consider the problem in a general way with the aid of some simple model. For only in this way can we expect to obtain some physical insight to the problem by combining theoretical studies with observational ones. Otherwise, theoretical calculations can never meet the challenge of observations.

Since the resisting (or accelerating) medium is intimately related to the change in mass of component stars, we shall come back to this problem in the following section.

3.5 Change in Mass of the Component Stars

The effect of mass loss of component stars on the orbital elements has been discussed by Krat (1950) and Wood (1950). The following discussions, however, will be based on Huang's (1963) presentation.

Because the orbital elements—period, P, semi-major axis, a, and eccentricity, e, — are related to the masses of the two components, angular momentum, h, and dynamical energy, E, per unit reduced mass of the system by the well-known formulae in the two-body problem, it can be easily shown that

$$\frac{1}{a}\frac{da}{dt}=\frac{1}{M_1+M_2}\frac{d(M_1+M_2)}{dt}-\frac{1}{E}\frac{dE}{dt}, \tag{3.45}$$

$$\frac{1}{p}\frac{dp}{dt}=\frac{1}{M_1+M_2}\frac{d(M_1+M_2)}{dt}-\frac{3}{2}\frac{1}{E}\frac{dE}{dt}, \tag{3.46}$$

$$\frac{e}{1-e^2}\frac{de}{dt}=\frac{1}{M_1+M_2}\frac{d(M_1+M_2)}{dt}-\frac{1}{2}\frac{1}{E}\frac{dE}{dt}-\frac{1}{h}\frac{dh}{dt}, \tag{3.47}$$

where E and h are given respectively by equations (3.32) and (3.39) or (3.42). Thus, the changes in orbital elements are not uniquely determined by the rate of mass variation but are dependent also upon dE/dt and dh/dt. This illustrates clearly the importance of the mode with which the variation of mass takes place because both dE/dt and dh/dt depend critically upon the mode of mass variation. Since the mass of the component star can lose or gain in an infinite variety of ways, this makes the problem somewhat uncertain.

The quantity, h, is related to the angular momentum per unit mass, h_0 by

$$h_0=\frac{M_1M_2}{(M_1+M_2)^2}h \tag{3.48}$$

and a similar relation exists between E and E_0 the latter being dynamical energy per unit mass Eliminating dE/dt from equations (3.46) and (3.47) and combining the resulting equation with

equation (3.48) we obtain

$$\frac{1}{p}\frac{dp}{dt}=\frac{4}{M_1+M_2}\frac{d(M_1+M_2)}{dt}$$
$$-3\left(\frac{1}{M_1}\frac{dM_1}{dt}+\frac{1}{M_2}\frac{dM_2}{dt}\right)$$
$$+3\frac{1}{h_o}\frac{dh_o}{dt}+\frac{3e}{1-e^2}\frac{de}{dt} \qquad (3.49)$$

This is a very useful relation because dE/dt (or equivalently dE_o/dt) which is unknown in most cases do not appear in it.

As in any problem that has multitudinous choices of possibilities we may idealize a few simple modes of mass variation that have a physical significance and then examine their effect on the orbital elements. In this way three modes of mass loss from the component stars have been isolated: (1) slow mode, (2) intermediate mode and (3) fast (or Jean's) mode. We shall discuss these three modes here under the simplified assumption that there is no interchange of angular momentum between axial rotation and orbital revolution of the component stars. A treatment of coupling between these two kinds of motion can be found in Huang's (1963b) paper.

1. Slow Mode

This represents an extreme case of losing mass. When the ejection velocities from the stars are low, the ejected matter will not be expected to escape from the binary system. Consequently, the total angular momentum of the system will be conserved but the total dynamical energy of the binary can vary in either way.

We may subdivide this mode into two cases: (a) the particles ejected fall back either to the original star or to its companion and (b) the particles ejected from the less massive component form, after many collisions among themselves, a rotating ring around the more massive component. The second case is proposed in accordance with observational results (Joy 1942, 1947; also Sahade 1960a). Formation of a rotating ring around the less massive component from matter ejected from the more massive component does not seem common because not a single case has been observationally established without dispute. Theoretically it is difficult to form and maintain such a ring (Huang and Struve 1953; Huang 1957a) but it is by no means prohibitive.

For a long time scale, the gaseous rings are expected to dissipate since they are continuously perturbed by the less massive component of the system. Hence, case (b) represents only a transient stage of mass loss. In either case we may set as a first approximation

$$\frac{d(M_1+M_2)}{dt}=0 \qquad (3.50)$$

In adopting this approximation we may regard the rings in case (b) as a part of the more massive component.

In case (a) we have

$$\frac{dh_o}{dt}=0 \qquad (3.51)$$

which yields, after its combining with equations (3.49) and (3.50),

$$\frac{1}{p}\frac{dp}{dt}=-3\left(\frac{M_2-M_1}{M_1M_2}\right)\frac{dM_1}{dt}+\frac{3e}{1-e^2}\frac{de}{dt}. \qquad (3.52)$$

For binaries of small eccentricities, the second term on the right-hand side of equation (3.52) may be neglected. Thus, we arrive at the conclusion that a transfer of mass from the more massive component to the less massive component (i.e., $M_1>M_2$, $dM_1/dt<0$) makes the period decrease with time and a transfer of mass in the reverse direction (i.e., $M_2>M_1$, $dM_1/dt<0$) results in an increasing period. Woolf has privately pointed out that the increase in period of β Lyrae may be precisely due to this mechanism, supporting the proposition (Huang 1962, Woolf 1962) that the primary component of this system is less massive than the secondary component.

More elaborate calculation of the variations of orbital elements with mass exchange has been recently carried out by Piotrowski (1964). He has considered separately the effect on the orbital elements of ejection, of flight and of landing of each mass element. While the calculations are mathematically interesting, it is hard to evaluate their significance in connection with observed data because of the infinite varieties of ways that an individual mass element can be transferred from one component to another. Also the results will

be seriously upset by collisions of particles during the flight (Cf., §6).

In case (b) h_o is no longer constant because a part of angular momentum is absorbed in the rotating rings. If we approximate the ring (formed out of matter ejected by M_2) as rotating *directly* in a circular orbit of radius, a_t, around, and under the gravitational attraction of the more massive component, M_1, alone, we obtain

$$\frac{1}{p}\frac{dp}{dt}=\alpha\frac{1}{M_2}\frac{dM_2}{dt}+\frac{3e}{1-e^2}\frac{de}{dt} \qquad (3.53)$$

where

$$\alpha=3\left(\delta_1-1+\frac{M_2}{M_1}\right) \text{ and } \gamma_1^2=\frac{(M_1+M_2)a_t}{M_1a(1-e^2)} \qquad (3.54)$$

Numerical values of α for a few cases of a_t/a and M_2/M_1 have been given in Huang's (1963b) paper.

Finally, it is to be noted that (a) and (b) represent two physically different cases of idealized circumstances. When we mentioned that (a) represents the ultimate situation of the slow mode, we do not mean that case (b) will reduce to case (a) if the ring coalesces with the stellar surface, because that would result a transfer of angular momentum from orbital revolution to axial rotation—a situation precluded to the present discussion. In such cases we must include the coupling between orbital motion and axial rotation into our considerations.

2. Intermediate Mode

By the intermediate mode we mean that the ejection velocities are large enough to overcome the attraction of both components so that the ejected particles can penetrate the inner contact surface (Cf 5) of the system but not large enough to justify the neglect altogether of interaction between the ejected particles and the binary system. Actually the change of orbital elements under this mode of mass loss is most uncertain because we have no way to estimate the angular momentum carried away by escaped mass.

This mode may also be divided into two cases: (a) ejected particles escape to infinity directly and (b) particles form an envelope around the entire binary system, such as found in β Lyrae (Struve 1941, 1958). If the envelope is a quasi-stable rotating ring (or disk) around; and in the plane of, the binary system, the change in period may be evaluated in terms of the size of the ring (Huang 1963b). However, observationally there is no clear-cut evidence to conclude that such is actually the case.

Finally, we should again point out that case (b) represents only a transient stage of this mode since eventually the envelope will be dissipated.

3. Fast (or Jeans's) Mode

This is an extreme case, for it is assumed that there is no net reaction of the escaped mass on the binary itself. We call it Jeans's mode because Jeans (1924, 1925) was the first to treat this case when he examined the effect of radiation loss (which is equivalent to mass loss) on the binary system. Since then it has been applied to cases involving direct loss of mass (Huang 1956, Boersma 1961). The physical circumstance of actual mass ejection that may be approximated by Jeans's mode must satisfy two conditions: (1) the ejection of mass has a statistically spherical symmetry and (2) the velocities of ejection must be very high to insure a negligible interaction with the binary system. According to these conditions the loss of mass resulting from a supernova explosion (Blaauw 1961) and from ordinary nova outburst (Ahnert 1959) would closely approximate this mode. Mass loss due to corpuscular radiation may also fall into this mode if emission of particles is spherically symmetric. That corpuscular radiation may be important has been pointed out by Fesenkov (1949) and Masevich (1949) in their theory of stellar evolution. Tidman (1958) has proposed on the other hand a theory that suprathermal particles may be produced in the binary system if the expanding coronas collide each other, thus creating a region of violent motion favorable for accelerating charged particles. In any case, corpuscular radiation from component stars may be the reason, according to Huang (1958), why a luminosity anomaly exists in many binary stars.

It can be shown (Huang 1956) that Jeans's mode of ejection leads to

$$\frac{dh}{dt}=0 \text{ and } \frac{1}{E}\frac{dE}{dt}=\frac{2}{M_1+M_2}\frac{d(M_1+M_2)}{dt} \qquad (3.55)$$

which, when combined with equations (3.45)--(3.47), yield the following results due originally to Jeans:

$$\frac{1}{a}\frac{da}{dt}=-\frac{1}{M_1+M_2}\frac{d(M_1+M_2)}{dt},$$
$$\frac{1}{p}\frac{dp}{dt}=\frac{2}{M_1+M_2}\frac{d(M_1+M_2)}{dt} \tag{3.56}$$
$$-\frac{e}{1-e^2}\frac{de}{dt}=0$$

It follows from equation (3.48) and a similar equation between E and E_o, we may write equations (3.55) in the following form:

$$\left.\begin{aligned}\frac{1}{h_o}\frac{dh_o}{dt}&=\frac{1}{M_1}\frac{dM_1}{dt}+\frac{1}{M_2}\frac{dM_2}{dt}\\&\quad-\frac{2}{M_1+M_2}\frac{d(M_1+M_2)}{dt},\\\frac{1}{E_o}\frac{dE_o}{dt}&=\frac{1}{M_1}\frac{dM_1}{dt}+\frac{1}{M_2}\frac{dM_2}{dt}.\end{aligned}\right\} \tag{3.57}$$

Thus, in Jeans's mode the angular momentum per unit mass changes, although h maintains constant. Also, the dynamical energy per unit mass of the system has increased as a result of this mode of ejection.

An extreme case of Jeans's mode of ejecting matter is the sudden explosion either with a spherical symmetry or with an axial symmetry as found by Weaver (1958) in his study of Nova Aquilae (1918). If the mass ejected during the explosion exceeds a critical amount of $r(M_1+M_2)/(2a)$ where r represents the separation of the two stars at the instant when the instantaneous mass ejection takes place, it would result in a complete separation of the two component stars, as was pointed out by Blaauw (1961). Indeed he proposed that this is how the high-velocity O and B type stars sometimes observed in the galaxy had acquired their velocities from orbital motion.

So far we have discussed only the case of mass loss of component stars. It first appears that we may treat the problem of mass gain in a similar manner. Actually, the problem is much more complicated than the case of mass loss because first of all, we have to assign the state of motion of the medium from which stars may accret mass. In general, the problem of changes in orbital elements depends upon the following factors: (1) the relative motion of the center of mass of the binary system with respect to the medium, (2) the state of internal motion of the medium, such as rotation, turbulence, etc. and (3) the binary motion itself. From these factors we are supposed to evaluate the rate of accretion as well as the force (resisting or accelerating depending upon cases) that each component experiences. That this is a difficult task may be seen from a simple case that the medium possesses no net angular momentum with respect to any point inside. In such a case the total rate of accretion by both components may be roughly estimated by treating the system at large distances as a single body and thereby deriving the accretion column (McCrea 1953) of the entire system. However, how can we estimate the proportion at which the two components divide their loot in the accretion column? Since they move in the same region, the more massive component which is gravitationally stronger is expected to accret matter more effectively than the less massive one. On the other hand if we consider the size and location of the orbits we may expect the reversed situation, namely the less massive component is more effective. Here we see the intrinsic difficulty of the problem.

Observationally, we can determine only the period of a binary with any accuracy that justifies the theoretical investigations. Indeed, we have found variations in orbital period in many eclipsing binaries, although it should be remembered that the variations in period can be due to many causes other than the mass variations. For example, any perturbation due to the departure from the idealized binary composed of two mass points could produce complicated variation in observed period which is the time-interval between two consecutive principal (or secondary) eclipses. Actually, we have indeed found that periods of many binaries change in irregular ways, the same binary may have its period increasing and decreasing at different epochs in a seemingly unpredicted manner. This leads Wood (1950 see also Schneller 1962) to suggest that the changes may be due to ejection of gas jet by the component star sometime in the direction of its motion and at other times in the opposite direction. Wood's view has been criticized by Kopal (1959). The present writer inclines to regard

Wood's view to have its merit. We have found that the changes in period depend upon the modes of ejection as well as the rate of mass loss. Since there is a continuous range of possible modes and rates of ejection, it is apparent that there will be statistical fluctuations with respect to time, producing thereby corresponding fluctuations in period variations. Especially we should point out that if the velocities of ejection are high, mass could be thrown out of the star in any direction, not necessarily limited to the channel through the Lagrangian points L_1 and L_2. However, the cause of variations in period observed in eclipsing binaries is not uniquely determined. Most likely many factors are involved in producing the observed variations.

Extensive collections and analyses of observed period variations of eclipsing binaries have been given by Tung and Chang (1957), by Kwee (1958), by Prikhod'ko (1961), by Wood (1953) and by others (See references given by Kopal 1954, 1957; O'Connell 1960, 1962; and Merrill 1964).

4. RELATIVISTIC EFFECT

The advance of perihelion of the orbit of a planet around the sun resulting from the relativistic considerations provides one of the crucial tests of Einstein's theory of gravitation. Gravitation in the general theory of relativity is closely interwoven with the four dimensional space-time continuum. Using the language of geometry, we may state that the absence of gravitational field corresponds to a four-dimensional Euclidean (flat) space and the presence of any permanent gravitational field distorts the space-time continuum into a curved one. Indeed, the principles of relativistic mechanics are mainly contained in what is known as the field equation of Einstein (e.g. Eddington 1923), which connects the physical situation (i.e., the energy-momentum tensor which describes the distribution of matter and energy) with the geometry of space-time and which may be regarded as the relativistic analogue of Poisson's equation. Also, according to the theory, the space-time trajectory of a free particle is given by the geodesic, which, as is well known in geometry, directly depends upon the property of the space-time continuum and consequently on the gravitational field which shapes it. Since the property of space-time continuum corresponding to an empty space surrounding a gravitational point mass is known, the equations for a geodesic can be written down immediately and provides the basis for the treatment of the motion of planets around the sun. When these equations are reduced to an ordinary polar coordinate (r, φ) they yield the equation (e.g. Eddington 1923)

$$\frac{d^2u}{d\varphi^2}+u=\frac{GM}{(ch)^2}(1+3h^2u^2) \tag{4.1}$$

as compared with the equation of a Newtonian orbit

$$\frac{d^2u}{d\varphi^2}+\mu=\frac{GM}{(ch)^2} \tag{4.2}$$

where $u-1/r$, M is the mass of the central body, ch is a constant of integration, and c is the velocity of light in empty space.

The ratio, $3\,h^2u^2$, of the second term to the first term on the right hand side of equation (4.1) can be shown to be practically equal to three times the square of the transverse velocity of the planet measured in unit of the velocity of light. For example, the ratio for the earth is 0.000,000,03. Thus, the difference between the relativistic and Newtonian orbits must be slight. It can be shown (e.g. Eddington 1923) that the perturbation caused by the presence of this small term produces an advance of the perihelion equal to

$$\Delta\omega=\frac{6\pi GM}{c^2a(1-e^2)} \tag{4.3}$$

radians per orbital revolution.

The previous result is derived from the assumption that the planet has a negligible mass. The case of two bodies of comparable masses has been studied by Levi-Civita (1937) who shows that as a first approximation the same formula for the advance of periastron holds true in the case of binary stars as in the case of an infinitesimal planet moving around a central mass having the total mass of the binary system. Thus, except by replacing M by M_1+M_2 we have equation (4.3) for the advance of periastron in a binary system. If we write p' as the period that the periastron advances a complete cycle of 2π, equation

(4.3) can be written as

$$\frac{p'}{p}=\frac{c^2p^2(1-e^2)}{12\pi^2a^2} \tag{4.4}$$

If we now write $K=K_1+K_2$ where K_1 and K_2 are respectively the semi-amplitude of the velocity curves of the two components the previous equation can be expressed in terms of directly observable quantities, thus:

$$\frac{p'}{p}=\frac{1}{3}\left(\frac{c\sin i}{K}\right)^2. \tag{4.5}$$

Since K rarely exceeds 300 km/sec while c = 300,000 km/sec, p'/p is of the order of 3×10^5. Hence, the advance of periastron as a result of relativistic correction must be very small. Luyten, Struve and Morgan (1939) have computed the motion of apsides according to equation (4.5) and found that the computed values are less than the uncertainty in the observed values for all the seven binaries whose apsidal motion they have studied.

It is evident from (4.3) that apsidal motion due to relativistic correction should be most pronounced in massive binaries with short separations. After calculations based on various assumptions as to the dimensions and internal structure of the components, Rudkjbing (1959) has concluded that in the eclipsing binary, DI Hercules, the relativistic part of the apsidal motion may well be larger than the part due to deformation of the components. However, according to O'Connell (1962) "no indication of apsidal motion has in fact been observed as yet in this system."

Levi-Civita (1937) has furthermore pointed out that aside from the slow precession of the relative orbit, the general theory of relativity also predicts "the absolute motion in the sky of any double star system." This result comes naturally from the fact that general relativity does not include, as a rigorous law, the principle of action and reaction which plays such an important role in Newtonian mechanics. Thus, one of the most important consequences of this principle, namely, the uniform motion of the center of mass of a system in absence of external forces, becomes invalid.

According to the calculations by Levi-Civita, the secular acceleration of the center of mass of a binary is directed along the major axis towards the periastron of the more massive component, and the increase ΔV of the velocity of the center of mass in a century is given by

$$\Delta V=12.55\ \mu(1-\mu)(1-2\mu)\frac{e}{(1-e^2)^{3/2}}\frac{M_1-M_2}{M_\odot}\frac{1}{p_d^2}\ \text{km/sec}. \tag{4.6}$$

where p_d is the orbital period in days, $M_\odot$ the mass of the sun and μ is the mass ratio defined by (2.4). Because of the factor $\mu(1-\mu)(1-2\mu)$, Δv has a maximum with respect to the mass ratio at $\mu=(1-3^{-\frac{1}{2}})/2$, which roughly corresponds to two stars respectively containing $\frac{1}{4}$ and $\frac{3}{4}$ of the total mass of the system. The maximum value of $\mu(1-\mu)(1-2\mu)$ is about 0.1. Since e is usually small for close binaries, the change in velocity, even in a century, is too minute to be detected by present means of spectroscopic study. Actually, I doubt that this effect could be detected in ordinary cases even after we have observed a close binary for tens of centuries, because the apsidal motion resulting from the rotational and tidal distortion obliterates this effect completely.

As two stars revolve around each other in a binary, it is obvious that the gravitational field they produce varies with their motion. According to Newtonian theory, the effect is instantaneously felt everywhere. Consequently, the notion of gravitational waves never arises. As the relativity theory presupposes that no casual effect can travel faster than light, we might anticipate that the change in the gravitational field travels out into space with the speed of light. A moving disturbance thus propagated out may be called the gravitational wave. Such is the physical reason for the existence of gravitational waves (e.g. Eddington 1923; Synge 1960). A realistic evaluation however could be made only after empirical detection (which is now actually attempted) should be successful.

We may expect that the emission and absorption of gravitational waves carry, to be sure, very small amounts of energy. Hu (1947) has computed the very small damping forces due to the emission of these waves. Actually, the rate of energy radiated in the form of gravitational wave by a binary star (in circular motion) as given by

Landau and Lifshitz (1951) is

$$\frac{dE}{dt}=-\frac{32}{5}\frac{G^4}{c^5}\frac{(M_1+M_2)M_1^2M_2^2}{a^5} \qquad (4.7)$$

From equation (4.7) Kraft, Mathews and Greenstein (1962) have found that for the repeating nova WZ Sagittae (1913, 1946) the rate is in the range of $10^{32}-10^{35}$ ergs/sec if their mass ratio μ is between 0.1 and 0.5. However, they have pointed out that the gravitational flux reaching the earth is quite small and its detection very difficult. But they went on to suggest that the period change due to energy dissipation by gravitational wave is probably detectable. As we have seen in §3, variations in period may arise from many causes. Thus, even if the period change in this star were detected, it would remain a difficult problem to prove that the variation is indeed due to dissipation of energy through gravitational waves. This kind of uncertainty is intrinsic to observational science and strongly contrasts to the high precision of controllable investigations in experimental science, such as physics and chemistry. But nowhere in the entire field of astronomy do we encounter more frustration than in the attempt to detect relativistic effects in binary systems as we have seen in this section.

5. THE ROCHE MODEL

5.1 Stellar Surfaces According to the Roche Model

The shape of single stars without axial rotation must be a sphere since there is no one particular direction that is physically different from the other. When it rotates, the shape will no longer be spherical because of the existence of a preferred direction, namely the axis of rotation. As a result, we would expect that a rotating body, like the earth, Jupiter, etc., will be flattened to become an oblate spheroid.

In general, the component star (in the close binary) rotates in synchronism with its orbital revolution (Swings 1936, Plaut 1959). Hence, the centrifugal force does play a role in shaping the component star. However, a more serious complication arises from the force field due to its companion. What would be the shape of stars thus influenced by its close companion? We have investigated this problem in §3.1. But the result is too involved to be useful for the present purpose. Here we shall consider a simple model such that the gravitational potential due to each component star may be regarded as equivalent to a mass-point. Then the potential field of the entire system can be easily calculated and the shape of star determined. Such is the Roche model of close binary stars.

Let us assume that the two component stars in a system revolve around each other in circular orbits. Furthermore, we choose as the unit of mass the total mass of the system, i.e.,

$$M_2=\mu \qquad M_1=1-\mu \qquad (5.1)$$

Also, we choose the separation between two components as the unit of length. If we take $P/2\pi$ as the unit of time (i.e., the angular velocity of orbital motion is one), we shall have unity for the gravitational constant. This can be easily seen from equation (2.5).

We may easily write down the equations of motion of a test particle in the gravitational field of the binary system. It is equivalent to the restricted three-body problem in celestial mechanics and can be most advantageously expressed in a coordinate system rotating with the component stars (e.g., Moulton 1914, Plummer 1918). The rotating coordinate system (x, y, z) may be so chosen that its origin locates at the center of mass of the binary, that its z-axis is perpendicular to the orbital plane and that its x-axis coincides with the line joining the two component stars. If we denote by r the radius vector from the origin to the test particle (x, y, z) and by r_1 and r_2 the respective distances of the test particle from the $1-\mu$ component at $(-\mu, o, o)$ and from the μ component at $(1-\mu, o, o)$, the equation of motion becomes

$$\frac{d^2\vec{r}}{dt^2}=-2\vec{k}\times\frac{d\vec{r}}{dt}+\text{grad } U \qquad (5.2)$$

where

$$U=\frac{1}{2}(x^2+y^2)+\frac{1-\mu}{r_1}+\frac{\mu}{r_2} \qquad (5.3)$$

and $\vec{k}$ represents a unit vector in the positive z-direction. Taking the scaler product of $d\vec{r}/dt$ and equation (2) and integrating the resulting equation

over t we obtain an integral for the test particle

$$2U-\left|\frac{d\vec{r}}{dt}\right|^2=C \tag{5.4}$$

where C is a constant of integration and is therefore a characteri. tic value for the test particle of a given set of initial conditions. In celestial mechanics it is often called the Jacobian integral.

The equation (5.4) relates the kinetic energy with the coordinate of the test particle in the rotating frame of reference. Consequently negative U may be regarded as a potential function in this reference system and we may plot a series of equipotential surfaces by setting $2U=C$, namely

$$x^2+y^2+\frac{2(1-\mu)}{r_1}+\frac{2\mu}{r_2}=C \tag{5.5}$$

where C now serves as a parameter labeling the surfaces. However, it may be noted that because of the first term on the right side of equation (5.2), U does not behave exactly like a potential in an inertial system. It is for this reason that the surfaces defined by equation (5.5) are not called the equipotential surfaces in celestial mechanics.

For a given value of C, which is fixed by the initial conditions, equation (5.4) determines the speed $|dr/dt|$ of the test particle at different points in space. Therefore, we can find all the points at which the speed will vanish. The locus of these points then define the so-called zero-velocity surface in celestial mechanics. It is obvious from equation (5.4) that the zero-velocity surface thus defined is identical to the one given by equation (5.5) which we have previously called the equipotential surface.

In order to see the importance of the zero-velocity surface let us consider $|d\vec{r}/dt|^2$ of a test particle of a given value of C as a function of the coordinate according to equation (5.4). On the zero-velocity surface this function vanishes everywhere. Hence, in general, the function would be positive on one side and negative on the other. Since $|d\vec{r}/dt|^2$ cannot be negative, it only shows that the test particle cannot penetrate the zero-velocity surface into the other side. Because of this property, it is useful to know zero-velocity surfaces corresponding to different values of C (or different initial conditions). Finally, we may add that in some special cases, the values of C may be positive on both sides of a zero-velocity surface. This happens when the function i.e., $|d\vec{r}/dt|^2$ in the present case, has an extremum equal to zero at some points in space. Indeed, such an occurrence gives rise to a few interesting and critical cases of zero-velocity surface that will be discussed later.

It is to be noted that we have used C sometimes to characterize the condition of the test particle according to equation (5.4) and at other times to label the zero-velocity surfaces according to equation (5.5). Thus, a particle of a certain value of C cannot penetrate the zero-velocity surface labelled by the same value of C.

Let us now examine the general behavior of the zero-velocity surfaces defined by equation (5.5). For very large values of C, say C_l, we have three possibilities: (1) large x^2+y^2, (2) small r_1 and (3) small r_2, corresponding to three separate surfaces for a single value of C. If x^2+y^2 is large, other terms on the left side of equation (5.5) may be neglected. Consequently, the surface represents a large circular cylinder with its axis coinciding with the z-axis. Similarly, when r_1 (or r_2) is small, other terms become negligible. We then obtain a small sphere around the 1-μ component (or around the μ component). If we do not neglect small terms, we will obtain three surfaces slightly distorted from what have been described. Therefore, space around the binary can be divided into four regions, one outside the nearly cylindrical surface, two inside the two closed surfaces respectively around the two stars, and the fourth between these three surfaces. It is easy to see from equation (5.4) that the test particle whose C value is equal to C_l can be either inside the two small closed surfaces or outside the large cylindrical surface. But it is forbidden to enter space between these three surfaces. Thus, the test particle (of C_l) can move in any one of the three permitted regions but cannot jump from one to the other.

In a close system the shape of each component star that satisfies the Roche model (i.e., a highly centrally condensed star with an envelope of a negligible mass) is determined by the closed zero-velocity surface around it, since, as we have seen in equation (3.8), the equal pressure surfaces in a star must follow the equipotential surfaces as a result of the hydrostatic equilibrium under which negative U will now behave exactly like a potential function in the rotating coordinate system.

As C decreases, the two closed surfaces around the two stars increase in size and the outer cylindrical-like surface shrinks. Eventually, these surfaces will come into contact. Actually, the contact occurs first between the two closed surfaces at a certain value of C which will hereafter be denoted by C_1. This critical zero-velocity surface is often called the innermost contact surface (e.g., Kuiper 1941), or the S_1 surface as we shall call it later. The point of contact L_1 is one of the Lagrangian points (or double points), which are special solutions of the restricted three-body problem. The contact point illustrates the case we have mentioned earlier that in some special cases $|d\vec{r}/dt|^2$ can be positive on both sides of a zero-velocity surface. This is true at L_1, say along the x-axis.

In Figure 1 we have illustrated several zero-velocity surfaces including the critical ones for the case $\mu=\frac{1}{3}$. Two cross-sections for each surface have been drawn, one in the xy-plane (in the lower diagram) and the other in the xz-plane (in the upper diagram). Because of symmetry we have shown here only one half of each cross-section, namely for positive y and positive z only. The case $C=4.1$ represents the situation before surfaces come into contact. Thus, this value of C corresponds to three distinct surfaces mentioned before and all shown in the figure. The case $C=C_1=3.946$ corresponds to the innermost contact (S_1) surface. It consists of two lobes which we shall call the primary lobe (around the $1-\mu$ component) and the secondary lobe (around the μ component). A large distorted cylindrical surface associated with the same C_1 is also shown in the figure.

The importance of the contact point is the fact that it forms a channel through which a test particle with C only slightly less than C_1 may move from the permitted region around one star to that around the other. As a result of this property we can conclude that the two lobes of the S_1 surface provides respectively an upper limit to the size of individual component stars because if either of them touches the surface, the matter will leak out at the point of contact into the other lobe. We may find examples for this situation in the Algol-type eclipsing binary systems in which one star (the less massive secondary) has usually a size comparable with the corresponding lobe of the S_1 surface while the other star (the more massive primary) is invariably small compared with its lobe of the same surface.

In reality the star must be smaller than the corresponding lobe of the contact surface as the particles (atoms, ions, electrons) in the surface

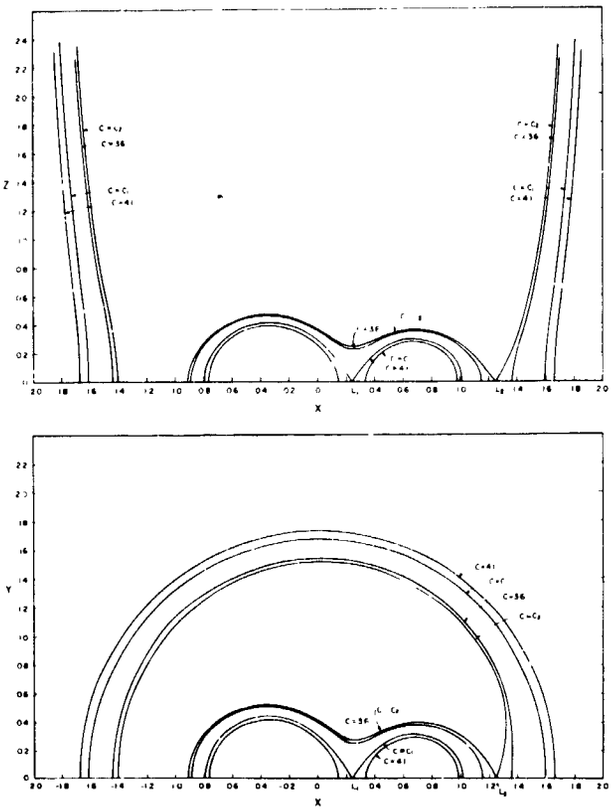


FIGURE 1.—Zero-velocity surfaces corresponding to $\mu=1/3$. The two cross-sections of several zero-velocity surfaces—one in the xy plane (below) and one in the xz plane (above) are illustrated here. The two contact surfaces are marked by C_1 (inner) and C_2 (outer) respectively with $C_1=3.95$ and $C_2=3.55$. It can be seen that for $C>C_1$ illustrated by the case $C=4.1$ each surface has three separate branches. For $C_2<C<_1$ (illustrated by the case $C=3.6$) there are two separate branches.

layer of a star are exercising all kinds of motion, thermal, turbulent or else. Therefore, they can escape the inner contact surface even before the stellar surface has reached the contact surface. However, the velocities in these motions (except of course, corpuscular radiation) may in general be regarded as small in the present consideration because the unity velocity in the present unit system is equal to $[G(M_1+M_2)/a]^{\frac{1}{2}}$ and amount to a few hundred km/sec in close binaries. Therefore, a velocity of a few tens of km/sec can be neglected

in equation (5.4), thus making the S_1 surface the actual limit of individual stars in a close system.

When C decreases from C_1, the two lobes of the innermost contact surface coalesce into one single dumb-bell like surface (corresponding to $C=3.6$ in Figure 1). They represent the appearance of two stars in actual contact such as in *W UMa* stars. Such a binary, though consisting of two stars, has a common envelope. A series of the contact configuration is possible, corresponding to different values of C. The smaller the value of C, the greater is the size of this dumb-bell like surface. On the other hand, the outer cylindrical like surface shrinks with decreasing C. Thus, C cannot decrease indefinitely without encountering another drastic change of the behavior of the zero-velocity surfaces. This time it is the contact between the dumb-bell like surface and the outer cylindrical-like surface. Kuiper has called it the outermost contact surface (or the S_2 surface) whose C value will be hereafter denoted by C_2 as is shown in Figure 1. The point of contact L_2 is another Lagrangian point. With this second contact, the inner region is connected to the outer region and the test particle with C only slightly less than C_2 can move without restriction due to energy. Thus, particles with C between C_1 and C_2 can move either inside the dumb-bell like surface or outside the distorted cylindrical surface. This leaves a forbidden region between the two surfaces. In Figure 1, this situation is illustrated by the case $C=3.6$. When $C<C_2$, there will not be any forbidden region for the particle from the consideration of the energy integral. Therefore, the S_2 surface represents the maximum size of a stable contact configuration of a close binary. For if the contact binary has reached this size, mass will continuously flow out of the system through L_2 and the binary is no longer stable.

With the aid of the S_1 surface we can now classify close binaries into three groups. If both components are well inside the S_1 surface, their system is said to be a detached one. If one component fills up the corresponding lobe of S_1 surface but the other is not, the two form a semi-contact (or semi-detached) binary. In the third group, i.e., the contact binaries, two components are in physical contact. More elaborate classification of close binaries may be found in Kopal's (1959) and Sahade's (1960 a and b) paper.

Finally, it may be noted that since the zero-velocity surfaces are defined in the rotating frame of reference, the Roche model predicts that axial rotation of stellar envelope and orbital revolution are synchronized, i.e., the two component stars revolve in their respective orbits face to face. This result appears to be valid in most close binaries (Swings 1936, Plaut 1959 Struve 1950). In a few cases, such as β Lyrae (Kopal 1959), where one or both component stars undergo secular expansion as a result of internal evolution, this synchronization could be temporarily violated.

We can now summarize the result briefly in the following way. The closed zero-velocity surfaces around two stars for $C>C_1$ represents the shape of two components that are detached from each other. The smaller the radius, the close to a sphere is the surface. What we are interested in, however, is only when the star approaches the size of the inner contact surface and is strongly distorted from a spherical shape. Thus, at $C=C_1$ each lobe of the S_1 surface represents the limiting size of the individual component. The star in a binary that is found actually in this limiting configuration is usually ejecting mass through L_1 towards its companion if the latter is still small compared with its own lobe of contact surface. For $C_1>/C>C_2$ the dumb-bell like surface of equation (5.5) represents the configuration of two component stars in physical contact. The limiting case of the contact configuration is given by the outer contact surface ($C=C_2$). Binary stars that are in this limiting configuration lose mass to outer space through the point, L_2. It can be seen from Figure 1 that the range of sizes of the contact systems (from the S_1 to S_2 surface) is quite small. Therefore, any binary that is in physical contact is usually losing mass through L_2. The problem of losing mass from a star through the points L_1 and L_2 was first extensively discussed by Kuiper (1941) (Cf. 6). Such an instability at the surface of the close binary stars gives rise to a number of observable phenomena which Martynov (1957 also Krat 1960) has summed up in a review together with examples of stars that show these various phenomena.

Because of their importance in the study of close binaries both S_1 and S_2 surfaces for different values of μ have been computed by Kuiper (1941); by Kopal, whose result was first published in 1954

and is now included in his book (1959); and by Kuiper and Johnson (1956) and more recently by Szebehely and Williams (1964). Table 2 gives C_1 and C_2 for 9 values of μ taken from Kuiper and Johnson's paper. Table 3 and 4 gives the intersection of the two contact surfaces with the xy-plane (first and second column) and with the xz-plane (first and third column) for those values of μ listed in Table 2. They are partly taken also

TABLE 2.—*Labelling Constants for the Contact Surfaces*

μ	C_1 (inner)	C_2 (outer)
.5	4	3.4567962
.4	3.9809086	3.5189346
1/3	3.9455706	3.5474582
2/7	3.9074840	3.5589342
.25	3.8706588	3.5611940
.20	3.8046532	3.5523932
.10	3.5969532	3.4666844
.05	3.4204164	3.3543942
.02	3.2523262	3.2257332

from Kuiper and Johnson's paper but partly computed especially for the use of the present occasion by Mr. C. Wade, Jr., of Goddard Space Flight Center. The case of $\mu=\frac{1}{3}$ is furthermore illustrated in Figure 1 as we have described. For diagrammatic illustrations of the contact surfaces corresponding to other values of μ we refer the reader to Kuiper's (1941) paper.

Finally, it should be noted that the basic size of a star, whether single or in a close binary, is determined by its internal structure. What we have said previously is only about the external shape of its envelope. That the envelope is all that we can observe of a star lies significance of the zero-velocity surfaces in the study of close binaries.

The Roche model of an infinitely centralized star represents only one end of a series of stellar models. At the other end there is the liquid model of a homogeneous density. Darwin (1906) and Jeans (1919) have studied the binary configurations and their stability based on the latter idealization. Darwin's and Jeans' results have recently been examined by Chandrasekhar (1964).

5.2 Departure from the Roche Model

An actual star is built neither on the infinite centralization nor on the homogeneous distribution of density. We have noted in 3.1 that the rate of apsidal motion decreases with the degree of central condensation. For component stars built on the Roche model the binary would show no apsidal motion at all. The fact that the apsidal motions in many binaries have been observed in relatively a short time of a few decades only shows how approximate the model is. However, for a prediction of the stellar surfaces the Roche model perhaps gives a good approximation. Indeed, according to a recent estimate by Plavec (1958), the departure of stars from the infinite central condensation of the Roche model does not produce a serious modification of the zero-velocity surfaces.

On the other hand, Plavec has pointed out in the same paper, that a deviation from the synchronization of axial rotation and orbital revolution would cause an important change in the shape of the contact surface. He has treated the problem of non-synchronization by simply modifying that term in U (given by equation [5.3]) which corresponds to the centrifugal force. In this way he has found that it is very simple to compute the modified contact surfaces. Unfortunately Plavec's process may not be regarded as legitimate because the centrifugal force arising from orbital motion and that arising from axial rotation do not have the same axis. This same problem has later been studied by Limber (1963) and Kruszewski (1963). The works by these two investigators are parallel and suffer the same difficulties.

In order to see the difficulties let us assume that axial rotation, say of the $1-\mu$ component, is not synchronized to the orbital motion, although its axis is still assumed to be perpendicular to the orbital plane. In considering the motion of a test particle in the envelope of the $1-\mu$ component we first translate the origin of the (xyz) coordinate system along the x-axis to the center of the $1-\mu$ component star by making the simple transformations:

$$x'=x+\mu,\ y'=y,\ z'=z, \tag{5.6}$$

and then rotate the $(x'y'z')$ system by a second transformation so that in the end it will be in

TABLE 3.—*Innermost Contact Surfaces for Different Values of* μ

x	$\pm y$	$\pm z$
	$\mu=0.5$	
±0.1	0.1456	0.1370
±0.2	0.2520	0.2388
±0.3	0.3205	0.3050
±0.4	0.3596	0.3426
±0.5	0.3740	0.3561
±0.6	0.3647	0.3465
±0.7	0.3288	0.3112
±0.8	0.2549	0.2398
±0.9	0.0595	0.0555
±0.9050	0.0000	0.0000
	$\mu=0.4$	
−0.8415	0.0000	0.0000
−0.8	0.1773	0.1650
−0.7	0.3065	0.2871
−0.6	0.3712	0.3494
−0.5	0.4037	0.3814
−0.4	0.4122	0.3902
−0.3	0.3987	0.3777
−0.2	0.3625	0.3431
−0.1	0.2996	0.2827
0.0	0.2025	0.1900
0.1	0.0649	0.0606
0.1416	0.0000	0.0000
0.2	0.0872	0.0819
0.3	0.2045	0.1941
0.4	0.2797	0.2670
0.5	0.3221	0.3081
0.6	0.3376	0.3227
0.7	0.3272	0.3120
0.8	0.2872	0.2727
0.9	0.2001	0.1878
0.9696	0.0000	0.0000
	$\mu=1/3$	
−0.8013	0.0000	0.0000
−0.8	0.0329	0.0303
−0.7	0.2772	0.2574
−0.6	0.3656	0.3415
−0.5	0.4144	0.3886
−0.4	0.4370	0.4110
−0.3	0.4380	0.4126
−0.2	0.4182	0.3940
−0.1	0.3759	0.3538
0	0.3068	0.2878
0.1	0.2028	0.1893
0.2	0.0591	0.0550
0.2374	0.0000	0.0000
0.3	0.0919	0.0865
0.4	0.2022	0.1925
0.5	0.2701	0.2585
0.6	0.3051	0.2925
0.7	0.3123	0.2990
0.8	0.2912	0.2779
0.9	0.2337	0.2216
1.0	0.0831	0.0781
1.0120	0.0000	0.0000
	$\mu=2/7$	
−0.7743	0.0000	0.0000
−0.7	0.2482	0.2288
−0.6	0.3578	0.3318
−0.5	0.4187	0.3901
−0.4	0.4513	0.4220
−0.3	0.4619	0.4328
−0.2	0.4524	0.4243
−0.1	0.4224	0.3960
0	0.3691	0.3453
0.1	0.2867	0.2672
0.2	0.1663	0.1543
0.3	0.0114	0.0106
0.3072	0.0000	0.0000
0.4	0.1283	0.1214
0.5	0.2205	0.2107
0.6	0.2729	0.2618
0.7	0.2939	0.2810
0.8	0.2861	0.2738
0.9	0.2453	0.2335
1.0	0.1455	0.1373
1.0415	0.0000	0.0000
	$\mu=.25$	
−0.7554	0.0000	0.0000
−0.7	0.2215	0.2027
−0.6	0.3502	0.3228
−0.5	0.4208	0.3899
−0.4	0.4609	0.4286
−0.3	0.4784	0.4460
−0.2	0.4761	0.4444
−0.1	0.4543	0.4241
0	0.4112	0.3834
0.1	0.3426	0.3185
0.2	0.2405	0.2225
0.3	0.0976	0.0903
0.3607	0.0000	0.0000
0.4	0.0587	0.0551
0.5	0.1738	0.1655
0.6	0.2422	0.2323
0.7	0.2751	0.2642
0.8	0.2776	0.2662
0.9	0.2482	0.2369
1.0	0.1704	0.1613
1.0628	0.0000	0.0000

TABLE 3.—*Innermost Contact Surfaces for Different Values of μ — Continued*

x	$\pm y$	$\pm z$	x	$\pm y$	$\pm z$
	$\mu = .2$			$\mu = .05$	
−0.7320	0.0000	0.0000	−0.7178	0.0000	0.0000
−0.7	0.1758	0.1590	−0.7	0.1510	0.1256
−0.6	0.3388	0.3090	−0.6	0.3733	0.3152
−0.5	0.4235	0.3885	−0.5	0.4859	0.4149
−0.4	0.4740	0.4367	−0.4	0.5595	0.4818
−0.3	0.5008	0.4628	−0.3	0.6085	0.5271
−0.2	0.5079	0.4702	−0.2	0.6385	0.5555
−0.1	0.4965	0.4600	−0.1	0.6524	0.5689
0	0.4659	0.4314	0	0.6511	0.5684
0.1	0.4136	0.3822	0.1	0.6347	0.5537
0.2	0.3339	0.3075	0.2	0.6023	0.5243
0.3	0.2178	0.1997	0.3	0.5515	0.4781
0.4	0.0619	0.0571	0.4	0.4778	0.4118
0.4381	0.0000	0.0000	0.5	0.3724	0.3187
0.5	0.0883	0.0831	0.6	0.2189	0.1875
0.6	0.1867	0.1784	0.7	0.0257	0.0232
0.7	0.2398	0.2304	0.7152	0.0000	0.0000
0.8	0.2583	0.2482	0.8	0.1024	0.0971
0.9	0.2444	0.2340	0.9	0.1554	0.1494
1.0	0.1895	0.1801	1.0	0.1572	0.1508
1.0909	0.0000	0.0000	1.1	0.1002	0.0947
			1.1412	0.0000	0.0000
	$\mu = .1$			$\mu = .02$	
−0.7050	0.0000	0.0000	−0.7585	0.0000	0.0000
−0.7	0.0758	0.0659	−0.7	0.2858	0.2231
−0.6	0.3332	0.2930	−0.6	0.4534	0.3616
−0.5	0.4434	0.3931	−0.5	0.5563	0.4508
−0.4	0.5125	0.4572	−0.4	0.6273	0.5143
−0.3	0.5559	0.4982	−0.3	0.6765	0.5593
−0.2	0.5797	0.5211	−0.2	0.7085	0.5891
−0.1	0.5862	0.5278	−0.1	0.7255	0.6053
0	0.5762	0.5192	0	0.7288	0.6088
0.1	0.5493	0.4945	0.1	0.7184	0.5997
0.2	0.5034	0.4522	0.2	0.6939	0.5776
0.3	0.4344	0.3887	0.3	0.6539	0.5415
0.4	0.3343	0.2977	0.4	0.5956	0.4893
0.5	0.1906	0.1698	0.5	0.5136	0.4174
0.6	0.0147	0.0134	0.6	0.3973	0.3187
0.6090	0.0000	0.0000	0.7	0.2224	0.1789
0.7	0.1151	0.1092	0.8	0.0058	0.0052
0.8	0.1819	0.1747	0.8035	0.0000	0.0000
0.9	0.2037	0.1960	0.9	0.0990	0.0945
1.0	0.1855	0.1775	1.0	0.1197	0.1149
1.1	0.1072	0.1012	1.1	0.0671	0.0360
1.1349	0.0000	0.0000	1.1258	0.0000	0.0000

TABLE 4.—*Outermost Contact Surfaces for Different Values of* μ

x	$\pm y$	$\pm z$
	$\mu=0.5$	
0	0.3261	0.2911
±0.1	0.3488	0.3129
±0.2	0.3948	0.3571
±0.4	0.4674	0.4247
±0.6	0.4819	0.4329
±0.8	0.4325	0.3763
±1.0	0.3019	0.2431
±1.1	0.1873	0.1382
±1.1984	0.0000	0.0000
	$\mu=0.4$	
−0.9300	0.0000	0.0000
−0.9	0.2705	0.2292
−0.8	0.3761	0.3271
−0.6	0.4794	0.4298
−0.4	0.5016	0.4561
−0.2	0.4580	0.4175
0	0.3544	0.3200
0.1	0.3046	0.2736
0.2	0.3000	0.2709
0.3	0.3387	0.3082
0.5	0.4120	0.3792
0.7	0.4255	0.3873
0.9	0.3678	0.3235
1.1	0.2137	0.1708
1.2	0.0670	0.0470
1.2308	0.0000	0.0000
	$\mu=1/3$	
−0.9165	0.0000	0.0000
−0.9	0.1251	0.1059
−0.8	0.3174	0.2754
−0.7	0.4108	0.3625
−0.5	0.5021	0.4526
−0.3	0.5150	0.4688
−0.1	0.4612	0.4200
0.1	0.3427	0.3092
0.2	0.2841	0.2556
0.3	0.2765	0.2510
0.4	0.3143	0.2889
0.6	0.3820	0.3538
0.8	0.3823	0.3493
1.0	0.3027	0.2646
1.1	0.2224	0.1850
1.2	0.0954	0.0717
1.2490	0.0000	0.0000

x	$\pm y$	$\pm z$
	$\mu=2/7$	
−0.8714	0.0000	0.0000
−0.8	0.2597	0.2246
−0.7	0.3829	0.3367
−0.5	0.5007	0.4497
−0.3	0.5314	0.4824
−0.1	0.4945	0.4500
0.1	0.3878	0.3505
0.2	0.3132	0.2816
0.3	0.2582	0.2328
0.4	0.2668	0.2440
0.5	0.3085	0.2857
0.6	0.3438	0.3199
0.8	0.3610	0.3332
1.0	0.2962	0.2634
1.2	0.1085	0.0848
1.2597	0.0000	0.0000
	$\mu=0.25$	
−0.8404	0.0000	0.0000
−0.8	0.2016	0.1737
−0.7	0.3581	0.3138
−0.6	0.4443	0.3941
−0.4	0.5294	0.4770
−0.2	0.5386	0.4887
0	0.4806	0.4359
0.2	0.3498	0.3146
0.3	0.2712	0.2434
0.4	0.2497	0.2177
0.5	0.2708	0.2501
0.7	0.3365	0.3141
0.9	0.3271	0.3004
1.1	0.2227	0.1928
1.2	0.1145	0.0916
1.2659	0.0000	0.0000
	$\mu=.2$	
−0.8014	0.0000	0.0000
−0.8	0.0395	0.0338
−0.7	0.3179	0.2706
−0.5	0.4933	0.4386
−0.3	0.5569	0.5011
−0.1	0.5504	0.4974
0.1	0.4761	0.4292
0.3	0.3238	0.2892
0.4	0.2393	0.2145
0.5	0.2208	0.2019
0.6	0.2589	0.2410
0.8	0.3106	0.2907
1.0	0.2747	0.2507
1.2	0.1169	0.0963
1.2710	0.0000	0.0000

TABLE 4.—*Outermost Contact Surfaces for Different Values of μ -- Continued*

x	$\pm y$	$\pm z$	x	$\pm y$	$\pm z$
	$\mu=.1$		0.1	0.6569	0.5657
			0.3	0.5766	0.4932
−0.7452	0.0000	0.0000	0.5	0.4087	0.3444
−0.7	0.2311	0.1954	0.6	0.2734	0.2300
−0.6	0.3966	0.3401	0.7	0.1340	0.1179
−0.4	0.5536	0.4837	0.8	0.1390	0.1295
−0.2	0.6147	0.5424	0.9	0.1759	0.1667
0	0.6104	0.5402	1.0	0.1788	0.1689
0.2	0.5416	0.4778	1.1	0.1418	0.1306
0.4	0.3897	0.3403	1.2	0.0464	0.0393
0.5	0.2739	0.2388	1.2281	0.0000	0.0000
0.6	0.1708	0.1526		$\mu=0.02$	
0.7	0.1772	0.1643			
0.8	0.2175	0.2051	−0.7717	0.0000	0.0000
0.9	0.2355	0.2223	−0.7	0.3177	0.2446
1.0	0.2248	0.2101	−0.6	0.4739	0.3735
1.1	0 1815	0.1653	−0.4	0.6420	0.5212
1.2	0.0926	0.0791	−0.2	0.7213	0.5945
1.2597	0.0000	0.0000	0.0	0.7410	0.6139
			0.2	0.7066	0.5832
	$\mu=.05$		0.4	0.6102	0.4967
			0.6	0.4188	0.3322
−0.7424	0.0000	0.0000	0.7	0.2562	0.2030
−0.7	0.2347	0.1913	0.8	0.0849	0.0746
−0.6	0.4136	0.3429	0.9	0.1133	0.1069
−0.5	0.5170	0.4342	1.0	0.1306	0.1241
−0.3	0.6328	0.5404	1.1	0.0934	0.0861
−0.1	0.6745	0.5805	1.1801	0.0000	0.0000

synchronization with the axial rotation of the star. Thus:

$$\begin{aligned} x' &= \xi \cos \omega t - \eta \sin \omega t, \\ y' &= \xi \sin \omega t + \eta \cos \omega t, \\ z' &= \zeta, \end{aligned} \tag{5.7}$$

where ω represents the rotational angular velocity of the star in the $(x'y'z')$ coordinate system. Therefore, the (ξ,η,ζ) coordinate system follows both the orbital revolution and axial rotation of the $1-\mu$ component. The equations of motion of the test particle in the (ξ,η,ζ) system can be easily derived by applying successively two transformations given by equations (5.6) and (5.7) to equation (5.2). If we denote by $\vec{\omega}$ the rotational angular velocity-vector, by $\vec{k}$ a unit vector in the same direction and by $\vec{r}_1$ the radius vector of the test particle from the origin o' the (ξ,η,ζ) system, the final result may be given by

$$\frac{d^2\vec{r}_1}{dt^2}+2(\vec{\omega}+\vec{k})\times\frac{d\vec{r}_1}{dt}=\text{grad } U_1+\mu\vec{s}, \tag{5.8}$$

where

$$U_1=\tfrac{1}{2}(\omega+1)^2(\xi^2+\eta^2)+\frac{1-\mu}{r_1}+\frac{\mu}{r^2}, \tag{5.9}$$

and $\vec{s}$ represents a time-dependent unit vector which has the following components·

$$-\cos \omega t, \sin \omega t, o. \tag{5.10}$$

It is now apparent that non-synchronization cannot be properly treated by simply replacing U in equation (5.2) by U_1 in equation (5.9) because equation (5.8) is not identical in form to equation (5.2). Because of the time-dependence of the

vector $\vec{s}$ and of r_2 which is now given by

$$r_2^2=(\xi-\cos\omega t)^2+(\eta+\sin\omega t)^2+\zeta^2, \quad (5.11)$$

we cannot derive a simple relation like equation (5.4) in the present coordinate system. Consequently, the potential function U_1 given by equation (5.9) predicts neither the equilibrium surfaces of the star in general nor its limiting surface in particular. A detailed discussion on this point and a proper treatment of non-synchronization will be given in a later paper.

Actually, the non-synchronization represents only one of several possibilities that will destroy the existence of zero-velocity surfaces. Indeed, under no circumstance can we define zero-velocity surface if the orbital motion of the binary is not circular or if a third star is present in the neighborhood (Huang 1964). In these cases the envelope of the component stars become unstable and stars themselves may not be expected to have stationary surfaces.

6. GASEOUS MOTION IN THE CLOSE BINARY SYSTEM

As a result of spectroscopic observations by Struve (1941, 1945 et seq. 1949a, 1950), Abhyankar (1959a, 1960) and many others, gaseous flow originating from stellar surfaces in the close binary system has been fully established. In the meantime Kuiper (1941) has examined the gaseous flow in the binary system from a theoretical point of view and introduced the idea of ejection of mass by the component stars through the Lagrangian points L_1 and L_2 discussed in the previous section. We may therefore, conclude that the components of some close binaries have touched the S_1 surface and their atmospheres become unstable such that gaseous particles are flowing out of the star either at L_1 or L_2. The conclusion is further strengthened by the photometric observations (Wood 1946, 1957; also Kopal 1959) which indicates that the relative radii of component stars in such binaries are indeed comparable to what would be expected from the two lobes of the S_1 surface for their respective mass ratios.

The establishment of gaseous flow in close binaries opened up a new field of theoretical study. Several approaches have since been advanced, namely (1) the orbital approach, (2) the hydrodynamic approach and (3) a rudimentary statistical approach. Needless to say, if the binary stars possess magnetic fields, a magnetohydrodynamic approach to the problem would be in order. However, unless we clearly know the nature of the magnetic field, we cannot proceed with our study. Consequently, no one has seriously treated the gaseous flow in the binary system as a magnetohydrodynamic problem although magnetic activities have been suggested rather quantitatively to account for the peculiar behavior of some binaries (Huang 1959). However, one of the effects of magnetic activities may be predictable and that is the braking of binary motion. Several mechanisms have been proposed for braking stellar axial rotation (e.g. Huang and Struve 1960) but the one proposed recently by Schatzman (1962) may be regarded as most effective. According to him, the angular momentum of axial rotation of a star is transferred outwards as mass is ejected out during strong magnetic activities. Similarly, we may propose that the same mechanism can dissipate the orbital angular momentum. If so, we wonder whether the many close binaries are indeed the result of braking due to magnetic activities in the early phase of their evolution.

6.1 The Orbital Approach

The motion of a particle in the binary system is computed according to equation (5.2), hence, the particle is treated as if it were a celestial body in classical mechanics. Kuiper (1941) studied the problem of gaseous motion in this way and his paper gives a simple but illuminating discussion that has often been referred to in the literature.

Kuiper first emphasizes the symmetry of the equation of motion with respect to the xy-plane. At least this symmetry will not be disturbed if an initial symmetry in the distribution of matter is assumed. Consequently the principal current should be symmetrical with respect to the plane. Furthermore, since two symmetrical currents in the z-direction would lead to a dissipation of energy not present in currents parallel to the xy-plane, Kuiper concludes that very probably the latter are the most important currents in the binary system. In what follows we shall consider only the motion in the xy-plane which incidently

is also the plane in which observations may be made in eclipsing binaries.

Kuiper next points out that because of the Coriolis force (the first term on the right side of equation (5.2)), a flow near L_1 in the positive x-direction will be deflected in the $-y$ direction. This can be seen easily from the vector product if we set the initial velocity $d\vec{r}/dt$ to be pointing in the $+x$ direction. The current will therefore go principally in the direction of $\alpha,\beta,\ldots$ as is shown in Figure 2.

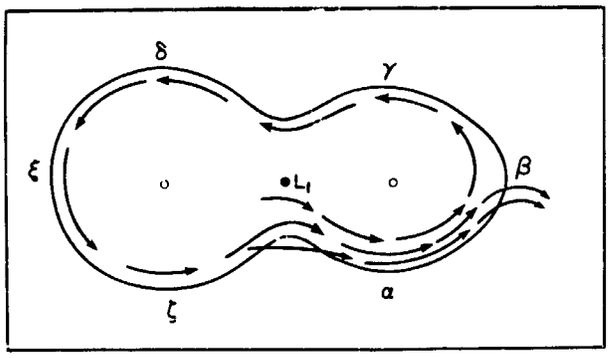


FIGURE 2.--Gaseous flow in the orbital plane expected in a close binary system (adopted from Kuiper's 1941 paper).

Numerical integrations of gaseous motion in a contact binary by Kuiper furthermore "show that if, the speed is below a certain value, the matter will go around the companion in the sense, $\alpha,\beta,\gamma\ldots$. If, however, the speed is large, the matter will fly off near β. The returning stream (in the $-x$ direction) will experience a Coriolis force in the $+y$ direction." The current would continue in the direction of δ,ξ,ζ as is shown in the figure with perhaps decreasing strength. Similarly, if the current starts in the negative x direction near L_1, the same flow pattern will result. Such a flow pattern is indeed consistent with what has been observed, say in β Lyrae.

Now if only one component of the binary is in contact with the S_1 surface, the gaseous flow pattern following the ejection of matter at L_1 will be affected by the Coriolis force in the same way as we have described for the contact binary. With frequent collisions among the ejected particles, such a tendency of motion may bring them to form a rotating ring or rings around the other component star as Kuiper has suggested. Indeed Kuiper's prediction was soon verified by Joy's (1942) discovery of a rotating gaseous ring around the primary component in the R W Tauri system —one of the Algol-type eclipsing binaries. Joy unaware of Kuiper's investigation, deduced the existence of the gaseous ring from the emission lines. This discovery induced Struve (e.g. 1950) and others to initiate an extensive study of the emission structure of many Algol type binaries. Emission rings have been since discovered in many systems. A detailed discussion of rings has been given by Struve and Huang (1957a) and a up-to-date list of such binaries may be formed in Sahade's (1960a) paper.

If the ejection velocity in the rotating system is small, the motion of particles near the point of ejection (namely L_1 and L_2) may be treated by linearized theory for the stability of orbits in the neighborhood of the Lagrangian points (Moulton 1914). Physically, we may imagine the region near the ejection point as a nozzle where the density is high and collisions are frequent. Consequently, we would not expect that any result derived from such a consideration will have any practical significance. If the ejection velocity is large, only direct numerical integration of equation (5.2) is possible. Abhyankar (1959b) has studied the motion in the neighborhood of these points in this way.

Kuiper has also investigated the motion of a particle farther away from the ejection points L_1 and L_2 by numerical integration. His pioneer work in this direction has since been followed by Kopal (1956, 1957b, 1959), Gould (1957, 1959) and many others after the introduction of the high speed digital computer. They have computed extensive series of orbits according to equation (5.2). Many orbits thus obtained are difficult to interpret because of their seemingly erratic behavior with loops, cusps, etc. which obviously will all be erased by collisions. A few, however, yield some interesting results. For example, some of the computed orbits indicate quite convincingly that particles ejected from the less massive component may indeed coalesce into a rotating ring if we properly take into account in our mental process the effect of collisions among the particles themselves.

6.2 The Hydrodynamic Approach

In spite of some success in the orbital approach, the motion of gaseous stream in a binary system

is not faithfully described by equation (5.2) since it does not include collisions and pressure. A more realistic treatment of this problem should at least take the collisions into consideration. Prendergast (1960) has examined the problem along this line. The hydrodynamic equation for the flow velocity $\vec{u}$ in space takes the following form in the rotating (x, y, z) coordinate system:

$$\frac{\partial \vec{u}}{\partial t}+\vec{u}\cdot\text{grad}\,\vec{u}+2\vec{k}\times\vec{u}=-\frac{1}{\rho}\text{grad}\,p+\text{grad}\,U \quad (6.1)$$

where U is given by equation (5.3) while p and ρ are respectively pressure and density in the gaseous medium. The unit vector $\vec{k}$ has the same meaning as in equation (5.2). The hydrodynamic equation of motion is supplemented by the equation of continuity.

$$\frac{\partial \rho}{\partial t}+d\omega(\rho\vec{u})=o \quad (6.2)$$

A comparison of equations (5.2) and (6.1) is helpful in our understanding of the difference between the two approaches. In the first place we have now introduced in equation (6.1) a new term grad p/ρ due to the pressure, which does not appear in the orbital study. Secondly, we now consider a continuous velocity field $\vec{u}(x,y,z,t)$ instead of $d\vec{r}/dt$ of individual particles. Thus, $d\vec{r}/dt$ is replaced by the first two terms in equation (6.1). We may take the continuity of the velocity field as a consequence of collisions among particles. This illustrates how by introducing the pressure and collisions we derive the hydrodynamic equation from equation (5.2). This point of view helps explain the statistical point of view of the problem that will be discussed in §6.3.

By considering a steady state ($\partial\vec{u}/\partial t=\partial\rho/\partial t=o$) and by neglecting the pressure term, Prendergast has obtained an approximate solution for the two dimensional case (in the xy-plane). He has assumed that the velocity component at right angles to the zero-velocity curve is small and to be neglected whenever convenient. Therefore, his solution is a very special one. But because of this assumption he can advantageously choose a coordinate system based on the zero-velocity curves themselves. Thus, he has found that the solution predicts circulatory currents around each of the two components inside the S_1 surface and around both components outside the S_2 surface. In this way it has been shown that the gaseous rings around the primary component of many Algol-type binaries is indeed a solution of the hydrodynamic equation.

6.3 A Statistical Point of View

In the present section we shall present a point of view from a statistical consideration. Its scope lies in the middle between the orbital approach (which neglects both pressure and collisions) and the hydrodynamic approach (which includes both). It may also be given some physical insight to the problem that cannot be seen by simply integrating equation of motion—be it equation (5.2) or (6.1). Also, it may develope into a stochastic method that treats the gaseous particles in the binary system like stars in a cluster (Chandrasekhar 1943) and eddies in turbulence (e.g., Chandrasekhar 1949). However, this is only a remote possibility.

At present, we can only give some elementary properties of gaseous particles in the binary system according to a recent paper by Huang (1965).

Let us consider n particles of mass, m_i ($i=1, 2...n$) moving in a binary system. Furthermore, the particle m_i corresponds at a given instant to a definite value of C_i for the Jacobian constant defined in equation (5.4). Because of collisions the individual C_i's of the particles change continually. However, since the particles are of atomic and subatomic sizes, the colliding particles at the instant of collision may be regarded as occupying the same point in space. If the total kinetic energy of the colliding particles is furthermore conserved during the collisions, it follows from the definition of C given by equation (5.4) that the new Jacobian constants denoted by C'_i for these particles after collisions should satisfy the following equation

$$\sum_{i=1}^{n} m_i C'_i=\sum_{i=1}^{n} m_i C_i \quad (6.3)$$

Thus, if we define an average $<C>$, we have

$$<C>=\frac{\sum_{i=1}^{n} m_i C}{\sum_{i=1}^{n} m_i}=\text{constant} \quad (6.4)$$

under the processes of elastic collisions. However, it may be noted that the dispersion of C_i's from their average value will in general change with collisions.

For inelastic collisions an equation connecting C_i's and C'_i's can always be obtained from the energy consideration if we know the detailed process of the collision. In the present consideration we shall assume that the collisions that take place among atomic particles in the binary system are statistically elastic, i.e., endoergic collisions balancing exoergic ones.

As a result of the constancy of $<C>$ during collision, the problem of gaseous flow is considerably simplified because we have now a macroscopic quantity, $<C>$, to deal with instead of following the courses of numerous particles in the system. This situation resembles the introduction of the concept of temperature and pressure which simplifies the study of the chaotic motion of molecules in gases in free space. Therefore, whatever is the nature of ejection that occurs on the stellar surface, the mean value of C_i's of ejected particles and their dispersion may serve as two of the most characteristic indices for the physical mode of ejection as regards the course of their subsequent motion.

It should be noted, however, that although the gaseous particles maintain a constant $<C>$, the mean flow does not follow the orbit derived from equation (5.2).

In this way, we can treat the problem of gaseous streams as a statistical problem of C_i's. One of the possibilities for further investigations along this line of thought is to link the statistical properties of C_i's from the theoretical consideration with the strengths of spectral lines arising from the gaseous streams observed at different phases of the binary motion.

Our immediate purpose is, however, to show that an intermediate approach between the orbital and hydrodynamic one may be obtained from a statistical consideration of C_i's as well as of the angular momentum (the z-component) per unit mass, h, of each particle in the system. Since a particle is moving in a two-centered field of force, its h, which is given by

$$h=x^2+y^2+x\frac{dy}{dt}-y\frac{dx}{dt}, \tag{6.5}$$

varies with time. However, it can be shown from equations (6.5) and (5.2) that

$$\frac{dh}{dt}=\mu(1-\mu)y\left(\frac{1}{r_1^3}-\frac{1}{r_2^3}\right), \tag{6.6}$$

which is a function of coordinates of the particle only, being independent of its velocity. What is more important is that the total angular momentum is conserved among the colliding particles if the collision may be regarded as taking place instantaneously.

We can now derive the properties of a continuous gaseous medium in the binary system from those derived from a consideration of discrete particles. Let us illustrate it by the two-dimensional case. Since the average values of C and h do not change by collision, we may write in a steady state

$$\vec{u}\cdot\nabla C=o \tag{6.7}$$

from equation (6.4) and

$$\vec{u}\cdot\nabla h=\mu(1-\mu)\left(\frac{1}{r_1^3}-\frac{1}{r_2^3}\right) \tag{6.8}$$

from equation (6.6) if we follow the streamline. While we have obtained these two equations by physical arguments, they can be easily derived from the hydrodynamic equation by first neglecting the pressure term. Naturally, equations (6.7) and (6.8) yield the same approximate solution of circulatory motions as Prendergast has obtained.

From equations (6.7) and (6.8) we can also see clearly the range of validity of the approximate solutions. As expected, the solution gives a good approximation for the circulatory motion close to each of the component stars. It has been illustrated by numerical calculation that wherever the solution represents a good approximation, the predicted motion from equations (6.7) and (6.8) approaches the periodic orbits of the restricted three-body problem (Huang 1964a). Thus, the gaseous ring observed in many binaries may be regarded equivalently either as a hydrodynamic flow or as motions of particles in a continuous series of periodic orbits that exist around each of the component stars.

Since we have emphasized the two physical parameters, h and C it is interesting to note that at some points in space certain combinations of

values for these two quantities are incompatible. In other words, with a given pair of values for h and C a particle cannot go into a certain region of space which may be called the forbidden zone (Huang 1965). Since h is not a constant of motion in the restricted three-body problem, the forbidden zone is not as important as the zero-velocity surfaces. However, it may serve some useful purposes in depicting the trend of motion in general, especially when it is applied in combination with equation (6.6).

7. THE ORIGIN OF CLOSE BINARIES

Essential difficulties encountered in various hypotheses of the formation of binary systems in general, have been discussed by Hynek (1951). Many of his arguments can be applied to the close binaries and are indeed followed in the present discussion. From a general ground we may state that the close binaries can be formed by one or more of the three possibilities. They were formed (1) as a result of the fission process in rapidly rotating single stars, (2) from distant binaries and (3) directly as close binaries in the beginning. We shall briefly discuss the plausibility of these three processes.

Jeans (1919, 1928, 1944) has spent many years in studying the fission process of rapidly rotating stars. He has found that when the rotational velocity is small, the shape of a flattened spheroid is common to all rotating bodies whether they are composed of gases, liquids or plastic material. However, if the rotational velocity is high, their shape depends greatly upon their internal constitution and especially the degree of their central condensation. According to him fission would occur in massive liquid bodies in which there is no appreciable central concentration of mass. But for a gaseous body of an extreme central condensation (i.e., the Roche model), rapid rotation only makes it flatten more and more with an accompanying loss of mass at its equator. No fission will result. Jeans further shows that all bodies having less than a certain critical degree of central condensation behave very much like those made of incompressible liquid while all bodies having more than this critical amount of central condensation behave very much like those of an infinite central condensation. Since actual stars have a central condensation much higher than the critical value, fission does not occur in them (Jeans 1944).

The fission theory encounters another difficulty in that it cannot explain the existence of binaries of large separations. Because of the conservation of angular momentum, it can be easily seen that two components of a binary resulting from the fission process must be very close together. Thus, in order to account for binaries of wider separations other mechanism of formation has to be derived. While it is not prohibitive to have two different mechanisms for the formation of one kind of objects, the smooth distribution of separations and other statistical behavior of binaries throws serious doubt, according to Kuiper (1935a and b), on any theory that cannot explain the formation of all binaries, except perhaps those really wide ones with separations of the order of interstellar distances (e.g. Van Biesbroeck 1957), by a single mechanism.

Next, let us examine the possibility of converting distant binaries into close ones. In proposing this possibility, the presence of distant binaries is assumed. Then two components are supposed to drift together due to a resisting medium such as we have seen in 3.4 or due to other energy dissipating mechanisms. In order to make the theory complete we must first of all answer the question of how these distant binaries are formed. It has often been suggested that they may be formed by star capture. However, stars are far apart in space and are moving with an average velocity of about 10–20 km/sec with respect to the local centroid. Thus, if one star by chance approaches another, each will move on an hyperbolic orbit. Consequently, the two will recede from each other to large distances unless their energy can be dissipated during the encounter. The agency which absorbs the dissipated energy can be either a third star that happens to be in the immediate neighborhood or a resisting medium. But the chance of a three-star encounter in the galaxy is vanishingly small and the interstellar medium is in general too tenuous to be effective in reducing stellar velocities (McCrea 1953; Dodd 1954).

For the same reason of low interstellar densities and high orbital velocities, it is very doubtful that the resisting force of an ordinary interstellar cloud

can greatly reduce the separation of a binary in a time scale, say of 10^{10} years. Therefore, the possibility of first forming binaries by star capture and then converting them into close ones is also ruled out.

This leaves us only the third possibility, i.e., the binaries—both close and distant—were formed as they are. It appears that we have not found any serious objection to this possibility. On the contrary many observed facts are consistent with it. For example, according to Kopal (1959) the relative frequency of close binaries found in the general galactic field is about 0.1 percent and is of about the same order of magnitude as that found in galactic clusters and stellar associations where stars are supposed to be formed. Indeed many eclipsing binaries have been identified to be members of galactic clusters (e.g., Sahade and Frieboes 1960) and of stellar associations (e.g., Kraft and Landolt 1959; Semeniuk 1962). If binaries of different separations were formed as they are, their high abundance in the galaxy strengthens the suggestion made frequently (e.g., Roberts 1957) that stars (at least Pop. I stars) are formed in groups. As a result of his statistical study Batten (1960) gives reasons to suggest that many binaries may have formed together in space. However, statistics also give us some puzzling results. Jaschek and Jaschek (1957, 1959) have found that although the frequency occurrence of spectroscopic binaries is uniform among the young stars (about 20 percent) whether they are in associations, in clusters or in the general field, the percentages of spectroscopic binaries in old groups of stars (ages from 3×10^9 to 6×10^9 years) decreases with age.

From the standpoint of star formation in general we also find the emergence of binaries quite natural. Two condensations just happened to be formed near together in the primeval medium and they will evolve to become a binary if their relative velocity is not large enough to escape from each other. Since the two adjacent condensations evolve separately on their own, they will become two stars as if they were chosen from a random sample. This prediction is again consistent with the observed result (Kuiper 1935 a and b). Here we may mention the interesting argument in favor of the simultaneous formation of two components of a close binary advanced by Krat (1952). He observed that if their formation did not occur at the same time, then the second component could never have been formed in the neighborhood of the first one because of the latter's strong tidal force and intense radiation both of which incline to disrupt the gaseous condensation that is to become the second star.

If the binary systems were formed as they are, it is inevitable to conclude that two components of a binary must have the same age. Previously this consequence was regarded as a difficulty, because one might argue, as with Jeans (1944) for example, that components of binaries like Sirius, Procyon, and others show signs of very different ages. Both Sirius and Procyon contain a white dwarf as one component together with an early-type main-sequence companion. Since the white dwarf is an old object while the early-type main-sequence star is regarded as young, it is difficult to believe that they could be coeval. However, it should be noted that whether a star is young or old refers only to its own evolutionary sequence and has nothing to do with time in the absolute measure. For example, a massive star of $10M_\odot$ or more passes through all of its evolutionary stages perhaps in a fraction of time that a star of $1M_\odot$ remains on the main sequence. Thus with a loss of mass at late stages either continuously (Deutsch 1956) or cataclysmically, there is no reason to object why one star has reached the white dwarf while the other of the same age is still at the main-sequence stage. This is a fortiori true for two components in a close binary, because, as we have seen, one component can accret mass from its companion. Since we may regard mass accretion as a rejuvenation process while mass loss as an aging process, such an exchange of mass will enhance the apparent age difference between two components in a close binary. More will be said about mass exchange and its effect on evolution of close binary stars in the next section.

Although we have concluded that binaries were formed in the way they now appear, we still do not know for sure the detailed mechanism how they actually emerged from the primeval nebula, as the study of formation of binaries, like that of single stars, is still at the speculative stage. Here we shall review briefly three suggestions regarding the physical processes that shape the binaries in general and close binaries in particular. As we

will see presently, these three suggestions are not mutually exclusive because each deals only one aspect of the many faceted problem of binary formation.

Dodd (1954) has advanced a model of binary formation after he has found that star capture in the present state of interstellar space is untenable. He takes for granted that stars are forming around various centers of condensation in the originally homogeneous medium where there are no local variations in the gravitational field. If two condensations happen to be nearby, the material in the cylindrical volume between the two condensations, according to Dodd, will collapse on the axis joining them mainly under the gravitational influence of the condensations themselves. Because the collapse on to the axis is cylindrical, energy dissipation is much more effective than would be the case of a spherical collapse. This cylindrical collapse results in a column of comparatively small cross-section extending between two condensations. When the latter moves towards each other under their gravitation, they sweep the mass between them. In acquiring mass in this way the condensations also meet resistance which dissipates the dynamical energy of the system. In the end two condensations become two component stars in a binary. It appears that Dodd has assumed the condensations to be at rest in the medium in the beginning. They form a binary instead of a single star only by a further assumption that the condensations will be deflected from the straight collision course by perturbation due to neighboring stars. Perhaps by making these two assumptions, Dodd has proposed an unrealistic model since without initial relative motions among the various condensations, one can only expect the collapse of all condensations into one big mass.

Kuiper (1935b, 1957) has studied the problem of angular momentum of binaries both observationally and theoretically. If one considers the proto-star as a single hydrodynamical unit, its Reynolds number is high and consequently motion inside it must be turbulent. Thus, its angular momentum may be computed from the random motion of large eddies as Kuiper has done. Since the motion is random, each of the three components of the total angular momenta of stars must be given by the Gaussian curve. In other words the magnitudes of the total angular momenta of proto-stars follow the Maxwellian distribution of velocity magnitudes, from which Kuiper is able to derive a theoretical distribution for the separations between two components in binaries. He has found that the theoretical distribution curve indicates a less dispersion than the observed one but has given several convincing reasons to show why this should be expected.

Parenthetically it may be noted that unaware of Kuiper's investigation, McCrea (1959 see also Struve and Zebergs 1962) has used practically the same arguments to derive the angular momenta of stars.

Finally, Huang (1957b) has suggested that solid prestellar nuclei may serve as the basis for stellar condensations. If a medium possesses a large amount of angular momentum and is composed of only gas, its contraction will lead to a rotational break-up at the equator. It will not form a binary as we have already mentioned in connection with the fission problem. However, the situation will be different if there are large solid bodies embedded in the gaseous medium. These large bodies must move very slowly in comparison with gaseous molecules for the same reason that the Brownian movement is slow compared with molecular motions. Consequently, they are gravitationally less stable than the gaseous substratum. It could therefore happen that these bodies coalesce into two revolving nuclei which serve later as two centers of condensation. As a result, a binary instead of a rapidly rotating single star is formed.

All these ideas are qualitative in nature and function no more than some vague suggestions. Much work needs to be done before any of them will develop into a complete and consistent theory. Furthermore, other ideas may come up in the future. Therefore, we cannot avoid the impression that the problem of binary formation is a wide open field for further investigations.

Let us now put aside the question of physical processes of formation and turn our attention to the phenomenological side of the problem. In his study of separations of binaries Kuiper (1935a and b) has realized that the distances of the major planets from the sun fall nearly in the mid-range of the binary separations from 10^{-2} to 10^{5} A.U. with a median value at about 20 A.U. Therefore,

Kuiper (1951) believes that the binaries and the planetary system form a single uniform group. In his way he has argued that the number of planetary systems in the galaxy must be considerable because binaries are numerous. His view is now generally accepted.

At the same time Struve (1949b, 1950; also Struve and Huang 1958) has raised some interesting points on the relation between binaries and planetary systems. Especially, he has suggested from a consideration of angular momentum that planetary systems may result from the mass dissipation of *W* Urase Majoris systems.

Whether Struve's intuition will turn out right we cannot tell at this moment. One point we are now certain is that the formation of close binaries and single stars are events with no intrinsic difference. We suggest that only a few varying parameters in the identical process of formation make all the different catagories of stars. One of the important parameters doubtless is the angular momentum. If a condensation has a low value of angular momentum, single stars neither rotating rapidly nor possessing planetary systems will be formed. If the angular momentum is high, rapidly rotating single stars, stars possessing planetary systems of close binaries will be formed. Which one of these possibilities will actually be followed depends presumably upon other parameters, among them we may mention density, total mass, turbulence, magnetic activities, and other (Huang 1965).

8. EVOLUTION OF CLOSE BINARY STARS

Concerning the evolution of close binary stars we are faced with many peculiar systems but with few tangible leads that may indicate their stages of evolution. Consequently, we have a wide choice of theories for the evolutionary scheme of close binary stars. While it is interesting to know the fascinating ideas about binary evolution advanced in the past decades, we shall follow a conservative course here in preparing this section not only because we do not have enough space but also because most of these ideas are tentative at best.

The difficulty in understanding the evolution of close binary stars can be easily seen from the fact that in spite of recent successes in the study of stellar evolution (e.g. Burbidge and Burbidge 1958; Schwarzschild 1958; Hayashi, Hōshi and Sugimoto 1962) we are still not certain about evolution of single stars after the red-giant stage. The reason for this uncertainty lies in the first place with the complications of various nuclear reactions that take place in different layers of the star. But a more serious one is the rate of loss of mass from the star at different phases at and after the red-giant stage. As a result of investigations by Deutsch (1956, 1960, 1961) there is no doubt that red giant or supergiant stars lose mass continuously but the exact amount is not certain. Coming back to the close binary stars we know even less about the rate of mass loss or gain at different stages.

Because of this uncertainty we shall discuss evolution of close binaries here only on some general ground. In the first place the component stars must follow the natural sequence of evolution of every star because of the continuous outflow of energy which is derived either from their gravitational potentials or from thermonuclear reactions. Since two components of the same binary are coeval as we have concluded in the previous section, the more massive component departs from the zero-age main sequence at a faster rate because of its higher luminosity than the less massive companion. According to Smak (1959) this is the reason why the primary components in close binaries are systematically greater than the secondaries of the same mass. However, unlike single stars we would expect that stars even at the main-sequence stage will lose mass if they are components of close binaries. This is true not only for the components in contact and semi-contact binaries but also for those in completely detached systems if prominence activities like those taking place on the solar surface are active there. Indeed the variations in period found in the detached and semi-detached systems (Kwee 1958) may well be due to this kind of mass variations.

When the star is well inside the S_1 surface, the loss of mass is slow and its internal structure will not be seriously different from that of a single star of the equivalent mass. Therefore, its structure may be derived by linear perturbation as has been performed by Morton (1960). The situation becomes quite different when one or both components reach the size of the corresponding lobe of the S_1 surface, such as in several important

categories of close binaries like β Lyrae variables, *W* Ursae Majoris variables, Algol-type variables, etc. In these cases, the rate of mass variation is presumably high and the structure of the overflowing component star at any time perhaps will find no parallel among single stars. For example, it is doubtful that the subgiant component of the Algol-type binaries has any resemblance in their internal structure to the single subgiant stars (Struve 1954). Reddish (1957) has suggested some special models for red giants in order to account for the process of mass loss.

If we cannot derive the structure of component stars of contact and semi-contact systems by applying perturbation to the normal single stars, can we calculate directly their sequence of evolution by assuming a given rate of mass loss? The prospect of doing so is not good. There are several intrinsic difficulties involved in this kind of calculations. In the first place the mass loss has an effect not only on the star itself but also on the separation and the contact surfaces which limit externally the radius of the overflowing component star. In other words, the structure of the component stars is coupled to the orbit of the binary itself. Now the contact surface is not uniquely determined by the amount of mass lost from the star. It depends critically on the mode of ejection that can vary greatly from case to case as we have already seen in 3.5. Thus, starting from one configuration of contact or semi-contact we do not know exactly what will be the configuration at the next moment even if we assume a definite rate of mass loss.

We also encounter difficulty basic to the calculation of stellar structure itself. The effect of mass loss produces not only a pressure inbalance but also a thermal inbalance in the interior. The time of adjustment to the equilibrium condition is therefore measured by the Kelvin scale of gravitational contraction instead of the much shorter time scale of pulsation (Crawford and Kraft 1957; Morton 1960; Schwarzschild 1962). In other words we no longer have the simple relation of energy loss (at the surface through radiation) balanced exactly by energy production in the interior by thermonuclear reactions. It follows that the stellar structure in the present case becomes a time-dependent problem instead of a problem of equilibrium.

Or we may put our arguments differently. According to the Vogt-Russell theorem (e.g., Chandrasekhar 1939), if the pressure, the opacity and the rate of generation of energy are functions of the local values of the density, the temperature and the chemical composition only, then the structure of a star is uniquely determined by the mass of the star and its chemical compositions. In other words, among other structural properties of a star, the radius is determined internally by the mass and the chemical composition. Now in the contact and semi-contact binaries, the radius of the overflowing component is further limited by the contact surface which is imposed externally. Therefore, in general, we cannot find a stellar model that can fulfill both internal and external condition for its radius. This is equivalent to say that no solution can be obtained from the consideration of equilibrium. This explains why we cannot derive the evolutionary sequence of models by computation for the overflowing component as we have done for the single stars.

After what has been said it becomes quite apparent that discussions on the evolution of binary stars are necessarily qualitative and highly speculative. We should always remember that these discussions represent only the preliminary probing of the possibilities but not the final verdict.

Since the secondary components of many *W* Ursae Majoris stars are over-luminous with respect to their masses by more than two magnitudes on the average, Kitamura (1959, 1960) has suggested that the extra energy radiated away is derived from the gravitational contraction of mass that is being accumulated on the secondary surface at the expense of the primary. Accordingly, he has built a series of gravitationally contracting stellar models that are accreting mass. However, he has not given any reason why the secondary captures mass from the primary instead of the reversed course, namely the primary captures mass from the secondary, for after all the gases are flowing around both components of these contact binaries. Therefore, the present writer personally inclines to regard Kuiper's (1948) interpretation of the departure from the mass-luminosity relation of *W* Ursae Majoris stars still worthy of our attention.

Because of the slowness of energy transfer in the star, the loss of mass at the surface will change

the internal structure and the luminosity only slowly (on the Kelvin time scale as we have said) Thus, we arrive as with Krat (1957) at the conclusion that the star evolves under the conditions of a constant luminosity and decreasing mass. Using a simple model Krat has shown that the process of mass ejection under these conditions will be a self-accelerating process. Hence, he went on to suggest that the ejection velocity will increase. In this way he has tried to understand the high velocities of ejected matter from the Wolf-Rayet stars. According to him "the evolution of massive stars proceeds from the stage of stable hot supergiants through the stage of Wolf-Rayet stars and red giants to the main sequence (thus approaching solar-type stars)" as a result of mass loss. It appears that Krat's view may have overstretched the idea of mass loss. However, he is not alone in putting the Wolf-Rayet stars before the main sequence (e.g., Sahade 1958).

Among the many interesting views concerning the evolution of close binary stars one that is appealing is about mass exchange between two components and the consequence as regards their evolution, proposed independently by Crawford (1955) and by Kopal (1955) for the Algol-type binaries. Consider in a detached close binary two components of different masses. Since the luminosity is proportional to about 4th power of the mass (Russell and Moore 1940), the more massive one evolves faster and consequently reaches the stage of hydrogen-exhaustion in the central core earlier than its companion. According to the current understanding of stellar evolution, once the hydrogen has been exhausted in the central core, its envelope expands. Eventually, it will reach the S_1 surface. Mass that flows out from the expanding component at near the point L_1 because of the contract condition will be collected at least partly by the other component. In this way the more massive component loses while the less massive one gains mass. According to Crawford this process can proceed until the originally more massive component becomes the less massive of the two. In this way, he explains why in the Algol-type binaries it is the less massive component that now fills up one lobe of the S_1 surface while the more massive component is small and stable.

While this is an appealing idea there remain some difficulties in applying it to the Algol-type binaries as was initially suggested by Crawford and Kopal. Let us first consider the difficulty raised by Struve and Huang (1958). If M_1 and M_2 are respectively the present masses of the primary and secondary component and if in their evolution, mass of an amount ΔM has been transferred from the secondary to the primary and mass of an amount δM has been dissipated into outer space by the secondary, the following inequality can be easily established from the condition that the initial mass of the secondary must be greater than that of the primary:

$$2\Delta M+\delta M>M_1-M_2 \qquad (8.1)$$

Now the primary of the Algol-type binaries is usually a normal main-sequence star of spectral type A and may be assigned a mass of $M_1=3M_\odot$. In some cases, M_2 has been found to be $0.2M_\odot$ or less (Sahade 1945, 1949). If we assume $M_2=0.2M_\odot$, we find from the inequality (8.1) that $\Delta M>1.4M_\odot$ if all mass dissipated by the secondary has been collected by the primary and $\Delta M=\delta M>14/15M_\odot$ if equal amounts are collected by the primary and escaped to outer space. Thus, eight-tenths to nine-tenths of the initial mass of the secondary must have been lost. If we remember that when the envelope of a star expands during its evolution after the main sequence the central core of more than one-tenth of the total mass of the star contracts, it is hard to see a star to lose eight-tenths to nine-tenths of its mass by the mechanism described above.

Indeed Kopal (1959) has finally rescinded this interpretation. In addition to the mass consideration, he has also called attention to the fact that there is a complete lack of binaries whose more massive component fills up the lobe of S_1 surface. However, the present writer does not regard this as a serious difficulty because the Kelvin time scale is short and the chance of discovery is small (Morton 1960). Kopal has further pointed out that there is not enough energy to lift the mass from the hydrogen-depleted central core of the secondary to the level of the primary surface, if more than eight-tenths of the total mass is to be transported. Therefore, his second objection arises from the difficulty of too much mass to be transferred which was discussed before. Kopal

has not provided a new explanation for the Algol-type binaries. However, he presumes on the observational ground that the less massive secondaries in these systems for some unknown reason begin to expand at a certain stage of their evolution before the more massive primaries will do so.

From these considerations this idea of mass transfer between two components does not appear promising for the interpretation of the Algol-type binaries. On the other hand, the mechanism is a plausible one because, as we have seen in 3.5, it is a necessary consequence of the slow mode of mass ejection. Indeed, we now have reasons to believe that it may be working in β Lyrae (Huang 1963a).

Thus far we have discussed only evolution of component stars in close binaries. There remains the question of evolution of the binary system itself. One can easily see that the two problems are closely related. Theoretically we cannot attack the problem of evolution of the binary system without first understanding evolution of its component stars. The difficulty encountered in the latter makes a theoretical investigation of the former impractical. Observationally, data are too scanty to lead us to any definite and plausible scheme of binary evolution. Struve (1950) once made an interesting suggestion along this line and Sahade (1960b) has recently rediscussed some of Struve's original ideas of deriving *W* Ursae Majoris stars from early-type (B or A) systems.

Another fascinating problem connected with close binaries concerns the nature of various peculiar objects in the galaxy. Struve often expressed the view that Wolf-Rayet stars, symbiotic stars, and many other peculiar objects may all be binaries and moreover derive their peculiarities from being binaries. As a result of recent investigations by Abt (1961) the metallic line stars may now be added to the list of peculiar objects which appear to exist only in binaries. Although Struve's suggestions have not been accepted by all astrophysicists, they have inspired many to look at the problem. After the discovery of the binary nature of Nova DQ Hercules (1934) by Walker (1954, 1956), the question has again been raised (Struve 1955) as to whether the novae and nova-like objects are always components of close binaries. This question has led to many interesting studies about the relationship between these exploding stars and binaries (Crawford and Kraft 1956; Huang 1956; Kraft 1962, 1964; Schatzman 1958). A review of extensive works in this field is beyond the scope of the present article.

Finally I would like to express my sincere thanks to Mr. Clarence Wade, Jr. for performing on the IBM 7094 computer some calculations used in this paper.

REFERENCES

Abhyankar, K. D., 1959a, Ap. J. Suppl., **4**, 157. 1959b, A.J., **64**, 163. 1960, Observatory, **80**, 26.

Abt, H. A., 1961, Ap. J. Suppl., **6**, 37.

Ahnert, P., 1959, A.N., **285**, 191.

Ambarzumian, V., 1937, A.J., U.S.S.R., **14**, 207.

Batten, A. H., 1960, Mém. Soc. Roy. Sci. Liège (5) **3**, 63.

Blaauw, A., 1961, B.A.N., **15**, 265.

Boersma, J., 1961, B.A.N., **15**, 291.

Brooker, R. A. and Olle, T. W., 1955, M.N., **115**, 101.

Brouwer, D., 1946, A.J., **52**, 57.

Brouwer, D. and Clemence, G. M., 1961, "Methods of Celestial Mechanics" (New York: Academic Press), chap. 11.

Brown, E. W., 1936, M.N., **97**, 56, 62, 116. 1937, Ibid., **97**, 388.

Burbidge, G. R. and Burbidge, E. M., 1958, "Handbuch der Physik," ed. S. Flügge, **51**, 134.

Chandrasekhar, S., 1933, M.N., **93**, 449. 1939, "An Introduction of the Study of Stellar Structure" (Chicago: University of Chicago Press). 1943, Rev. Mod. Phys., **5**, 1. 1944a, Ap. J., **99**, 25. 1944b, Ibid., **99**, 54. 1949, Ibid., **110**, 329. 1950, Radiative Transfer (Oxford: Clarendon Press). 1964, Ap. J., **140**, 599.

Chandrasekhar, S. and Münch, G., 1952, Ap. J., **115**, 103.

Cowling, T. G., 1938, M.N., **98**, 734.

Crawford, J. A., 1955, Ap. J., **121**, 71.

Crawford, J. A. and Kraft, R. P., 1956, Ap. J., **123**, 44.

Darwin, G. H., 1906, Phil. Trans. R. Soc. London A, **206**, 161, or Scientific Papers **3**, (Cambridge: Cambridge University Press 1910), p436.

Deutsch, A. J., 1956, Ap. J., **123**, 210. 1960, "Stellar Atmospheres", ed. J. L. Greenstein (Chicago: University of Chicago Press), chap. 15. 1961, "Aerodynamic Phenomena in Stellar Atmospheres", ed. R. N. Thomas (I.A.U. Symposium No. 12), p. 238.

Dodd, K. N., 1954, M.N., **114**, 664.

Eddington, A. S., 1923, "The Mathematical Theory of Relativity" (Cambridge: Cambridge University Press).

Eggen, O. J., 1948, Ap. J., **108**, 1.

Fesenkov, V. G., 1949, A.J. U.S.S.R., **26**, 67. 1956, ibid., **33**, 614.

Gould, N. L., 1957, Pub. A.S.P., **69**, 541. 1959, A.J., **64**, 136.

Härm, R. and Rogerson, J., 1955, Ap. J., **121**, 439.

Hayashi, C., Hōshi, R. and Sugimoto, D., 1962, Progress of Theoretical Phys. Suppl., No. 22.

Hu, N., 1947, Proc. Roy. Irish Acad., **A51**, 87.
Huang, S. -S., 1956, A.J., **61**, 49. 1957a, J.R.A.S. Canada, **51**, 91. 1957b, Pub. A.S.P., **69**, 427. 1958, Ibid., **70**, 473. 1959, Annales d'Ap., **22**, 527. 1962, Ap. J., **136**, 903. 1963a, Ibid., **138**, 342. 1936b, Ibid., **138**, 471, 481. 1964a, 1964b, 1965.
Huang, S. -S. and Struve, O., 1956, A.J., **61**, 300. 1960, "Stellar Atmospheres", ed. J. L. Greenstein (Chicago: University of Chicago Press), chap. 8.
Hynek, J. A., 1951, Astrophysics (New York: McGraw-Hill Book Co.), chap. 10.
Irwin, J. B., 1952, Ap J., **116**, 211. 1959, A.J., **64**, 149. 1962, Astronomical Technique, ed. W. A. Hiltner (Chicago: University of Chicago Press) chap. 24.
Jaschek, C. and Jaschek, M., 1957, Pub. A.S.P., **69**, 546. 1959, Zs. f. Ap., **48**, 263.
Jeans, J. H., 1919, "Problems of Cosmogony and Stellar Dynamics" (Cambridge: Cambridge University Press). 1924, M.N., **85**, 2. 1925, Ibid., **85**, 912. 1928, "Astronomy and Cosmogony" (Cambridge: Cambridge University Press). 1944, "The Universe Around Us," 4th ed. (Cambridge: Cambridge University Press).
Jeffreys, H., 1918, M.N., **78**, 424.
Joy, A. H., 1942, Pub. A.S.P., **54**, 35. 1947, Ibid., **59**, 171.
Keller, G., 1948, Ap. J., **108**, 347.
Kiang, T., 1963, M.N., **123**, 359
Kitamura, M., 1959, Pub. A.S. Japan, **11**, 216. 1960, Ibid., **12**, 1.
Kopal, Z., 1954, Trans. I.A.U., **8**, 631. 1955, Mém. Soc. Roy. Sci. Liège, **15**, 684. 1956, Ann. d'Ap., **19**, 298. 1957a, Trans I.A.U., **9**, 599. 1957b, "Non-Stable Stars," ed. G. H. Herbig, (Cambridge: Cambridge University Press), chap. 17. 1959, "Close Binary Systems" (London: Chapman and Hall).
Kopal Z. and Shapley, M. B., 1956, Jodrell Bank Ann., **1**, 139.
Kraft, R. P., 1962, Ap. J., **135**, 408. 1964, Ap. J., **139**, 457.
Kraft, R. P. and Landolt, A. U., 1959, Ap. J., **129**, 287.
Kraft, R. P., Matthews, J. and Greenstein, J. L., 1962, Ap. J., **136**, 312.
Krat, V. A., 1950, Pulkovo Bull. **18**, No. 4, 1. 1952, Ibid., **19**, No. 2, 1. 1957, "Non-Stable Stars" (I.A.U. Symposium No. 3), ed. G. H. Herbig, 159. 1960, Veriable Star Bull., **11**, 359.
Kruszewski, A., 1963, Acta Astronomica, **13**, 106.
Kuiper, G. P., 1935a, Pub. A.S.P., **47**, 15. 1935b, Ibid., **47**, 121. 1941, Ap. J., **93**, 133. 1942, Ap. J., **95**, 201. 1948, Ibid., **108**, 541. 1951, Astrophysics, ed. J. A. Hynek, (New York: McGraw-Hill Book Co.) chap. 8. 1955, Pub. A.S.P., **67**, 387.
Kuiper, G. P. and Johnson, J. R., 1956, Ap. J., **123**, 90.
Kukarkin, B. V. and Parenago, P. P., 1963, "Basic Astronomical Data", ed. K. Aa. Strand (Chicago: University of Chicago Press), chap. 18.
Kuskwaha, R. S., 1957, Ap. J., **125**, 242.
Kwee, K. K., 1958, B.A.N., **14**, 131.
Landau, L. and Lifspitz, E., 1951, "The Classical Theory of Fields" (Cambridge: Addison-Wesley Press), chap. 11.
Leimanis, E. and Minorsky, N., 1958, "Dynamics and Non-linear Mechanics" (New York: J. Wiley & Sons, Inc.).
Levi-Civita, T., 1937, Am. J. Math., **59**, 225.
Limber, D. N., 1963, Ap. J., **138**, 1112.
Luyten, W. J., Struve, O. and Morgan, W. W., 1939, Pub. Yerkes Obs., **7**, 251.
Lyttleton, B. A., 1934, M.N., **95**, 42.
McCrea, W. H., 1953, M.N., **113**, 162. 1959, Modèles d'Ésoiles et Évolution Stellaire (Comm. Coll. Intern. d'Ap. Liege), p. 332.
McLaughlin, D. B., 1934, Pub. Obs. U. Michigan, **6**, 3.
Martynov, D. Y., 1938, Izvestia Engelhardt Obs. Kazan, No. 25. 1957, "Non-Stable Stars" (I.A.U. Symposium No. 3), p. 138.
Masevich, A. G., 1949, A.J. U.S.S.R., **26**, 207.
Merrill, J. E., 1964, I.A.U. Draft Reports, 635.
Morton, D. C., 1960, Ap. J., **132**, 146.
Motz, L., 1941, Ap. J., **94**, 253. 1950, Ibid., **112**, 434. 1952, Ibid., **115**, 562. 1953, Ibid., **118**, 147.
Moulton, F. R., 1914, "An Introcution to Celestial Mechanics" 2nd ed. (New York: Macmillan), chap. 8.
O'Connell, D. J. K., 1960, Trans. I.A.U., **10**, 619. 1962, Ibid., **11A**, 477.
Osterbrock, D. E., 1952, Ap. J., **116**, 164.
Petrie, R. M., 1962, "Astronomical Techniques," ed. W. A. Hiltner, (Chicago: University of Chicago Press), chap. 23.
Pike, J. N., 1955, Ap. J., **122**, 202.
Piotrowski, S. L., 1964, Acta Astronomica **14**, 000.
Plaut, L., 1950, Pub. Kapteyn Astr. Lab. Groningen No. 54. 1953, Ibid., No. 55. 1959, Pub. A.S.P., **71**, 167.
Plavec, M., 1958, Liège Symposium No. 8 on Emission-line Stars, 411. 1960, Bull. Astr. Inst. Csl., **11**, 154.
Plummer, H. C., 1918, "An Introductory Treatise on Dynamical Astronomy" (Cambridge: Cambridge University Press).
Poincare, H., 1911, "Leçons sur les Hypothèses Cosmogoniques" (Paris: A. Herman et fils), p. 117.
Pendergast, K. H., 1960, Ap. J., **132**, 162.
Prikhod'ko, A. E., 1961, A.J. U.S.S.R., **38**, 927.
Reddish, V. C., 1957, Observatory, **77**, 202.
Roberts, M. S., 1957, Pub. A.S.P., **69**, 59.
Rudkjobing, M., 1959, Ann. d'Ap., **22**, 111.
Russell, H. N., 1928, M.N., **88**, 641.
Russell, H. N. and Moore, C. E., 1940, 'The Masses of the Stars" (Chicago: University of Chicago Press).
Sahade, J., 1945, Ap. J., **102**, 474. 1949, Ibid., **109**, 439. 1958, Observatory, **78**, 79. 1960a, "Stellar Atmospheres" ed. J. L. Greenstein (Chicago: University of Chicago Press), chap. 12. 1960b, Mém. Soc. Roy. Sci. Liège (5), **3**, 76.
Sahade, J. and Frieboes, H., 1960, Pub. A.S.P., **72**, 52.
Schatzman, E., 1958, Ann. d'Ap., **21**, 1. 1962, Ibid., **25**, 18.
Schneller, H., 1962, Mitt. Sternw. Budapest, **53**, 28.
Schwarzschild, M., 1958, "Structure and Evolution of the Stars" (Princeton: Princeton University Press). 1962, Trans. I.A.U., **11B**, 137.
Semeniuk, I., 1962, Acta Astronomica **12**, 122.

SLAVENAS, P., 1927, Trans. Astr. Obs. Yale U., **6**, 35.
SMAK, J., 1959, Acta Astronomica, **9**, 13.
SPITZER, L., 1958, Ap. J., **127**, 17.
SPITZER, L. and SCHWARZSCHILD, M., 1951, Ap. J., **114**, 385.
STERNE, T. E., 1939a, M.N., **99**, 451. 1939b, Ibid., **99**, 662. 1939c, Ibid., **99**, 670.
STRUVE, O., 1941, Ap. J., **93**, 104. 1945, Ibid., **102**, 74. 1949a, M.N., **109**, 487. 1949b, Proc. Nat. Acad. Sci. (Washington), **35**, 161. 1950, "Stellar Evolution" (Princeton: Princeton University Press). 1954, Mém. Soc. Roy. de Liège, **14**, 236. 1955, Sky and Telescope, **14**, 275. 1958, Pub. A.S.P., **70**, 5.
STRUVE, O. and HUANG, S. -S., 1957, Occasional Notes R.A.S., **3**, 161. 1958, "Handbuch der Physik," ed. S. Flügge, **50**, 243.
STRUVE, O and ZEBERGS, V., 1962, "Astronomy of the 20th Century" (New York: Macmillan Co.)
SWINGS, P., 1936, Zs. f. Ap., **12**, 40.
SYNGE, J. L., 1960, "Relativity: The General Theory" (Amsterdam: North-Holland Publ. Co.).
SZEBEHELY, V. and WILLIAMS, C. A., 1964, A. J., **69**, 460.
TAKASE, B., 1953, Am. Tokyo Astr. Obs. (2), **3**, 192.
TIDMAN, D. A., 1958, Phys. Rev., **112**, 1759.
TISSERAND, F., 1891, Traité de la Mecanique Céleste (Paris: Gauthier-Villars 1960) Vol. 2, chap. 13–18.
TUNG, F. and CHANG, Y. C , 1957, Acta Astr. Sinica **5**, 239.
VAN BIESBROECK, G., 1957, J.R.A.S. Canada, **51**, 35.
VARSAVSKY, C. M., 1962, "Symposium on Stellar Evolution" ed. J. Sahade (Astr. Obs. La Plata, Argentina) p. 173.
WALKER, M. F., 1954, Pub. A.S.P., **66**, 230. 1956, Ap. J., 123, 68.
WEAVER, H. J., 1958, Science, **128**, 1149.
WOLTJER, J., 1922, B.A.N., **1**, 93.
WOOD, F. B., 1946, Contr. Princeton U. Obs. No. 21. 1950, Ap. J., **112**, 196. 1957, "Non-Stable Stars" (I.A.U. Symposium No. 3), ed. G. H. Herbig, 144. 1963, "Basic Astronomical Data," ed. K. Aa. Strand (Chicago: University of Chicago Press), chap. 19.
WOOLF, N. J., 1962, M. N., **123**, 399.

EFFECT OF NON-SYNCHRONOUS ROTATION* ON CLOSE BINARY STARS

N66 37948

SU-SHU HUANG†

The effect of non-synchronous rotation of the component star in close binaries on its limiting surface is analyzed and the difficulties in the previous investigations are examined. It has been shown that no potential function exists in a coordinate system following the axial (non-synchronous) rotation of the component star. Consequently, no equilibrium surfaces can be obtained in this way. It is therefore suggested that we should retain the conventional coordinate system that rotates with the binary motion. Since a simple energy integral exists in this coordinate system, we can obtain a physical picture for the problem of the non-synchronously rotating star in the binary systems. Indeed we have found that the non-synchronously rotating star will modify its rotation gradually and become eventually synchronous with its orbital revolution.

In this multifarious activities and achievements that make him one of the greatest astronomers in the modern time, Professor Otto Struve has never forsaken the binary stars. After having spent some years of successful researches in other fields of astronomy now and then he always returned to this field of his early interest. Therefore, few astronomers have observed spectroscopically as many binary stars as he has done. Fewer still have left such a distinct mark as his in the history of double stars.

In a close association and collaboration with him for a decade during which I am proud to have received his confidence and learned the process of his mental perceptions and responses as a result of our daily afternoon meetings in cafes outside the North Gate of the Berkeley Campus of the University of California, I have always been impressed by his open mindedness to new ideas, his enthusiasm for and devotions to astronomy. Because of his open mindedness he saw any problem in its multi-faceted angles, thus making him not only an outstanding scientist but also a great leader. Because of his enthusiasm, he enlisted many astronomers into the fields of his interest. Our daily contact also induced me to become a novice in the study of binary stars in the mid fifties: It is therefore befitting for me ten years later to write this article on binary stars in this memorial volume as my attribute to a great leader as well as a dear friend whom I have both respected and admired.

I. A CRITICAL REVIEW OF THE PREVIOUS INVESTIGATIONS

The shape of stellar surfaces in the close binary system has been given as a first approximation by what has been predicted by the Roche Model (e.g. Struve and Huang 1957. According to this model, the axial rotation of component stars and their orbital revolution are synchronized. While the majority of close binaries indeed obey the rule of synchronization (Swings 1936, Struve 1950, Plaut 1959) there are some exceptions to which belongs β Lyrae—a peculiar binary system that Professor Struve (1941, 1958) had a life-long interest. Therefore, in recent years attempts have been made to predict the stellar surfaces, especially the limiting surface, when the component star does not rotate in synchronization with its orbital revolution (Kopal 1956, Plavec 1958, Kruszewski 1963, and Limber 1963).

In the restricted three-body problem one can define a potential function U in a coordinate system that rotates with the binary's orbital motion

*Prepared for a Memorial Volume for Prof. Otto Struve.

†Goddard Space Flight Center and Catholic University, Washington, D. C.

(e.g., Moulton 1914). Also an energy integral can be obtained in the form that the sum of the potential energy and the kinetic energy is constant. It is the existance of the energy integral in this form that enables us to predict the stellar surfaces of close binaries because only then the density and the pressure should be constant on any equi-potential surface. Thus, we should emphasize that the mere existence of a potential function does not necessarily warrant a prediction of the equilibrium surface of a star. Indeed, this is the reason that the calculations by previous investigators have to be regarded as unsatisfactory.

Kopal (1956) first studied the limiting surface of the non-synchronously rotating star, using a potential function which Kruszewski (1963) has since pointed out to be incorrect. Later Plavec (1958), Kruszewski (1963) and Limber (1963) have all derived a potential function and obtained the limiting surface. However, even though their potential function is correct, their result is still open to debate as we may see in the following way.

Let us first choose a dimensionless system of units of measurement with the total mass of the binary as the unit of mass, with the separation between the two components as the unit of length and with $P/2\pi$ as the unit of time, where P stands for the orbital period. Thus, if we denote $1-\mu$ as the mass of one component, μ will be the mass of the other. Let us further choose a rotating (x,y,z) system such that the origin is at the center of the $1-\mu$ component, the x-axis points always towards the μ component, and the xy plane coincides with the orbital plane. We have the equations of motion for a test particle in the (x,y,z) system:

$$\frac{d^2x}{dt^2}-2\frac{dy}{dt}=\frac{\partial U_1}{\partial x},\quad \frac{d^2y}{dt^2}+2\frac{dx}{dt}=\frac{\partial U_1}{\partial y},\quad \frac{d^2z}{dt^2}=\frac{\partial U_1}{\partial z}, \tag{1}$$

where

$$U_1(x,y,z)\equiv\frac{1}{2}[(x-\mu)^2+y^2]+\frac{1-\mu}{r_1}+\frac{\mu}{r_2} \tag{2}$$

and r_1 and r_2 are respectively the distances of the test particles from the two component stars. Note that U_1 is a function of time only through the coordinates (x,y,z) of the test particle.

We now introduce a new rotating (ξ,η,ζ) system such that it is rigidly fixed to the $1-\mu$ component, sharing the latter's axial rotation. Although the axial rotation of the $1-\mu$ component is not synchronous with the orbital motion, its rotating axis (chosen as the ζ-axis) is still assumed to be perpendicular to the orbital plane. Thus, if the (ξ,η,ζ) system is rotating with an angular velocity, ω, with respect to the (x,y,z) system, it rotates with an angular velocity, $1+\omega$, with respect to the rest frame of reference. The transformation equations between (x,y,z) and (ξ,η,ζ) are as follows:

$$x=\xi\cos\omega t-\eta\sin\omega t,$$
$$y=\xi\sin\omega t+\eta\cos\omega t,\ z=\zeta \tag{3}$$

The equations of motion in the (ξ,η,ζ) system are

$$\frac{d^2\xi}{dt^2}-2(\omega+1)\frac{d\eta}{dt}=-\mu\cos\omega t+\frac{\partial U_2}{\partial\xi} \tag{4}$$

$$\frac{d^2\eta}{dt^2}+2(\omega+1)\frac{d\xi}{dt}=\mu\sin\omega t+\frac{\partial U_2}{\partial\eta} \tag{5}$$

$$\frac{d^2\zeta}{dt^2}=\frac{\partial U^2}{\partial\zeta} \tag{6}$$

where

$$U_2(\xi,\eta,\zeta)\equiv\frac{1}{2}(\omega+1)^2(\xi^2+\eta^2)+\frac{1-\mu}{r_1}+\frac{\mu}{r_2} \tag{7}$$

may be regarded as the negative value of the potential function. Acutally, it is a misleading name as we shall see presently.

$U_2(\xi,\eta,\zeta)$ as given by equation (7) does not give the limiting surface of the non-synchronous $1-\mu$ component. In the first place, we encounter the time-dependent terms $\mu\cos\omega t$ and $\mu\sin\omega t$ in equations (4) and (5). Secondly r_2, which is given by

$$r_2^2=(\xi-\cos\omega t)^2+(\eta+\sin\omega t)^2+\zeta^2, \tag{8}$$

is now an explicit function of time. Therefore $U_2(\xi,\eta,\zeta)$, unlike U_1 in the (x,y,z) system, is no longer an implicit function of time through the space coordinates alone but involves time explicitly. For brevity we write $U_2(t)$ for $U_2(\xi,\eta,\zeta)$. Because of the two complications, we can no longer derive from equations (4)–(6) the simple result that the sum of the potential energy and the kinetic energy is a constant of motion. Consequently, the surfaces defined by equation (7) do not represent equilibrium surfaces of the star. We can see this point analytically in the following way.

Multiplying equations (4)–(6) respectively by $d\xi/dt$, $d\eta/dt$ and $d\zeta/dt$ and adding the resulting equations together, we obtain after integration with respect to time that

$$\frac{1}{2}[V^2(t)-V^2(t_o)]-[U_2(t)-U_2(t_o)]$$
$$=\int_{t_o}^{t}\left[\mu\left(\frac{d\eta}{dt}\sin\omega t-\frac{d\xi}{dt}\cos\omega t\right)-\frac{\partial U_2}{\partial t}\right]dt, \qquad (9)$$

where $V(t)$ and $V(t_o)$ denote respectively the velocity of the test particle at time t and at the initial time t_o. $\partial U_2/\partial t$ may be evaluated from equations (7) and (8). When the result is substituted into equation (9), we obtain finally

$$\frac{1}{2}[V^2(t)-V^2(t_o)]-[U_2(t)-U_2(t_o)]$$
$$=\mu\int_{t_o}^{t}\left[\left(\frac{d\eta}{dt}-\frac{\omega\xi}{r_2^3}\right)\sin\omega t-\left(\frac{d\xi}{dt}+\frac{\omega\eta}{r_2^3}\right)\cos\omega t\right]dt. \qquad (10)$$

It is obvious from equation (10) that the sum of the potential and kinetic energy is no longer constant. Since the integral in equation (10) actually depends upon the path of the test particle, $-U_2$ does not behave exactly like a potential.

The previous investigators have introduced

$$U_3(\xi,\eta,\zeta)=U_2(\xi,\eta,\zeta)+\mu(\eta\sin\omega t-\xi\cos\omega t). \qquad (11)$$

In doing so, they can replace the right hand sides of equations (4)–(6) respectively by $\partial U_3/\partial\xi$, $\partial U_3/\partial\eta$, and $\partial U_3/\partial\zeta$. Because of the transformation relation (3), $U_3(\xi,\eta,\zeta)$ can now be expressed in terms of x, y and z. In this new expression for U_3, time appears only implicitly through x,y,z but does not enter explicity. Consequently, they have independently derived the limiting surface in the (x,y,z) system by considering U_3 as the potential function.

However, we should remember that it is the (ξ,η,ζ) system that is rigidly fixed to the rotating star. Any stationary surface of the non-synchronous rotating star ($1-\mu$ component) must be expressed as a function of ξ, η and ζ alone without the explicit appearance of time. It does not appear that an expression in x, y, and z can represent the stationary surfaces of the non-synchronous component because in the (x,y,z) system the component has a net rotation. Since, as we have seen before, we cannot obtain a potential function in the $(\xi\eta,\zeta)$ system, we have no means to compute the stationary surfaces of a non-synchronous rotating component star, a conclusion contrary to the previous investigators.

II. PHYSICAL INTERPRETATION

Because of the difficulties we have just observed, it is advisable to retain the (x,y,z) coordinate system that rotates with the binary motion. In this coordinate system we have for the test particle a simple energy integral

$$2U_1-V^2=C \qquad (12)$$

where $U_1(x,y,z)$ is given by equation (2).

We should now recall that the zero-velocity surfaces are labelled by C (e.g., Moulton 1914). Let the C value that is associated with the innermost contact surface S_1 (Kuiper 1941) be C_1. Therefore, all particles inside the S_1 surface with $C<C_1$ could penetrate the S_1 surface. As a result, these particles will gradually be lost from the star. Now the $1-\mu$ component is rotating with an angular velocity ω in the x,y,z coordinate system along the z-axis. Particles in the stellar surface layers have velocities

$$V=(x^2+y^2)^{\frac{1}{2}}\omega \qquad (13)$$

if thermal motion or other kinds of motion are neglected. Hence, all particles in the $1-\mu$ component star that are above the surface (called the R surface hereafter) given by

$$2U(x,y,z)-\omega^2(x^2+y^2)=C_1 \qquad (14)$$

and consequently have C values less than C_1 may be regarded as unstable and could easily escape out of the S_1 surface. Figure 1 illustrates the cross sections in both the xy plane and the xz plane of a few R surfaces for the case $\mu=0.4$ and $\omega=0$, 2, 4 and 8. The case $\omega=0$ is simply the S_1 surface for $\mu=0.4$. The elongated shape of the cross sections in the xz plane is easy to understand because those particles near the equatorial plane that have the highest linear velocities due to non-synchronous rotation are the easiest to be ejected out of the S_1 surface.

However, it should be noted that the R surfaces represent neither the limiting nor equilibrium surface. Particles below the R surface for a

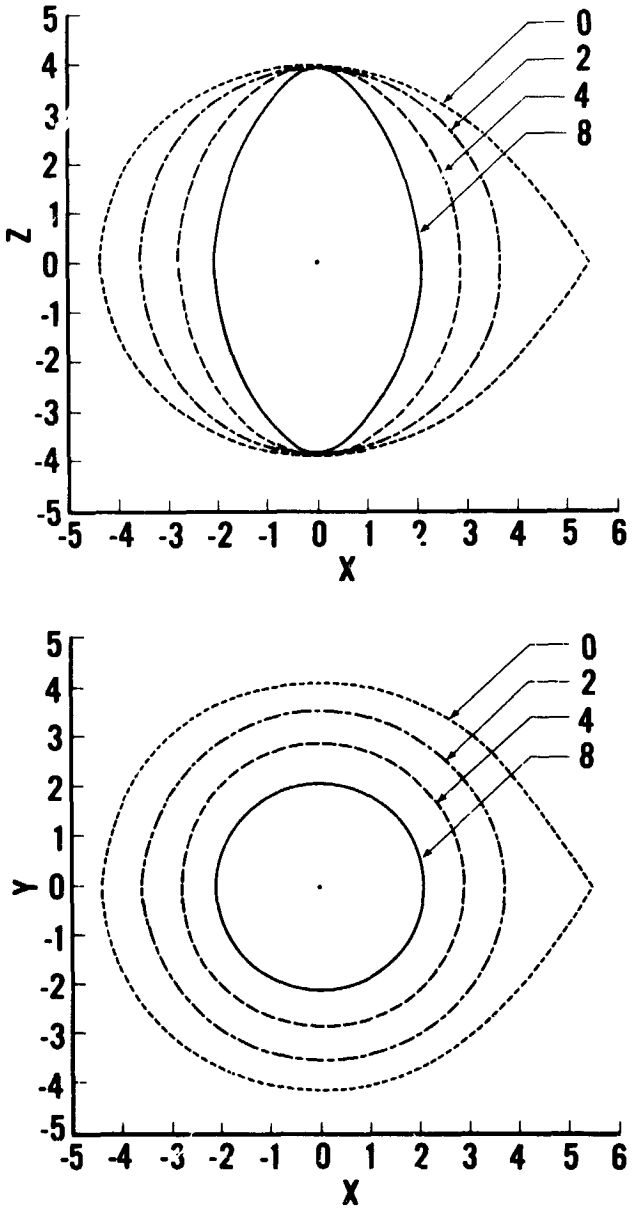


FIGURE. 1—The R surfaces for different values of ω (0, 2, 4, 8) for $\mu = 0.4$. The R surfaces are defined by equation (14). The case $\omega = o$ represents simply the innermost contact surface (denoted by S_1). When the rotation of a component star is not synchronous with its orbital motion, the gaseous particles above the R surface may escape the S_1 surface and are replaced by gaseous particles moving out from below the R surface. This process will go on until the star finally becomes synchronized to its orbital motion.

given value of ω can move up although they cannot escape through the S_1 surface without collisions. Statistically, we can state that most particles that are originally located above the R surface are lost and are replaced by particles coming from below the R surface. These new particles will naturally fill up the entire lobe of the S_1 surface. Therefore, the limiting surface of the $1-\mu$ component is still given by the S_1 surface. Since the angular momentum per unit mass due to the axial rotation of the $1-\mu$ component decreases downward in the star if it rotates as a rigid body, the new material above the R surface will rotate less rapidly (with respect to the xyz system) than the old material before the letter's escape. The simultaneous transfer of mass from below the R surface to above the R surface and from inside the S_1 surface to outside the S_1 surface will continue until the orbital revolution and the axial rotation become synchronize. Perhaps this axial rotation become synchronized. Perhaps this is one of the most effective mechanisms for synchronizing orbital revolution and axial rotation of close binary stars. Therefore, non-synchronization observed in those close binaries must be a temporary phenomenon triggered by rapid evolution of the component star itself as is suggested by Kopal (1959). But even at the time the two kinds of motion are temporarily out of step, the shape of the component is still given by the S_1 surface according to the present analysis.

Finally, one may question our argument on the ground that it has not proven that the particles are necessarily flying away from the star. If all particles on some surface of a non-synchronously rotating star should tend to move inward, this surface could be regarded as a stable surface for the star. This is however not true because when the particles move inward they collide with other particles. Since the average C value (the C value of each particle being defined by Eq. [12]) of all particles participating in a collision remains constant (Huang 1965) and since directions of the velocities will be modified after the collision, the chance of escape from the surface increases with time whatever are the directions of their initial velocities (in the xyz system). Therefore, our conclusion about the non-synchronously rotating star is valid without qualification for the Roche model.

It is my pleasure to note my thanks to Mr. Clarence Wade, Jr. who has performed the computation involved in obtaining diagrams in Figure 1.

REFERENCES

HUANG, S. -S., 1965, Ap. J., **141**,

KOPAL, Z., 1956, Ann. d'Ap., **19**, 298.

KOPAL, Z., 1959, "Close Binary Systems" (London: Chapman and Hall).

KRUSZEWSKI, A., 1963, Acta Astronomica, **13**, 106.

KUIPER, G. P., 1941, Ap. J., **93**, 133.

LIMBER, D. N., 1963, Ap. J., **138**, 1112.

MOULTON, F. R., 1914, "An Introduction to Celestial Mechanics" 2nd ed. (New York: Macmillan) chap. 8.

PLAUT, L., 1959, Pub. A.S.P. **71**, 167.

PLAVEC, M., 1958, Liège Symposium No. 8 on Emission-line Stars, p. 411.

STRUVE, O., 1941, Ap. J., **93**, 104.

STRUVE, O., 1950, "Stellar Evolution" (Princeton: Princeton University Press).

STRUVE, O., 1958, Pub. A.S.P., **70**, 5.

STRUVE, O. and HUANG, S. -S., 1957, Occasional Notes R.A.S., **3**, 161.

SWINGS, P., 1936, Zs. f. Ap., **12**, 40.

AN INTERPRETATION OF ϵ AURIGAE*

SU-SHU HUANG†

A model of ϵ Aurigae has been proposed in order to explain the contradictory behavior of this star found in photometric and spectroscopic observations. This model which consists of a rotating gaseous disk that appears opaque when viewed edge-on resembles the one we have suggested for β Lyrae (Huang 1963). The success of this simple model to explain these two peculair stars which have defied other interpretations for so long, together with the fact that rotating gaseous rings are frequently associated with the primary component of the Algol-type binary systems (Joy 1942, also Sahade 1960) leads us to a belief that rotating gaseous rings or disks are the result of natural development of gases that are injected into the binary system by its component stars. This belief is further strengthened by our knowledge that such a rotating ring or disk is dynamically or hydro-dynamically feasible (Prendergast 1960, Huang 1964).

I. INTRODUCTION

The binary ϵ Aurigae whose period is 27.1 years—one of the longest among eclipsing systems—gives every indication in its light curve of undergoing periodically total eclipses. But the spectrum of the eclipsed component, a F2 supergiant, can be observed persistently during the totality. These two seemingly irreconcilable phenomena have puzzled astronomers since the turn of the century. A serious attempt to resolve this incompatibility in the observational results may perhaps be traced to a paper by Kuiper, Struve and Strömgren (1937). They proposed an idea of making the huge eclipsing body—referred to as the I component—so tenuous as to be optically transparent to optical light. However, they argued that the ultraviolet radiation of the F-type supergiant would ionize the crescent-like thin layer of the I component that faces the F-type primary. The scattering by free electrons produced by ionization in this layer would then cut down the light from the F-type primary when the I component is in front, producing the phenomenom of eclipse. Since electron-scattering is frequency independent, the opacity in the crescent-like layer reduces the overall flux of the F component but does not change the spectral nature of the light. In this way, they are able to explain the difficulty we have mentioned before.

While this model has indeed resolved the long-standing difficulty, it introduces several new ones of equal, if not more, serious nature. Some of them have been mentioned in a recent work by Struve and Zebergs (1962).

Although the model by Kuiper *et al.* has been criticized, the idea of electron scattering as the cause of eclipse has persisted in the latter models by Struve (1956) and more recently by Hack (1961). This same idea has also strongly influenced the directions in which spectroscopic investigations of the star has been made (Kraft 1954).

Both Struve and Hack models may be regarded as modified versions of the original one by Kuiper *et al.* Struve introduced the idea that both components are surrounded by nebulous gases that fill the respective lobe of the inner contract surface of the system. These nebulous gases are supposed to cause gas streaming from one component to another and vice versa. Struve attributed the apparent splitting of lines observed before and after total eclipse to the presence of these gaseous streams. Except for this, Struve's interpretation of this peculiar system follows the same line of thought as in their previous paper. Consequently, the basic difficulties of eclipse by electron scattering are not removed.

*Published in *The Astrophysical Journal* 141:976, April 1965.

†Goddard Space Flight Center and Catholic University, Washington, D. C.

Hack's model does introduce a new concept. She proposed a hot secondary star of an effective temperature reaching 20,000°K. According to her, it is this hot star that is responsible for the ionization of gases in the shell or ring around it. The phenomenon of eclipse is then attributed to the intervention of this shell or ring between the primary F2 component and the observer.

It is more convincing to have an O or a B star around to do the work of ionization than an F2 star. But the introduction of a hot secondary meets other difficulties as we shall see in the following ways.

The depth of the eclipse during the apparent totality is about 0.8 magnitudes, which corresponds to an optical thickness of 0.74 over the entire spectral frequencies, if we assume that the eclipse is due to obscuration of the entire F component by the shell or ring. It requires 10^{24} of electrons in the column of unit cross-section and of a length equal to the thickness of the shell or ring. Since hydrogen is the dominant constituent, there will be approximately 10^{24} atoms in each column. Now, how could one explain that these atoms in their ionized states do not impress some additional spectroscopic feature on the light that passes through them? With ionizing radiation coming from a star of 20,000°K one would expect to observe spectral lines arising from such atoms as neutral or ionized helium, ionized oxygen, nitrogen, etc. that correspond to those found in the early-type supergiant stars. None have been observed. What have been actually observed, such as doubling and strengthening of lines in certain phases during eclipse, only show that the excitation level of the absorbing medium associated with the eclipsing body is not greatly different from that of the primary F2 atmosphere. Indeed, it has been specifically pointed out in the paper by Kuiper *et al.* that the lines associated with the secondary component and observed during eclipse are approximately, though not identically, the same as the normal lines of the F2 star.

Although it may be argued that compared with the F2 supergiant component, the hot star proposed by Hack is too faint to be seen in the visible region even during eclipse, it is difficult to comprehend why it has not been detected in the ultraviolet region. All these questions impair the otherwise attractive proposal by Hack.

From what has been said before, we cannot help but conclude that in spite of various modifications and refinements made since its inception. the electron scattering theory of the eclipse of ϵ Aurigae cannot explain this binary system in an internally consistent way.

Other theories of eclipse have been suggested in the meantime, but according to Hack (1961), none of them can explain the spectroscopic behavior of the system. However, it should be noted that Kopal's (1955) theory of attributing the opacity to solid particles did introduce a new conception of a ring structure which has been followed by Hack herself.

II. A PROPOSED MODEL

The success of our interpretations of β Lyrae by the introduction of an opaque gaseous disk rotating around its secondary component (Huang 1963) has led us to examine whether the same kind of model may be used for ϵ Aurigae, because in many respects the two peculiar binary systems show a similar behavior. Their similarities are: (1) The light from the primary component can be seen during the entire duration of eclipse while the spectrum of the secondary component itself has never been observed at any phase. (2) Additional lines appear during eclipse. These lines are, in both system, displaced towards the red end before the principal mid-eclipse and towards the violet end after mid-eclipse. (3) Light shows fluctuations, especially during eclipse, and (4) both show emission features.

However, there are also differences between these two systems. In the first place, the magnitudes of wavelength shifts of those additional lines observed during eclipse are different. In the case of β Lyrae, the radial velocities corresponding to the shifts are of the order of 100–300 km/sec, while in the case of ϵ Aurigae the velocities are perhaps of the order of 30–50 km/sec. This difference actually confirms our conviction that the two systems are similar, if we remember that the period of β Lyrae is only about 13 days while that of ϵ Aurigae is about 27.1 years. Thus, for the purpose of a similarity consideration, we must compare the stream velocities in terms of the respective orbital velocities of the component stars. Then the stream velocities in the two systems are of the same order of magnitude.

Another difference between these two systems comes from the fact that β Lyrae shows a secondary eclipse but ϵ Aurigae does not. This difference can be readily explained on the basis of our model of an opaque disk, as we shall see presently.

Figure 1 shows the model we propose for ϵ Aurigae. The obscuration of the primary component by the rotating gaseous disk causes an eclipse which would look, in the light curve, like

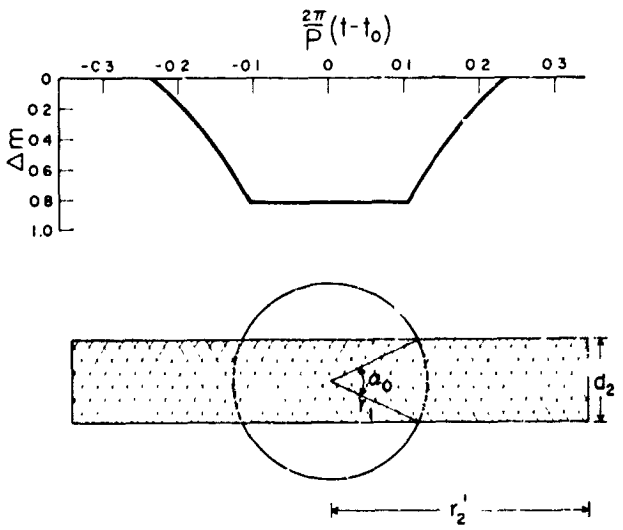


FIGURE 1.—A schematic diagram of our model for ϵ Aurigae and its resulting light curve during eclipse. It is assumed that we observe this sytem edge-on. Consequently, the rotating gaseous disk around the secondary component will appear like a dark rectangle which obscures the primary component during eclipse. The light curve at the top of the figure is derived by assuming a uniform stellar disk.

total but the light from the primary will continuously be seen even at the apparent totality. We suggest that the inclination, i, of this system is very near to 90°. Thus, we see only the edge of the disk; the secondary component itself cannot be seen because it is hidden in the disk. Because of this inclination neither do we receive any radiation from the primary reflected by the disk. As a result there will be no secondary eclipse. On the other hand, we have suggested in the previous paper that β Lyrae has an inclination which differs appreciably from 90° so that we can see the secondary star itself as well as the light reflected by the rotating gaseous disk. Consequently, we are able to observe a secondary eclipse.

Because of the difference in inclination, the shape of minimum differs in these two systems too. In the case of β Lyrae, the projected area of the disk will be an elongated ellipse, resulting in a curved minimum in the light variation. In the case of ϵ Aurigae, the projected area would be a rectangle (as shown in Figure 1), producing a flat minimum.

The sizes of the primary component and the opaque disk may be determined from the light curve. Let us, for the sake of illustration, assume that the stellar disk of the primary component is uniform in brightness and the edge of the disk we actually face is completely dark. We can now determine the radii, denoted respectively by r_1 and r_2' of both the primary component and the disk around the secondary component as well as the thickness of d_2' of the disk. All these quantities will be measured in terms of the mean separation between the two components of the system.

Different investigators gave different values for the duration of eclipse, D, and the of totality, d. We shall follow Plaut's (1950) values $D=0.0760$ and $d=0.340$, which yield after a simple calculation,

$$r_1=0.065 \text{ and } r_2'=0.171. \tag{1}$$

In order to determine d_2' we must use the observed depth of eclipse during the apparent totality. If we denote by α_0 the angle subtended at the center of the primary by the two points of intersection, all projected on the celestial sphere, between the boundary lines of the opaque disk and the primary component during eclipse (see Figure 1), the maximum light L_1 and the minimum light L_2 of the system are related by

$$\frac{L_2}{L_1}=\frac{\pi-\alpha_0-\sin\alpha_0}{\pi} \tag{2}$$

where

$$\sin\left(\frac{\alpha_0}{2}\right)=\frac{1}{2}\frac{d_2'}{r_1} \tag{3}$$

Since the observed depth of minimum is 0.81 mag., equation (2) gives $\alpha_0=0.881$ which in turns yields

$$d_2'=.055 \tag{4}$$

III. INTERPRETATION OF THE LIGHT CURVE

As an illustration that the present model can predict a light curve very much like the one actually observed, we shall consider a central eclipse ($i=90°$) of a uniformly bright stellar disk of the primary component by the dark projection of the rotating disk, shown schematically in Figure 1. If we denote by δ the apparent distance between the center of the primary and the center of the opaque disk and if we let

$$\delta_1=r_1+r_2', \qquad \delta_2=r_1\cos\tfrac{1}{2}\alpha_o+r_2', \qquad (5)$$
$$\delta_3=r_2'-r_1\cos\tfrac{1}{2}\alpha_o, \qquad \delta_4=r_2'=r_1,$$

where r_1, r_2' and α_o have been determined in the previous section, the light curve predicted on this model is given by

$$\frac{L_2}{L_1}=1-\frac{1}{2\pi}(\alpha_1-\sin\alpha_1), \quad \delta_2\leq|\delta|\leq\delta_1 \qquad (6)$$

$$\frac{L_2}{L_1}=1-\frac{1}{2\pi}(\alpha_o-\sin\alpha_o)-\frac{d_2'}{\pi r_1^2}(\delta_2-\delta),$$
$$\delta_3\leq|\delta|\leq\delta_2 \qquad (7)$$

and

$$\frac{L_2}{L_1}=1-\frac{1}{\pi}(\alpha_o+\sin\alpha_o)+\frac{1}{2\pi}(\alpha_2-\sin\alpha_2),$$
$$\delta_4\leq|\delta|\leq\delta_3 \qquad (8)$$

where

$$\cos\frac{1}{2}\alpha_1=1-\frac{\delta_1-\delta}{r_1} \text{ and } \cos\frac{1}{2}\alpha_2=1-\frac{\delta-\delta_4}{r_1}$$

For $o\leqslant|\delta|\leq\delta 4$, L_2/L_1 has already been given in equation (2).

The light curve according to these equations has been computed with the numerical values given by equations (1) and (4) and is plotted at the top of Figure 1. There is a general agreement between this predicted curve and the actually observed one (Güssow 1936).

It may be noted parenthetically that this interpretation of the light curve of ϵ Aurigae requires only that the rotating disk is opaque when viewed along the edge. Whether it is transparent or opaque in its vertical direction, does not matter in our interpretation. Indeed, we have no observational means to ascertain the optical depth in its vertical direction.

As has been pointed out by Fredrick (1960), the light curve shows a slight asymmetry. On the basis of our model, the asymmetry in the light curve reflects either a corresponding asymmetry in the thickness of the rotating disk or an additional absorption by gaseous stream above and below the rotating disk. Physically, it is difficult to envisage a permanent asymmetry in the thickness of the disk. Therefore it is most probable that the asymmetry is caused by the gaseous stream. If we assume that there is a gaseous stream flowing out from the primary component through the Lagrangian point into the secondary lobe of the innermost contact surface (Kuiper 1941), the gaseous stream will circulate around the secondary component above and below the main body of the rotating disk. As the gaseous streams circulate around the secondary component in the same sense as the rotation of the disk, which is in turn supposed to be rotating in the same sense as the binary motion, we would expect that they will gradually coalesce into the disk itself as a result of collisions. It is then not difficult to see that at any moment more gas would be found in the rear side than in the front side of the secondary component, giving rise to an asymmetry in absorption in agreement with the observed result. In this respect the behavior of the gaseous streams outside the main body of the rotating disk resembles closely what has been suggested for β Lyrae, which shows an asymmetric primary eclipse, with the decline steeper than the rise. We have attributed this asymmetry also to obscuration of gases just streaming out from the primary component, this obscuration making the eclipse last longer, and consequently show slower decline than would be the case of intervention of a simple disk. According to this interpretation, the light curves of both ϵ Aurigae and β Lyrae before mid-eclipse are perhaps less distorted by gaseous streams outside the rotating disk than those after mid-eclipse.

Other interesting facts found in this system are, according to Güssow (1936) and Fredrick (1960), that the light fluctuation far away from eclipse does not exceed 0.1 mag., that in 3–5 years before and after eclipse the variation may get as large as, but seldom 0.2 mag. and that in totality, the fluctuation may reach 0.3 mag. or even slightly larger. Such a manner of variations

in light follows also quite naturally from our model. While the intrinsic variation of light far away from eclipse is expected from the supergiant F2 primary itself, the greater fluctuation in light near and during eclipse only indicates the unsteadiness of gaseous streams on both sides of the rotating disk.

IV. INTERPRETATION OF SPECTROSCOPIC RESULTS

That ϵ Aurigae shows an H_α emission has long been known (Adams and Sanford 1930). More recently, Wright and Kushwaha (1958) have made an extensive study of the structure of this H_α line both inside and outside the last eclipse. According to them, the two emission wings are almost equal in intensity outside eclipse. Moreover, the central absorption gives the same radial velocity as the absorption lines of *Fe*II, *Ba*II, and *Y*II in all phases outside eclipse. This result suggests that the material that produces the emission feature moves with the primary.

Moreover, Wright and Kushwaha have found that just before and during ingress the emission wing towards violet is usually stronger than that towards red and during egress, the reverse is true. At the time of totality, the absorption becomes dominant.

The H_α emission is produced by gases moving with the primary component, but we still do not know how they are distributed in space around the primary component. The line profile outside eclipse appears to suggest a rotating gaseous ring like the one observed in many an Algol-type binary system (Joy 1942; also Sahade 1960). However, the behavior of this emission line found in ϵ Aurigae during eclipse differs from what has been observed in the Algol-type variables. In the latter, the emission wing toward violet decreases in strength when the system enters into eclipse, but in ϵ Aurigae, the same wing increases in strength. Similarly, in egress, the two cases show variation in the relative intensity of the two emission wings in opposite directions. Two hypotheses as regards the distribution of emitting gases may be proposed to explain the behavior of H_α emission found in ϵ Aurigae in and out of eclipse. (1) The emission indeed comes from a gaseous ring rotating around the primary but the sense of rotation is opposite to the orbital revolution of components in the system. (2) Emission is produced by expanding gases ejected from the primary component. It is interesting to note that Adams and Sanford (1930) did find that H_α showed then as an emission line with an absorption border on its violet edge, indicating outward motion of gases. Consequently, Beals (See Hack 1961) classified it as a ρ Cygui star.

According to Wright and Kushwaha the weak emission feature that occurs during totality seems to correspond to the velocity of the system and could be due to the tenuous gases enveloping the entire binary.

White the H_α emission mainly comes from gases associated in whatever way with the primary component, the complicated absorption feature of this line observed during eclipse is due to gases spilled above and below the main body of the disk around the secondary component.

It is reasonable to assume that the distribution of gases around the secondary component is not confined to the disk itself. However, for the same reason that the material density in the galactic plane falls off rapidly in the polar directions, the density around the secondary component must also decrease rapidly with the perpendicular distance. At places not too far away from the main body of the disk, the gaseous distribution would be so rare that it is no longer opaque to the continuous radiation but it may still produce line absorption. Therefore, we should expect to observe it through spectroscopic study. Indeed, according to our model, it is this gas distribution that produces the additional absorption that modifies the profile of H_α continuously with phase during eclipse. Thus, during ingress, additional absorption which occurs in the red side of the normal position makes the emission wing toward violet appear stronger than the other. A similar argument leads to the reverse conclusion at the time of egress. During totality the light from the primary passes through a large amount of gases, resulting in a dominantly absorption line as observed. In this way Wright and Kishwaha's results can be satisfactorily explained on the present model.

Because of the appearance of H_α emission in its spectrum, it has been proposed (Wright and Kishwaha 1958) that the excitation might be due

to a hot secondary, since under normal conditions, no emission at H_α appears in the spectra of a star as late as spectral type F of the primary component. As we have seen before, Hack has incorporated this suggestion in her theory in order to explain the ionization of gases in the shell or ring around the secondary. In section I, we have seen that a hot secondary introduces some new difficulties. Now we have found that the H_α emission comes from gaseous distribution associated with the primary instead of the secondary. Hence, the hypothesis of a hot secondary made in order to explain H_α emission becomes even less satisfactory than the case if the emission were associated with the secondary.

If the gases spilled over the disk give rise to absorption that cuts into H_α emission, we would expect during eclipse the appearance of lines belonging to other elements than hydrogen. Indeed, the indication of such an additional absorption long been known. According to Kuiper *et al*, the spectral lines of many elements become asymmetric during ingress by the presence of a strong core on the red side. The degree of asymmetry increases, reaches a maximum at second contact and then diminishes until mid-eclipse when each line becomes symmetric as it is outside eclipse. However, the intensities of lines at mid-eclipse are somewhat stronger than those outside eclipse. After mid-eclipse, the lines become asymmetrical again, but with a strong core on the violet side this time. This asymmetry increases rapidly until it reaches third contact. The asymmetry is obviously due to the presence of a double structure (Adams and Sanford 1930) and indicates the additional line absorption by gases spilled over from the disk. The sense of the Doppler shift of these lines agrees again with what one would expect from gaseous streams rotating in the same sense as the orbital revolution.

In the previous section we have attributed the asymmetry of the light curve also to gaseous streams which are located perhaps closer to the disk than those giving rise to absorption lines. We have suggested that the gaseous streams must be more complicated and extend a larger region in the rear side than in the front side of the secondary. Accordingly, we would expect the structure of absorption lines observed after mid-eclipse to be more complicated than that observed before mid-eclipse. Struve and Pillans' (1957) observational results appear to bear this prediction out. They have found that at the very end of totality, the absorption lines show a remarkable amount of structure not previously observed in the star. Some lines are triple, while others show a double structure, indicating gaseous streams, just coming out from the primary before collision which would have erased the velocity differences of the streams.

V. THE F2 SUPERGIANT ATMOSPHERE, ITS MASS EJECTION AND TURBULENT VELOCITIES

Finally, we may say a few words about the atmosphere of the F2 supergiant primary. Whatever is the mass ratio of the system, it is reasonably certain that the size of the primary component derived in Section II must be small compared with the innermost contact surface (e.g. Kuiper 1941), which will be referred to hereafter as the S_1 surface. Indeed, the light curve outside eclipse does not indicate any distortion due to ellipticity. On the other hand, both spectroscopic and photometric results indicate that mass is continuously flowing out of the primary lobe of the S_1 surface into the secondary lobe. It follows that at the photospheric surface, the star must steadily eject matter to keep the flow of gas from the primary lobe to the secondary lobe of the S_1 surface. Since the photosphere is well below the S_1 surface, the effect of its comparison must be small. Consequently, mass ejection at the stellar surface (i.e. photosphere) must be intrinsic to this F2 supergiant primary and should not be attributed to the interaction within a binary.

Thus, between the photosphere and the corresponding lobe of the S_1 surface, the vast volume must be filled with a tenous atmosphere. Its extension must be several times the stellar radius; through the extended atmosphere the primary is losing its mass. The picture thus derived agrees completely with what Deutsch (1956, 1960) has found in many red giant and supergiant stars and confirms the general belief that a supergiant—whatever is its spectral type—is always ejecting mass.

Violent motions have always been found in atmosphesre of supergiant stars (e.g. Huang and Struve 1960). This is especially true in the primary of ϵ Aurigae (Wright and van Dien 1949;

Hack 1959). However, the nature of the observed motion has never been clearly understood, although frequently but rather vaguely we have attributed it to some kind of prominence activities. With the conception of mass ejection by the supergiant star reasonably established, we can now understand why violent motion should always be associated with its atmosphere. Indeed, the mass ejection from the stellar surface, presumably due to convective currents below, acts as a stirring mechanism of the atmosphere over the photosphere. The ejection creates a velocity field in the atmosphere which also makes the latter very extended.

REFERENCES

ADAMS, W. S., and SANFORD, R. F., 1930, Pub. A.S.P., **42**, 203.

DEUTSCH, A. J., 1956, Ap. J., **123**, 210. 1960, "Stellar Atmospheres," ed. J. L. Greenstein, (Chicago: University of Chicago Press), Chap. **15.**

GÜSSOW, M., 1936, Veröff Berlin-Babelsberg, **11**, No. 3.

HACK, M., 1959, Ap. J., **129**, 291. 1961, Mem. Soc. Astr. italiana, **32**, No. 4.

HUANG, S. -S., 1963, Ap. J., **138**, 342. 1964, to be published.

HUANG, S. -S., and STRUVE, O., 1960, "Stellar Atmospheres' ed. by J. L. Greenstein (Chicago: University of Chicago Press), Chap. **8.**

JOY, A. H., 1942, Pub. A.S.P., **54**, 35.

KOPAL, Z., 1955, Comm. Coll. Intern. d'Astrop. Liége **6**, 241.

KRAFT, R. P., 1954, Ap. J., **120**, 391.

KUIPER, G. P., 1941, Ap. J., **93**, 133.

KUIPER, G. P., STRUVE, O. and STRÖMGREN, B., 1937, Ap. J., **86**, 570.

PLAUT, L., 1950, Pub. Kapteyn Astr. Lab. Groningen No. 54.

PRENDERGAST, K. H., 1960, Ap. J., **312**, 162.

SAHADE, J., 1960, "Stellar Atmospheres," ed. J. L. Greenstein (Chicago: University of Chicago Press), Chap. **12.**

STRUVE, O., 1956, Pub. A.S.P., **68**, 27.

STRUVE, O. and PILLANS, H., 1957, Pub. A.S.P., **69**, 169.

STRUVE, O. and ZEBERGS, V., 1962, "Astronomy of the 20th Century," (New York: Macmillan Co.)

WRIGHT, K. O. and DIEN, E. VAN, 1949, J.R A.S. Canada **43**, 15.

WRIGHT, K. O. and KUSHWAHA, R. S., 1958, Comm. Coll. Intern. d'Astrop. Liége **8**, 421.

THE MOTION OF GASEOUS STREAMS IN THE BINARY SYSTEM*

SU-SHU HUANG†

We have studied several problems in connection with the gaseous flow in the close binary system. First, the statistical property of the Jacobian constants of colliding particles in the system is examined and a conversation law for their mean value established. Then the velocity-independent nature of the rate of change in angular momentum of a particle moving in the binary system is pointed out. These properties prompt us to derive for the gaseous flow a set of differential equations that provide a point of view lying in the middle between the orbital approach (which neglects both pressure and collision) and the hydrodynamic approach (which includes both) because our equations take into account the collision but not the pressure. The equations have been solved under the same approximation as Prendergast (1960) has assumed and have been found to yield a similar result as was obtained by him from the hydrodynamic equations.

Because of our emphasis on the Jacobian constant and angular momentum in the treatment of gaseous flow we have called attention to the fact that some combinations of these two physical quantities are incompatible in a certain region of space which we have called the forbidden zone.

Finally, the formation, the evolution and the significance of rotating rings observed in many Algol-type eclipsing binaries are discussed in an effort to understand the mode of ejection of matter from the secondary surface.

I. INTRODUCTION

The problem of gaseous streams observed in some binary systems has been studied theoretically either as individual bodies moving in orbits independently of each other (Kopal 1959, Gould 1959) or as an aerodynamical flow (Prendergast 1960). Both approaches encounter difficulties, though of entirely different nature. In the present paper, we shall call the attention to a few general properties of the motion of gaseous particles in the binary systems, which lead us to a theory that somewhat reconciles these two fundamentally different approaches and thereby makes the flow problem easier to comprehend.

II. STATISTICAL PROPERTY OF THE JACOBIAN CONSTANTS DURING A COLLISION OF PARTICLES

One of the differences between the two approaches mentioned in Section I concerns the collision of particles. While the neglect of this important process makes the orbital approach unrealistic, some results obtained in celestial mechanics of the motion of an infinitesimal body in a gravitational field of two revolving components has its physical significance. This is because of the statistical property of the Jacobian constants that we will discuss in this Section.

Let us assume that the two stars are revolving around each other in circular orbits. This is generally true for close binaries (Struve 1950). Thus, a motion of a particle in such a system is identical to what is treated in the restricted three-body problem in celestial mechanics (e.g., Moulton 1914; Brouwer and Clemence 1961). Following the standard treatment of the problem we shall choose as the unit of length, the separation between the two components, as the unit of mass, the total mass of the two components, and as the unit of time, the perior of the orbital motion of the two components divided by 2π. In such a unit system the gravitational constant is one. Let us now denote by μ the mass of the secondary component. Thus, the mass of the primary will be $1-\mu$. If, furthermore, a rotating coordinate system xyz is

*Published in *The Astrophysical Journal*, 141(1):201–209, January 1, 1965.

†Goddard Space Flight Center and Catholic University, Washington, D. C.

so chosen that its origin is at the center of mass of the two components, its x-axis coincides at all times with the line joining the two components and its z-axis is perpendicular to the orbital plane of the stars, then the coordinates of the primary will be $x_1=-\mu, y_1=o, z_1=o$ and those of the secondary will be $x_2=1-\mu, y_2=o, z_2=o$ in consistency with the adopted unit system.

The equations of motion of the third infinitesimal body in the restricted three-body problem admits an integral, frequently known as the Jacobian integral, as follows:

$$C=2U-(\dot{x}^2+\dot{y}^2+\dot{z}^2) \tag{1}$$

where

$$U=\frac{1}{2}(x^2+y^2)+\frac{1-\mu}{r_1}+\frac{\mu}{r_2} \tag{2}$$

Here r_1 and r_2 are respectively the distances of the infinitesimal body from the primary $(1-\mu)$ and the secondary (μ) component, while (x,y,z) the three vector components of $\vec{r}$, are the coordinates of the infinitesimal body. The dot represents as usual the time derivative. The integration constant C, is known as the Jacobian constant.

We can transform equation (1) into a stationary system. Let us now consider a collision of n particles of mass, $m_i(i=1, 2 \ldots n)$. Since we have particles of atomic sizes, the coordinates may be regarded as the same for all colliding particles at the instant of collision. Moreover, the total kinetic energy of the colliding particles conserves during an elastic collision. It follows from these considerations as well as the definition of C by equation (1) that

$$\sum_{i=1}^{n} m_i C_i' = \sum_{i=1}^{n} m_i C_i \tag{3}$$

where C_i and C_i' are respectively the Jacobian constant of the i-th particle before and after the collision. Thus, if we define an average C, such that

$$<C>\sum_{i=1}^{n} m_i = \sum_{i=1}^{n} m_i C_i, \tag{4}$$

$<C>$ will be an invariant under the physical processes of elastic collisions. However, it may be noted that the dispersion of C's from their average value will in general change after each collision.

For inelastic collisions an equation connecting various C_i and C_i' can always be obtained from the energy consideration if we know the detailed process of the collision. We shall assume in the present paper that the collisions that take place among particles in the binary system are statistically elastic, i.e., endoergic collisions balancing exoergic ones.

As a result of the constancy of $<C>$ during collision, the problem of gaseous flow is considerably simplified because we have now a macroscopic quantity, $<C>$, to deal with instead of following the courses of numerous particles in the system. Thus, the gaseous particles must maintain a constant value of $<C>$ in their stream motion. Indeed, Prendergast (1960) has shown that C is a constant along a stream line. This situation resembles the introduction of the concept of temperature and pressure which simplifies our study of the chaotic motion of molecules in gases in free space. Therefore, whatever is the nature of ejection that occurs on the stellar surface, the mean value of C's of ejected particles and their dispersion serve as two of the most characteristic indices of the mode of ejection as regards the course of their subsequent motion.

It should be noted, however, that although the gaseous particles maintain a constant $<C>$, the mean flow does not follow the orbit derived from the equations of motion of the three-body problem. It is physically obvious that all those loops, cusps, sudden reversal in the direction of motion, and other erratic behavior found in the orbits of the three-body problems must be completely erased by collisions.

III. THE RATE OF CHANGE OF ANGULAR MOMENTUM

The angular momentum (the z-component) per unit mass of the third body with respect to the center of mass of the binary system will be denoted by h. It is given by:

$$h=x^2+y^2+x\frac{dy}{dt}-y\frac{dx}{dt} \tag{5}$$

and varies with time because the third body is continuously interacting with the two revolving component stars.

Although h varies with time, there are two points which make it physically significant. First, the total angular momentum is conserved among the colliding particles if the collision takes place rapidly. Secondly, it can be easily shown from the equations of motions in the restricted three-body problem that

$$\frac{dh}{dt}=\mu(1-\mu)y\left(\frac{1}{r_1^3}-\frac{1}{r_2^3}\right). \tag{6}$$

The significance of equation (6) derives from the fact that $^{dh}/_{dt}$ is a function of coordinates of the third body only, being independent of its velocity. Moreover, it is anti-symmetric with respect to the x-axis and to the line bisecting the separation segment between the two finer bodies. Thus, it vanishes on these two lines. Here we see physically why five Lagrangian points all lie on either of these two lines. It follows that any steady flow in a closed curve must cross either one or both of these lines so that h will recover to its original value after the completion of the circuitous flow. Indeed, this is the case of rotating gaseous rings frequently observed around the primary component in many an Algol-typebinary system (Joy 1942, 1947; also Sahade 1960).

In the three dimensional case, $^{dh}/_{dt}$ vanishes in the XZ plane and in the plane bisecting the line joining the two components. Thus, if we divide space into four regions by these two planes the sign of $^{dh}/_{dt}$ is positive in the two regions and negative in the other two.

IV. GASEOUS FLOW DERIVED FROM THE C AND h CONSIDERATIONS

In the two dimensional case, a knowledge of C and h at every point defines completely the flow pattern. Therefore, a velocity vector field of gaseous motion in a binary system can be defined by two scalar fields of h and C. Since the average values of C and h do not change by collision, we may write

$$\frac{\partial C}{\partial t}+\vec{u}\cdot\nabla C=o \tag{7}$$

from the constancy of C and

$$\frac{\partial h}{\partial t}+\vec{u}\cdot\nabla h=\mu(1-\mu)\left(\frac{1}{r_1^3}-\frac{1}{r_2^3}\right) \tag{8}$$

from equation (6) when we follow the stream lines. In writing these equations where u denotes velocity at points (x, y) we have made an additional assumption that C and h are continuous over the plane. In this way, we have derived two flow equations from the results of celestial mechanics.

By imposing the continuity condition of C and h over space we are able to take advantage of the result derived from celestial mechanics but at the same time to discard as meaningless the seemingly erratic and infinitely varied forms of orbits that one may actually obtain by a straight integration of the equations of motion in the three-body problem. Thus, equations (7) and (8) are proposed here purely from physical arguments of conservation laws in collisions. A mathematical derivation from hydrodynamic equations is given in the Appendix.

Since the pressure is not included in equations (7) and (8), the present formulation of the flow problem lies between the orbital approach (which neglects both pressure and collision) and the bona fide hydrodynamic equations of flow (which include both). Thus, the present treatment is mathematically equivalent to the hydrodynamic approach when the pressure is neglected. In presenting the problem in this manner we gain a better physical insight because both C and h are physical quantities.

Actually an inclusion of pressure would make the problem very difficult. Indeed, Prendergast (1960) who started directly from hydrodynamic equations also neglected the pressure term when he came to the stage of solving the equations. Thus, the equations (7) and (8) should be equivalent to what Prendergast has used, although the basic approach is different.

What we will show in the following is that we can obtain a similar solution as obtained previously by Prendergast. Also by following the present derivation we can see very clearly the conditions under which the solution will be valid.

Following Prendergast, we shall consider the two-dimensional case, neglect the velocity component at right angles to the zero-velocity curves and choose a right-handed orthogonal curvilinear coordinate system (ξ, η, z) where ξ is the label of the zero-velocity curves, namely

$$U=\xi, \tag{9}$$

U is given by equation (2) and η is an angular measure along the zero-velocity curve. If we now denote Q_ξ and Q_η the metric coefficients corresponding to this coordinate system, we can express equations (7) and (8) in this new system as follows:

$$\frac{u_\xi}{Q_\xi}\left[1-\left(u_\xi\frac{\partial u_\xi}{\partial \xi}+u_\eta\frac{\partial u_\eta}{\partial \xi}\right)\right] -\frac{u_\eta}{Q_\eta}\left(u_\xi\frac{\partial u_\xi}{\partial \eta}+u_\eta\frac{\partial u_\eta}{\partial \eta}\right)=o \quad (10)$$

and

$$\frac{u_\xi}{Q_\xi}\left(\frac{\partial r^2}{\partial \xi}+r_\xi\frac{\partial u_\eta}{\partial \xi}+\frac{\partial u_\xi}{\partial \xi}u_\eta-r_\eta\frac{\partial u_\xi}{\partial \xi}-\frac{\partial r_\eta}{\partial \xi}u_\xi\right) +\frac{u_\eta}{Q_\eta}\left(\frac{\partial r^2}{\partial \eta}+r_\xi\frac{\partial u_\eta}{\partial \eta}+\frac{\partial r_\xi}{\partial \eta}u_\eta-r_\eta\frac{\partial u_\xi}{\partial \eta}-\frac{\partial r_\eta}{\partial \eta}u_\xi\right) =\mu(1-\mu)y\left(\frac{1}{r_1^3}-\frac{1}{r_2^3}\right) \quad (11)$$

if we assume a steady state of flow. Here the subscripts ξ and η denote respectively the components of the vector in the ξ and η direction.

If we neglect u_ξ we obtain

$$\frac{\partial u_\eta}{\partial \eta}=o \quad (12)$$

from equation (10) and

$$\frac{u_\eta}{Q_\eta}\frac{\partial r^2}{\partial \eta}+\frac{u_\eta}{Q_\eta}\left(r_\xi\frac{\partial u_\eta}{\partial \eta}+\frac{\partial r_\xi}{\partial \eta}u_\eta\right)=\mu(1-\mu)y\left(\frac{1}{r_1^3}-\frac{1}{r_2^3}\right) \quad (13)$$

from equation (11). Combining equations (12) and (13) we obtain a second degree algebraic equation for u_η

$$\frac{u_\eta}{Q_\eta}\frac{\partial r^2}{\partial \eta}+\frac{u_\eta^2}{Q_\eta}\frac{\partial r_\xi}{\partial \eta}=\mu(1-\mu)y\left(\frac{1}{r_1^3}-\frac{1}{r_2^3}\right) \quad (14)$$

By neglecting u_ξ in one of the two hydrodynamic equations of flow a u_ξ^2 in the other, Prendergast has also obtained and second degree algebraic equation for u_η. While the coefficients in the equation here (which involves $\partial r^2/\partial \eta$ and $\partial r_\xi/\partial \eta$) are completely different from those in Pendergast's equation, (which involves $\partial Q\eta/\partial_\xi$), both results are valid under the same approximation. Now it appears from equation (10) that in order to neglect u_ξ, $\partial u_\eta/\partial \eta$ must be small. This is the condition for the validity of our solution.

Physically we have started by assuming that the flow follows the zero-velocity curves, but once we have found u_η, according to equation (14) we immediately see that u_ξ does not vanish because $\partial u_\eta/\partial \eta$ does not. So the flow cannot exactly follow the zero-velocity curves. Whether we can find a converging field of velocities for this steady state by an iterating process, we do not know. Neither has Prendergast commented on this possibility.

However, near the two finite bodies, the variations of velocity with η is small, so the flow pattern obtained here represents a good approximation. This explains why gaseous rings are frequently observed around the primary component of many an Algol-type eclipsing binary.

In order to complete the analogy of the present calculation with Prendergast's, we may write equation (10) by neglecting u_ξ^2 and higher order terms, as follows

$$u_\xi=\left(\frac{Q_\xi}{Q_\eta}\right)u_\eta^2\frac{\partial u_\eta}{\partial \eta}\left(1-u_\eta\frac{\partial u_\eta}{\partial \xi}\right)^{-1}$$

We can now examine the asymptotic behavior of u_η for small values of r_1 (or similarly of r_2). In the immediate neighborhood of the $1-\mu$ component, we may take the position of this component as the origin and use the polar coordinate system (r_1, φ) where φ is the angle that the radius vector $\vec{r}_1$ makes with the x-axis, being counted, as usual, positive in the counter-clockwise direction from the positive x-axis. The zero-velocity curves in the immediate neighborhood can be approximated by circles. Thus, we have

$$\xi=U\to\frac{1-\mu}{r_1},\ Q_\eta=r_1 \quad (15)$$

and

$$\eta=-\varphi \quad (16)$$

since (ξ,η,z) is a right-handed coordinate system.

It can be easily shown that equation (14) reduces to

$$u_\eta^2-2r_1u_\eta-\frac{1-\mu}{r_1}=o, \quad (17)$$

under the approximation given by equations (15) and (16). We take the negative sign before the square root in the solution of this quadratic equation, because as Prendergast has pointed out, the velocity should vanish if the force vanishes. Retaining the dominant terms, we find the solution for small r_1

$$u_\eta = r_1 - \left(\frac{1-\mu}{r_1}\right)^{\frac{1}{2}}. \tag{18}$$

This asymptotic expression represents the Keplerian velocity in the neighborhood of the $1-\mu$ component in the rotating coordinate system.

In order to study the asymptotic behavior at large distances from both the components, we use the polar coordinates (r, θ) and expand all quantities in terms of $1/r$, θ being now the angle between the radius vector $\vec{r}$ and the x-axis. It can be shown that the angle that the normal to the zero-velocity curve makes with the radius vector $\vec{r}$ decreases as $1/r^5$. Consequently, we may take the radius vector as the normal to the curve as a first approximation for zero-velocity curves at large distances. Similarly, we can show that the radius of curvature may be set equal to r if we neglect terms of $1/r^5$ and higher orders. With these approximations we can easily derive from equation (14)

$$u_\eta = -r, \tag{19}$$

if we remember that in this asymptotic case

$$\eta = \theta. \tag{20}$$

According to equation (19) the gas remains at rest in the stationary frame of reference. Thus, the asymptotic behavior in both cases is, as it should be, identical to what Prendergast has obtained.

The numerical evaluation of u_η is not simple but it can be done. We shall illustrate it by computing values of u_η at those two points in the flow around the $1-\mu$ component where a stream line intersects with the x-axis. At these two points we can take advantage of the simplification arising from the symmetry of the zero-velocity curves with respect to the x-axis.

Let point $A(x_o, o)$ be the one of these points on the right side of the $1-\mu$ component and point $B(x_o, o)$ the one on the other side. Thus, $x_o > -\mu$ at A and $x_o < -\mu$ at B. The relation between x_o, r_1 and r_2 at both A and B can be easily obtained.

Since Q_η denotes the radius of curvature of the zero-velocity curve at A or B then we have in the immediate neighborhood of A or B

$$r^2 = Q_\eta^2 + (x_o \mp Q_\eta)^2 + 2Q_\eta(x_o \mp Q_\eta)\cos\varphi \tag{21}$$

and

$$r_\xi = -Q_\eta - (x_o \mp Q)\eta\cos\varphi, \tag{22}$$

where φ represents the angle which the normal to the zero-velocity curve at any point near A or B makes with the positive x-axis. When two signs appear together, the upper one corresponds to point A and its neighborhood and the lower one points to B and its neighborhood. Again $\eta = -\varphi$ as in the first case of asymptotic expansions. With the aid of equations (21) and (22) we may reduce equation (14) to

$$u_\eta^2 - 2Q_\eta u\eta + kQ_\eta^2 = o \tag{23}$$

where

$$k = \frac{\mu(1-\mu)}{x_o \mp Q_\eta}\left(\frac{1}{r_1^3} - \frac{1}{r_2^3}\right) \tag{24}$$

for point A and B respectively according to whether we take the upper or lower sign in equation (24). The solution of equation (23) is

$$u_\eta = Q_\eta[1 - (1-k)^{\frac{1}{2}}], \tag{25}$$

the minus sign before the parenthesis has been chosen in order to agree with the asymptotic behavior found previously in equation (18). The radius of curvature Q_η can be derived from equation (9). We shall omit its long expression here.

We have computed u_η according to equation (25) for several cases of x_o with $\mu = 0.2$. It appears evident from the results of computations that velocities thus obtained are very near to those found in the periodic orbits which we may obtain either by the numerical process of successive approximation (Huang and Wade 1963) or by the series solution (Huang 1964). In the second and third column of Table 1 we have a few velocities of a particle as it crosses the x-axis obtained by equation (25) and from the periodic solutions respectively. Needless to say, the disagreement in sign between $\dot{y}$ and u_η in one half of

the cases in Table 1 arises purely from the difference in the coordinate system. We should compare only the magnitudes between the second and third column. As would be expected, the agreement between two kinds of computations becomes better and better as we approach more and more to the star. Thus, the gaseous rings found observationally in many binary systems may be regarded equivalently either as a hydrodynamic flow or as motions of particles in a continuous series of periodic orbits that exist around the component.

TABLE 1.—*A Comparison of Velocities at Points on the x-axis Obtained from Equation (14) and from Periodic Solutions*

x_0	u, From Eq. (14)	$\dot{y}$ From Periodic Solutions
.054472	−1.492	1.525
−.450000	−1.529	−1.556
.001841	−1.775	1.791
−.400000	−1.794	−1.809
−.049380	−2.148	2.154
−.350000	−2.156	−2.164
−.099858	−2.724	2.726
−.300000	−2.727	−2.730
−.149986	−3.949	3.949
−.250000	−3.950	−3.950

V. THE FORBIDDEN ZONE

Since our approach to the problem of gaseous flow in the binary system emphasizes the two phvsical quantities C and h, it is interesting to point out that at some points in space, a certain combination of values for these two quantities is incompatible. In other words, with a given value of C and h, sometimes the particle cannot into a certain region of space which we shall call the forbidden zone. It can be easily seen as follows. We may express equations (1) and (5) simply as

$$\dot{x}^2+\dot{y}^2+\dot{z}^2 A, \tag{26}$$

and

$$x\dot{y}-y\dot{x}=B, \tag{27}$$

where A and B are functions of x, y, z, C and h and can be easily found from equations (1) and (5). Now in the (x, y, z) velocity space, equation (26) represents a sphere with the center at the origin and with a radius equal to $A^{1/2}$, while equation (27) represents a plane. It is then obvious that for any given combination of A, B, x, y, z (or equivalently C, h, x, y, z) the two surfaces may or may not intersect with each other. If they do not, no real velocity components (x, y, z) will satisfy both equations. This means that the particle with the given values of C and h cannot reach the point (x, y, z). In other words, the point (x, y, z) lies in the forbidden zone associated with the given C and h values.

The forbidden zone can be easily calculated from the condition that the distance of the origin from the plane given by equation (27) in the velocity place is greater than the radius, $A^{1/2}$ of equation (26). Explicitly, the forbidden zone is given by

$$B^2>A(x^2+y^2) \tag{28}$$

We shall illustrate the forbidden zone only in the x, y plane. It can be obtained by plotting the curve defined by the following equation

$$f(r, \theta)\equiv r^2+r\left[r^2+\frac{2(1-\mu)}{r_1}+\frac{2\mu}{r_2}-C\right]-h=0. \tag{29}$$

where θ which enters into the expression for r_1 and r_2 denotes the angle between the radius vector $\vec{r}$ of the third body and the positive x-axis.

Although h is a physical quantity, it is not a constant of motion in the restricted three-body problem. Consequently, the forbidden zone is not as important as the zero-velocity curves in depicting the motion of particles. However, combined with the property of dh/dt discussed in the previous section, the forbidden zone may serve some useful purpose of excluding certain modes of gaseous flow in the binary system.

Figure 1 illustrates three forbidden zones in the x, y plane for $C=3.5$ and for the three values of h. The areas that include the origin are forbidden to particles having the assumed C and h values. Because of the symmetry with respect to the x-axis, only one half of the zone is drawn in each case. However, the signs of dh/dt are marked in the figure in all four quadrants at the corners.

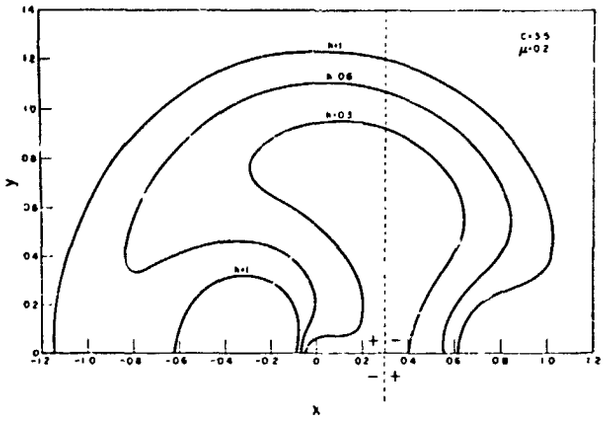


FIGURE 1.—The Forbidden Zone—Particles with given C and h at any instant cannot be found in a certain region of space, called the forbidden zone. Three forbidden zones in the xy plane are illustrated here for three pairs of (C, h). They are (3.5, 1), (3.5, 0.6) and (3.5, 0.3). Because of symmetry with respect to the x-axis, only one-half of each zone is shown here. The forbidden zone corresponding to each of the other two pairs of (C, h) lies inside a single closed curve. The sign of dh/dt is marked at each corner of the four quadrants formed by $x=o$ and $y=0.5$.

VI. SOME REMARKS CONCERNING THE ROTATING GASEOUS RINGS OBSERVED IN THE BINARY SYSTEM

1. The Chance of Ring Formation

Let us denote C_1 and C_2 as the value of C that corresponds respectively to the innermost and outermost contact surface (Kuiper, 1941). The latter will be hereafter called, for the sake of brevity, the S_1 and S_2 surface. Both C_1 and C_2 have been computed by Kopal (e.g. 1959) and Kuiper and Johnson (1956).

According to the result obtained in the restricted three-body problem, those ejected particles whose C values are greater than C_1 cannot penetrate the S_1 surface, and those whose C values are greater than C_2 cannot penetrate the S_2 surface. It follows that those particles whose C values are less than C_2 could escape from the system and those whose C values are larger than C_1 will remain inside the S_1 surface.

The quantity,

$$\Delta C = C_1 - C_2, \tag{30}$$

measures the closeness of the two critical zero-velocity surfaces S_1 and S_2 and may have an important effect on the flow of matter ejected by the secondary component into the primary lobe of the S_1 surface. We do not mean that only those particles with C values between C_1 and C_2 will penetrate into this volume, since any particle with $C<C_1$ can move into it. But the amount of accumulation of matter inside this volume at any given time perhaps increases with the increase of ΔC. It follows from this reasoning that formation of gaseous rings around the primary $(1-\mu)$ component favors large values of μ, as we can easily see, for example, from Kuiper and Johnson's Table, that ΔC increases with μ.

On the other hand, we have pointed out (Huang and Struve, 1956) that from the consideration of available space for their ring formation around primary component, gaseous rings have a better chance to exist in binaries of small μ. From the two arguments we may conclude that perhaps formation of gaseous rings has its highest chance in binaries with μ neither near the maximum end of 0.5 nor near the minimum end of approaching zero. Observationally gaseous rings have been found in binaries with μ around 0.2. While this result agrees with the prediction from the previous simple arguments, it may also be caused by the effect of observational selection, since it is extremely difficult to measure μ when it is much less than 0.1

In passing, it may be noted that following the argument of available space we have predicted a few eclipsing binaries in which gaseous ring may be expected but not yet observationally detected (Huang and Struve, 1956). Among these predicted stars is β Per. (Algol) whose emission feature was later discovered by Struve and Sahade, 1957. While they have concluded that the emission feature does not indicate a ring structure, it nevertheless reveals an accumulation of gases in the system. And the accumulation of gases is a necessary condition for the ring formation.

2. Ring Formation, Mass Dissipation and Ejection Velocities

It is evident from observations that the gaseous particles flowing in the binary system come from the component stars themselves (e.g. Wood 1950; Kopal 1959; Sahade 1960). In fact, it is usually the less massive component that is losing mass. Accordingly, we will assume the injection of particles into the system by the μ component.

Let $\vec{V}$ be the velocity of ejection with respect to that point of the stellar surface from which the

particle is ejected. If the secondary component is rotating axially as a rigid body with an angular velocity $\vec{\omega}$ with respect to a stationary frame of reference, any particle that is attached to the surface rotates with a velocity $\vec{\omega}\times\vec{R}_2$ where $\vec{R}_2$ is the radius vector of a point on the secondary surface from its center. Since the center of the secondary revolves with a unit angular velocity $\vec{k}$ in the z-direction in its orbit, the ejection velocity in the xyz coordinate system is given by

$$\frac{d\vec{r}}{dt}=\vec{V}+(\vec{\omega}-\vec{k})\times\vec{R}_2. \tag{31}$$

In order to compute the C values of ejected particles, we can take advantage of the fact that the surface of the secondary coincides with the secondary lobe of the S_1 surface. Thus, it follows from the equation (1) that

$$C-C_1-\left(\frac{d\vec{r}}{dt}\right)^2 \tag{32}$$

If axial rotation and orbital revolution of the secondary (μ) component are synchronized

$$C=C_1-V^2.$$

While the particles ejected from the secondary component has values always less than C_1 according to equation (32), the C values corresponding to those periodic orbits close to the $1-\mu$ component are greater than C_1. It becomes evident that before the ejected particles accumulate to form gaseous rings close to the primary component, they must have collided one another many times such that C's of some particles have been increased to the necessary values to make the ring formation possible. Because of the conservation law given by equation (3) we may expect that C's of other particles must have been reduced as a result of collisions. Since particles of small C correspond to high velocities, they will easily escape from the system. It can, therefore, be concluded that the formation of gaseous rings of small radii around the primary component must be accompanied by dissipation of mass from the system.

If V should be very large, it would be doubtful whether the velocities of an appreciable amount of particles can be reduced by collisions to make ring formation possible. Therefore, we would suggest that the ejection velocities from the secondary are in general, small if gaseous rings are observed around the primary component. That is why we have classified ejection leading to the ring formation as a slow mode (Huang 1963).

3. Evolution and Physical Significance of Rotating Gaseous Rings

The gaseous rings around the primary component formed by the material from the secondary component cannot be permanent. Because of the tidal friction, the rapidly rotating ring will gradually lose the angular momentum to its original source of orbital motion. As the angular momentum of rotating particles decreases, they fall into the primary. Therefore, without other disturbances, the rotating rings represent only an intermediate step in the transfer of mass from the secondary to the primary component. If the ring will be dissipated easily, its presence can only indicate an active secondary at the epoch of observation.

We know solar surface activities because they provide us with a disk to observe. In the case of stars, little can be learned about their surface conditions because they appear to us as point sources, although eclipsing binaries have revealed some of the secrets of the stellar surface. Now from the observable behavior of gaseous rings we can derive, according to the present idea of ring formation, the mode of ejection of mass from stellar surface. Thus, if the ring can maintain its existence only when matter is continually supplied to it by the secondary component, its fluctuations in intensity or even its disappearance and re-emergence, which have been actually observed (Wyse 1934; Joy 1947; McNamara 1951) can only reflect the manner in which matter from the secondary is ejected.

Rings may disappear when the secondary ceases to eject matter. In this case, their disappearance would be gradual. Rings may also disappear when the secondary suddenly ejects a large number of particles of high velocities. The latter simply sweeps all rotating particles off their orbit. In this case, the disappearance of rings would most likely occur suddenly. Perhaps the fluctuation of light intensity of gaseous rings and sometimes their total disappearance actually observed are due to the second cause.

In any case, from what has been observed of the gaseous rings the ejection of matter from the

stellar surface does not resemble a continuous steady process such as the evaporation from a liquid surface. If there is ever a steady background flow out of the secondary, it is superimposed by intermittent bursts like the prominence activities on the solar surface. Thus, by observing the variation in intensity and structure of the emission lines that are produced by the rotating rings we will be able to learn something about the manner of how a component star loses its mass when its evolutionary stage of expansion brings it to touch the S_1 surface. Since such an empirical knowledge is unlikely to be found elsewhere, the importance of observing gaseous emission in binary systems cannot be exaggerated.

There remains the question whether a rotating ring or disk can be formed around the secondary (less massive) component when the primary component is losing mass. It is obvious that the ring would be less stable around the secondary than around the primary because of a larger perturbation in the first case. Also the available volume for ring formation is smaller in the first than the second case. But there is no a priori reason to believe that rings cannot be formed around the secondary component. However, observationally we have never found a gaseous ring around the secondary component.

Actually we have found few binaries whose more massive component has filled the primary lobe of the S_1 surface while those of less massive components remain small compared with the secondary lobe of the S_1 surface. Therefore, the impending question is not why we have not found any gaseous ring around the less massive component, but rather why systems whose more massive component injects particles into the secondary lobe should be so rare. Presumably, some selection effect plays a role here, but it is unlikely that this is the sole cause.

ACKNOWLEDGMENT

It is my pleasure to note my sincere thanks to Mr. Clarence Wade, Jr. for performing on the IBM 7094 computer some numerical calculations involved in this paper.

REFERENCES

BROUWER, D. and CLEMENCE, G. M., 1961, "Methods of Celestial Mechanics," Variation of Parameters, Chap. 11, New York Academic Press.
GOULD, N. L., 1959, A. J., **64**, 136.
HUANG, S. -S., 1963, Ap. J., **138**, 471.
HUANG, S. -S., 196x, to be published.
HUANG, S. -S. and STRUVE, O., 1956, A.J., **61**, 300.
HUANG, S. -S. and WADE, C., JR., 1963, A. J., **68**, 388.
JOY, A. H., 1942, Pub. A.S.P., **54**, 35.
JOY, A. H., 1947, Pub. A.S.P., **59**, 171.
KOPAL, Z., 1959, "Close Binary Systems," London: Chapman and Hall.
KUIPER, G. P., 1941, Ap. J., **93**, 133.
KUIPER, G. P. and JOHNSON, J. R., 1956, Ap. J., **123**, 90.
MCNAMARA, D. H., 1951, Ap. J., **114**, 513.
MOULTON, F. R., 1914, "An Introduction to Celestial Mechanics" 2nd Ed., New York: Macmillan, Chap. 8.
PRENDERGAST, K. H., 1960, Ap. J., **132**, 162.
SAHADE, J., 1960, "Stellar Atmospheres" Ed. J. L. Greenstein, Chicago: University of Chicago Press, Chap. 12.
STRUVE, O., 1950, "Stellar Evolution," Princeton: Princeton University Press.
STRUVE, O. and JAHADE, J., 1957, Pub. A.S.P. **69**, 41.
WOOD, F. B., 1950, Ap. J., **112**, 196.
WYSE, A. B., 1934, Lick Obs. Bull. **17**, 37 (No. 464).

APPENDIX

DERIVATION OF EQUATIONS (7)–(8) FROM HYDRODYNAMIC CONSIDERATIONS

Equation (7) has been derived already by Prendergast (1960), while equation (8) can be obtained as follows: if we neglect the pressure term, the hydrodynamic equations in the orbital x–y plane become:

$$\frac{\partial u_x}{\partial t}+u_x\frac{\partial u_x}{\partial x}+u_y\frac{\partial u_x}{\partial y}-2\mu_y=\frac{\partial U}{\partial x} \tag{1}$$

$$\frac{\partial u_y}{\partial t}+u_x\frac{\partial u_y}{\partial x}+u_y\frac{\partial u_y}{\partial y}+2u_x=\frac{\partial U}{\partial y} \tag{2}$$

Multiplying equation (1) by y and equation (2) by x and subtracting the two resulting equations we obtain

$$x\frac{\partial u_y}{\partial t}+x\left(u_x\frac{\partial u_y}{\partial x}+u_y\frac{\partial u_y}{\partial y}\right)+2xu_x-y\frac{\partial u_x}{\partial t}-y\left(u_x\frac{\partial u_x}{\partial x}+u_y\frac{\partial u_x}{\partial y}\right)+2yu_y=x\frac{\partial U}{\partial y}-y\frac{\partial U}{\partial x} \tag{3}$$

Since h may be written now as

$$h=x^2+y^2+xu_y-yu_x, \tag{4}$$

it can be easily shown that equation (3) becomes identical to equation (8) in the main test.

See N65-29437

ROTATIONAL BEHAVIOR OF THE MAIN-SEQUENCE STARS AND ITS PLAUSIBLE CONSEQUENCES CONCERNING FORMATION OF PLANETARY SYSTEMS*

SU-SHU HUANG†

A phenomenological theory for the behavior of axial rotation of main-sequence stars is proposed here by considering the effect of braking. It does not specify the physical process of braking but does provide a statistical model by which histograms of observed rotational velocities of stars of different spectral types can be explained in terms of the braking strength. It shows that rotation of all stars including those of O and B type has been braked in various degrees during their course of evolution.

As a result of braking, the stellar angular momentum is transferred outwards to the surrounding nebula that may be regarded as the remnant of star formation. This leads plausibly to the formation around the star of a planetary system from the nebula. If the angular momentum is not further dissipated from the nebula, we can estimate the size of the planetary system and its probability of occurrence. It shows that the size of a planetary system around any star is critically dependent upon the mass in the nebula as compared to that of the star itself.

I. A THEORY FOR EXPLAINING THE OBSERVED DISTRIBUTION OF STELLAR ROTATIONAL VELOCITIES

Axial rotation of the main-sequence stars stops quite abruptly at about F5 (Struve 1930). It has frequently been attributed to the braking effect of various mechanisms reviewed in a previous paper (Huang and Struve 1960). A more efficient braking mechanism than those previously proposed has recently been suggested by Schatzman (1962) as due to mass loss through stellar magnetic activities.

In the present paper we do not intend to discuss the physical process of how the braking of stellar rotation is actually affected. Rather, we will show phenomenologically that the observed distribution of rotational velocities of stars belonging to different spectral types can indeed be understood in terms of braking strengths.

As the result of a statistical study (Struve 1945; Huang and Struve 1954) it has been concluded that the rotational axes show no preference in their directions in space. This indicates that the rotating stars have not acquired their angular momentum from the galactic rotation of the pre-stellar gaseous medium. Instead, their angular momenta must have been derived from a random process, such as collisions of interstellar clouds, turbulence, etc. (Huang and Struve 1954). Consequently, we may expect that the angular momenta per unit mass, h, of stars are originally distributed according to the Maxwellian law. Finally we would like to call attention here to the fact that this observational consequence has not been fully appreciated. Many theoreticians still take the rotational angular momenta of stars as due ultimately to the galactic rotation of the tenuous pre-stellar medium (e.g. Edgeworth 1946; Hoyle 1960).

For a given spectral type the geometrical radius and the radius of gyration do not vary greatly. Hence, we may take the equatorial rotational velocity, v, to be in direct proportion to h. If we set $x = v/v_m$, where v_m denotes the most probable

*Published in The Astrophysical Journal 141: 985, April 1965.

†Goddard Space Flight Center and Catholic University, Washington, D. C.

equatorial rotational velocity, the original distribution of x in a given spectral range should be given by

$$f_o(x) = \frac{4\gamma}{\sqrt{\pi}} x^2 e^{-x^2}, \quad x \leq x_c \tag{1}$$

according to our assumption. Here x_c is the upper limit of rotational velocities because of rotational instability and γ is a numerical factor to normalize the function $f_o(x)$. Now if we assume that the star rotates as a rigid body, x is obviously related to the angular momentum per unit mass, h, of the star as follows:

$$h = Rk^2 v_m x \tag{2}$$

where R is the geometrical radius and Rk the radius of gyration of the star. We may define $h = h_c$ by equation (2) when $x = x_c$. Then interstellar clouds with $h > h_c$ could very possibly lead to the formation of close binary systems. Since at present we are interested mainly in the behavior of the observed distribution at its small-velocity end which is not significantly affected by the high-velocity tail, we shall set, for simplicity of calculation, $x_c > \infty$ and consequently $\gamma = 1$.

As a simple model, we propose that the effect of braking is to reduce h of all stars in a given spectral range by a constant amount, h_1. Define $x = x_1$ when $h = h_1$ is substituted in equation (2). It is then obvious that x_1 measures the degree of braking of axial rotation of the stars within a spectral type, namely, the stronger the braking mechanism, the greater the value of x_1. In this way we are able to obtain the distribution of rotational velocities after the braking mechanism has been applied.

Needless to say, this statistical model is very idealized and represents an extreme case. Most likely h_1 is not a constant even within a spectral sub-type but increases with the original value of h. However, since the observed distribution is not well determined, we do not see any advantage at this moment to build an elaborate model which usually would involve many adjustable parameters. On the other hand, because of this over-simplified model, the quantitative results obtained in this paper represent only a very rough estimate.

If we accept this simplified model, the distribution of equatorial rotational velocities after braking has taken its toll becomes

$$f(x, x_1) = \psi_o(x_1)\delta(x-o) + \frac{4}{\sqrt{\pi}}(x+x_1)^2 e^{-(x+x_1)^2} \tag{3}$$

where

$$\psi(x) = \frac{2}{\sqrt{\pi}}\left(\int_o^x e^{-t^2} dt - x e^{-x^2}\right) \tag{4}$$

and $\delta(x-o)$ denotes the delta function.

If $\varphi(y, x_1)$ denotes the distribution of observed velocities, y, we have (Kuiper 1935; Chandrasekhar and Münch 1950)

$$\varphi(y, x_1) = y \int_y^\infty \frac{f(x, x_1)dx}{x(x^2-y^2)^{\frac{1}{2}}} \tag{5}$$

Substituting equation (3) into equation (5) we obtain

$$\varphi(y, x_1) = \psi_o(x_1).\delta(y-o) + \psi(y, x_1) \tag{6}$$

where

$$\psi(y, x_1) = \frac{4}{\sqrt{\pi}} \int_o^{\pi/2} (y \sec\theta + x_1)^2 e^{-(y \sec\theta + x_1)^2} d\theta \tag{7}$$

The observed distributions are given in the form of histograms, i.e.,

$$\Phi(y, x_1) = \int_{(n-1)\Delta y}^{n\Delta y} \varphi(y, x_1)dy, \qquad \gamma(n-1)\Delta y < y < n\Delta y. \tag{8}$$

In Figure 1 we have shown 11 computed histograms of $\Phi(y, x_1)$ according to equations (5)–(7) for / / values of x_1 (0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 1.0, 1.4, 1.8, 2.6) with $\Delta y = 0.2$. This sequence of histograms illustrates the effect of the braking strength on the distribution of observed rotational velocities of stars in a spectral type. The case $x_1 = 0$ corresponds to that of no braking, while the case $x_1 = 2.6$ leads practically to a complete stop of axial rotation by braking.

These diagrams should be compared with the observed histograms. Unfortunately, as we can see in our previous paper (Huang 1953), the two sets of observed diagrams derived from two uncombinable sets of observational data show slight differences. The two uncombinable sets of observational data are respectively from the three-prism and two-prism spectrograms that have

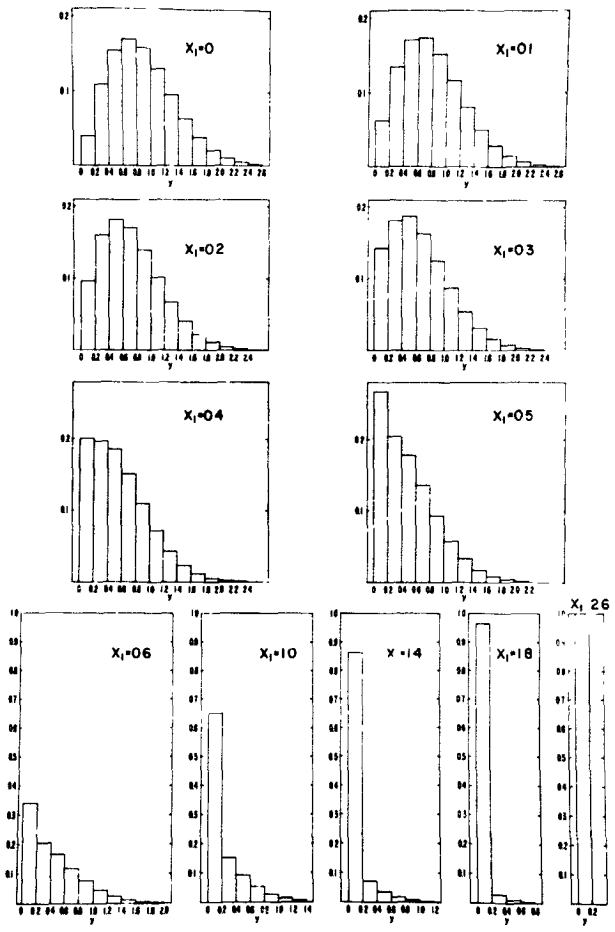


FIGURE 1.—Computed histograms of projected rotational velocities of groups of stars that have suffered different degrees of braking. We have plotted here 11 histograms for 11 values of X, which, according to our model, measures the strength of braking of stellar axial rotation.

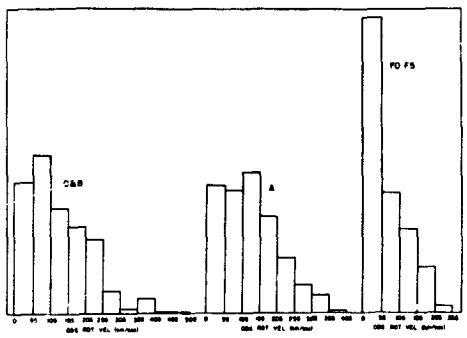


FIGURE 2—.Observed histograms of projected rotational velocities of three groups of stars, O and B, A and F0–F5, taken from a previous paper. The histogram of projected rotational velocities for later type stars will be consisted of only a single column of small velocities. These histograms should be compared with those in Figure 1. The trend of increasing braking strengths from O and B, through A and F0–F5 to later-type stars is unmistakable. From the comparison we may make a rough estimate of x_1 for each of these three groups of stars.

been taken during the years by astronomers at the Lick Observatory. There are 1103 stars for which rotational velocities were measured from three-prism spectrograms and 445 stars for which rotational velocities were measured from two-prism spectrograms. Therefore, the histograms of rotational velocities derived from the three-prism spectrograms are definitely more reliable than those from the two-prism spectrograms from a consideration of both the dispersion and the number of stars studied. For this reason we shall compare the computed histograms in Figure 1 with the observed ones obtained from three-prism histograms, the latter being reproduced here in Figure 2. Finally, it may be noted that there are available better observational data for rotational velocities (Slettebak 1949 et seq.; Cf. Huang and Struve 1960) than what was adopted here. However, our crude measurements of stellar rotational velocities provide the only uniform set of data which cover such a large number of stars that make the histograms in Figure 2 statistically significant.

In Table 1 we give the results of our determination of x_1 and v_m for each of three observed histograms after a comparison with those in Figure 1, x_1 being determined by the shape and v_m by the scales of abscissas of both observed and computed diagrams. It should once more be emphasized that these values are very tentative as both theory and observational data (especially the conversion of measured widths of the spectral line to rotational velocities) are crude. However, there is no doubt at all that x_1 is least (but definitely non-vanishing) for O and B-type stars, increases only slightly for A-type stars and becomes much larger for F0–F5 stars. Since few single stars of spectral types later than F5 are found to be rotating with a measurable velocity, we must conclude that x_1 for these late-type stars is greater than 2.6.

While we cannot determine x_1 accurately because of the uncertainty in observational data, the general behavior does indicate that our simple statistical model for the braking effect represents very well the gradual change of the distribution of observed rotational velocities from O and B through A and F0–F5, and finally to later-type

stars. Especially it explains the long puzzled fact that a large number of O, B, A stars do not rotate appreciably, although their average rotational velocities as group are high.

The previous result induces us to conclude that contrary to our previous belief, braking is not limited only to stars of spectral types later than F5 but extends to O, B, A and early F stars as well. It suggests that along the main sequence all stars show evidence of having been braked. However, the braking mechanism is weak at the upper branch of the main sequence and increases its strength rapidly after F0. It becomes so strong after F5 that rotation of all stars is practically stopped by it.

We have mentioned before that by assuming $x_c \to \infty$ we introduce an approximation. While this approximation does not seriously modify the distribution function at the lower end of observed rotational velocities, it makes an appreciable difference at the upper end. Thus, when we compare histograms (those of $x_1 = 0.3$, 0.4 and 0.8) in Figure 1 with those in Figure 2, we can immediately see a systematic difference arising from this approximation, since in all three histograms obtained from observed data, the distribution function drops much more rapidly than what have been calculated on the assumption of $x_1 \to \infty$. Thus, a cut-off of rotational velocities at their upper end is evident in all three cases in Figure 2. This shows that rotational instability has played a role not only in O- and B-type stars but also in F- and perhaps even later-type stars sometime during the course of their evolution. This result may be significant to our search of clues for star formation.

The strength of braking varies with the moment of braking force and the duration in which braking acts. If indeed the braking is due to magnetic activities which are in turn induced by convection and rotation itself in the early phase of the star's evolution as Schatzman (1962) has suggested, we would expect from the present study that magnetic activities have occurred in stars of all masses in their early phases of evolution, differing only in the strength and duration of such activities. This conclusion is in agreement with Pvoeda's (1964) interpretation of flare stars based on Hayashi's (1961) results of evolution on one hand and Fowler, Greenstein and Hoyle's (1962) theory of formation of light elements—Li, Be, B—in the early phase of the solar system on the other.

Thus far, we have followed the reasoning of both Schatzman and Poveda that the magnetic field is generated in the star as a result of convective motion and differential rotation. However, it may be noted that it would not modify our conclusion if we assume that the strong magnetic field is derived from accretion of interstellar magnetic lines of force during the period of star formation. If so, the strong magnetic activities in the early phase would represent the dissipation because of convection, of magnetic energy that is already there.

II. PROBABILITY OF FORMATION OF PLANETARY SYSTEMS AROUND STARS

It has often been suggested (e.g. Huang 1959) that the disappearance of rotation after F5 may indicate the emergence of planetary systems because it is difficult to understand how the original distribution of angular momenta of the late-type stars should be so radically different from that of early ones. This is especially true if we follow the reasoning in the previous section where the original distribution of angular momenta per unit mass of stars is assumed to follow the Maxwellian distribution (which may be truncated at the high-value end because of rotational instability). Indeed, if angular momenta that have been dissipated from the stars remain in their neighborhood, formation of planetary systems would appear inevitable. Following this line of reasoning we shall make in this section an estimate, though a very crude one, not only of the probability among single stars that will possess planetary systems but also of their general behavior with respect to the spectral type of their parent stars.

Before we shall understand why our planetary system was formed in a state as it is, we are not expected to derive, simply from a consideration of the angular momentum, the detailed structure of planetary systems around stars. In order to obtain a general behavior of any planetary system around a star without going into the detail structure we may define the concept of an "equivalent planet," which is a fictitious planet that would be moving in a circular orbit with an angular momentum per unit mass, h', equal to the average value

of the entire mass outside of, but belonging to, the star. Thus, in the case of our solar system, we have (Allen 1955) the total angular momentum of the planetary system equal to 3.15 x 10^{50} gm cm^2/sec, while its total mass is 2.68 x 10^{30} gm. Consequently, its h', denoted hereafter by h_s' is

$$h_s = 1.176 \times 10^{20} \text{ cm}^2/\text{sec}. \tag{9}$$

Then it is easy to find that the equivalent planet of our own planetary system is located at a distance

$$a_s = 1.042 \times 10^{14} \text{ cm} \tag{.10}$$

from the sun (i.e. between Jupiter and Saturn as would be expected) and revolves with a linear velocity

$$V_s = 11.3 \text{ } km/\text{sec}. \tag{11}$$

around the sun. It is obvious from our definition of the equivalent planet that it does give us a measure of the extent of any planetary system although it conveys no idea about its total mass and mass distribution in the system.

Let us now assume that the gaseous remnant around a star after its formation is m and the mass, radius and radius of gyration of the star itself are respectively M, R, and Rk. Therefore, the total angular momentum of the star is

$$\Omega(v) = MRk^2v \tag{12}$$

After braking, all of its angular momentum is transported into the surrounding medium (of mass m) according to our theory if $v \leq v_1 (= x_1 v_m)$ and a constant part (equal to $\Omega_1 = MRk^2 v_m x_1$ of it is transported out if $v > v_1$. However, we should realize that the angular momentum in the nebula that will evolve to become a planetary system could be different from what has been fed into it by the star, because some angular momentum, say $\vec{\Omega}_0$, may be originally associated with it and has never gone into the star. Since both the star and its surrounding nebula are supposed to have been formed from the same interstellar medium, we may reasonably assume that the vector, $\vec{\Omega}_0$, points to the same direction as the stellar angular momentum does. Hence, we can consider their magnitudes only and the angular momentum associated with the nebula (of mass m) after the completion of braking process has the following value:

$$\begin{aligned} \Omega(x) &= MRk^2 v_m x + \Omega_0 \text{ for stars with } x \leq x_1 \\ &= MRk^2 v_m x_1 + \Omega_0 \equiv \Omega_1 \text{ for stars with } x > x_1 \end{aligned} \tag{13}$$

For stars with slow rotation which indicates smallness of h of the pre-stellar medium, Ω_0 must be small and may be neglected. However, for stars of rapid rotation and especially for those which have passed through the stage of rotational instability in the course of their evolution, h of the prestellar medium must be large and Ω_0 large too.

If we include the possibility that a part of the angular momentum in the nebula may be dissipated away (for example through the loss of mass), Ω_0 may be even negative. Thus, Ω_0 cannot be estimated from the present consideration. However, as we shall see, this uncertainty can be formally circumvented in the following calculation.

Since

$$h' = \frac{\Omega}{m} = (GMa)^{\frac{1}{2}}, \tag{14}$$

where a is the radius of the equivalent planet's orbit around a star and

$$h_s' = a_s V_s = (GM_\odot a_s)^{\frac{1}{2}} \tag{15}$$

for the planetary system of the sun, we may introduce a new variable

$$\xi = \left(\frac{Ma}{M_\odot a_s}\right)^{\frac{1}{2}} \tag{16}$$

to measure the size of the equivalent planet's orbit around any star. It follows from equations (13)–(16) that ξ extends from

$$\xi_0 = \frac{\Omega_0}{ma_s V_s} \quad \text{to} \quad \xi_1 = \frac{\Omega_1}{ma_s V_s} \tag{17}$$

if Ω_0 is positive and zero to ξ_1, if Ω_0 is negative. Therefore, the lower and upper limit of the radius of the equivalent planet's orbit denoted respectively by a_0 and a_1 are given by

$$\xi_0 = \left(\frac{Ma_0}{M_\odot a_s}\right)^{\frac{1}{2}} \quad \text{and} \quad \xi_1 = \left(\frac{Ma_1}{M_\odot a_s}\right)^{\frac{1}{2}} \tag{18}$$

Now x and ξ are related by

$$x=\lambda\xi-\frac{\Omega_o}{MRk^2v_m} \tag{19}$$

where

$$\lambda=\frac{m}{M}\frac{a_s}{R}\frac{V_s}{v_m}\frac{1}{k^2} \tag{20}$$

It is easy to derive from equations (1) and (19) that the distribution function $\delta(\xi)$ with respect to ξ.

If we further let

$$t=\xi-\frac{\Omega_o}{ma_sV_s} \tag{21}$$

and

$$t_1=\xi_1-\frac{\Omega_o}{ma_sV_s}=\frac{x_1}{\lambda} \tag{22}$$

we obtain from $\delta(\xi)$ that

$$g\left(t+\frac{\Omega_o}{ma_sV_s}\right)=\frac{4}{\sqrt{\pi}}\lambda^3t^2e^{-\lambda^2t^2}\delta(t-t_1)[1-\psi_o(x_1)] \tag{23}$$

Since from t to ξ it involves only a horizontal translation, the shape of the distribution function does not depend upon Ω_0. If Ω_0 is positive, we simply translate the entire curve to the right, thereby increasing systematically the size of the expected orbits. If Ω_0 is negative, we translate the curve to the left and set the probability to zero everywhere that is on the left side of the origin. In this way we are able *formally* to take into account any value of Ω_0. Physically, however, we should remember that Ω_0 likely varies from case to case and moreover we do not know how it does vary (say, with respect to M, h, m, h^1, etc.). Hence, it is one of the uncertain factors that prevent us to derive the probability of occurrence of a planetary system of a given size.

Since we can obtain easily the distribution with respect to ξ from that with respect to t, we shall consider hereafter only the latter which, as we see from equation (24), depends upon two parameters, λ and t^1 or equivalently λ and x, because of equation (23). While x_1 has been determined from observed data, t, involves an unknown parameter, namely m/M.

From the values of x_1 and v_m in Table 1 we can compute t_1 (or λ) for each of the three cases given there. We shall use the mean radius $R/R_\odot=3.9$ and mean mass $M/M_\odot=5.9$ for O and B stars, $R/R_\odot=1.74$ and $M/M_\odot=2$ for A stars and $R/R_\odot=1.35$ and $M/M_\odot=1.4$ for F0–F5 stars. In all cases we adapt $k^2=0.05$. Then it follows from Table 1 and equations (20) and (23) that mt_1/M

TABLE 1.—*Values of x_1 and v_m determined from a Comparison and Observed and Calculated Histograms of Rotational Velocities*

Sp. Type	x_1	v_m(km/sec)
O and B	0.3	~200
A	0.4	~200
F0–F5	0.8	~200
later than F5	>-2.6	----------

is equal to 6.9×10^{-4} for O and B stars, equal to 4.1×10^{-4} for A stars, and equal to 6.4×10^{-4} for F0–F5 stars. It is quite evident that the error involved in the determination of x_1 and v_m is much greater than the differences we have found in these three expressions of t_1. Consequently, we may roughly set

$$t_1=6\times10^{-4}\frac{M}{m} \tag{24}$$

as representing all stars from 0 to F5. If $=0$, equation (24), when combined with equations (17), (18) and (23), gives

$$\frac{a_1}{a_s}=(6\times10^{-4})^2\frac{MM_\odot}{m^2} \tag{25}$$

which states that the maximum size of the planetary system is directly proportional to the mass of its parent star and inversely proportional to the square of its total mass. It should be noted that the numerical coefficient in equation (24) does not apply to stars later than F5 for which we do not have any observed data for making the calculation.

We have seen how critically does the size of a planetary system depend upon the mass in the

nebula. A further complication is the dissipation of mass in the nebula and accompanying dissipation of angular momentum before the formation of a planetary system of finite bodies. The difficulty of this point can be seen from the fact that even in our own planetary system there is no consensus concerning the amount of mass in the solar nebula in the beginning, as som einvestigators advocate a value of $m/M=0.1$ while others use considerably smaller values.

Since the actual size of any planetary system must be equal to or less than the upper limit, we have for our own planetary system $a_1 \geq a_s$. It then follows from equations (17) and (18) that

$$\frac{m}{M_\odot} \leq \frac{\Omega_o}{a_s V_s M_\odot} + \left(\frac{R_\odot}{a_s}\frac{v_m}{V_s}k^2\right)x_1 \qquad (26)$$

The ratio m/M for the solar system at present is equal to 1.3×10^{-3}. In the early days it must be much greater than this value. If we now take v_m to be the same as that determined in Table 1 (200 Km/sec), we have

$$\frac{R_\odot}{a_s}\frac{v_m}{V_s}k^2 = 5.9\times10^{-4} \qquad (27)$$

Therefore, it follows from equation (26) that if $\Omega_0=0$, $x_1>2.2$. Equation (1) shows that the probability of having $x>2.2$ is rather small. This would put our planetary system as an unusual case, namely its chance of occurrence is small. The other alternative is that Ω_0 is appreciable. This means that a significant portion of the angular momentum that is now found in our planetary system has never belonged to the sun. The present consideration cannot decide which alternative is the more plausible one.

Thus, we have seen many uncertainties concerning planetary systems around the stars in general. They cannot be resolved for the time being. What we can predict from the present simple theory is that the size of a planetary system increases with a decrease of its total mass and that the distribution of sizes (measured by ξ of the equivalent planet's orbit) behaves like a Maxwellian distribution of velocity magnitudes at the lower end and is truncated at the upper end. For O, B and A stars the truncation occurs at relatively small values. Therefore unless Ω_0 has a large variation among individual systems, their sizes do not vary greatly from one to the other. Most of them are crowded at the upper limit. The truncation of systems around F-type stars occurs at relatively larger values. Consequently, the spread of sizes is also larger, although there is still an appreciable number of them at the upper limit of their size. On the other hand, the sizes of planetary systems around late-type (later than F) stars are expected to be distributed like a complete Maxwellian curve with a wide spread of values. They should not show the tendency to accumulate at any particular size like systems around early-type stars do. All these consequences would be modified if Ω_0 varies greatly from one system to another; in such cases the distribution of planetary sizes can be derived only when we know the variation of Ω_0.

Finally, we would like to ask if it is possible that there is no mass left after the actual processes of star formation (i.e., $m=0$). In that case no planet would be formed. Indeed from the point of view of star formation we do not have any compelling reason to assure us that there must be some mass left behind to form planets after the star itself is formed. On the other hand, the rotational behavior of stars does suggest that there must be some mass left behind since axial rotation could not be effectively braked without the presence of such mass around the star, although we cannot at present estimate the amount of mass that is there. It is therefore the main point of this paper, if nothing else, to show the behavior of axial rotation of stars indicates the high probability of occurrence of planetary systems around stars.

The present paper perhaps represents the first serious attempt to link the problem of occurrence of planetary systems in the stellar universe with some observational facts. The uncertainties in both theory and observation force us to adapt an over-simplified model. Thus, it should be regarded as a pioneer exploration of a nebulous field of learning rather than a legitimate treatment of something that is clearly understood. As a result we emphasize only the qualitative conclusions that have been derived here.

It is my pleasure to express my thanks to Mr. Clarence Wade, Jr. for the numerical integrals in equations (7) and (8) on the IBM 7094 digital computer at Goddard.

REFERENCES

ALLEN, C. W., 1955, Astrophysical Quantities (London: Athlone Press).

CHANDRASEKHAR, S. and MÜNCH, G. 1950, Ap. J., **111**, 142.

EDGEWORTH, K. E., 1946, M.N. **106**, 470, 476 and 484.

FOWLER, W. A., GREENSTEIN, J. L. and HOYLE, F., 1962, Geophys. J. R.A.S. **6**, 148.

HAYASHI, C., 1961, Pub. Astr. Soc. Japan, **13**, 450.

HOYLE, F., 1960, Quarterly J.R.A.S. **1**, 28.

HUANG, S. -S., 1953, Ap. J., **118**, 285.

HUANG, S. -S., 1959, P.A.S.P. **71**, 421.

HUANG, S. -S., and STRUVE, O., 1954, Ann. d'Ap. **17**, 85.

HUANG, S. -S. and STRUVE, O., 1960, "Stellar Atmospheres" Ed. J. L. Greenstein (Chicago, Chicago University Press) Chap. 8. p. 321.

KUIPER, G. P., 1935, P.A.S.P. **47**, 15 and 121.

POVEDA, A., 1954, Nature **202**, 1319.

SCHATZMAN, E., 1962, Ann. d'Ap. **25**, 18.

SLETTEBAK, A., 1949, Ap. J. **110**, 498.

STRUVE, O., 1930, Ap. J. **72**, 1.

STRUVE, O., 1945, Pop. Astr. **53**, 201 and 259.

SEQUENCE OF EVENTS IN THE EARLY PHASE OF THE SOLAR SYSTEM*

SU-SHU HUANG†

The purpose of this note is to construct a sequence of events in the early phase of the solar system by bringing together several recent theories in different fields of learning. First we show that these theories are mutually salutary. Thus, when we put them together, a reasonably clear picture of the formation of the planetary system emerges.

In a recent investigation by Fowler, Greenstein and Hoyle (1) (2), it has been assumed that there was strong magnetic activities on the surface of the primeval sun. The light elements, lithium, beryllium and boron were then produced according to them by spallation processes when high-energy particles (mainly protons), which had been accelerated by the electromagnetic force on the solar surface, bombarded the dense material in the solar nebula. Furthermore, they have concluded from observed fact that at the time of being bombarded the planetesimals in the solar nebula must be on the average of 12 meters in radius if their shape was spherical.

In the meantime Hayashi (3) (4) has shown that the premain sequence stars must be in a convective equilibrium. He and his associates have calculated evolutionary tracks in the H-R diagram for these stars of several masses including the solar mass. Their results differ from the previous model based on the radiative equilibrium (5) (6) by the high luminosities in the early phase of the evolution before the main sequence.

In a paper by Faulkner, Griffiths and Hoyle (7), a question as regards the consistency between Fowler's and Hayashi's theory has been raised, because in Fowler's theory, it is required that the temperature in the solar nebula must be low—a requirement that appears to contradict the high luminosity obtained by Hayashi. However, in a recent article (8) we have argued that the temperature in the solar nebula can be low even when the luminosity of the primeval sun is high if the solar nebula is opaque enough in the directions in the plane of disk (9) (10). Since the charged particles which follow the magnetic lines do not necessarily travel in the plane of disk, the high density in the disk does not prevent charged particles to bombard the planetesimals in the disk.

More recently, Poveda (11) has presented a theory of flare stars. The success of his theory for explaining the location of these stars in the H-R diagram indicates the sound reasoning of his arguments. Now, if we follow the same reasoning as he does in his investigation of flare stars, we can immediately conclude that not only Hayashi's theory presents no difficulty to Fowler's theory, but the two are mutually conducive. In order to see this point, we have to describe briefly Poveda's theory.

The flare stars show very rapid and non-periodic change in brightness (12). They are always dwarfs stars of late spectral types. It has long been suggested that the flare stars obtain their variations of luminosity in a similar manner as the solar flares do (13) (14). Since it is generally known that solar flares are a result of magnetic activities which are in turn caused by convective motion and differential rotation (15), Poveda argues that flares must be very active when stars are undergoing evolution where convection dominates. From the calculated results of Hayashi and his associates (4), Poveda is able to plot a

*Published in *the Astronomical Society of the Pacific* 77:42, February 1965.

†Goddard Space Flight Center and Catholic University, Washington, D.C.

curve in the H-R diagram on which convection stops to be a dominant factor. Thus the flare stars should all lie on that side of the curve where convection dominates and none on the other side. Indeed, that is just what has been observed.

Then we have Schatzman's theory (16) of braking axial rotation of stars due to magnetic activities. Schatzman also invokes convective motion in the early phase of stars as the cause for magnetic activities and thereby explains why axial rotation of the main-sequence stars stops at about F5 (17). In the case of the solar system, it is very difficult to understand the present distribution of angular momentum with its high concentration in major planets as representing the original state. Schatzman's theory provides an effective mechanism for braking the rotation of the primeval sun.

Thus, we have seen three theories (Fowler's, Poveda's and Schatzman's) which predict results in agreement with observed facts and at the same time all consistently require intensive magnetic activities in the primeval sun and one theory (Hayashi's) which satisfactorily explains the loci of new stars in the H-R diagram and, at the same time, provides the clue why there should be intense magnetic activities. All these four theories not only predict observed phenomena but also have strong theoretical base. Hence, the appearance of magnetic activities in the primeval sun may be regarded as an established fact.

It follows that we should inquire whether there are indeed magnetic activities in *T* Tauri stars and flare stars that are now evolving toward the main sequence. It appears that no such stars are on the list of magnetic stars discovered by Babcock (18). However, it should be noted that the magnetic field of a star has a good chance to be discovered only when it is systematic like that of a dipole field. When the magnetic activities in stars are chaotic as are envisaged by the previous theories, the lines of force are oriented at random. Hence, they are difficult to be discovered by polarization measurements.

Hayashi and his associates (4) have given the time at which the primeval sun stopped to be completely convective to be less than 10^6 years from the time of its initial condensation and the time at which the "hydrogen burning" began to be about 25×10^6 years. If we follow Poveda that the solar surface activities of the primeval sun became reduced to the present level when the evolutionary track based on the convective model meets that of the radiative model, the time of intensive magnetic activities proposed by Fowler and others would be confined to the first 8×10^6 years of the formation of the system. This is the time that has generally been accepted as the gravitationally contracting time of the sun (1). If the intensive magnetic activities occurred only when the sun was completely convective, the formation of planetesimals of an average radius of 12 meters must have taken place in the first 10^6 years after the sun attended hydrostatic equilibrium. It is conceivable that the magnetic activities were more intense at the very early stage when the luminosity was high and convection complete than at later stages. If so, we may indeed expect that the formation of the planetesimals occurred in the first few 10^6 years.

Another supporting fact for the shorter time scale for the formation of planetesimals comes from a recent study by Hunger (19), who claims that contrary to the previous understanding (20) (21), T Tauri stars do not rotate rapidly because he has found many sharp stellar lines in three of these stars. He attributes the broad features previously believed to be due to axial rotation now to the blending of lines. It follows from his conclusion that axial rotation has already been reduced by magnetic braking before the evolving star becomes a T Tauri star.

We can now reconstruct the sequence of events in the early phase of the solar system from the previous and other theories. The early phase of evolution of the sun (or for that matter, any star) has often been divided into two stages: (1) the condensing or collapsing stage characterized by hydrodynamic inflow of matter and (2) the stellar stage characterized by hydrostatic equilibrium. According to recent studies (22) (23), the first stage is catastrophic if we neglect the effect of angular momentum and magnetic field. Gaustad (23) gives a time scale of 5×10^5 years for complete free-fall from interstellar densities to stellar conditions. The presence of a net angular momentum in the cloud prolongs somewhat the time of collapse. But from a consideration of the average angular momentum of the entire solar system observed at present we find that the effect of angular momentum on the time scale is small. Hence,

we may take one half to one million years as the time that the sun underwent the first stage. The time scale of the second stage follows results given by Hayashi and his associates (4).

The early phase of evolution of the planetary system can be divided into two corresponding stages. The transition occurred when the evolution of the sun itself was at its second stage.

As masses were falling into the protosun, accumulation of mass by direct capture of non-volatile matter (24) (25) (26) to form small local condensations far away from the primeval sun took place in what may be regarded as the primeval solar nebula which was then distributed in a spherical symmetry with respect to the sun. We may regard it as the outermost layers of infalling material at low temperatures.

According to a recent investigation by Donn and Sears (27), the particles first formed in the solar nebula are expected to be filaments and thin platelets which they call whiskers. When the whiskers are collected together they form loosely compacted instead of solidly packed condensations. Thus, the condensations in the solar nebula would have a structure resembling the lint-balls under beds or balls made of tumbleweeds that roll in the wind over the prairie under the fall sky.

Regarding the mass of these local condensations we may take the clue from Fowler's planetesimals which may be estimated from their radii to be 10^{10} gm on the average. For various reasons, a density of 10^{-9} gm/cm^3 has often been assumed for the flattened solar nebula. With this density the rate of growth of condensations by direct accumulation of mass can be calculated (24) (25). It grows about 1 cm in radius in ⅓ to 30 years, independent of the size of the body itself (25). In the spherical distribution before flattening, the density must be less than 10^{-9} (say 10^{-11} to 10^{-13}) and the rate of growth would be correspondingly slower. However, it may be noted that the previous rate was obtained (25) by assuming that the accumulating body is solidly packed. For the porus body the rate of growth in radius is faster by a factor inversely proportional to the ratio of the over-all density to the mass density of the porus body. Thus, we estimate the time of formation of condensations of mass comparable to Fowler's planetesimals to be about 10^6 years. The formation took place when the solar nebula was spherical distributed. This time scale is consistent with the time scale of solar evolution. It also agrees with the fact that the orbits of comets now observed are randomly oriented, indicating that the local condensations took place when the solar nebula was still spherical.

When strong magnetic activities appeared in the sun in its second stage of evolution before reaching the main sequence, the transfer of angular momentum from the sun to the solar nebula induced inevitably the collapse of the solar nebula from a spherical distribution to a disk one. This marks the transition of the evolution of the solar nebula from the first to the second stage. Therefore, the transition from the first stage to the second stage for the sun and that for the solar nebula are not supposed to occur at the same time but differ by an interval which covers the time for developing strong magnetic fields in the sun from hydrodynamic motion and for transporting angular momentum outward.

The local condensations might be temporarily heated up and perhaps melt during the collapse, or a rapid accumulation of matter in the process might fill their porus matrix. In any case the local condensations must have lost their porus nature and became planetesimals that Fowler envisaged when they were settled down in the rotating disk to be bombarded by high-energy particles from the sun. The collapse of the solar nebula from a spherical to a disk distribution being about 100 years if the nebula extended not far beyond Pluto's orbit, we may regard that these planetesimals received all their dosage of bombardment when they were already in the rotating disk.

The solar nebula has a life time of 2×10^8 years (25). Therefore, the formation of planets from the planetesimals must have taken place in a time scale less than this value. By considering first direct capture and then gravitational accretion, Kuiper (25) has found that this time scale is only barely enough to form planets. However, we are inclined to suggest that the formation of planets from planetesimals may take a much shorter time than this. Our reasoning rests on the fact that the planetesimals are gravitationally unstable. One can easily visualize this instability by imagining a large number of planetesimals floating in

space. A slight increase in density at one point (due to statistical fluctuations) will easily cause a rapid inflow of these bodies to that point, thereby producing a condensation of the planet size. If so, the time of formation is simply the free-falling time. Since the free falling started from a density of 10^{-9} which is more than 10^{10} times the interstellar density, the free-falling condensation of proto-planets would take only a few years.

Two points should be noted here however. First, this kind of gravitational instability is not what is known as Jeans instability which applies to a gaseous medium (28). Secondly, the present instability would be damped when gas and dust are present together with planetesimals. Therefore, it might take a longer time to form major planets. In any case, it is very likely that when the sun reached the main sequence stage (i.e., when energy dissipated is completely balanced by energy produced by thermonuclear reactions of converting hydrogen into helium), the solar system was practically in the same state as it is now found.

We now propose that because of the difficulty of transporting angular momentum to great distances, the collapse of the solar nebula into a disk occurred only in the solar neighborhood, perhaps not far beyond the orbit of Pluto. Local condensations within this limit must all have fallen into the disk. For even if they survived the initial collapse, their later crossings of the disk, which reduce their vertical velocity component, would force them to follow the general motion of gas and dust in the disk.

The local condensations contained in that part of the spherical distribution that did not undergo the collapse continued their accretion of matter by direct capture until the remnant of the solar nebula was completely dissipated. We have mentioned that at the time of formation of the disk, the local condensations had an average mass of about 10^{10} gm each. Further accretions made these condensations to reach the comet masses of $10^{15}-10^{17}$ gm each. Since these condensations have not suffered catastrophic collapse and have remained far away from the sun all the time, they maintain the porous nature till today as comet nuclei (29).

It follows from the above considerations that comets must have been much more numerous in the early days of the solar system, because there was a large volume (corresponding to the uncollapsed portion of the solar nebula) which contained these cometary nuclei. Gradually, however, the cometary nuclei were perturbed by planets either to the vicinity of the sun and were then disintegrated or to large distances from the sun and were thereby survived. The latter forms the reservoir of comets at large distances from the sun from which the present observable comets come as a result of stellar perturbation. Except by putting the original formation beyond the orbit of Jupiter, the present picture follows what has been proposed by Oort (30).

We have seen that events in the early phase of the solar system formed a natural sequence one necessarily leading to the next. Consequently, we have now good reasons to expect that the existence of planetary systems around main-sequence stars, especially of spectral type later than F5 where axial rotation stops, must be common in the universe as has been heuristically suggested before (31).

It is a pleasure to acknowledge my sincere thanks to Dr. A. Poveda for letting me read his paper before publication. It is his paper that induced me to prepare this note.

REFERENCES

1. W. A. Fowler, Science, **135**, 1037 (1962).
2. W. A. Fowler, J. L. Greenstein and F. Hoyle, Geophys. J.R.A.S., **6**, 148 (1962).
3. C. Hayashi, publ. Astr. Soc. Japan, **13**, 450 (1961).
4. C. Hayashi, R. Hoshi and D. Sigimoto, Progress. Theor. Phys. Suppl. No. 22 (1962).
5. R. D. Levée, Astrop. J. 117, 200 (1953).
6. L. G. Henyey, R. LeLevier and R. D. Levée, publ. Astr. Soc. Pacific, **67, 154** (1955).
7. F. Faulkner, K. Griffiths and F. Hoyle, Mon. Notices R.A.S. **126,** 1 (1963).
8. S. -S. Huang, Sky and Telescope, **28,** 13 (1964).
9. B. Y. Levin, "The Origin of the Earth and the Planets" (Foreign Languages publ. House, Moscow ed. 2, 1958).
10. E. J. Öpik, Icarus **1, 200** (1962).
11. A. Poveda, pre-publ. print (1964).
12. G. Haro, "Non-stable stars" (Cambridge University Press, Cambridge, 1957) p. 26; "Symposium of Stellar Evolution" (Astr. Obs. La Plata, 1962).
13. G. R. Burbidge and E. M. Burbidge, Observatory, **75, 212** (1955).
14. L. Bierman and R. Lüst, "Stellar Atmospheres" (University of Chicago Press, Chicago, 1960) p. 260.

15. E. Parker, Astrop. J. **121**, 491; **122**, 293 (1955).
16. E. Schatzman, Ann. d'Astrop., **25**, 18 (1962).
17. O. Struve, Astrop. J., **72**, 1 (1930).
18. H. W. Babcock, Astrop. J. Suppl., **3**, 141 (1953).
19. K. Hunger, Zeits. f. Astrop. **56**, 285 (1963).
20. G. H. Herbig, J. Roy. Astr. Soc. Canada, **46**, 222 (1952); Astrop. J. **125**, 612 (1957); "Symposium on Stellar Evolution" (Astr. Obs. La Plata, 1962) p. 23.
21. S. -S. Huang and O. Struve, "Stellar Atmospheres" (University of Chicago Press, Chicago, 1960) p. 321.
22. A. G. W. Cameron, Icarus, **1**, 13 (1962).
23. J. E. Gaustad, Astrop. J., **138**, 1050 (1963).
24. S. Chandrasekhar, Rev. Mod. Phys., **18**, 94 (1946).
25. G. P. Kuiper, "Astrophysics" (McGraw-Hill Book Co., New York, 1951) p. 357.
26. H. E. Urey, "The Planets" (Yale University Press, New Haven, Conn. 1952).
27. B. Donn and G. W. Sears, Science, **140**, 1208 (1963).
28. J. H. Jeans, "Astronomy & Cosmogony" (Cambridge University Press, Cambridge 1928).
29. F. L. Whipple, Astrop. J., **111**, 375 (1950).
30. J. H. Oort, Bull. Astr. Netherlands **11**, 91 (1950).
31. S. -S. Huang, publ. Astr. Soc. Pacific **71**, 421 (1959).

PLASMA NEUTRINO EMISSION FROM A HOT, DENSE ELECTRON GAS*

CULLEN L. INMAN† AND MALVIN A. RUDERMAN‡

Neutrino pair emission from coherent electrone excitations (transverse plasmons) in a hot, dense stellar plasma is calculated for a regime of temperatures and densities relevant to stellar evolution. Detailed numerical results are presented for temperatures in the range 4×10^7–5×10^8 °K and densities in the range 10^4–10^8 gm/cm³.

I. INTRODUCTION

A coupling of electron and neutrino pairs is implied by almost all models which describe the weak Fermi interactions of elementary particles. Although there has not yet been any direct experimental detection of such an electron-neutrino coupling, it is a necessary consequence of the apparent existence (Brookhaven 1963) of an intermediate heavy charged boson in β-decay. Moreover the form and magnitude of the resulting electron-neutrino interaction are unambiguously determined. The strength of the coupling is characterized by the small Fermi constant

$$g\sim3.08\sim10^{-12}\frac{\hbar^3}{m_e{}^2c},$$

where m_e is the electron mass; its form is analogous to the interaction of electromagnetic radiation with the electron current. An accelerated electron can then radiate a neutrino (ν)–antineutrino ($\bar{\nu}$) pair with the same matrix element as that for electromagnetic radiation but with greatly reduced probability. For example in an atomic transition of energy E the probability $R_{\nu\bar{\nu}}$ for radiation of $\nu\bar{\nu}$, relative to R_γ, that for radiating a photon, is only

$$\frac{R_{\nu\bar{\nu}}}{R_\gamma}\sim\frac{g^2E^4}{e^2\hbar^5c^5}\sim10^{-21}\left(\frac{E}{m_ec^2}\right)^4. \tag{1}$$

The probability for emitting a $\nu\nu$ pair rises rapidly with energy (in this case, like E^7) but even an electron-positron pair will annihilate into $\nu\nu$ rather than a pair of gamma rays only about once in 10^{22} times.

An extremely strong energy dependence is characteristic of all mechanisms for radiating $\nu\nu$ pairs: an electron moving with frequency ω in a classical circular orbit of radius R radiates electromagnetic waves, gravitational waves, and neutrino pairs with powers, P, that follow from simple dimensional considerations:

$$P_\gamma\sim\frac{e^2}{c^3}R^2\omega^4, \tag{2}$$

$$P_{\text{Grav}}\sim\frac{Gm_e{}^2R^4\omega^6}{c^5}\quad\text{(quadrupole)}, \tag{3}$$

and

$$P_{\nu\nu}\sim\frac{g^2R^2\omega^8}{hc^8}. \tag{4}$$

At low frequencies neutrino pair processes are entirely negligible. For a double star in which each member has a different electron molecular weight, the center of electron charge rotates about the fixed center of mass, but the neutrino pair radiation is smaller by perhaps eighty orders of

*Published in *The Astrophysical Journal*, 140(3): 1025–1040, October 1, 1964. This work was supported in part by the National Science Foundation.

†Goddard Space Flight Center and New York University.

‡New York University, on leave from the University of California, Berkeley, California.

magnitude than the small quadrupole radiation of gravitational waves. High frequencies ($\omega \gtrsim 10^{18}$ ~1 keV), sufficient for significant $\nu\nu$ radiation by electrons, are available only on the microscopic scale where electrons are strongly accelerated by photons or intimate collisions.

Mechanisms for the radiation of $\nu\nu$ pairs by electrons within a star include the following:

(a) $\gamma+e\rightarrow e+\nu+\nu$, (photoneutrinos) (Ritus 1962; Ida and Vahara; Chiu and Stabler 1961)

(b) $e^-+e^+\rightarrow\nu+\nu$, (pair annihilation neutrinos) (Chiu and Morrison 1960; Chiu 1961)

(c) e+Coulomb field$\rightarrow e+\nu+\nu$, (neutrino bremsstrahlung) (Gandel'man and Pineau 1960)

(d) γ+Coulomb field$\rightarrow\nu+\nu$, (photonuclear neutrinos) (Rosenberg 1963; Matinyan and Tsilosani 1962)

(e) $\gamma+\gamma\rightarrow\gamma+\nu+\nu$, (van Hieu and Shabalin 1963)

(f) plasma excitation$\rightarrow\nu+\nu$. (plasma neutrinos) (Adams, Ruderman, and Woo 1963).

Processes (a), (b), and (f) are the dominant ones in those regimes of density and temperature typical of stellar interiors (Reeves 1963). Photo and pair annihilation neutrino emissions have exact analogues in Compton scattering and the two-photon annihilation of an electron-positron pair. The main contribution to process (d) arises from the electron current associated with a transverse (E perpendicular to k) electromagnetic wave moving through the plasma.

II. PLASMA NEUTRINOS

Plasma neutrino emission has been calculated by Adams, Ruderman, and Woo (1963) for any medium whose dielectric constants are known functions of frequency and wave-number. The relevant dielectric constants have also been evaluated for an electron gas which may be relativistic and degenerate. For a plasma frequency ω_o such that $(\hbar\omega_o)^2 \ll 4(m_e c^2)^2$ the relevant transverse dielectric constant is well approximated by

$$\epsilon^t = 1-\frac{\omega_o^2}{\omega^2}, \tag{5}$$

with

$$\omega_o^2=\frac{4\pi e^2}{m_e}\int dp f(p)\left(1-\frac{p^2c^2}{3E_p^2}\right)\frac{m_e c^2}{E_p}. \tag{6}$$

Here $f(p)$ is the momentum distribution function for the electrons and

$$E_p=(p^2c^2+m_e^2c^4)^{1/2}. \tag{7}$$

The dielectric constant of equation (5) leads to a greatly simplified neutrino pair emissivity Q_t(erg/sec/cm^3):

$$Q_t=\frac{g^2\omega_o^6}{12\pi^4c^2\hbar^5}\int_0^\infty k^2dk(e^{\omega\beta}-1)^{-1}, \tag{8}$$

with $\omega=(\omega_o^2+k^2c^2)^{1/2}$ and $\beta=(k_BT)^{-1}$. The function Q_t computed from equations (6) and (8) is given as a function of density and temperature in Section III.

The form of equation (8) can be inferred in a rather straightforward way from equation (5). For a transverse electromagnetic wave in a plasma with the dielectric constant the dispersion relation is

$$\omega^2=k^2c^2+(1-\epsilon)\omega^2=k^2c^2+\omega_o^2. \tag{9}$$

Therefore such waves, when quantized, behave as if they were relativistic particles of mass $\hbar\omega_o/c^2$, and they are energetically unstable against the decay into a neutrino pair. If r is the decay rate for such a particle at rest, its decay rate when moving is decreased by the usual relativistic time dilatation to $(r\omega_o/\omega)$ and its rate of production of $\nu\nu$ becomes $(r\omega_o/\omega)\hbar\omega=r\hbar\omega_o$. Thus the total neutrino emissivity per unit volume is

$$Q_t\sim r\hbar\omega_o 8\pi\int\frac{k^2dk}{(2\pi)^3}(e^{\beta\omega}-1)^{-1}. \tag{10}$$

The integral is just the total number of "photons" of mass per unit volume in a canonical Einstein-Bose distribution. We need only calculate the $\nu\nu$ emission $r\hbar\omega_o$ caused by the oscillating electric field with $\omega=\omega_o$ and infinite wavelength ($k=0$). When $k=0$ this mode is identical to the usual plasma oscillation; the accelerated electrons oscillating in phase with frequency ω_o, amplitude x, and acceleration ω_o^2x coherently radiate neutrino pairs. But because of the finite neutrino wavelength $\lambda_\nu\sim c/\omega_o$ all of the electrons in a large volume Ω can not radiate as if they constituted a

single highly charged particle. Rather only those electrons in a cube of volume $\sim(\lambda_\nu)^3$ effectively radiate coherently. Then from a large volume Ω with electron density n, the total rate of neutrino energy emission P can be estimated from equation (4):

$$P\sim\frac{g^2\omega_o{}^8x^2}{hc^8}(n\lambda_\mu{}^3)^2\frac{\Omega}{(\lambda_\nu)^3}, \tag{11}$$

with

$$\lambda_\nu=\frac{c}{\omega_o}. \tag{12}$$

The additional kinetic energy of the oscillating electrons is $\frac{1}{2}m\omega_o{}^2x^2n\Omega$. This energy together with an equivalent average potential energy must total $h\omega_o$ when the oscillation amplitude is quantized to correspond to a single quantum of excitation (plasmon). Thus

$$m\omega_o{}^2x^2n\Omega=h\omega_o. \tag{13}$$

The combination of equations (11), (12), and (13) yields P for a single plasmon at rest, i.e., $rh\omega_o$:

$$h\omega_o r\sim\frac{g^2}{c^5}\omega_o{}^4\frac{n}{m}. \tag{14}$$

But

$$\omega_o{}^2\sim4\pi ne^2/m, \tag{15}$$

so that

$$h\omega_o r\sim\frac{g^2}{c^2}\frac{\omega_o{}^6}{c^5}. \tag{16}$$

Equations (16) and (10) yield equation (8).

III. EXACT CALCULATION OF THE TRANSVERSE EMISSIVITY

(a) Introduction

In Section II it was shown that the rate of loss of energy in neutrino pairs due to decay of transverse plasmons is

$$Q_t=2g^2(3\pi e^2)^{-1}(2\pi)^{-3}\omega_o{}^6\sum_1^\infty\int^\infty \exp(-n\beta\omega)k^2dk. \tag{17}$$

The units are $h=m_e=c=1$. Here $g=3.08\times10^{-12}$ is the weak coupling constant; ω is given by the dispersion relation for the transverse plasmons, $\omega^2=\omega_o{}^2+k^2$; and

$$\omega_o{}^2=4e^2p_F{}^3/3\pi E_F) \tag{18}$$

is the plasma frequency. Of course $\beta=1/k_BT$). We take $\mu_e=2$.

We define a function $\mathfrak{F}(x)$ of x alone by

$$\mathfrak{F}(x)=\sum_1^\infty\int_0^\infty \exp[-nx\cosh\xi]\sinh^2\xi\cosh\xi d\xi. \tag{19}$$

By considering the dispersion relation and equation (17) one sees that

$$Q_t=2g^2(3\pi e^2)^{-1}(2\pi)^{-3}\omega_o{}^9\mathfrak{F}(\beta\omega_o), \tag{20}$$

and we may note that the expression for the number density of a gas of bosons at zero chemical potential is

$$\frac{N}{V}=\frac{1}{\pi^2}\left(\frac{mc}{\hbar}\right)^3\mathfrak{F}(\beta mc^2). \tag{21}$$

Apart from care in handling the conversion of units we need then only discuss the calculation of $\mathfrak{F}(x)$.

(b) The Function $\mathfrak{F}$ (x)

1. *Small ($x\leq0.5$) values of the argument.*—It is shown in the Appendix that for $x<2\pi$ there exists an expansion of $x^3\mathfrak{F}(x)$; note

$$\zeta(3)=\sum_{n=1}^\infty(1/n^3).$$

Truncating it,

$$\mathfrak{F}(x)\doteq\frac{1}{x^3}\left[\left[2\zeta(3)+\frac{1}{2}x^2\ln x-\frac{1}{4}(2\ln2+1)x^2+\frac{1}{96}x^4\ln x-\frac{1}{96}\left\{\ln2-\frac{1}{4}+\ln2\pi+\left[-\frac{\zeta'(2)}{\zeta(2)}\right]\right\}x^4\right]\right]. \tag{22}$$

For $x=0.5$ this gives a result which differs from that obtained by summing the Hankel series (see subsection 2 below) by one part in 10^6. It will also be seen that by taking only the constant term in the square brackets, and using equation (20), we get just the equation preceding equation (28) in Adams *et al.* (1963).

2. *Intermediate and large values of the argument ($x\geq0.5$).*—Chandrasekhar (1957) gives a result (p. 398, eq.[252] and [248] that, omitting the

minus signs and Λ, since we have bosons at zero chemical potential, reads

$$\mathfrak{F}(x)=\sum_{n=1}^{\infty}\frac{K_2(nx)}{nx}\doteq\sum_{n=1}^{M}\frac{K_2(nx)}{nx}, \qquad (23)$$

where $K_2(z)$ is the modified Bessel function of the second kind, of order 2. The criterion for the choice of M is discussed in subsection 3. The series equation (23) is a Hankel series, similar to a Dirichlet series, and converges for all x, albeit very slowly for small x: (see Greenwood [1941]). Note that for small ξ, $K_2(\xi)\approx 2/\xi^3$, so that substituting this expression in equation (23) we get again

$$\mathfrak{F}(x)\approx\frac{2}{x^3}\sum_{n=1}^{\infty}\left(\frac{1}{n^3}\right)=\frac{2\zeta(3)}{x^3},$$

for small x, just what one obtains by taking only the *zero*th-order term in braces in equation (22).

For large ξ, on the other hand, $K_2(\xi)\approx(\pi/2\xi)^{1/2}e^{-\xi}$, and if we take only the first term of equation (23), using this expression as an approximation for it, we get $\mathfrak{F}(x)\approx(\pi/2)^{1/2}x^{-3/2}e^{-x}$, for large x.

If one uses this last plus equation (20) one gets just the equation preceding equation (29) in Adams *et al.* (1963).

3. *Error terms.*—For equation (22), the small argument form, it is shown in the Appendix that the lowest order neglected term is $O(x^6\ln x)$. The coefficient multiplying this term will be quite small.

For equation (23), the intermediate and large form, we break off the sum when

$$\frac{K_2(Mx)}{Mx}<10^{-6}\frac{K_2(x)}{x}.$$

By approximating, for large M, the remainder term by

$$\sum_{n=M+1}^{\infty}\frac{K_2(nx)}{nx}\approx\frac{1}{x}\int_{M+1/2}^{\infty}\frac{K_2(nx)}{n}dn$$

$$=\frac{1}{x}\frac{K_1[(M+\tfrac{1}{2})x]}{(M+\tfrac{1}{2})x}<\frac{1}{x}\frac{K_2(Mx)}{Mx},$$

we find that the error is less than $(1/x)\times10^{-6}$ times the first term. This is for the range 0.5–5, since for larger values we have no need of many terms.

(c) Conversion of Units

The dimensions of Q_t are erg/cm^3/sec. In the units in which equation (20) is given, units of energy, length, and time are m_ec^2, $\hbar/m_ec$, and $\hbar/m_ec^2$, respectively. Thus Q_t in c.g.s. units is

$$Q_t=\frac{m_ec^2}{(\hbar/m_ec)^3}\frac{1}{\hbar/(m_ec^2)}\left[\frac{2}{3\pi}\times137.04\times(3.08\times10^{-12})^2\frac{1}{(2\pi)^3}\right]+\left(\frac{\hbar\omega_o}{m_ec^2}\right)^9\mathfrak{F}(\beta\hbar\omega_o)=$$

$$=1.228\times10^{22}\left(\frac{\hbar\omega_o}{m_ec^2}\right)^9\mathfrak{F}(\beta\hbar\omega_o), \qquad (24)$$

where now all dimensioned quantities are in c.g.s. units. In c.g.s. units

$$p_F{}^3=3\pi^2N_{\mathrm{AVO}}\frac{1}{\mu_e}\rho=\frac{3}{2}\pi^2N_{\mathrm{AVO}}\,\rho.$$

Thus in c.g.s. units

$$\hbar^2\omega_o{}^2=2\pi(\hbar c)^2\left[1+\left(\frac{\hbar}{m_ec}\right)^2\left(\frac{3}{2}\pi^2N_{\mathrm{AVO}}\,\rho\right)^{2/3}\right]^{-1/2}N_{\mathrm{AVO}}\frac{e^2}{m_ec^2}\rho$$

or

$$\hbar\omega_o=3.265\times10^{-11}(1+6.413\times10^{-5}\rho^{2/3})^{-1/4}\rho^{1/2}, \qquad (25)$$

where now all dimensioned quantities are in c.g.s. units.

In c.g.s. units

$$\beta=\frac{7.244\times10^8}{T_7}, \qquad (26)$$

where T_7 is the temperature in units of 10^7 °K.

Using equations (25) and (26) in equation (24) we can compute Q_t for given values of ρ and T. Of course

$$q_t=\frac{Q_t}{\rho}.$$

Table 1 gives q_t as a function of ρ and T.

TABLE 1.—*Neutrino Emissivity by Transverse Plasmons*

T	q sub T	T	q sub T	T	q sub T
log rho = 4.00					
$.40000\times10^{8}$	$.30491\times10^{-2}$	$.20000\times10^{9}$	$.43162\times10^{0}$	$.36000\times10^{9}$	$.25327\times10^{1}$
$.60000\times10^{8}$	$.10963\times10^{-1}$	$.22000\times10^{9}$	$.57531\times10^{0}$	$.38000\times10^{9}$	$.29796\times10^{1}$
$.80000\times10^{8}$	$.26678\times10^{-1}$	$.24000\times10^{9}$	$.74776\times10^{0}$	$.40000\times10^{9}$	$.34762\times10^{1}$
$.10000\times10^{9}$	$.52820\times10^{-1}$	$.26000\times10^{9}$	$.95157\times10^{0}$	$.42000\times10^{9}$	$.40251\times10^{1}$
$.12000\times10^{9}$	$.92010\times10^{-1}$	$.28000\times10^{9}$	$.11894\times10^{1}$	$.44000\times10^{9}$	$.46289\times10^{1}$
$.14000\times10^{9}$	$.14687\times10^{0}$	$.30000\times10^{9}$	$.14638\times10^{1}$	$.46000\times10^{9}$	$.52902\times10^{1}$
$.16000\times10^{9}$	$.22001\times10^{0}$	$.32000\times10^{9}$	$.17774\times10^{1}$	$.48000\times10^{9}$	$.60116\times10^{1}$
$.18000\times10^{9}$	$.31405\times10^{0}$	$.34000\times10^{9}$	$.21328\times10^{1}$	$.50000\times10^{9}$	$.67958\times10^{1}$
log rho = 4.20					
$.40000\times10^{8}$	$.71282\times10^{-2}$	$.20000\times10^{9}$	$.10627\times10^{1}$	$.36000\times10^{9}$	$.62533\times10^{1}$
$.60000\times10^{8}$	$.26293\times10^{-1}$	$.22000\times10^{9}$	$.14174\times10^{1}$	$.38000\times10^{9}$	$.73579\times10^{1}$
$.80000\times10^{8}$	$.64689\times10^{-1}$	$.24000\times10^{9}$	$.18432\times10^{1}$	$.40000\times10^{9}$	$.85854\times10^{1}$
$.10000\times10^{9}$	$.12883\times10^{0}$	$.26000\times10^{9}$	$.23466\times10^{1}$	$.42000\times10^{9}$	$.99421\times10^{1}$
$.12000\times10^{9}$	$.22520\times10^{0}$	$.28000\times10^{9}$	$.29340\times10^{1}$	$.44000\times10^{9}$	$.11435\times10^{2}$
$.14000\times10^{9}$	$.34029\times10^{0}$	$.30000\times10^{9}$	$.36119\times10^{1}$	$.46000\times10^{9}$	$.13069\times10^{2}$
$.16000\times10^{9}$	$.54058\times10^{0}$	$.32000\times10^{9}$	$.43868\times10^{1}$	$.48000\times10^{9}$	$.14853\times10^{2}$
$.18000\times10^{9}$	$.77254\times10^{0}$	$.34000\times10^{9}$	$.52651\times10^{1}$	$.50000\times10^{9}$	$.16792\times10^{2}$
log rho = 4.40					
$.40000\times10^{8}$	$.16221\times10^{-1}$	$.20000\times10^{9}$	$.25971\times10^{1}$	$.36000\times10^{9}$	$.15346\times10^{2}$
$.60000\times10^{8}$	$.61954\times10^{-1}$	$.22000\times10^{9}$	$.34674\times10^{1}$	$.38000\times10^{9}$	$.18061\times10^{2}$
$.80000\times10^{8}$	$.15476\times10^{0}$	$.24000\times10^{9}$	$.45125\times10^{1}$	$.40000\times10^{9}$	$.21077\times10^{2}$
$.10000\times10^{9}$	$.31071\times10^{0}$	$.26000\times10^{9}$	$.57483\times10^{1}$	$.42000\times10^{9}$	$.24412\times10^{2}$
$.12000\times10^{9}$	$.54582\times10^{0}$	$.28000\times10^{9}$	$.71908\times10^{1}$	$.44000\times10^{9}$	$.28081\times10^{2}$
$.14000\times10^{9}$	$.87607\times10^{0}$	$.30000\times10^{9}$	$.88559\times10^{1}$	$.46000\times10^{9}$	$.32100\times10^{2}$
$.16000\times10^{9}$	$.13174\times10^{1}$	$.32000\times10^{9}$	$.10760\times10^{1}$	$.48000\times10^{9}$	$.36485\times10^{2}$
$.18000\times10^{9}$	$.18858\times10^{1}$	$.34000\times10^{9}$	$.12918\times10^{1}$	$.50000\times10^{9}$	$.41251\times10^{2}$
log rho = 4.60					
$.40000\times10^{8}$	$.35602\times10^{-1}$	$.20000\times10^{9}$	$.62850\times10^{1}$	$.36000\times10^{9}$	$.37354\times10^{2}$
$.60000\times10^{8}$	$.14256\times10^{0}$	$.22000\times10^{9}$	$.84023\times10^{1}$	$.38000\times10^{9}$	$.43975\times10^{2}$
$.80000\times10^{8}$	$.36355\times10^{0}$	$.24000\times10^{9}$	$.10946\times10^{2}$	$.40000\times10^{9}$	$.51335\times10^{2}$
$.10000\times10^{9}$	$.73812\times10^{0}$	$.26000\times10^{9}$	$.13956\times10^{2}$	$.42000\times10^{9}$	$.59471\times10^{2}$
$.12000\times10^{9}$	$.13055\times10^{1}$	$.28000\times10^{9}$	$.17471\times10^{2}$	$.44000\times10^{9}$	$.68424\times10^{2}$
$.14000\times10^{9}$	$.21049\times10^{1}$	$.30000\times10^{9}$	$.21529\times10^{2}$	$.46000\times10^{9}$	$.78231\times10^{2}$
$.16000\times10^{9}$	$.31753\times10^{1}$	$.32000\times10^{9}$	$.26170\times10^{2}$	$.48000\times10^{9}$	$.88931\times10^{2}$
$.18000\times10^{9}$	$.45557\times10^{1}$	$.34000\times10^{9}$	$.31432\times10^{2}$	$.50000\times10^{9}$	$.10056\times10^{3}$

TABLE 1.—*Neutrino Emissivity by Transverse Plasmons—Continued*

T	q sub γ	T	q sub T	T	q sub T
log rho = 4.80					
$.40000\times10^{8}$	$.74499\times10^{-1}$	$.20000\times10^{9}$	$.15011\times10^{2}$	$.36000\times10^{9}$	$.89949\times10^{2}$
$.60000\times10^{8}$	$.31782\times10^{0}$	$.22000\times10^{9}$	$.20105\times10^{2}$	$.38000\times10^{9}$	$.10594\times10^{3}$
$.80000\times10^{8}$	$.83355\times10^{0}$	$.24000\times10^{9}$	$.26232\times10^{2}$	$.40000\times10^{9}$	$.12372\times10^{3}$
$.10000\times10^{9}$	$.17184\times10^{1}$	$.26000\times10^{9}$	$.33485\times10^{2}$	$.42000\times10^{9}$	$.14338\times10^{3}$
$.12000\times10^{9}$	$.30679\times10^{1}$	$.28000\times10^{9}$	$.41959\times10^{2}$	$.44000\times10^{9}$	$.16501\times10^{3}$
$.14000\times10^{9}$	$.49773\times10^{1}$	$.30000\times10^{9}$	$.51749\times10^{2}$	$.46000\times10^{9}$	$.18871\times10^{3}$
$.16000\times10^{9}$	$.75411\times10^{1}$	$.32000\times10^{9}$	$.62947\times10^{2}$	$.48000\times10^{9}$	$.21457\times10^{3}$
$.18000\times10^{9}$	$.10854\times10^{2}$	$.34000\times10^{9}$	$.75649\times10^{2}$	$.50000\times10^{9}$	$.24269\times10^{3}$
log rho = 5.00					
$.40000\times10^{8}$	$.14647\times10^{0}$	$.20000\times10^{9}$	$.35234\times10^{2}$	$.36000\times10^{9}$	$.21359\times10^{3}$
$.60000\times10^{8}$	$.67973\times10^{0}$	$.22000\times10^{9}$	$.47315\times10^{2}$	$.38000\times10^{9}$	$.25169\times10^{3}$
$.80000\times10^{8}$	$.18510\times10^{1}$	$.24000\times10^{9}$	$.61862\times10^{2}$	$.40000\times10^{9}$	$.29409\times10^{3}$
$.10000\times10^{9}$	$.38953\times10^{1}$	$.26000\times10^{9}$	$.79101\times10^{2}$	$.42000\times10^{9}$	$.34098\times10^{3}$
$.12000\times10^{9}$	$.70433\times10^{1}$	$.28000\times10^{9}$	$.99257\times10^{2}$	$.44000\times10^{9}$	$.39259\times10^{3}$
$.14000\times10^{9}$	$.11524\times10^{2}$	$.30000\times10^{9}$	$.12256\times10^{3}$	$.46000\times10^{9}$	$.44914\times10^{3}$
$.16000\times10^{9}$	$.17565\times10^{2}$	$.32000\times10^{9}$	$.14922\times10^{3}$	$.48000\times10^{9}$	$.51087\times10^{3}$
$.18000\times10^{9}$	$.25392\times10^{2}$	$.34000\times10^{9}$	$.17948\times10^{3}$	$.50000\times10^{9}$	$.57800\times10^{3}$
log rho = 5.20					
$.40000\times10^{8}$	$.26574\times10^{0}$	$.20000\times10^{9}$	$.80867\times10^{2}$	$.36000\times10^{9}$	$.49794\times10^{3}$
$.60000\times10^{8}$	$.13778\times10^{1}$	$.22000\times10^{9}$	$.10898\times10^{3}$	$.38000\times10^{9}$	$.58734\times10^{3}$
$.80000\times10^{8}$	$.39434\times10^{1}$	$.24000\times10^{9}$	$.14290\times10^{3}$	$.40000\times10^{9}$	$.68679\times10^{3}$
$.10000\times10^{9}$	$.85295\times10^{1}$	$.26000\times10^{9}$	$.18314\times10^{3}$	$.42000\times10^{9}$	$.79683\times10^{3}$
$.12000\times10^{9}$	$.15688\times10^{2}$	$.28000\times10^{9}$	$.23024\times10^{3}$	$.44000\times10^{9}$	$.91798\times10^{3}$
$.14000\times10^{9}$	$.25962\times10^{2}$	$.30000\times10^{9}$	$.28475\times10^{3}$	$.46000\times10^{9}$	$.10508\times10^{4}$
$.16000\times10^{9}$	$.39891\times10^{2}$	$.32000\times10^{9}$	$.34717\times10^{3}$	$.48000\times10^{9}$	$.11958\times10^{4}$
$.18000\times10^{9}$	$.58015\times10^{2}$	$.34000\times10^{9}$	$.41806\times10^{3}$	$.50000\times10^{9}$	$.13534\times10^{4}$
log rho = 5.40					
$.40000\times10^{8}$	$.43556\times10^{0}$	$.20000\times10^{9}$	$.18037\times10^{3}$	$.36000\times10^{9}$	$.11346\times10^{4}$
$.60000\times10^{8}$	$.26082\times10^{1}$	$.22000\times10^{9}$	$.24426\times10^{3}$	$.38000\times10^{9}$	$.13399\times10^{4}$
$.80000\times10^{8}$	$.79688\times10^{1}$	$.24000\times10^{9}$	$.32150\times10^{3}$	$.40000\times10^{9}$	$.15685\times10^{4}$
$.10000\times10^{9}$	$.17871\times10^{2}$	$.26000\times10^{9}$	$.41333\times10^{3}$	$.42000\times10^{9}$	$.18214\times10^{4}$
$.12000\times10^{9}$	$.33617\times10^{2}$	$.28000\times10^{9}$	$.52100\times10^{3}$	$.44000\times10^{9}$	$.21001\times10^{4}$
$.14000\times10^{9}$	$.56484\times10^{2}$	$.30000\times10^{9}$	$.64573\times10^{3}$	$.46000\times10^{9}$	$.24056\times10^{4}$
$.16000\times10^{9}$	$.87732\times10^{2}$	$.32000\times10^{9}$	$.78876\times10^{3}$	$.48000\times10^{9}$	$.27393\times10^{4}$
$.18000\times10^{9}$	$.12861\times10^{3}$	$.34000\times10^{9}$	$.95131\times10^{3}$	$.50000\times10^{9}$	$.31023\times10^{4}$

TABLE 1 —*Neutrino Emissivity by Transverse Plasmons—Continued*

T	q sub T	T	q sub T	T	q sub T
log rho = 5.60					
.40000×10⁸	.62929×10⁰	.20000×10⁹	.38829×10³	.36000×10⁹	.25132×10⁴
.60000×10⁸	.45341×10¹	.22000×10⁹	.52917×10³	.38000×10⁹	.29727×10⁴
.80000×10⁸	.15074×10²	.24000×10⁹	.70009×10³	.40000×10⁹	.34846×10⁴
.10000×10⁹	.35435×10²	.26000×10⁹	.90385×10³	.42000×10⁹	.40517×10⁴
.12000×10⁹	.68650×10²	.28000×10⁹	.11433×10⁴	.44000×10⁹	.46767×10⁴
.14000×10⁹	.11766×10³	.30000×10⁹	.14211×10⁴	.46000×10⁹	.53624×10⁴
.16000×10⁹	.18537×10³	.32000×10⁹	.17402×10⁴	.48000×10⁹	.61116×10⁴
.18000×10⁹	.27463×10³	.34000×10⁹	.21033×10⁴	.50000×10⁹	.69271×10⁴
log rho = 5.80					
.40000×10⁸	.78021×10⁰	.20000×10⁹	.80068×10⁰	.36000×10⁹	.53805×10⁴
.60000×10⁸	.71030×10¹	.22000×10⁹	.11003×10⁴	.38000×10⁹	.63779×10⁴
.80000×10⁸	.26307×10²	.24000×10⁹	.14655×10⁴	.40000×10⁹	.74902×10⁴
.10000×10⁹	.65691×10²	.26000×10⁹	.19025×10⁴	.42000×10⁹	.87235×10⁴
.12000×10⁹	.13219×10³	.28000×10⁹	.24174×10⁴	.44000×10⁹	.10084×10⁵
.14000×10⁹	.23248×10³	.30000×10⁹	.30165×10⁴	.46000×10⁹	.11578×10⁵
.16000×10⁹	.37307×10³	.32000×10⁹	.37060×10⁴	.48000×10⁹	.13211×10⁵
.18000×10⁹	.56035×10³	.34000×10⁹	.44919×10⁴	.50000×10⁹	.14990×10⁵
log rho = 6.00					
.40000×10⁸	.80711×10⁰	.20000×10⁹	.15694×10⁴	.36000×10⁹	.11072×10⁵
.60000×10⁸	.98284×10¹	.22000×10⁹	.21800×10⁴	.38000×10⁹	.13161×10⁵
.80000×10⁸	.41696×10²	.24000×10⁹	.29289×10⁴	.40000×10⁹	.15494×10⁵
.10000×10⁹	.11240×10³	.26000×10⁹	.38294×10⁴	.42000×10⁹	.18085×10⁵
.12000×10⁹	.23737×10³	.28000×10⁹	.48949×10⁴	.44000×10⁹	.20946×10⁵
.14000×10⁹	.43140×10³	.30000×10⁹	.61386×10⁴	.46000×10⁹	.24090×10⁵
.16000×10⁹	.70879×10³	.32000×10⁹	.75739×10⁴	.48000×10⁹	.27532×10⁵
.18000×10⁹	.10835×10⁴	.34000×10⁹	.92139×10⁴	.50000×10⁹	.31283×10⁵
log rho = 6.20					
.40000×10⁸	.67726×10⁰	.20000×10⁹	.29029×10⁴	.36000×10⁹	.21789×10⁵
.60000×10⁸	.11768×10²	.22000×10⁹	.40878×10⁴	.38000×10⁹	.25994×10⁵
.80000×10⁸	.59073×10²	.24000×10⁹	.55529×10⁴	.40000×10⁹	.30699×10⁵
.10000×10⁹	.17521×10³	.26000×10⁹	.73263×10⁴	.42000×10⁹	.35933×10⁵
.12000×10⁹	.39311×10³	.28000×10⁹	.94360×10⁴	.44000×10⁹	.41722×10⁵
.14000×10⁹	.74457×10³	.30000×10⁹	.11910×10⁵	.46000×10⁹	.48094×10⁵
.16000×10⁹	.12603×10⁴	.32000×10⁹	.14775×10⁵	.48000×10⁹	.55075×10⁵
.18000×10⁹	.19700×10⁴	.34000×10⁹	.18059×10⁵	.50000×10⁹	.62693×10⁵

TABLE 1.—*Neutrino Emissivity by Transverse Plasmons—Continued*

T	q sub T	T	q sub T	T	q sub T
log rho = 6.40					
.40000×10⁹	.44834×10⁰	.20000×10⁹	.50338×10⁴	.36000×10⁹	.40850×10⁵
.60000×10⁸	.11950×10²	.22000×10⁹	.72106×10⁴	.38000×10⁹	.48954×10⁵
.80000×10⁸	.73637×10²	.24000×10⁹	.99315×10⁴	.40000×10⁹	58048×10⁵
.10000×10⁹	.24563×10³	.26000×10⁹	.13254×10⁵	.42000×10⁹	.68187×10⁵
.12000×10⁹	.59398×10³	.28000×10⁹	.17234×10⁵	.44000×10⁹	.79425×10⁵
.14000×10⁹	.11843×10⁴	.30000×10⁹	.21928×10⁵	.46000×10⁹	.91817×10⁵
.16000×10⁹	.20801×10⁴	.32000×10⁹	.27392×10⁵	.48000×10⁹	.10901×10⁶
.18000×10⁹	.33433×10⁴	.34000×10⁹	.33681×10⁵	.50000×10⁹	.12028×10⁶
log rho = 6.60					
.40000×10⁸	.22771×10⁰	.20000×10⁹	.81367×10⁴	.36000×10⁹	.72753×10⁵
.60000×10⁸	.10087×10²	.22000×10⁹	.11905×10⁵	.38000×10⁹	.87684×10⁵
.80000×10⁸	.79523×10²	.24000×10⁹	.16681×10⁵	.40000×10⁹	.10450×10⁶
.10000×10⁹	.30589×10³	.26000×10⁹	.22579×10⁵	.42000×10⁹	.12330×10⁶
.12000×10⁹	.81050×10³	.28000×10⁹	.29709×10⁵	.44000×10⁹	.14420×10⁶
.14000×10⁹	.17209×10⁴	.30000×10⁹	.38183×10⁵	.46000×10⁹	.16729×10⁶
.16000×10⁹	.31636×10⁴	.32000×10⁹	.48110×10⁵	.48000×10⁹	.19270×10⁶
.18000×10⁹	.52622×10⁴	.34000×10⁹	.59597×10⁵	.50000×10⁹	.22051×10⁶
log rho = 6.80					
.40000×10⁸	.86165×10⁻¹	.20000×10⁹	.12199×10⁵	.36000×10⁹	.12289×10⁶
.60000×10⁸	.69370×10¹	.22000×10⁹	.18317×10⁵	.38000×10⁹	.14916×10⁶
.80000×10⁸	.73257×10²	.24000×10⁹	.26213×10⁵	.40000×10⁹	.17887×10⁶
.10000×10⁹	.33420×10³	.26000×10⁹	.36104×10⁵	.42000×10⁹	.21225×10⁶
.12000×10⁹	.98873×10³	.28000×10⁹	.48205×10⁵	.44000×10⁹	.24944×10⁶
.14000×10⁹	.22657×10⁴	.30000×10⁹	.62727×10⁵	.46000×10⁹	.29069×10⁶
.16000×10⁹	.44028×10⁴	.32000×10⁹	.79878×10⁵	.48000×10⁹	.33618×10⁶
.18000×10⁹	.76364×10⁴	.34000×10⁹	.99864×10⁵	.50000×10⁹	.38611×10⁶
log rho = 7.00					
.40000×10⁸	.23503×10⁻¹	.20000×10⁹	.16879×10⁵	.36000×10⁹	.17069×10⁶
.60000×10⁸	.38023×10¹	.22000×10⁹	.26157×10⁵	.38000×10⁹	.24081×10⁶
.80000×10⁸	.56626×10²	.24000×10⁹	.38406×10⁵	.40000×10⁹	.29160×10⁶
.10000×10⁹	.31620×10³	.26000×10⁹	.54035×10⁵	.42000×10⁹	.34673×10⁶
.12000×10⁹	.10670×10⁴	.28000×10⁹	.73445×10⁵	.44000×10⁹	.41108×10⁶
.14000×10⁹	.26795×10⁴	.30000×10⁹	.97028×10⁵	.46000×10⁹	.48169×10⁶
.16000×10⁹	.55669×10⁴	.32000×10⁹	.12518×10⁶	.48000×10⁹	.55985×10⁶
.18000×10⁹	.10157×10⁵	.34000×10⁹	.15827×10⁶	.50000×10⁹	.64591×10⁶

TABLE 1.—*Neutrino Emissivity by Transverse Plasmons—Continued*

T	q sub T	T	q sub T	T	q sub T
log rho = 7.20					
$.40000\times10^{8}$	$.44441\times10^{-4}$	$.20000\times10^{9}$	$.21436\times10^{5}$	$.36000\times10^{9}$	$.29799\times10^{6}$
$.60000\times10^{8}$	$.16197\times10^{1}$	$.22000\times10^{9}$	$.34507\times10^{5}$	$.38000\times10^{9}$	$.36871\times10^{6}$
$.80000\times10^{8}$	$.36048\times10^{2}$	$.24000\times10^{9}$	$.52266\times10^{5}$	$.40000\times10^{9}$	$.44975\times10^{6}$
$.10000\times10^{9}$	$.25530\times10^{3}$	$.26000\times10^{9}$	$.75453\times10^{5}$	$.42000\times10^{9}$	$.54178\times10^{6}$
$.12000\times10^{9}$	$.10066\times10^{4}$	$.28000\times10^{9}$	$.10480\times10^{6}$	$.44000\times10^{9}$	$.64546\times10^{6}$
$.14000\times10^{9}$	$.28188\times10^{4}$	$.30000\times10^{9}$	$.14102\times10^{6}$	$.46000\times10^{9}$	$.76145\times10^{6}$
$.16000\times10^{9}$	$.63435\times10^{4}$	$.32000\times10^{9}$	$.18483\times10^{6}$	$.48000\times10^{9}$	$.89040\times10^{6}$
$.18000\times10^{9}$	$.12300\times10^{5}$	$.34000\times10^{9}$	$.23693\times10^{6}$	$.50000\times10^{9}$	$.10330\times10^{7}$
log rho = 7.40					
$.40000\times10^{8}$	$.55558\times10^{-3}$	$.20000\times10^{9}$	$.24808\times10^{5}$	$.36000\times10^{9}$	$.42668\times10^{6}$
$.60000\times10^{8}$	$.52006\times10^{0}$	$.22000\times10^{9}$	$.41804\times10^{5}$	$.38000\times10^{9}$	$.53475\times10^{6}$
$.80000\times10^{8}$	$.18484\times10^{2}$	$.24000\times10^{9}$	$.65728\times10^{5}$	$.40000\times10^{9}$	$.65975\times10^{6}$
$.10000\times10^{9}$	$.17286\times10^{3}$	$.26000\times10^{9}$	$.97880\times10^{5}$	$.42000\times10^{9}$	$.80285\times10^{6}$
$.12000\times10^{9}$	$.81841\times10^{3}$	$.28000\times10^{9}$	$.13955\times10^{6}$	$.44000\times10^{9}$	$.96524\times10^{6}$
$.14000\times10^{9}$	$.26067\times10^{3}$	$.30000\times10^{9}$	$.19201\times10^{6}$	$.46000\times10^{9}$	$.11481\times10^{7}$
$.16000\times10^{9}$	$.64504\times10^{3}$	$.32000\times10^{9}$	$.25653\times10^{6}$	$.48000\times10^{9}$	$.13525\times10^{7}$
$.18000\times10^{9}$	$.13447\times10^{5}$	$.34000\times10^{9}$	$.33435\times10^{6}$	$.50000\times10^{9}$	$.15796\times10^{7}$
log rho = 7.60					
$.40000\times10^{8}$	$.43345\times10^{-4}$	$.20000\times10^{9}$	$.25926\times10^{5}$	$.36000\times10^{9}$	$.57571\times10^{6}$
$.60000\times10^{8}$	$.12127\times10^{0}$	$.22000\times10^{9}$	$.46135\times10^{5}$	$.38000\times10^{9}$	$.73281\times10^{6}$
$.80000\times10^{8}$	$.74286\times10^{1}$	$.24000\times10^{9}$	$.75856\times10^{5}$	$.40000\times10^{9}$	$.91665\times10^{6}$
$.10000\times10^{9}$	$.96082\times10^{2}$	$.26000\times10^{9}$	$.11725\times10^{6}$	$.42000\times10^{9}$	$.11293\times10^{7}$
$.12000\times10^{9}$	$.56359\times10^{3}$	$.28000\times10^{9}$	$.17250\times10^{6}$	$.44000\times10^{9}$	$.13728\times10^{7}$
$.14000\times10^{9}$	$.20884\times10^{4}$	$.30000\times10^{9}$	$.24380\times10^{6}$	$.46000\times10^{9}$	$.16491\times10^{7}$
$.16000\times10^{9}$	$.57807\times10^{4}$	$.32000\times10^{9}$	$.33333\times10^{6}$	$.48000\times10^{9}$	$.19603\times10^{7}$
$.18000\times10^{9}$	$.13133\times10^{5}$	$.34000\times10^{9}$	$.44326\times10^{6}$	$.50000\times10^{9}$	$.23084\times10^{7}$
log rho = 7.80					
$.40000\times10^{8}$	$.19678\times10^{-5}$	$.20000\times10^{9}$	$.24173\times10^{5}$	$.36000\times10^{9}$	$.72854\times10^{6}$
$.60000\times10^{8}$	$.19624\times10^{-1}$	$.22000\times10^{9}$	$.45895\times10^{5}$	$.38000\times10^{9}$	$.94488\times10^{6}$
$.80000\times10^{8}$	$.22635\times10^{1}$	$.24000\times10^{9}$	$.79592\times10^{5}$	$.40000\times10^{9}$	$.12017\times10^{7}$
$.10000\times10^{9}$	$.42712\times10^{2}$	$.26000\times10^{9}$	$.12862\times10^{6}$	$.42000\times10^{9}$	$.15027\times10^{7}$
$.12000\times10^{9}$	$.32175\times10^{3}$	$.28000\times10^{9}$	$.19648\times10^{6}$	$.44000\times10^{9}$	$.18513\times10^{7}$
$.14000\times10^{9}$	$.14235\times10^{4}$	$.30000\times10^{9}$	$.28677\times10^{6}$	$.46000\times10^{9}$	$.22509\times10^{7}$
$.16000\times10^{9}$	$.44951\times10^{4}$	$.32000\times10^{9}$	$.40311\times10^{6}$	$.48000\times10^{9}$	$.27051\times10^{7}$
$.18000\times10^{9}$	$.11302\times10^{5}$	$.34000\times10^{9}$	$.54916\times10^{6}$	$.50000\times10^{9}$	$.32172\times10^{7}$

TABLE 1.—*Neutrino Emissivity by Transverse Plasmons—Continued*

T	q sub T	T	q sub T	T	q sub T
log rho = 8.00					
$.40000\times10^{8}$	$.47800\times10^{-7}$	$.20000\times10^{9}$	$.19809\times10^{5}$	$.36000\times10^{9}$	$.85874\times10^{6}$
$.60000\times10^{8}$	$.20858\times10^{-2}$	$.22000\times10^{9}$	$.40604\times10^{5}$	$.38000\times10^{9}$	$.11391\times10^{7}$
$.80000\times10^{8}$	$.50212\times10^{0}$	$.24000\times10^{9}$	$.75007\times10^{5}$	$.40000\times10^{9}$	$.14781\times10^{7}$
$.10000\times10^{9}$	$.14708\times10^{2}$	$.26000\times10^{9}$	$.12779\times10^{6}$	$.42000\times10^{9}$	$.18817\times10^{7}$
$.12000\times10^{9}$	$.14833\times10^{3}$	$.28000\times10^{9}$	$.20417\times10^{6}$	$.44000\times10^{9}$	$.23558\times10^{7}$
$.14000\times10^{9}$	$.80734\times10^{3}$	$.30000\times10^{9}$	$.30965\times10^{6}$	$.46000\times10^{9}$	$.29063\times10^{7}$
$.16000\times10^{9}$	$.29750\times10^{4}$	$.32000\times10^{9}$	$.44998\times10^{6}$	$.48000\times10^{9}$	$.35392\times10^{7}$
$.18000\times10^{9}$	$.84264\times10^{4}$	$.34000\times10^{9}$	$.63103\times10^{6}$	$.50000\times10^{9}$	$.42601\times10^{7}$

APPENDIX

CALCULATION OF $x^3\mathfrak{F}(x)$ FOR SMALL x

Put $\eta = x\cosh\xi$ in equation (19); then

$$x^3\mathfrak{F}(x) = \int_x^\infty \frac{1}{e^\eta - 1}\left(1-\frac{x^2}{\eta^2}\right)^{1/2}\eta^2 d\eta. \tag{27}$$

By splitting the integral

$$x^3\mathfrak{F}(x) = \int_0^\infty \frac{1}{e^\eta-1}\eta^2 d\eta - \int_0^x \frac{\eta}{e^\eta-1}\eta d\eta - \frac{1}{2}x^2\int_x^\infty \frac{1}{e^\eta-1}d\eta + \int_x^\infty \frac{1}{e^\eta-1}\left[\left(1-\frac{x^2}{\eta^2}\right)^{1/2} - 1 + \frac{1}{2}\frac{x^2}{\eta^2}\right]\eta^2 d\eta,$$

or

$$x^3\mathfrak{F}(x) = 2\zeta(3) - a_1(x) + \tfrac{1}{2}x^2\ln(1-e^{-x}) + a_2(x). \tag{28}$$

Now

$$a_1(x) = \int_0^x\left[1-\frac{1}{2}\eta+\frac{1}{12}\eta^2-\frac{B_2}{4!}\eta^4+O(\eta^6)\right]\eta d\eta = \frac{1}{2}x^2-\frac{1}{6}x^3+\frac{1}{48}x^4-O(x^6) \tag{29}$$

for $x < 2\pi$.

An elementary calculation gives

$$\tfrac{1}{2}x^2\ln(1-e^{-x}) = \tfrac{1}{2}x^2\ln x - \tfrac{1}{4}x^3 + \tfrac{1}{48}x^4 - O(x^6). \tag{30}$$

The calculation of $a_2(x)$ is quite lengthy; we begin by splitting the integral again:

$$a_2(x) = -\frac{1}{8}x^4\int_A^\infty \frac{1}{e^\eta-1}\frac{d\eta}{\eta^2} + O(x^6) + \int_x^A \frac{1}{e^\eta-1}\left[\left(1-\frac{x^2}{\eta^2}\right)^{1/2} - 1 + \frac{1}{2}\frac{x^2}{\eta^2}\right]\eta^2 d\eta$$

$$= -\frac{1}{8}\left(\int_A^\infty \frac{1}{e^\eta-1}\frac{d\eta}{\eta^2}\right)x^4 + a_3(x) + O(x^6). \tag{31}$$

Put $n\xi = x$; then

$$(e^{\eta}-1)^{-1}\eta^2 d\eta = x^2\frac{x}{\xi}(e^{x/\xi}-1)^{-1}\frac{d\xi}{\xi^3}$$

and

$$a_3(x) = x^2\int_{x/A}^{1}\frac{x/\xi}{e^{x/\xi}-1}[(1-\xi^2)^{1/2}-1+\frac{1}{2}\xi^2]\frac{d\xi}{\xi^3} = x^2\int_{x/A}^{1}\left[1-\frac{1}{2}\frac{x}{\xi}+\frac{1}{12}\frac{x^2}{\xi^2}-\frac{B_2}{4!}\frac{x^4}{\xi^4}+\ldots+(-)^{n+1}\frac{B_n}{(2n)!}\frac{x^{2n}}{\xi^{2n}}\right.$$

$$\left.+\ldots\right]\cdot[(1-\xi^2)^{1/2}-1+\frac{1}{2}\xi^2]\frac{d\xi}{\xi^3},$$

where we require

$$\frac{x}{\xi}<2\pi \quad \text{or} \quad A<2\pi.$$

Integrating term by term, we get

$$a_3(x) = x^2 b_o(x)-\frac{1}{2}x^3 b_1(x)+\frac{1}{12}x^4 b_2(x)-\frac{B_2}{4!}x^6 b_4(x)+\ldots+(-)^{n+1}\frac{Bn}{(2n)!}x^{2n+2}b_{2n}(x)+\ldots, \tag{32}$$

where

$$b_m(x) = \int_{x/A}^{1}\left[(1-\xi^2)^{1/2}-1+\frac{1}{2}\xi^2\right]\frac{d\xi}{\xi^{3+m}}$$

These integrals are elementary; in particular

$$b_o(x) = \frac{1}{4}-\frac{1}{2}\ln 2+\frac{1}{16}(x/A)^2+O(x^4), \tag{33}$$

$$b_1(x) = -\frac{1}{6}+\frac{1}{8}\frac{x}{A}+O(x^3), \tag{34}$$

$$b_2(x) = \frac{1}{8}\ln\frac{x}{A}+\frac{1}{32}-\frac{1}{8}\ln 2+O(x^2), \tag{35}$$

$$b_4(x) = -\frac{1}{16}(x/A)^{-2}+O(\ln x); \tag{36}$$

and for $n>2$

$$b_{2n}(x) = -\frac{1}{8(2n-2)}\left(\frac{x}{A}\right)^{2-2n}+O(x^{4-2n}). \tag{37}$$

Substituting equations (33)–(37) into equation (32), we get

$$a_3(x) = \left(\frac{1}{4}-\frac{1}{2}\ln 2\right)x^2+\frac{1}{12}x^3+\frac{1}{96}x^4\ln x+\left[\frac{1}{16}\frac{1}{A^2}-\frac{1}{16}\frac{1}{A}-\frac{}{96}\ln A+\frac{1}{12}\left(\frac{1}{32}-\frac{1}{8}\ln 2\right)\right.$$

$$\left.+\frac{B_2}{4!\,16}A^2+\ldots+(-)^n\frac{B_n}{(2n)!\,8(2n-2)}A^{2n-2}+\ldots\right]x^4+O(x^6\ln x). \tag{38}$$

Substituting equations (29), (30), and (31) into equation (27), and using equation (38), we get

$$x^3\mathfrak{F}(x) = 2\zeta(3)+\frac{1}{2}x^2\ln x-\frac{1}{4}(2\ln 2+1)x^2+\frac{1}{96}x^4\ln x-\frac{1}{8}q(A)x^4+O(x^6\ln x), \tag{39}$$

re

$$q(A)=\int_A^\infty \frac{1}{e^\eta-1}\frac{d\eta}{\eta^2}-\frac{1}{2}\frac{1}{A^2}+\frac{1}{2}\frac{1}{A}+\frac{1}{12}\ln A+\frac{1}{12}\left(\ln 2-\frac{1}{4}\right)$$
$$-\frac{B_2}{4!\,2}A^2+\cdots+(-)^{n+1}\frac{B_n}{(2n)!\,(2n-2)}A^{2n-2}+\cdots, \tag{40}$$

$q(A)$ is, of course, a constant; it is, however, a bit difficult to evaluate. We will use the relations

$$\frac{1}{t}\frac{1}{e^t-1}+\frac{1}{2}\frac{1}{t}=\frac{1}{2t}\coth\frac{t}{2}=\sum_{n=1}^\infty\frac{2}{4n^2\pi^2+t^2}+\frac{1}{t^2}. \tag{41}$$

Using the first of equations (41) we find on integrating by parts

$$-\int_A^\infty \ln t\left[\frac{d}{dt}\left(\frac{1}{2t}\coth\frac{t}{2}-\frac{1}{t^2}\right)\right]dt=\int_A^\infty\frac{1}{e^t-1}\frac{dt}{t^2}-\frac{1}{2}\frac{1}{A^2}+1/(2A)+(1/12)\ln A+O(A^2\ln A)$$
$$\therefore\ q(A)=q(0)=-\int_0^\infty \ln t\left[\frac{d}{dt}\left(\frac{1}{2t}\coth\frac{t}{2}-\frac{1}{t^2}\right)\right]dt+\frac{\ln 2-\frac{1}{4}}{12}. \tag{42}$$

Substituting the second of equations (41) into equation (42), and using the absolute and uniform convergence of the series (Knopp 1963) we obtain

$$q(0)-\frac{1}{12}\left(\ln 2-\frac{1}{4}\right)=4\int_0^\infty t\ln t\left(\sum_{n=1}^\infty\frac{1}{(4n^2\pi^2+t^2)^2}\right)dt=4\sum_{n=1}^\infty B(2n\pi), \tag{43}$$

where

$$B(\xi)\equiv\int_0^\infty\frac{x\ln x}{(x^2+\xi^2)^2}dx.$$

Putting $x^2=y$, we get (deHaan 1963)

$$B(\xi)=\frac{1}{4}\int_0^\infty\frac{\ln y}{(\xi^2+\nu)^2}d\nu=\frac{1}{4}\frac{1}{\xi^2}\ln(\xi^2)=\frac{1}{2\xi^2}\ln\xi.$$

By using this in equation (43),

$$q(0)-\frac{1}{12}\left(\ln 2-\frac{1}{4}\right)=2\sum_{n=1}^\infty\frac{1}{(2n\pi)^2}\ln(2n\pi)=\frac{1}{2\pi^2}\left(\ln 2\pi\sum_{n=1}^\infty\frac{1}{n^2}+\sum_{n=1}^\infty\frac{\ln n}{n^2}\right). \tag{44}$$

But

$$\sum_{n=1}^\infty\left(\frac{1}{n^2}\right)=\zeta(2)=\frac{\pi^2}{6},$$

and

$$\sum_{n=1}^\infty\frac{\ln n}{n^2}=\left[-\frac{d}{dz}\sum_{n=1}^\infty e^{-z\ln n}\right]_{z=2}=-\zeta'(2).$$

So

$$q(0)=\frac{1}{12}\left(\ln 2-\frac{1}{4}+\ln 2\pi-\frac{\zeta'(2)}{\zeta(2)}\right). \tag{45}$$

Using this in equation (39)

$$x^3\mathfrak{F}(x)=2\zeta(3)+\frac{1}{2}x^2\ln x-\frac{1}{4}(2\ln 2+1)x^2+\frac{1}{96}x^4\ln x$$
$$-\frac{1}{96}\left(\ln 2-\frac{1}{4}+\ln 2\pi-\frac{\zeta'(2)}{\zeta(2)}\right)x^4+O(x^6\ln x). \tag{46}$$

REFERENCES

ADAMS, B., RUDERMAN, M., and WOO, C. 1963, *Phys. Rev.* **129**, 1383.

Brookhaven. 1963, Brookhaven Conference on Fundamental Aspects of Weak Interactions, September, 1963 (to be published).

CHANDRASEKHAR, S. 1957, *Stellar Structure* (New York: Dover Publications), p. 398.

CHIU, H. -Y. 1961, *Phys. Rev.* **123**, 1040.

CHIU, H. -Y., and MORRISON, P. 1960, *Phys. Rev. Letters*, **5**, 573.

CHIU, H. -Y., and STABLER, R. 1961, *Phys. Rev.*, **122**, 1317.

GANDEL'MAN, G., and PINEAU, V. 1960, *Soviet Phys.—J.E.T.P.*, **10**, 764.

GREENWOOD, R. 1941, *Ann. Math.*, **42**, 778–805.

HAAN, B. DE. 1963, *Nouvelles Tables d'Intégrales Définies* (Hafner), p. 196.

HIEU, N. VAN, and SHABALIN, E. P. 1963, *Soviet Phys.—J.E.T.P.*, **17**, 681.

IDA, M., and VAHAKA, M., unpublished.

KNOPP, K. 1963, *Theory and Application of Infinite Series* (Hafner), p. 535.

MATINYAN, S., and TSILOSANI, N. 1962, *Soviet Phys.—J.E.T.P.*, **14**, 1195.

REEVES, H. 1963, *Ap. J.*, **138**, 79.

RITUS, V. 1962, *Soviet Phys.—J.E.T.P.*, **14**, 915.

ROSENBERG, L. 1963, *Phys. Rev.*, **129**, 2786.

N65-84089

N66 37949

MAGNETOGRAVITATIONAL INSTABILITY OF ANISOTROPIC PLASMA WITH HALL EFFECT*

G. L. KALRA† AND S. P. TALWAR

1. INTRODUCTION

The problem of the magnetogravitational instability of interstellar matter is of considerable importance in connection with protostar and star formation in magnetic dust clouds. The initial analysis of gravitational instability in an infinite homogenous medium is due to Jeans (1929) who derived the expression for the maximum size of a uniform gravitating mass which is stable to small fluctuations in density. This analysis has been extended in recent years to include a variety of other effects, e.g., isotropic turbulence (Chandrasekhar, 1951), rotation (Chandrasekhar, 1954; Bel and Schatzman, 1958) and uniform magnetic field (Chandrasekhar and Fermi, 1953). It has been shown that a uniform magnetic field and rotation (or even both together) leave the Jeans' criterion for fragmentation unchanged. Furthermore, it is also found that the transport processes like finite electrical conductivity and viscosity do not modify Jeans' formula (Pacholczyk and Stadolkiewicz, 1960).

In all investigations on magnetogravitational stability reported so far (Pacholczyk, 1963), the scalar pressure approximation has been used. The interstellar medium, is, however, very dilute and the interparticle collisions are too infrequent so that in the presence of the interstellar magnetic field, the plasma pressure is likely to assume a tensorial character, with unequal magnitudes parallel and perpendicular to the magnetic field. If this anisotropy becomes large enough, it may lead to instability in an otherwise stable system. Further Ware (1961) pointed out that an important factor which has hitherto been ignored in various stability investigations is the Hall effect. He suggested, without, however, solving any definite problem that when the Hall effect is included, a growing wave instability (overstability) should replace the pure instability as predicted by the ideal magnetohydrodynamic equations. It is felt that the Hall effect assumed importance in a low density plasma such as interstellar space.

With this viewpoint, we investigate, in the present note, the problem of the magnetogravitational instability of a static, unbounded, inviscid, homogenous, gravitating plasma including finite conductivity and Hall effect. We assume the plasma to be carrying a uniform magnetic field so that the physical conditions prevalent in the medium require the gas pressure to be anisotropic. We shall be making use of the "Double-Adiabatic" hydromagnetic equations (Chew, Goldberger and Low, 1956).

2. PERTURBATION EQUATIONS AND DISPERSION RELATION

Neglecting heat flow tensor, we may write the linearized equations for the problem under considerations as,

$$\zeta_0 \frac{\partial v}{\partial t} = -\nabla \cdot \delta p + \rho_0 \nabla \delta U + \frac{L}{4\pi}(\nabla \times h) \times H. \tag{1}$$

$$\frac{\partial}{\partial t}\delta\rho = -\rho_0 \nabla \cdot v \tag{2}$$

*Goddard Space Flight Center Document X-640-64-40, January 1964.
†Delhi University, India.

$$\nabla^2\delta U = -4\pi G\delta\rho \tag{3}$$

$$\frac{\partial L}{\partial t} = \nabla\times(v\times H_o)+\eta\nabla^2 h-\frac{c}{4\pi ne}\nabla\times[(\nabla\times h)\times H_o]+\frac{c^2}{4\pi\omega\rho^2}\frac{\partial}{\partial t}\nabla^2 h \tag{4}$$

$$\nabla\cdot h = o \tag{5}$$

$$\frac{\delta p_\parallel}{p_\parallel} = 3\frac{\delta\rho}{\rho_o}-2\frac{h_z}{H_o} \tag{6}$$

$$\frac{\delta p_\parallel}{p_\parallel} = \frac{\delta\rho}{\rho_o}+\frac{h_z}{H_o} \tag{7}$$

Here $\eta = c^2/4\pi\sigma$ (σ being the electrical conductivity) and $\omega\rho = \sqrt{\frac{4\pi ne^2}{m}}$ (plasma frequency). Further V, $\delta\rho$, δU, h, $\delta p_\parallel$, $\delta p_\perp$ denote respectively the velocity vector, perturbations in density, gravitational potential, magnetic field vector, plasma pressure parallel and perpendicular to the initial magnetic field H_0. $p_\parallel$ and $p_\perp$ represent the constant components of pressure (along and perpendicular to the magnetic field respectively) of the pressure tensor $\underset{\leftrightarrow}{p}$ given by,

$$\underset{\leftrightarrow}{p} = p_\parallel n\cdot n+p_\perp(\underset{\leftrightarrow}{i}-n\cdot n) \tag{8}$$

where n is a nuit vector along the magnetic field and l is lunit dyadic.

Choosing the orientation of the coordinate axes such that $H_0 = (0,0H_0)$ and seeking solutions which are independent of y and whose dependence on x, z, and t is of the form exp. $(ik_x x+ik_z z+\omega t)$, we rewrite the equations (1) to (5) as,

$$\rho_0\omega v_x = -ik_x\delta p_\perp-ik_z(p_\parallel-p_\perp)\frac{h_x}{H_0}+ik_x\rho_0\delta U+\frac{H_0}{4\pi}(ik_x h_x-ik_x h_z) \tag{9}$$

$$\rho_0 u v_y = -ik_z(p_\Delta-p_\perp)\frac{hy}{H_0}+ik_z\frac{H_0}{4\pi}h_y \tag{10}$$

$$\rho_0\omega v_z = -ik_z\delta p_\parallel+ik_z(p_\parallel-p_\perp)\frac{h_z}{H_0}+ik_z\rho_0\delta U \tag{11}$$

$$\omega\delta\rho = -\rho_0\nabla\cdot v \tag{12}$$

$$\delta U = -\frac{4\pi G\rho_0}{k^2}\nabla\cdot v \tag{13}$$

$$\left\{\omega\left(1+\frac{c^2}{4\pi\omega\rho^2}k^2\right)+\eta k^2\right\}h_x = ik_z H_0 v_x-\frac{c}{4\pi ne}k_z^2 H_0 hy \tag{14}$$

$$\left\{\omega\left(1+\frac{c^2k^2}{4\pi\omega\rho^2}\right)+\eta k^2\right\}h_y = ik_z H_0 v_y-\frac{c}{4\pi ne}ik_z H_0(ik_z h_x-ik_x h_z) \tag{15}$$

$$\left\{\omega\left(1+\frac{c^2k^2}{4\pi\omega\rho^2}\right)+\eta k^2\right\}h_z = -H_0(\nabla\cdot v-ik_z v_z)-\frac{c}{4\pi ne}H_0 ik_x k_z h_y \tag{16}$$

$$ik_x h_x-ik_z h_z = o \tag{17}$$

Taking divergence of equation (1), and using equation (3), we get

$$\rho_0\omega\nabla\cdot v = -\nabla\cdot(\nabla\cdot\delta\underset{\leftrightarrow}{p})-4\pi G\rho_0\delta\rho+\frac{H_0}{4\pi}k^2 h_z \tag{18}$$

where

$$k^2 = k_x^2 + k_z^2.$$

The first term on the right hand side of equation (18) is written with the help of equations (6) and (7) as,

$$-\nabla\cdot(\nabla\cdot\delta\underset{\leftrightarrow}{p}) = -\frac{\mu_z}{H_0}\left[2k_z^2(2p_\parallel - p_\perp) - k_x^2 p_\perp\right] - \frac{\nabla\cdot v}{\omega}\left[k_x^2 p_\perp + 3k_z^2 p_\parallel\right] \tag{19}$$

Through a process of elimination wherein we use the equations (9), (11), (14), (15) and (17) we obtain,

$$h_z\left[A + \left(\frac{c}{4\pi ne}\right)^2 \frac{H_0^2 k_x^2 k_z^2}{\Lambda + \frac{k_z^2}{\rho_0\omega}\left(\frac{H_0^2}{4\pi} - p_\parallel + p_\perp\right)}\right] = -H_0(\nabla\cdot v - ik_z v_z) \tag{20}$$

where

$$\Lambda = \omega\left(1 + \frac{c^2k^2}{4\pi\omega\rho^2}\right) + \eta k^2 \tag{21}$$

Substituting equations (12), (19) and (20) in equation (18), we get

$$\frac{\nabla\cdot v}{\omega}\left[\rho_0\omega^2 + \frac{\omega\left\{\frac{k^2H_0^2}{4\pi} - 2k_z^2(2p_\parallel - p_\perp) + k_z^2 p_\perp\right\}}{A + \left(\frac{c}{4\pi ne}\right)^2 \frac{H_0^2 k^2 k_z^2}{A + \frac{p_z^2}{\rho_0\omega}\left(\frac{H_0^2}{4\pi} - p_\parallel + p_\perp\right)}} + (3p_\parallel k_z^2 + k_x^2 p_\perp - 4\pi G\rho_0^2)\right]$$

$$= ik_z v_z\left\{\frac{k^2H_0^2}{4\pi} - 2k_z^2(2p_\parallel - p_\tau) + k_z^2 p_\perp\right\}\left\{A + \left(\frac{c}{4\pi ne}\right)^2 \frac{H_0^2 k^2 k_z^2}{A + \frac{k_z^2}{\rho_0\omega}\left(\frac{H_0^2}{4\pi} - p_\parallel + p_\perp\right)}\right\}^{-1} \tag{22}$$

The expression for $\nabla\cdot v$ in terms of v_z is now obtained from equation (11), using equations (6), (7) and (20). We get

$$\nabla\cdot v = v_z \frac{\left[\rho_0\omega + \frac{k_z^2(3p_\parallel - p_\perp)}{A + \left(\frac{c}{4\pi ne}\right)^2 \frac{H_0^2 k^2 k_z^2}{A + (H_0^2/4\pi - p_\parallel + p_\tau)k_z^2/\rho_0\omega}}\right]}{ik_z\left[\frac{3p_\parallel - 4\pi G\rho_0^2/k^2}{\omega} \frac{(3p_\parallel - p_\perp)}{A + \left(\frac{c}{4\pi ne}\right)^2 \frac{H_0^2 k^2 k_z^2}{A + \frac{k_z^2}{\rho_0\omega}\left(\frac{H_0^2}{4\pi} - p_\parallel + p_\perp\right)}}\right]} \tag{23}$$

From equations (22) and (23) we finally obtain the dispersion relation as,

$$\left[\rho_0\omega^2 + 3p_\parallel k_z^2 + k_x^2 p_\perp - 4\pi G\rho_0^2\right]\left[\omega\rho_0\left\{A + \left(\frac{c}{4\pi ne}\right)^2 \frac{H_0^2 k^2 k_z^2}{A + \frac{k_z^2}{\rho_0\omega}\left(\frac{H_0^2}{4\pi} - p_\parallel + p_\perp\right)}\right\} + k_z^2(3p_\parallel - p_\perp)\right]$$

$$+ \left[\frac{k^2H_0^2}{4\pi} - 2k_z(2p_\parallel - p_\perp) + k_z^2 p_\perp\right]\left[\omega\rho_0 - k_z^2\left(3p_\parallel - \frac{4\pi G\rho_0^2}{k^2}\right)\right] = 0 \tag{24}$$

The equation (24) reduces without the Hall effect to the one obtained in an earlier paper (Tandon and Talwar, 1963) with $\Omega=0$.

3. DISCUSSION OF THE DISPERSION RELATION

The dispersion relation (24) is a sixth order equation in ω, the parameter to decide the question of the stability of the configuration. We shall, however, simplify our discussion by considering the following special cases.

Case (i) Transverse Propagation

In the case of transverse propagation ($k_x=k$, $k_z=0$), the dispersion relation assumes the simple form,

$$\omega^2+\frac{\omega}{A}[V^2+S_\perp{}^2]k^2+(k^2S_\perp-4\pi G\rho_0)-0 \tag{25}$$

where $V^2=H_0^2/4\pi P_0$, $S_\perp{}^2=p_\perp/P_0$ and the parameter A is given by equation (21).

The equation (25) is rewritten as,

$$\omega^3\left(1+\frac{c^2k^2}{4\pi\omega\rho^2}\right)+\eta\omega^2k^2+k^2\omega\left[V^2+S_\perp{}^2+\left(1+\frac{c^2k^2}{4\pi\omega\rho^2}\right)\left(S_\perp{}^2-\frac{4\pi G\rho_0}{R^2}\right)\right]+\eta k^2(k^2S_\perp{}^2-4\pi G\rho_0)=0 \tag{26}$$

For plasma with infinite electrical conductivity ($\eta=0$) and $\omega\rho\to\infty$, we conclude that the configuration is unstable for transverse propagation vector less than a certain critical value k^* given by

$$k_*^2=\frac{4\pi G\rho_0}{V^2+2S_\perp{}^2} \tag{27}$$

On the other hand if the conductivity were finite the equation (27) gives,

$$k_*=\sqrt{\frac{4\pi G\rho_0}{S_\perp{}^2}} \tag{28}$$

as the critical wave number below which any perturbation increases its amplitude monotonically till the linear theory breaks down. Further the equation (26) admits two complex conjugate roots with positive real part. It therefore follows that irrespective of the fact that the system is self-gravitating or not, there are present, for transverse propagation in a finitely conducting medium alone, growing wave instability (overstability).

Case (ii) Parallel Propagation

With $k_x=k$, $k_z=0$ and restricting ourselves to infinitely conducting configuration but retaining the Hall current term, we obtain the dispersion relation as,

$$[\omega^2+3k^2S_\parallel{}^2-4\pi G\rho_0]\left[\{\omega^2+k^2(V^2+S_\perp{}^2-S_\parallel{}^2)\}^2+\left(\frac{c}{4\pi ne}\right)^2H_0^2k^4\omega^2\right]=0 \tag{29}$$

This gives a gravitational mode uncoupled with the other two modes which involve the Hall effect term. The gravitational mode leads to instability provided $k<k^*$ where k^* is given by,

$$k^*=\sqrt{\frac{4\pi G\rho_0}{3S_\parallel{}^2}} \tag{30}$$

For a non-gravitating plasma, the instability could result from the second factor in equation (29). The non-gravitating configuration shall be stable if $(S_\parallel{}^2-S_\perp{}^2-V^2)\leqslant 0$. On the other hand if

$(S_\parallel^2 - S_\perp^2 - V^2) > 0$, the configuration without Hall term is unstable for all modes of disturbance (Lüst, 1960). In our present context there is no instability for wave numbers greater than a certain critical value k_c given by,

$$k_c^2 = \frac{2(S_\parallel^2 - S_\perp^2 - V^2}{(c/4\pi ne)^2 H_0^2} \tag{31}$$

We, therefore, conclude that the mode of disturbance (in range k_c to ∞) which leads to instability in a configuration devoid of Hall effect, is stabilized by the Hall effect, athough the Hall effect is unable to suppress instability for all wavelengths.

REFERENCES

Bel, N. and Schatzmann, E., (1958), Rev. Mod. Phys. **30**, 1015.

Chandrasekhar, S. (1951), Proc. Roy. Soc. **A, 210**, 26.

Chandrasekhar, S. and Fermi, E. (1953) Ap. J. **118**, 116.

Chandrasekhar, S. (1954) Ap. J. **119**, 7.

Chew, G. F., Goldberger, M. and Low, F. (1956) Proc. Roy. Soc. A, **236**, 112.

Jeans, J. H. (1929) Astronomy and Cosmogony 313 (Camb. Univ. Press).

Lüst, R. (1959) Riso Report 18, Danish A.E.C.

Pacholczyk, A. G. and Stodolkiewicz (1960), Acta. Astronomica **10**, 1.

Pacholczyk, A. G. (1963) Soviet Astronomy **6**, 741.

Tandon, J. N. and Talwar, S. P. (1963) Nuclear Fusion **3**, No. 2, 75

THE DETECTION OF SOLAR NEUTRINOS*

HUBERT REEVES†

An old dream of nuclear physicists has now entered the realm of possibility. Raymond Davis of Brookhaven National Laboratory has the plans for an apparatus capable of detecting the flux of neutrinos from the sun. The Brookhaven Solar Neutrino Observatory, as this equipment is called, will be built in future years and buried thousands of feet underground. Then, for the first time a direct "view" into the center of the sun will be possible.

Neutrino astronomy is quite new, but physicists have in recent years become increasingly involved with these elementary particles. The story of neutrinos goes back to the early days of nuclear physics (about 1935), when a number of disturbing facts came to light.

For instance, the beta disintegration (emission of a negative electron) of certain nuclei showed no sign of conserving energy and momentum. An important case (discovered later) was the observed decay of a neutron into a proton and an electron. Analysis of this process showed that the sum of the energies of the proton and electron (including the rest mass) was *smaller* than the total energy of the neutron. Further, this sum was not constant but varied according to the particular neutron observed. It was also discovered that the two emerging particles did not move in opposite directions (as would be expected from a neutron at rest), but were directed in many possible ways, as in Fig. 1.

To overcome these difficulties, W. Pauli postulated the existence of a third, unobserved particle emitted during beta decay. This particle—neutral and massless, like a photon of light—carried off the balance of the energy and momentum; it was later named *neutrino*. According to the principles of quantum mechanics, it could be detected by a reaction inverse to the one responsible for its birth. For example, a neutron could absorb a neutrino and give rise to a proton and an electron. But calculations showed that the probability of capture was extremely small, far below the detection power of existing equipment. This explains why the presence of neutrinos had not been felt before in physics.

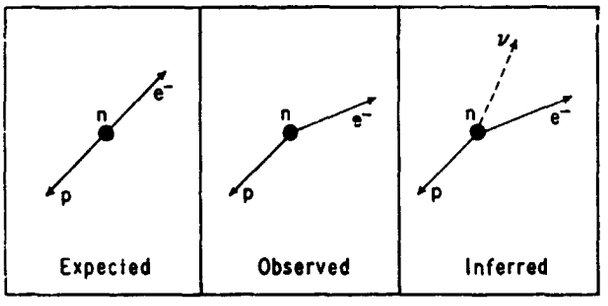


FIGURE 1.—The dynamics of neutron decay, in which the lack of symmetry gave rise to the prediction that an unknown particle was present. In the third section this particle, a neutrino, is symbolized by the Greek letter nu (ν).

Later it was shown that this particle existed in two brands: neutrinos and antineutrinos. The only difference between them that is important to us concerns their origins. Neutrino births accompany the beta decay of protons (in complex nuclei for energy considerations), and antineutrinos are born when neutrons undergo beta decay. Inversely, a neutrino can be captured only by a neutron, an antineutrino only by a proton.

As physicists developed beta-decay theory, close agreement between its predictions and the experiments gave abundant indirect evidence for the existence of neutrinos. It soon became clear that they had reality and were not mere artifices of thought; they became accepted in physics on the same footing as other elementary particles.

Experimental detection of neutrinos had to wait many years. But World War II gave tremendous

*Published in *Sky and Telescope*, 27(5), May 1964.

†Goddard Space Flight Center and University of Montreal.

impetus to nuclear physics technology, and created huge antineutrino sources in the form of nuclear reactors. Their production of neutron-rich isotopes roused hopes that finally materialized in 1958, when F. Reines and C. L. Cowan, Jr., succeeded in detecting antineutrinos coming from the Savannah River reactor. Although their detection apparatus was exposed to a flux of about 10^{13} antineutrinos per square centimeter per second, the counted number was, at best, a few per hour. The antineutrino capture probability was thus very small indeed, but it agreed with the prediction made in the early days of neutrino theory. Elusive as the neutrino is, belief in its existence has an excellent foundation both in experiment and theory.

Recently it has been shown that two other types of neutrinos exist, related to the decay of muon particles.

NEUTRINOS AND ASTRONOMY

Nuclear physics has brought an answer to a very old problem: Where does the sun find the energy that it has been radiating for so many eons? Around 1938, Hans Bethe showed convincingly that the centers of the sun and other stars are seats of intense thermonuclear reactions. Both the liberation of energy and the formation of new isotopes occur in these stellar nuclear furnaces. Analysis of the network of reactions taking place showed that interspersed among some purely nuclear reactions were a certain number of beta-decay processes—hence, neutrino-emitting reactions.

But even a star is not thick enough to absorb traveling neutrinos appreciably. Proper shielding would require a few billion stars lined up one after another! The sun is essentially transparent to neutrinos, which escape, carrying away their energy. Any proper study of stellar interiors has to take this escape into account: the solar furnace must be hotter to make up for the loss.

In the sun this effect is only a few percent. But its immediate consequence is that an appreciable neutrino flux reaches the earth, about 6.5×10^{10} neutrinos per square centimeter per second. The flux of antineutrinos from the sun is negligibly small, since most solar nuclear reactions are of the proton-capture type, forming proton-rich neutrino-emitting isotopes.

SOLAR NEUTRINOS

Basically, the source of the sun's energy is the conversion of four hydrogen atoms into one helium atom. Under stellar conditions, the transformation is accomplished through four main channels, the so-called PP branches illustrated in Fig. 2. Notice that each of these branches includes neutrino emission.

In each process, the numbers of neutrinos having different energies follow a specific pattern. As illustrated in Fig. 3 for neutrinos from hydrogen, most emitted particles belong to a broad hump, but there is also a sizable peak of particles all with about the same energy—some one million electron volts greater than the maximum energy in the hump. For certain nuclei, such as beryllium-7, *two* peaks of monoenergetic neutrinos are emitted, since the beta decay sometimes goes to an excited state of lithium-7.

The PP Chain

$$H^1 + H^1 \rightarrow D^2 + \beta^+ + \nu$$
$$D^2 + H^1 \rightarrow He^3 + \gamma \qquad \text{PP I}$$
$$He^3 + He^3 \rightarrow He^4 + 2\,H^1$$

or

$$He^3 + He^4 \rightarrow Be^7 + \gamma$$
$$Be^7 + e^- \rightarrow Li^7 + \nu + \gamma \qquad \text{PP II}$$
$$Li^7 + H^1 \rightarrow 2\,He^4$$

or

$$Be^7 + H^1 \rightarrow B^8 + \gamma$$
$$B^8 \rightarrow Be^8 + \beta^+ + \nu \qquad \text{PP III}$$
$$Be^8 \rightarrow 2\,He^4$$

The CNO Cycle

$$C^{12} + H^1 \rightarrow N^{13} + \gamma$$
$$N^{13} \rightarrow C^{13} + \beta^+ + \nu$$
$$C^{13} + H^1 \rightarrow N^{14} + \nu \qquad \text{PP IV}$$
$$N^{14} + H^1 \rightarrow O^{15} + \gamma$$
$$O^{15} \rightarrow N^{15} + \beta^+ + \nu$$
$$N^{15} + H^1 \rightarrow C^{12} + He^4$$

FIGURE 2.—Nuclear reactions responsible for energy generation in the sun. In some of the proton-proton series, later reactions depend on earlier ones. Greek letter symbols are $\beta+$ (positron), γ (gamma ray), and ν (neutrino).

Before we can evaluate the neutrino energy spectrum from the sun, and compute the mean energy, we must specify the conditions in the solar interior. Indeed, a first step is to find the contribution of each PP branch to the sun's total energy output.

Using the most recent data about nuclear energy generation rates and stellar opacities, we have computed a solar model whose evolution during the past 4½ billion years is described in Table I. For completeness, we include the changing luminosity, adius, surface and central temperatures, and ther

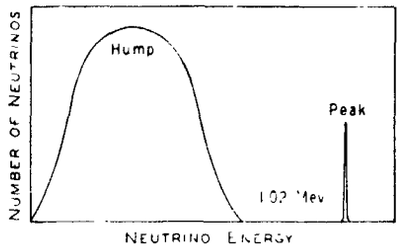


FIGURE 3.—The energy spectrum of neutrinos coming from hydrogen nuclei. The hump is the result of processes involving creation of an electron during the decay, while the peak indicates processes involving the capture of an orbital or a free electron by the decaying nucleus.

TABLE I.—*Evolution of the Sun*

Billions of years ago	4.5	3.6	2.7	1.8	0.9	Now
Luminosity	0.73	0.77	0.83	0.87	0.92	1.00
Radius	0.95	0.95	0.97	0.98	0.99	1.00
Surface temperature (degrees Kelvin)	5,400	5,500	5,500	5,600	5,600	5,800
Central temperature (millions of degrees)	14	15	15	15	16	16
Central density	89	98	110	120	140	170
Percent hydrogen	68	61	53	46	38	31
% PP I	86	82	79	71	64	56
% PP II	13	17	22	28	34	40
% PP III	0.004	0.008	0.01	0.02	0.03	0.05
% PP IV	0.4	0.06	0.9	1.3	2.0	3.2

Luminosity and radius are in terms of the present value for the sun, central density in grams per cubic centimeter. The percentage of hydrogen is in terms of the sun's mass. The last four lines give the percentage of the total energy provided by the four different reactions.

central density and central hydrogen abundance. But note particularly the last four lines of the table, which give the fractional contribution of each PP branch. We see that more than half the sun's energy has come from PP I during the past 4½ billion years. The PP II branch has been slowly increasing in importance until it now provides 40 percent of the sun's energy. On the other hand, the remaining two branches appear to be relatively unimportant.

Our computations have included a detailed calculation of how temperature, pressure, and other properties change from the center out to the surface of the present sun. This permitted a further computation that gave the relative contributions of the four PP branches at different levels in the solar interior. Finally, we could predict the neutrino energy spectrum of the sun, illustrated in Fig. 4.

From this chart we see the relative importance of different neutrino sources over the whole energy range. For instance, the hump marked H represents the reaction where two protons unite to form a deuteron, positron, and neutrino (see the first line of PPI in Fig. 2). At an energy of 0.3 million electron volts, it rises to a flux of about three neutrinos per square centimeter each second per million electron volts. Almost the same flux comes from the peak H′, for the combination of two protons and an electron to yield a deuteron and neutrino (not included in Fig. 2).

The most prominent features are the two peaks of beryllium-7, which rise to 47 and 344 but are very narrow. Although the neutrinos from boron-8 are very rare, they turn out to be the most important, because their high energies mean very large capture cross sections and hence easier detection. Since this chart has been normalized to

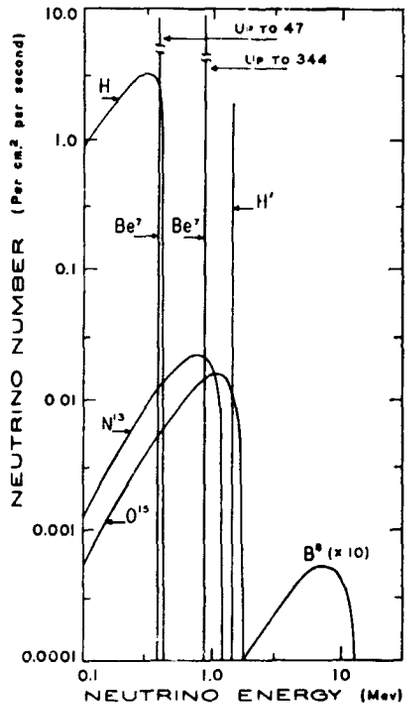


FIGURE 4.—The neutrino energy spectrum of the sun, normalized to one particle in the entire flux. The horizontal scale is in millions of electron volts. Capture of free electrons by beryllium-7 nuclei and by protons (H') results in very narrow, high peaks. Their heights depend on density and temperature in the sun's interior.

one neutrino over the whole energy range, to obtain the flux at the earth multiply ordinate values by 6.5×10^{10}.

DETECTING SOLAR NEUTRINOS

A neutrino, as we have seen, can be captured by a neutron (either free or in a complex nucleus). Since free neutrons are not readily available, a detection system must utilize some suitable isotope. A very good choice is chlorine-37, which can react with a neutrino to yield argon-37 and an electron.

For capture by chlorine-37 to take place, an incident neutrino must have an energy of at least 0.814 million electron volts. And above this limit, high-energy neutrinos have a much better chance of inducing the reaction, since the collision cross section is proportional to the square of the energy of the out-coming electron.

Looking back at Fig. 4 with these facts in mind, we see at once that most of the solar neutrinos are not capable of inducing this reaction. Indeed, we lose all the neutrinos from the hump of the proton-proton reaction, and all the neutrinos from the first beryllium-7 peak.

In fact, a recent calculation by J. N. Bahcall shows that of every 100 solar neutrinos captured by a chlorine-37 detector, about eight will come from the higher beryllium-7 peak and about 92 from the boron hump. Negligible contributions come from the other neutrino-producing processes. The total probability of a capture per target nucleus of chlorine-37 will be as low as 4×10^{-35} per second.

This means that any feasible detection apparatus must be enormous! Indeed, multiplying this rate by the number of chlorine atoms will give the probability of an event per second in the detector. Since practical considerations require a counting rate not much smaller than one a day, the detector must contain about 10^{30} chlorine atoms. At the present time, Dr. Davis considers the experiment perfectly feasible.

THE DETECTING EQUIPMENT AND ITS OPERATION

In his first attempt, Dr. Davis has worked with a 1,000-gallon and later a 3,000-gallon tank of carbon tetrachloride (CCl_4). Now he uses perchlorethylene (C_2Cl_4) as it is less toxic and has a lower vapor pressure. The estimate noted in this article has required him to consider a detecting tank holding 100,000 gallons of material (2.2×10^{30} atoms of chlorine-37). This amount of liquid would fill 40 tank trucks of the usual size one sees on the highway!

Before use, the filled tank has to be cleansed of all argon atoms already there from atmospheric contamination or other sources. The procedure is to sweep helium bubbles through the liquid. Then the tank is left to rest, exposed as always to the effect of solar neutrinos. Gradually, at a rate of about one per day, chlorine-37 atoms are transformed into argon-37 by the reactions. However, the argon-37 is naturally radioactive, and through beta decay changes back to chlorine-37 in about 35 days, so it does not pay to wait too long. In fact, after a few months, the concentration of argon-37 saturates and remains at the same level indefinitely.

Therefore, after about 100 days the tank will again be swept clear of argon. Then the helium

will be removed from the argon by another chemical process, and the argon atoms will be gathered in small containers. Among various other isotopes, the argon contains argon-37 produced by neutrino reactions. Through their beta disintegration, these atoms will produce a signal measurable by standard nuclear physics techniques, thus making their presence known.

As can well be imagined, the major problem is background. A large number of other processes may give exactly the same final result, and the experimentalist must make absolutely sure that the counts he registers in his counter are really from argon-37 generated by solar neutrinos. The tank is close to the impossible, and requires the utmost skill of the experimental nuclear chemist. For instance, to avoid competition by cosmic rays in producing this reaction, the tank and all the chemical equipment will have to be buried far beneath the surface of the earth. Present indications are that the tank is to be located in a mine at a depth of several thousand feet, in a room specially excavated to hold all the apparatus. So far, the whole experiment is on paper, except for the detection apparatus, which is partly set up at Brookhaven, New York. We hope that construction will start soon, and that there will be some results within a few years.

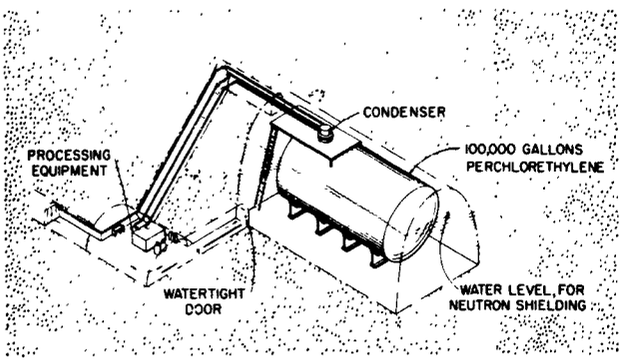


FIGURE 5.—A plan for the Brookhaven observatory that may detect neutrino emission from the sun. In this enormous tank, roughly 20 feet in diameter and 40 feet long, the nuclear chemist will have to fish for perhaps 100 atoms of radioactive argon–37 generated by chlorine-neutrino interactions. Illustration courtesy Raymond Davis.

WHAT CAN WE LEARN FROM SOLAR NEUTRINOS?

If the experiment is successfully carried through, one possibility is that we will find just what has been predicted. This in itself would be a major feat. All that we know directly about the sun has come from observations of its surface layers. We have been able to explore the interior only by applying point by point the differential equations of theoretical physics. Thus, we have only highly indirect evidence of conditions deep inside the sun, from which the nuclear reactions taking place there must be analyzed. The weight of indirect evidence should not be underestimated, but it would be very desirable to obtain *direct* information on the center of the sun, and to verify that our understanding of the processes there is correct. In particular, the rate of neutrino emission would provide a sensitive test of the adopted central temperature. There is also a rather remote possibility that some important physical factors, as yet unknown, are operating inside the sun. The history of science teaches us about many such upsets.

Is it possible that the neutrino flux from the sun varies with time, for instance with the solar cycle? Most solar experts consulted believe this unlikely. But we really cannot tell beforehand what will come from this and some other suggested problems for the solar neutrino observatory. In any event, it is a new channel of information for solar physics, and should be thoroughly explored. We are all waiting for the first clicks in Dr. Davis' immense tank.

ACCRETION OF INTERSTELLAR MATTER BY MASSIVE OBJECTS*

E. E. SALPETER†

N66 37950

Observations of quasi-stellar radio sources have indicated the existence in the Universe of extremely massive objects of relatively small size. The present note discusses the possible further growth in mass of a relatively massive object, by means of accretion of interstellar gas onto it, and the accompanying energy release. Although there is no evidence for (and possibly some evidence *against*) quasi-stellar radio sources occurring inside ordinary galaxies, for the sake of concreteness we consider the fate of an object of mass $M>10^6$ (masses in solar units throughout) in an ordinary spiral galaxy somewhat like ours.

We first re-examine the hypothetical problem of an object of mass M moving with velocity U (in km/sec) relative to a completely uniform gas medium of density n (expressed as H-atoms per cm^3) and thermal speed $U_{\rm th}$. We define (Hoyle and Lyttleton 1939) a characteristic length s_0 and express the rate of accretion in terms of a dimensionless parameter α to be determined,

$$s_0=GM/U^2=(M/U^2)\times 4.3\times 10^{-3}\ \text{pc},$$
$$dM/dt=2\pi\alpha s_0^2\, nU=\alpha M/t_0, \qquad (1)$$
$$t_0=(U^3/Mn)\times 3.3\times 10^{11}\ \text{years}.$$

We assume that the size of the object, as well as the collision mean free path in the gas, is very much smaller than s_0. For cases with large Mach number, $U\gg U_{\rm th}$, a standing (relative to the object) shock front develops behind the object, with a large pressure increase across the front. The shape of the shock front is conical about the accretion axis, as shown schematically in Figure 1, but the cone angles depend on the specific heat ratio γ of (and on any irreversible energy loss from) the gas. If $\gamma-1\ll 1$ (almost isothermal compression) then the density increases greatly across the shock front, the cone angles are everywhere small, and pressure gradients parallel to the accretion axis are small compared with gravitational forces. For such cases Bondi and Hoyle (1944) found that $1<\alpha<2$, no matter what the degree of mixing in the shocked region is. For the special case of purely laminar flow, the results are as though molecules moved without collisions except for giving up their transverse momentum when they reach the accretion axis. In this case all the gas with impact parameters less than $\sqrt{(2)}\ s_0$ is accreted and $\alpha=2$.

If $\gamma-1$ is appreciable, then the cone angles of the shock front are not very small, and the gas molecules lose less of their transverse momentum when crossing the shock front and also regain some longitudinal momentum during the re-expansion in the shocked region (see Ruderman and Spiegel [1964]). These effects result in α decreasing with

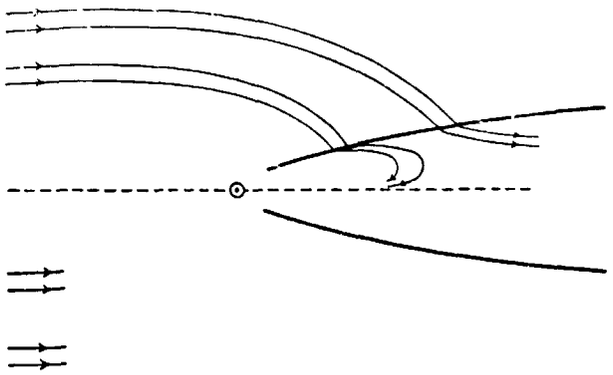


FIGURE 1.—A schematic view of the flow. The dotted line is the axis of accretion. The heavy line is the standing shock front; the thin lines are (laminar) flow lines.

*Published in *The Astrophysical Journal*, 140(2): 796–800. This work was supported in part by contract AFOSR-321-63 with Cornell University.

†Goddard Space Flight Center and Cornell University, Ithaca, New York.

increasing γ. Explicit calculations for $\alpha(\gamma)$ have not yet been carried out but plausibility arguments lead to the conjecture that α is non-zero for $\gamma<\frac{5}{3}$ and not much less than unity for $\gamma<\frac{4}{3}$. The α is also reduced slightly if U is only slightly larger than U_{th} (see Bondi [1952] for cases with $U<nU_{th}$).

McCrea (1953) has discussed the slowing down of an object during its passage through a gas cloud with velocity $U\ll U_{th}$. This retardation, analogous to that experienced by a charge moving through a plasma, can be expressed as

$$dU/dt=-\beta U/t_0, \qquad \beta\approx 1+\ln(b_{max}/s_0), \qquad (2)$$

where t_0 and s_0 are defined in equation (1). For a uniform medium of infinite extent the limiting impact parameter would be $b_{max}\sim U/\sqrt{(4\pi\rho G)}$ as in the plasma problem. In practice the limit in the gravitational problem is the extent of the uniform medium (e.g., thickness of the galactic disk). Consider a massive object in the "halo population" with $U\sim 100$ or 200 (relative to the galactic disk) which spends a fraction of about $(10/U)$ of its time in the disk. According to equations (1) and (2) the time required for slowing down is $(U10)\ t_0/4\beta$ (if mass increase is neglected and β assumed constant). Unlike accretion of mass, the retardation involves only distant gravitational encounters and does not require collisions by the "gas molecules," so that the stars in the disk also contribute to the retardation. We assume an average gas density $n\sim 1$ in the disk and $n_{tot}\sim 10n$ (including stars) and $\beta\approx 5$. A halo object with initial velocity U (relative to the disk) will slow down in a time less than 10^{10} years, roughly the age of our Galaxy, if its mass M exceeds a critical value of

$$M_{cr,\,sl.}\approx (U/100)^4\times 1.5\times 10^6. \qquad (3)$$

For a highly supersonic object retardation is more rapid than accretion ($4\beta n_{tot}\gg\alpha n$), an we need to consider accretion only after the object has slowed down to a velocity U not much bigger than U_{th}. For our large mass M the characteristic impact parameter s_0 exceeds the "spacing between interstellar gas clouds" (or "turbulence scale length") d, and we have to include the random cloud velocities in the thermal velocity. We assume $U_{th}\sim 10$ or 15, $U\sim 25$, $n\sim 1$. We thus cannot make use of the much lower thermal velocity and higher density available in individual H I region, as attempted in previous work for smaller masses. On the other hand, the difficulties encountered previously (Schatzman 1955) due to the heating and expansion of H I regions are absent wher $s_0>d$. The energies released by the compression at distances near s_0 can easily be radiated away at typical H II temperatures, so the compression is almost isothermal ($\gamma-1$ small). The presence of turbulent magnetic fields in the gas makes the compression more adiabatic and raises the effective γ toward $\frac{4}{3}$. The accretion time (t_0/α) is less than 10^{10} years if the mass M exceeds

$$M_{cr,\,acc.}\approx (U/25)^3\ (0.25/\alpha n)\times 2\times 10^6, \qquad (4)$$

and α probably lies between 0.1 and 1. The characteristic impact parameter is $s=\sqrt{(\alpha)}s_0$ (with $s\approx 7$ pc for $M=M_{cr,\,acc.}$) and the size of the object must be less than s. The mass of typical halo globular clusters in our own Galaxy is too small for retardation and accretion to become catastrophic by one order of magnitude or so (and their diameter would be sufficiently small).

So far we have only discussed the gravitational capture of gas atoms from the stream at radial distances r from the central object with $r\sim s$, where the energy released per H atom ($\sim GMH/S\sim\frac{1}{2}HU^2$) is only a few eV. If the original gas stream possessed no angular momentum about the accretion axis at all (and $\gamma<\frac{5}{3}$) the captured gas would immediately collapse toward the center, somewhat as in star formation but with a continuously increasing mass in the condensation. The evolution of massive stars takes less than 10^7 years and no zero-temperature spherical equilibrium models exist exceeding two or three solar masses. After a relatively short time, then, most of the accreted matter would have collapsed to the "Schwarzschild singularity" ($r=2GMc^2$). Although the gravitational redshift for radiation from "collapsed matter" is large, note that its "active gravitational" (as well as "passive inertial") mass is essentially undiminished. If the original star cluster (or other non-condensed object) exceeded both critical masses in equations (3) and (4), then the accreted collapsed matter eventually dominates as the cause for further accretion.

Theories for quasi-stellar (or other condensed) objects which invoke the gravitational collapse of an *isolated* mass or cluster usually encounter some difficulties in shedding original angular momentum. In our problem the incident "turbulent" gas stream will also carry some small angular momentum at any instant of time, and a captured gas "turbule" will not be able to penetrate to very small radial distances by itself. For our continuous gas stream, however, the angular momentum is only a statistical quantity, provided that the accretion impact parameter s is much smaller than characteristic dimensions of the galaxy. The angular momentum then changes sign over times of the order of s/U and in about that time a gas "turbule" can shed its angular momentum by collisions with other turbules and spiral in toward the Schwarzschild radius. In this turbulent manner gravitational energy is continually converted into bulk kinetic energy, compressional heating, and compression of turbulent magnetic fields. Densities $n(r)$ at radial distance $r \ll s$ must lie in the range between $(s/r)^{1.5}n$ and $(s/r)^3 n$ and radiating away energy (bremsstrahlung or synchrotron radiation via energetic electrons) at these high densities presents no great problems.

Each accreted gas atom eventually approaches the Schwarzschild radius $r=2r_0$ (with $r_0=GM/c^2 = (M/10^8)\times 4.7\ 10^{-6}$ pc), and we have to estimate the fraction $f_{\rm tot}$ of its rest-mass energy which is radiated away altogether. As a simplified model for the spiraling in via many collisions, consider a particle in circular orbits with diminishing radii caused by a slow drain of angular momentum and energy. Due to general relativistic effects (see the Appendix) the orbits become unstable to *spontaneous* spiraling in at $r=6r_0$. By this time a fraction $f=0.057$ of the rest-mass energy has escaped to infinity. If, at $r=6r_0$, the particle could be brought to rest (by the unlikely collision with a matched particle of opposite momentum) an additional fraction $\Delta f=0.126$ could in principle escape. After free fall to $r=3r_0$ the additional fraction is 0.22, and so on. However, the chances of arresting the free fall with appreciable efficiency seem small. In addition, even in the absence of any optical absorption, 50 per cent of all radiation (or highly relativistic particles) emitted isotropically at $r=3r_0$ would be prevented from escaping by general relativistic effects (but only 14.5 per cent at $r=6r_0$). We estimate $f_{\rm tot}$ (net energy escaping to infinity in all forms) to lie in the range of 0.05–0.20. Note that the gravitational mass of the collapsed source is increased by a fraction $1-f>0.8$ of the accreted mass, and we omit a corresponding correction factor which should be applied to equation (1).

The optical luminosity $L_{\rm opt}$ can be written in the form

$$L_{\rm opt}=\left(\frac{25}{U}\right)^3 n\frac{\alpha}{0.25}\,\frac{f_{\rm opt}}{0.03}\left(\frac{M}{10^7}\right)^2\times 2\times 10^9 L_\odot \quad (5)$$

If α (and n) remained constant, then M (and L) would increase to infinity after a finite time (t_0/α), according to equation (1). However, the effects of radiation pressure decrease the effective value of the gravitational constant appreciably as $(L_{\rm opt}/M)$ approaches a limiting value which depends on the average opacity coefficient. $(L_{\rm opt}/M)_{\rm lim}$ can be as low as 100 (in solar units) in H II-regions in our Galaxy, due to the presence of dust grains (van de Hulst 1955) and as large as 3×10^4 in pure ionized hydrogen if only Thomson scattering is important. Once the mass M has grown sufficiently for equation (5) to give a value of $(L_{\rm opt}/M)$ close to this limit, radiation pressure automatically lowers α in such a way as to keep $L_{\rm opt}/M$ almost constant. Subsequently, M (and hence $L_{\rm opt}$) increases with time only exponentially as $e^{t/\tau}$, where

$$\tau=(f_{\rm opt}/0.03)(10^3\, M/L_{\rm opt})\times 4\times 10^8 \text{ years}, \quad (6)$$

and the characteristic accretion impact parameter $s=\sqrt{(\alpha)}s_0$ increases only as $\sqrt{M}$. For instance, under typical conditions for our galactic disk with $(L_{\rm opt}/M)_{\rm lim}\sim 200$, the exponential growth would start at $M\sim 10^7$ and s would reach about 350 pc when $M\sim 10^9$ and $L_{\rm opt}\sim 2\times 10^{11}$.

To summarize the situation for a highly evolved spiral galaxy like ours (with only a few per cent of the mass remaining in the form of gas): Objects of the order of $10^6\ M_\odot$, more than typical globular cluster masses, are required to initiate a catastrophic accretion process. Unlike theories involving the gravitational collapse of *isolated* objects, the time scale of our process never becomes very short due to the self-limiting effects of radiation pressure. In our Galaxy, in fact, dust grains would keep the time scale well above 10^8 years,

which is much longer than required for quasi-stellar objects, and the luminosity too low ($<10^{12}\ L_\odot$). The situation is likely to be more favorable in systems less evolved than our Galaxy where (*a*) a larger fraction of the mass is still in the form of gas and (*b*) the relative abundance of heavier elements, and hence of dust grains, is lower. This leads to a shorter time scale and a larger limiting value for the luminosity-mass ratio for the accreting condensation.

I am indebted to Drs. R. P. Feynman, M. Ruderman, M. Schwarzschild, E. Spiegel, L. Spitzer, and L. Woltjer for helpful criticism and suggestions. In fact, they have contributed most of the positive ideas in this note without being responsible for any of the unwarranted conjectures. I am also grateful to the National Academy of Sciences for a senior postdoctoral fellowship.

APPENDIX

We use the usual Schwarzschild metric outside of a mass M (and put $c=1$),

$$ds^2\,(1-2u)dt^2-(1-2u)^{-1}dr^2-r^2d\theta^2, \qquad \text{(A1)}$$

where $u=GM/_r$, and we consider the motion of a single test particle of unit mass in this field. As discussed by Feynman in unpublished lecture notes, the following quantities are constants of the motion

$$k\equiv(1-2u)t^{(1)}, \qquad l\equiv r^2\theta^{(1)}, \qquad \text{(A2)}$$

where a superscript (1) denotes differentiation with respect to proper time. The terms l and k are generalizations of angular momentum and energy, respectively, and are also conserved in any *local* collision processes (any number of particles at the same u). The radial distance r satisfies the equation

$$[r^{(1)}]^2=k^2-1+2u-(1-2u)l^2u^2. \qquad \text{(A3)}$$

The circular orbit with the smallest radius which is stable is then at $u=\frac{1}{6}$ with $l=2\sqrt{3}$ and $k=\sqrt{\frac{8}{9}}$. A photon (or an extremely relativistic particle) can have an unstable circular orbit at $u=\frac{1}{3}$. A particle at rest at distance $r(r[1]=l=0)$ has $k=\sqrt{(1-2u)}$. Consider a particle which started at rest far away ($k=1$, $u=0$) and ends up at some finite distance from the source with some value of k. If the energy released during the change all escapes "to infinity" then the escaped energy (*after* all redshift corrections) is simply $(1-k)c^2$.

REFERENCES

Bondi, H. 1952, *M.N.*, **112**, 195.
Bondi, H. and Hoyle, F. 1944, *M.N.*, **104**, 273.
Hoyle, F., and Lyttleton, R. A. 1939, *Proc. Cambridge Phil. Soc.*, **35**, 405.
Huist, H. C. van de. 1955, *Mém. Soc. R. Sci. Liège*, **15**, 393.
McCrea, W. H. 1953, *M.N.*, **113**, 162.
Ruderman, M., and Spiegel, E. 1964, unpublished.
Schatzman, E. 1955, *Gas Dynamics of Cosmic Clouds* (I.A.U. Symposium No. 2 [Amsterdam: North-Holland Publishing Co.]), p. 193.

C_2 SWAN BANDS IN COMETS*

N66 37951

RALPH E. STOCKHAUSEN AND DONALD E. OSTERBROCK†

The relative populations of the vibrational levels of the X $^3\Pi$ and A $^3\Pi$ electronic levels are calculated assuming the fluorescence mechanism. Pure vibrational transitions are taken into account by making a rough estimate of the magnetic-dipole transition probabilities. Both the approximate method, Rosseland's theory of cycles applied to a three-level molecule, and the accurate solution of the equations of statistical equilibrium for a 10-level molecule, give similar results. The excitation temperatures derived from these relative populations agree satisfactorily with the observations of the Swan bands by McKellar and Climenhaga (1953) for various sun-comet distances. Finally, an estimate is made of the infrared radiation, due to pure vibrational transitions, expected from a bright comet. The expected amount of radiation is small and will be difficult to detect.

It is well known that the C_2 Swan bands observed in comets are excited by the mechanism of resonance fluorescence (Stawikowski and Swings 1960). Observations give excitation temperatures (defined by a Boltzmann distribution of the relative populations of vibrational levels of the X $^3\Pi$ ground electronic term) in the range $T_{ex} \approx 2000$–$3000°$ at sun-comet distances of 0.48 to 1.40 a.u. (McKellar and Climenhaga 1953). Other cometary molecules such as CN, CH, NH and OH are observed to have much lower excitation temperatures, and this is understood qualitatively to be a consequence of spontaneous downward radiative transitions (vibrational and pure rotational) within the ground electronic terms of these molecules (Wurm 1936). In fact, quantitative calculations using estimated transition probabilities approximately match the observed excitation temperature (Hunaerts 1953, 1957). However C_2 has no permanent electric dipole moment, and the downward radiative transitions within the ground term are therefore forbidden, so that large populations occur in highly excited vibrational levels (Wurm 1936).

Houziaux (1960) has calculated the expected populations of successive vibrational levels in C_2, taking account only of resonance fluorescence in the Swan bands, and assuming all transition probabilities within the ground X $^3\Pi$ term to be identically zero. However the calculations do not agree very well with observation in that the calculated excitation temperatures are higher than the observed excitation temperatures and furthermore vary considerably from one level to the next.

In fact, however, magnetic dipole transitions of the type $\Delta\Sigma = \pm 1$ can occur within the ground X $^3\Pi$ term of C_2, since all the selection rules for this type of transition are fulfilled (Van Vleck 1934). We can see the approximate effect of these transitions by applying Rosseland's theory of cycles to a simplified 3 level molecule, taking the lowest levels 1 and 2 as successive vibrational levels of X $^3\Pi$, and the high level 3 as any vibrational level of the upper A $^3\Pi$ term that is connected by strong radiative transitions with both levels 1 and 2. Then the population ratio is (Ambartsumyan 1958).

$$\frac{n_2}{n_1} = We^{-h\nu 21/kT} \frac{A_{31}e^{-h\nu_{32}/kT} + \frac{A_{21}}{A_{32}}(A_{31}+A_{32})}{WA_{31}e^{-h\nu_{32}/kT} + \frac{A_{21}}{A_{32}}(A_{31}+A_{32})} \equiv e^{-h\nu 21/kT_{ex}} \quad (1)$$

where W is the dilution factor, T is the radiation temperature and T_{ex} is the observed temperature as defined above.

*Published as *Goddard Space Flight Center document X-614-64-252*, August 1964.

†Washburn Observatory, Madison, Wisconsin.

In a crude first approximation $A_{31} \approx A_{32} \gg A_{21}$ so if $WA_{31} \exp(-h\nu_{32}/kT) \gg A_{21}$, that is if the transition 2→1 is strongly forbidden, then $T_{ex} \approx T \approx 5730°$. Alternatively, if $A_{31} \exp(-h\nu_{32}/kT) \gg A_{21}$, that is 2→1 can occur, then $[W] \exp(-h\nu_{21}/kT) \approx \exp(-h\nu_{21}/kT_{ex})$. For a comet at 1 a.u. distance from the sun the geometrical dilution factor is $W \approx 5 \times 10^{-6}$, $(h\nu_{21}/kT) \approx 0.4$, and this limit gives $T_{ex} \approx 180°$. The magnetic dipole transition probability is (Condon and Shortley 1951)

$$A_{vib} = \frac{64\pi^4 \sigma^3}{3h} (aMb)^2, \qquad (2)$$

where $\sigma = 1618$ cm^{-1} is the wave number between successive vibrational levels, and (aMb) is the matrix element of the magnetic dipole moment, which we can crudely estimate to be 1 Bohr magneton. We thus roughly estimate $A_{21}(=A_{vib}) = 10^{-1}$ sec^{-1}, which is of the same order of magnitude as $WA_{31}\exp(-h\nu_{32}/kT)$, but smaller than $A_{31}\exp(-h\nu_{32}/kT)$, since for the strongest transitions in the Swan bands $A_{31} \approx 5 \times 10^6$sec^{-1}. Accordingly, we expect the calculated excitation temperature to be between the two limits above but closer to the first, that is, to be of the order of a few thousand degrees.

Therefore accurate calculations of the statistical equilibrium of cometary C_2 were made, taking into account the lowest 5 vibrational levels of the ground X $^3\Pi$ term and also the lowest 5 vibrational levels of the upper A $^3\Pi$ term.

The relative populations were obtained by solving the equations of statistical equilibrium. Equations similar to those of Houziaux (1960) were used, with the addition of pure vibrational transitions within the electronic terms. Only those pure vibrational transitions with $\Delta v = \pm 1$ were considered. For compieteness, pure vibrational absorptions of solar infrared radiation were included, although they had negligible influence on the results. Following are the equations of statistical equilibrium, lower electronic term, i = 1, 5:

$$n_{i-1} B_{i-1\ i} \rho_{i-1\ i} - n_i(B_{i\ i-1} \rho_{i\ i-1} + A_{i\ i-1} + B_{i\ i+1} \rho_{i\ i+1} + \sum_{k=6}^{10} B_{ik} \rho_{ik}) + n_{i+1} (B_{i\ i+1} \rho_{i\ i+1} + A_{i+1\ i}) + \sum_{k=6}^{10} n_k (B_{ki} \rho_{ki} + A_{ki}) = 0 \qquad (3)$$

upper electronic term, j = 6, 10:

$$n_{j-1} B_{j-1\ j} \rho_{j-1\ j} - n_j[B_{j\ j-1} \rho_{j\ j-1} + A_{j\ j-1} + B_{j\ j+1} \rho_{j\ j+1} + \sum_{=1}^{5} (B_j \rho_j + A_j)] + n_{j+1} (B_{j+1\ j} \rho_{j+1\ j} + A_{j+1\ j}) + \sum_{=1}^{5} n B_j \rho_j = 0 \qquad (4)$$

The notation is essentially the same as used by Houziaux (1960).

The same values of A_{vib} were used for all vibrational transitions of both upper and lower electronic terms. For the pure vibrational absorptions, all upper levels were assumed to absorb at $\lambda = 5.70\mu$ while the lower absorbed at 6.18μ (these are wavelengths of the $v=0$ to $v=1$ transitions). The solar radiation at these wavelengths was obtained from Allen (1963). Values of the other parameters used in equations (3) and (4) are listed in Table 1. The columns headed m and n refer to the subscripts used in equations (3) and (4). The transition probabilities were computed from the overlap integrals and the oscillator strength using the same values of these quantities as Houziaux (1960). The radiation densities were derived from the "mean monochromatic intensities" in Table 1 of Minnaert (1953). The transition probabilities and radiation densities computed here are slightly different from those of Houziaux, but either set of values gives essentially the same results.

The equations of statistical equilibrium were solved with the aid of an electronic computer. For comparison, we show in Table 2 the results of our work for the case with $A_{vib} = 0$ and a sun-comet distance of 0.72 a.u. (which is probably the distance used by Houziaux) and Houziaux's results for the same case. We believe that there is some error in the latter. Table 3 shows the results for various values of A_{vib} and the sun-comet distance. In order to allow for uncertainties in the value of A_{vib}, calculations were made with the assumed values $A_{vib} = 1.0$, 0.1 and 0.01 sec^{-1}.

It can be seen that the results of Table 3 have the same behavior as the three-level case, i.e., decreasing vibrational temperature with increasing vibrational transition probability. It is also seen that transition probabilities between $A_{vib} = 0.1$ and 1.0 sec^{-1} give excitation temperatures in the 2000–3000° range observed by McKellar and Climenhaga.

TABLE 1.—*Parameters Used in the Equations of Statistical Equilibrium*

v'	v''	m	n	λ	$A_{v'v''}$ $\sec^{-1}$	$B_{v'v''}$ $\text{erg}^{-1}\text{cm}^3\sec^{-2}$	$\rho_{v'v''}$(0.72 a.u.) erg cm^{-3} sec
0	0	6	1	5165	$6.36\cdot 10^{6}$	$5.26\cdot 10^{18}$	$1.135\cdot 10^{-19}$
0	1	6	2	5635	$1.41\cdot 10^{6}$	$1.52\cdot 10^{18}$	1.257
0	2	6	3	6191	$2.12\cdot 10^{5}$	$3.02\cdot 10^{17}$	1.433
1	0	7	1	4737	$2.67\cdot 10^{6}$	$1.71\cdot 10^{18}$	1.020
1	1	7	2	5129	$3.23\cdot 10^{6}$	$2.61\cdot 10^{18}$	1.129
1	2	7	3	5585	$1.93\cdot 10^{6}$	$2.02\cdot 10^{18}$	1.263
1	3	7	4	6122	$4.60\cdot 10^{5}$	$6.34\cdot 10^{17}$	1.425
1	4	7	5	6764	$7.75\cdot 10^{4}$	$1.44\cdot 10^{17}$	1.530
2	0	8	1	4382	$3.42\cdot 10^{5}$	$1.73\cdot 10^{17}$	0.741
2	1	8	2	4715	$4.07\cdot 10^{6}$	$2.56\cdot 10^{18}$	1.009
2	2	8	3	5098	$1.47\cdot 10^{6}$	$1.17\cdot 10^{18}$	1.096
2	3	8	4	5540	$1.96\cdot 10^{6}$	$2.02\cdot 10^{18}$	1.223
2	4	8	5	6060	$6.68\cdot 10^{5}$	$8.93\cdot 10^{17}$	1.418
3	1	9	2	4371	$8.61\cdot 10^{5}$	$4.32\cdot 10^{17}$	0.737
3	2	9	3	4697	$4.68\cdot 10^{6}$	$2.92\cdot 10^{18}$	1.001
3	3	9	4	5071	$5.24\cdot 10^{5}$	$4.10\cdot 10^{17}$	1.119
3	4	9	5	5502	$1.89\cdot 10^{6}$	$1.89\cdot 10^{18}$	1.211
4	2	10	3	4365	$1.40\cdot 10^{6}$	$6.98\cdot 10^{17}$	0.734
4	3	10	4	4685	$4.98\cdot 10^{6}$	$3.07\cdot 10^{18}$	0.995

TABLE 2.—*Relative Populations*

THIS PAPER				HOUZIAUX (1960)			
Lower Electronic Term		Upper Electronic Term		Lower Electronic Term		Upper Electronic Term	
v''	$N(v'')/N(v''=0)$	v'	$N(v')/N(v''=0)$	v''	$N(v'')/N(v''=0)$	v'	$N(v')/N(v''=0)$
1	0.69	0	$9.4\cdot 10^{-8}$	1	0.95	0	$1.0\cdot 10^{-7}$
2	0.48	1	6.4	2	0.71	1	$8.7\cdot 10^{-8}$
3	0.33	2	4.3	3	0.61	2	7.9
4	0.23	3	2.9	4	0.37	3	4.5
		4	2.0			4	6.1

From the relative populations given in Table 3 it is also possible to estimate the expected infrared radiation of cometary C_2. We will make this calculation for the (1,0) band of the lower electronic term, which occurs at 6.18μ. Both the earth-comet and sun-comet distances are taken as 1 a.u.

According to Wurm (1963), the number of C_2 molecules in the ground state of a typical bright comet is

$$N(C) \approx 6\times 10^{32} \text{ molecules}$$

The luminosity in the (1,0) band is then given by the expression

$$L \approx \frac{N(v''=1)}{N(v''=0)} \cdot N(C_2) A_{vib}\, h\nu_{vib}$$

where $N(v''=1)/N(v''=0)$ is a relative population given in Table 3. With the assumption $A_{vib}=0.1$ $\sec^{-1}$ we obtain $L \approx 9\times 10^{18}$ ergs/sec, or a flux received at the earth, assuming the whole comet is in the field of the telescope, of about 3×10^{-9} ergs/cm^2 sec.

TABLE 3.—*Relative Populations and Boltzmann Temperatures*

Lower Level		T	Upper Level	
v''	$N(v'')/N(v''=0)$		v'	$N(v')/N(v'=0)$
		$R=0.4$ a.u. $A_{vib}=1.0$		
1	0.39	25×10^2	1	0.50
2	0.22	31	2	0.25
3	0.12	33	3	0.14
4	0.06	32	4	0.09
		$A_{vib}=0.1$		
1	0.65	53×10^2	1	0.62
2	0.44	56	2	0.42
3	0.29	56	3	0.28
4	0.20	56	4	0.19
		$A_{vib}=0.01$		
1	0.69	63×10^2	1	0.68
2	0.48	62	2	0.45
3	0.33	61	3	0.31
4	0.23	62	4	0.21
		$R=0.7$ a.u. $A_{vib}=1.0$		
1	0.20	14×10^2	1	0.39
2	0.09	19	2	0.13
3	0.04	21	3	0.06
4	0.01	21	4	0.03
		$A_{vib}=0.1$		
1	0.56	40×10^2	1	0.60
2	0.36	46	2	0.36
3	0.24	48	3	0.23
4	0.14	46	4	0.16
		$A_{vib}=0.01$		
1	0.68	60×10^2	1	0.67
2	0.47	61	2	0.45
3	0.32	60	3	0.30
4	0.22	60	4	0.20
		$R=1.0$ a.u. $A_{vib}=1.0$		
1	0.11	11×10^2	1	0.34
2	0.04	15	2	0.08
3	0.01	16	3	0.03
4	0.004	17	4	0.01
		$A_{vib}=0.1$		
1	0.47	31×10^2	1	0.55
2	0.28	37	2	0.30
3	0.17	39	3	0.18
4	0.09	38	4	0.12

TABLE 3.—*Relative Populations and Boltzmann Temperatures (Continued)*

Lower Level		T_{ex}	Upper Level	
v''	$N(v'')/N(v''=0)$		v'	$N(v')/N(v'=0)$
		$A_{vib}=0.01$		
1	0.66	57×10^2	1	0.66
2	0.45	58	2	0.43
3	0.30	58	3	0.29
4	0.21	58	4	0.20
		$R=1.4$ a.u. $A_{vib}=1.0$		
1	0.06	8×10^2	1	0.31
2	0.02	12	2	0.05
3	0.006	14	3	0.01
4	0.002	14	4	0.006
		$A_{vib}=0.1$		
1	0.35	22×10^2	1	0.48
2	0.19	28	2	0.22
3	0.11	31	3	0.12
4	0.05	30	4	0.08
		$A_{vib}=0.01$		
1	0.63	51×10^2	1	0.65
2	0.43	55	2	0.41
3	0.29	55	3	0.28
4	0.20	57	4	0.19

Rough calculations can be made as to the possibility of detecting this band. Assuming that the minimum power (i.e. Noise Equivalent Power) detectable by an infrared detector is about 10^{-11} watts; and that 0.1 of the power incident on the telescope reaches the detector, then it would require a telescope diameter of several hundred inches to observe this band (even above the atmosphere). Clearly, detection of the infrared lines of C_2 are not to be expected.

In summary, then, with the estimated pure vibrational transition probabilities and the assumption of the fluorescence mechanism, we have been able to derive relative populations of a 10-level C_2 molecule. These results are in agreement with those suggested by a simplified 3-level molecule and with published observations of comets.

It has also been possible to predict the infrared emission expected from a bright comet. This emission seems too weak to be observed. Similar calculations need to be performed on other cometary molecules to predict their infrared spectra.

REFERENCES

AMBARTSUMYAN, V. A. 1958, *Theoretical Astrophysics* (London: Pergamon Press), p 416.

CONDON, E. U. and SHORTLEY, G. H. 1951, *The Theory of Atomic Spectra* (Cambridge: Cambridge University Press).

HOUZIAUX, L. 1960, Ann d'ap., **23**, 1025.

HUNAERTS, J. 1953, Mem. Soc. R. Sci. Liege, **13**, 59.

HUNAERTS, J. 1957, Mem. Soc. R. Sci. Liege, **18**, 82.

MCKELLAR, A., and CHAMBENHAGA, J. L. 1953 *La Physique des Comètes* (Rept. of 4th Liege Symposium).

MINNAERT, M. 1953, *The Sun* (Chicago: University of Chicago Press), 92.

STAWIKOWSKI, A., and SWINGS, P. 1960, Ann d'ap., **23**, 585.

VAN VLECK, 1934, Ap. J., **80**, 161.

WURM, K. 1936, *Handbuch der Physik*, **7**, 306.

WURM, K. 1963, *The Moom, Meteorites and Comets* (Chicago: University of Chicago Press)

EVOLUTION OF O STARS. II. HYDROGEN EXHAUSTION AND GRAVITATIONAL CONTRACTION*

RICHARD STOTHERS

The evolution of a star of 30 $M_\odot$ is considered from the end of the stable phase of hydrogen burning to the onset of helium burning. Ten models are constructed for the hydrogen-exhaustion (E) phase, and six models for the gravitational contraction (G) phase.

The time scale of the E-phase is so short (8.8×10^4 years) that the shell source remains peaked at $q=0.34$ and undergoes little hydrogen depletion. Because radiation pressure remains strong in the core, convection does not vanish when the hydrogen content at the center falls to zero. Gravitational contraction of the core contributes more to the luminosity than shell burning, but since the total luminosity changes little, the structure of the envelope is hardly affected. On the H–R diagram, the evolutionary track turns back toward the main sequence when $X_c=0.03$, and does not turn away again until the shell source becomes important.

The G-phase begins when $L_{H,\,core}/L<0.001$, and lasts 9×10^3 years. Although the temperature in the shell increases, hydrogen depletion remains negligible because of the short time scale. However, the shell narrows considerably and its peak moves slightly inward for a while. The steep temperature gradient outside the shell causes the semiconvective zone to move inward to $q=0.48$; the hydrogen discontinuity attains a value $\Delta X=0.1$. As the luminosity and radiation pressure in the shell simultaneously increase, the envelope expands. The shell then behaves like a node, since the core continues to contract. The convective region near the center shrinks asymptotically to a value $q=0.06$, but at no time does the core approach an isothermal condition. The gravitational energy release is nearly uniform throughout the core in all phases. At the onset of helium burning, $T_c=1.5\times10^8$ °K and $\rho_c=270$ gm/cm³. Because of the brightening shell source, the stellar radius increases rapidly, bringing the evolutionary track to completion of the typical S-shaped curve on the H–R diagram.

When helium starts to burn, the spectral type is B3, and the star is expected never to return to the region of O stars during its active life.

I. INTRODUCTION

In massive stars when the central hydrogen content falls to a value of 0.03, the whole structure undergoes a drastic reorganization. Kushwaha (1957) first considered the subsequent early phase of hydrogen burning in a shell outside the convective core for a star of 10 $M_\odot$. Reiz (1963) has reconsidered and extended this work, but Hayashi and Cameron (1962) and Hayashi, Hoshi, and Sugimoto (1962) have pursued the study of a star of $15.6M_\odot$ into the entire hydrogen-exhaustion phase and then into the gravitational contraction phase preceding helium burning. These phases have also been considered for stars of intermediate mass, as follows: $3.89M_\odot$ (Hoyle 1960), $4M_\odot$ (Hayashi, Nishida, and Sugimoto 1962; Hayashi, Hoshi, and Sugimoto 1962), $5M_\odot$ (Polak 1962), and $7M_\odot$ (Hofmeister, Kippenhahn, and Weigert 1963). It is the purpose of the present paper to follow the evolution of a star of $30M_\odot$ during these phases to the onset of helium burning at the center. The previous hydrogen-burning stage has been computed in Paper I (Stothers 1963).

II. ASSUMPTION AND DEFINITIONS

a) Assumptions

The same general assumptions as in Paper I are made, except that nuclear-energy generation also occurs in a radiative shell around the core and the

*Published in *The Astrophysical Journal*, 140(2): 510–523, August 15, 1964. This paper and "Evolution of O Stars. I Hydrogen Burning" were based on the author's doctoral dissertation, Harvard University, 1963.

time rate of change of the physical variables must be taken into account during these fast evolutionary phases.

The opacity of stellar material is again assumed to be due only to electron scattering, and the equation of state is represented by the sum of the perfect gas pressure and radiation pressure.

b) Notation

In general, we apply the same notations for stellar zones as in Paper I, except here we divide the radiative intermediate zone into outer (Zone IIIa) and inner (Zone IIIb) subzones, separated by the radiative, hydrogen-burning shell (denoted by a subscript s).

To designate the successive evolutionary phases we introduce H (hydrogen burning), E (hydrogen exhaustion), and G (gravitational contraction) followed by a number indicating the particular model in the phase.

c) Chemical Composition

The outer radiative envelope retains the initial age-zero composition assumed in Paper I:

$$X_e=0.70,\ Y_e=0.27,\ Z_e=0.03,\ X_{\rm CNO}=Z_e/2. \quad (1)$$

The semiconvective zone constantly adjusts its composition to maintain convective neutrality. A discontinuity in X marks its interface with the inner radiative zone, which retains a constant gradient in X, left behind by the retreating, homogeneous convective core. The distribution of X for the last model of the hydrogen-burning phase is given by Figures 3 and 4 of Paper I. In the core of that model $X_c=0.07$.

The change of hydrogen content as an explicit function of time, τ, and mass fraction, q, is given by

$$\frac{\partial X}{\partial \tau}=-\frac{\epsilon_{\rm H}}{E_{\rm H}} \text{(radiative region)} \quad (2)$$

and

$$\frac{\partial X}{\partial \tau}=-\frac{1}{q_4 E_{\rm H}}\int_0^{q_4}\epsilon_{\rm H}dq \text{ (convective core)}, \quad (3)$$

where $\epsilon_{\rm H}$ is the energy generation due to hydrogen burning and $E_{\rm H}=6.0\times10^{18}$ erg/gm is the energy released per CNO cycle.

d) Nuclear Energy

The accurate expression for the rate of energy release due to the conversion of hydrogen into helium via the full CNO cycle has been given by Reeves (1962) in terms of the $N^{14}+H^1$ rate:

$$\epsilon_{\rm H}=7.94\times10^{27}f_{14,1}g_{14,1}\left(\frac{X_{14}}{X_{\rm CNO}}\right)X_{\rm CNO}X\rho T_6^{-2/3} \exp\left(-152.3T_6^{-1/3}\right) \text{ erg/gm sec} \quad (4)$$

where

$$f_{14,1}=1+1.75\,\rho^{1/2}T_6^{-2/3}, \quad (5)$$

$$g_{14,1}=1+0.0027\,T_6^{1/3}-0.0037\,T_6^{1/6}-0.00007T_6, \quad (6)$$

$$X_{14}/X_{\rm CNO}=0.99-0.00067\,T_6. \quad (7)$$

Here $f_{14,1}$ is the weak electron-screening factor, $g_{14,1}$ is the correction term to the zero-energy S factor, and $X_{14}/X_{\rm CNO}$ has been estimated from tables given by Reeves (1962) for temperatures in excess of 2.5×10^7 °K. We are assuming everywhere the full equilibrium abundance of oxygen, which, strictly speaking, is not attained in the cooler shell source because of the short time scale involved.

e) Gravitational Energy

The amount of energy involved during a gravitational contraction or expansion is expressed as the difference between the work done by the pressure and the change in thermal energy of the gas and radiation:

$$\epsilon_g=-P\frac{\partial v}{\partial s}-\frac{\partial u}{\partial \tau}=\frac{3kT}{2\mu H}\left[\frac{\partial}{\partial \tau}\ln\left(Ty^{-2/3}e^{-8y/3}\right)+\frac{5+8y}{3}\frac{\partial}{\partial \tau}\ln\mu\right], \quad (8)$$

where

$$v=\frac{1}{\rho},\qquad \frac{u}{v}=\frac{3}{2}\frac{k}{\mu H}\rho T-aT^4,\qquad y=\frac{1-\beta}{\beta}, \quad (9)$$

and the time derivative applies to a given mass fraction. As Hayashi and Cameron (1962) have noted, the logarithmic arguments in equation (8) are independent of position in convective regions of the star.

III. BASIC EQUATIONS

The basic equilibrium equations are given in Paper I. For Zones I–IIIa the equations were transformed and integrated, and the zones fitted to each other in the same way as before. Henceforth we shall call this region simply the *envelope*. It is specified only by the luminosity L (through the parameter C) since, down to the chosen termination point $q_f=0.500$, the composition in the radiative region has been fixed by the previous models.

In the *core* (Zones IIIb and IV) we employ q as the independent variable and logarithms of $r, L(r)$, P, and T as dependent variables. The core solutions are specified by β_c and T_c and are integrated from the center out through the shell to q_f, beyond which any energy sources are negligible. The customary series expansions near the center (Schwarzschild (1958), p. 114) have been modified to include radiation pressure and applied at $q=0\,002$. The step value was taken to be $\Delta q=0.001$.

Fitting of the envelope and core is made at q_f in U, V, and $L(r)=L$. Since L was expected to change little during the present evolutionary phases, it was sufficient to integrate a short series of envelopes at equally spaced intervals of L and then to interpolate values of U_f and V_f at finer intervals of L.

The change in $X(q)$ as a function of time has been given by equations (2) and (3). These may be approximated by difference equations, with the known run of (ϵ_H/X) as a function of q from the previous model. Letting ι refer to quantities of the previous model, we write in an obvious notation

$$\frac{\Delta \ln X}{\Delta \tau}=-\frac{1}{2E_H}\left[\left(\frac{\epsilon_H}{X}\right)^{\iota}+\left(\frac{\epsilon_H}{X}\right)\right]=-\Xi_r(q), \qquad (q\geq q_4^{\iota}) \quad (10)$$

$$\frac{\Delta \ln X}{\Delta \tau}=-\frac{1}{(q_4{}^{\iota}+q_4)E_H}\int_0^{q_4}\left[\left(\frac{\epsilon_H}{X}\right)^{\iota}+\left(\frac{\epsilon_H}{X}\right)\right]dq=-\Xi_c(q_4), \qquad (q_4\geq q>0) \quad (11)$$

$$\frac{\Delta \ln X}{\Delta \tau}=\Xi_r(q)+[\Xi_c(q_4)-\Xi_r(q_4)]\frac{q_4{}^{\iota}-q}{q_4{}^{\iota}-q_4}, \qquad (q_4{}^{\iota}\geq q\geq q_4) \quad (12)$$

for $q_4{}^{\iota}-q_4$ small. These approximations are slightly more accurate than those used by Haselgrove and Hoyle (1956) and by Hayashi and Cameron (1962). Outside the convective core, in which X_c is known (constant), an iterative procedure must be used to calculate X from equation (10) or equation (12), since (ϵ_H/X) depends implicitly on X through ρ. Using X from the previous model, we calculate a preliminary value of ρ and hence (ϵ_H/X). Insertion of (ϵ_H/X) into equations (10) or (12) yields a projected value of X. Repetition of this procedure yields a definitive value of X correct to the second order.

The method of producing an evolutionary step was varied according to the magnitude of the ratio of luminosity supplied by gravitational contraction to total luminosity, as follows.

a) $L_g/L<0.1$

The method of Hayashi and Cameron (1962) was adopted for this case only, although they seem to have used it up to $L_g/L\approx 0.6$. Here the part of the gravitational energy release inclosed in brackets in equation (8) is expressed as a quadratic function of q, with coefficients extrapolated from the previous models:

$$\epsilon_g=\frac{3}{2}\frac{kT}{\mu H}(a_0+a_1q+a_2q^2). \qquad (13)$$

The basic equations, where now $\epsilon=\epsilon_H+\epsilon_g$, are integrated from the center with input values of X_c, β_c, T_c, a_0, a_1, a_2, $q_4{}^{\iota}$, and the run of X^{ι} and $(\epsilon_H/X)^{\iota}$ as a function of q. At q_4, $\Delta\tau$ is determined by equation (11). When a self-consistent model is obtained, the quantity in brackets in equation (8) is computed exactly and improved values of a_0, a_1, and a_2 are obtained by least squares. Another model is integrated with the new values, but this second model is always sufficiently accurate for $L_g/L<0.1$.

b) $0.1<L_g/L<0.5$

In this case we specify X_c and estimate $\Delta\tau$ by extrapolation from the previous models. Hence ϵ_g can be computed exactly from equation (8) with the given $\Delta\tau$. In the course of integrations to find the correct β_c and T_c it is possible continually to improve $\Delta\tau$ by use of equation (11).

c) $L_g/L>0.5$

As L_g/L increases, L becomes more sensitive to ϵ_g and hence to $\Delta\tau$. To cope with the increasing difficulty of making $\Delta\tau$ converge, we choose a parameter that loses importance with increasing L_g/L, namely, X_c. In this case, we take the time step by specifying $\Delta\tau$ explicitly. With a value of X_c extrapolated from the previous models, it is possible to obtain a definitive value for X_c in the same way as for $\Delta\tau$ in Section III*b*. Eventually, when hydrogen burning ceases entirely in the shrinking convective core, we need only specify $\Delta\tau$.

IV. HYDROGEN-EXHAUSTION PHASE

The results of the computations for this phase are presented in Table 1. It is curious that the stellar radius starts to decrease, causing the star to turn back toward the main sequence in the H–R diagram, as soon as $X_c \approx 0.03$ for a broad range of stellar masses ($10 < M/M_\odot < 60$), even though the mass fractions contained in the convective core are very different (see Table 2). It is clear, however, why the radius will shrink when X_c becomes small. To maintain the luminosity, the central temperature must correspondingly increase; but since μ_c changes little, $R \sim T_c^{-1}$ from equation (10) of Paper I.

For a while hydrogen burning in the core is able, alone, to maintain the energy balance (Schwarzschild and Härm 1958). However, eventually an accelerated contraction of the core is necessary to supplement the deficient $L_{H,core}$ with a growing L_g. When the temperature becomes high enough, hydrogen burning in a shell outside the core also becomes important. It is interesting to note that in the above range of masses the peak of this shell source always coincides with the mass fraction of the convective core when $X_c \approx 0.07$, i.e., just *before* the radius starts to decrease. Table 2 shows that q_s, the peak of the shell source, is roughly half of q_0, the mass fraction of the core of the initial main-sequence model. Furthermore, q_0 itself increases by about 0.15 every time the stellar mass is doubled; this is a consequence of the increasing importance of radiation pressure (diminishing β), which extends the convective regions outward.

TABLE 1.—*Evolutionary Models of a Star of 30 $M_\odot$ during Hydrogen-Exhaustion (E) Phase*

	Models									
	1	2	3	4	5	6	7	8	9	10
log C	−3.272	−3.266	−3.264	−3.262	−3.260	−3.256	−3.252	−3.248	−3.244	−3.239
q_1	+0.717	+0.719	+0.720	+0.721	+0.721	+0.722	+0.723	+0.725	+0.728	+0.731
$q_2 = q_3$	+0.550	+0.547	+0.546	+0.545	+0.545	+0.543	+0.541	+0.539	+0.537	+0.534
X_2	+0.650	+0.647	+0.645	+0.644	+0.643	+0.640	+0.637	+0.634	+0.630	+0.627
X_3	+0.589	+0.584	+0.581	+0.579	+0.578	+0.573	+0.568	+0.564	+0.559	+0.554
q_s	+0.342	+0.342	+0.342	+0.342	+0.342	+0.342	+0.342	+0.342	+0.342	+0.342
β_s	+0.614	+0.608	+0.605	+0.605	+0.604	+0.604	+0.602	+0.600	+0.597	+0.595
log T_s	+7.436	+7.458	+7.487	+7.515	+7.542	+7.561	+7.571	+7.578	+7.584	+7.589
log $(r_s/R_\odot)$	+0.267	+0.240	+0.211	+0.181	+0.153	+0.130	+0.117	+0.107	+0.097	+0.089
q_4	+0.328	+0.320	+0.316	+0.311	+0.302	+0.282	+0.258	+0.236	+0.205	+0.182
log X_c	−1.523	−2.023	−2.523	−3.023	−3.523	−4.023	−4.523	−5.015	−5.882	−7.311
β_c	+0.564	+0.557	+0.554	+0.555	+0.555	+0.559	+0.563	+0.567	+0.572	+0.577
log T_c	+7.665	+7.696	+7.726	+7.756	+7.785	+7.808	+7.823	+7.833	+7.844	+7.854
log ρ_c	+0.701	+0.795	+0.886	+0.978	+1.066	+1.140	+1.193	+1.229	+1.272	+1.312
$L_{H,core}/L$	+0.999	+0.995	+0.985	+0.943	+0.853	+0.599	+0.319	+0.145	+0.028	+0.002
$L_{H,shell}/L$	.....	+0.001	+0.004	+0.014	+0.044	+0.094	+0.141	+0.175	+0.217	+0.249
L_g/L	+0.001	+0.004	+0.011	+0.043	+0.103	+0.307	+0.540	+0.680	+0.755	+0.749
log $(L/L_\odot)$	+5.439	+5.445	+5.447	+5.449	+5.451	+5.455	+5.459	+5.463	+5.468	+5.472
log $(R/R_\odot)$	+1.143	+1.136	+1.113	+1.090	+1.066	+1.057	+1.057	+1.062	+1.070	+1.079
log T_e	+4.548	+4.554	+4.565	+4.578	+4.590	+4.595	+4.596	+4.595	+4.592	+4.589
τ (10^4 years)	0.00	+5.91	+7.74	+8.33	+8.53	+8.60	+8.61	+8.67	+8.69	+8.72

It is for this reason, the very distance of the hydrogen-rich regions, that the temperature at $X \approx 0.07$ is low and shell burning is consequently delayed. (In massive stars, β is low not so much by virtue of high temperature as of low density.) In stars of lower mass the point at which $X_c \approx 0.07$ lies closer to the center, and shell burning begins to contribute significantly to the luminosity before the core contraction does. The critical mass for this to happen is about $20M_\odot$. If q_s is small enough that the core mass lies below the Schönberg-Chandrasekhar limit ($q \approx 0.1$), thus becoming isothermal or even degenerate, then L_ϱ may always be small. The approach to isothermality may be seen in a plot of log T against q for $5M_\odot$ (Polak 1962, Fig. 3), where $q_s = 0.11$. It may be inferred for $15.6M_\odot$ (Hayashi and Cameron 1962), where $q_s = 0.17$, since $L_{H,\,shell} > L_\varrho$ until $X_c < 0.001$. In stars with mass $\sim 1M_\odot$ (Schwarzschild and Selberg 1962), the shell source supplies practically all the luminosity outside a very large degenerate core. The critical stellar mass for the occurrence of degeneracy in the core prior to helium burning is $\sim 4M_\odot$ (Hayashi, Nishida, and Sugimoto 1962).

At $30M_\odot, q_s \gg 0.1$ and the core never even approaches an isothermal condition. Thus gravitational contraction proceeds unchecked to increase T_c in accordance with the virial theorem. The subsequent increase in ϵ_H reduces X_c to less than 1 percent. Thereafter the drop in X_c outruns the increase in T_c, and ϵ_H decreases, being supplanted by $\epsilon_\varrho \sim \partial T_c / \partial \tau$. However, hydrogen burning in the core still supplies over half the total luminosity even when X_c is as low as 10^{-4}.

The distribution of hydrogen in the central regions may be seen in Figure 1. It is clear that,

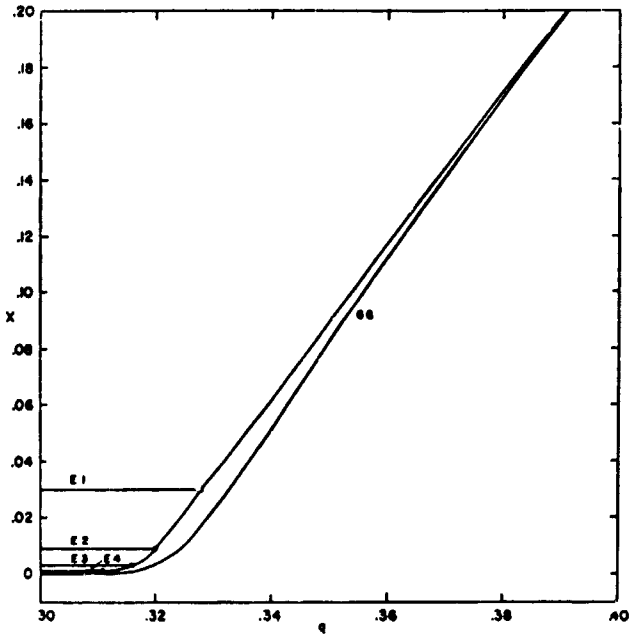


FIGURE 1.—Depletion of hydrogen as a function of mass fraction during the hydrogen-exhaustion (E) and gravitational-contraction (G) phases. Filled circles represent the boundary of the convective core in the E models.

because of the high central temperature and the luminosity requirements, X_c begins to decrease more and more rapidly with q_4. This has the effect of necessitating an even sharper increase of temperature, which consequently exhausts the central hydrogen and ignites a region which is still relatively hydrogen-rich but close enough to the center to feel the temperature rise. Thus Figure 1 shows why a shell source develops near $q = 0.34$ for $30M_\odot$. We expect this shell to be initially broader in massive stars than in stars of lower mass because the mean gradient of hydrogen throughout this region is gentler (larger $q_0 - q_s$; see

TABLE 2.—*Location of Critical Interfaces in Stars of Intermediate and High Mass When* $X_c \approx 0.03$

$M/M_\odot$	q_0	q_s	q_4	X_s	Reference
5	0.21	0.11	0.10	0.74	Polak (1962)
10	.24	.....	.07	.90	Kushwaha (1957)
15.6	.42	.17	.16	.90	Sakashita, Ōno, Hayashi (1959)
30	.60	0.34	.33	.70	Stothers (1963)
62.7	0.75	.....	0.41	0.75	Schwarzschild and Härm (1958)

Table 2). The subsequent growth of the shell source is shown in Figure 2.

In Figure 3 the relative contributions to the energy generation may be compared. The magnitude of ϵ_ϱ is very nearly constant throughout the core for all stages.

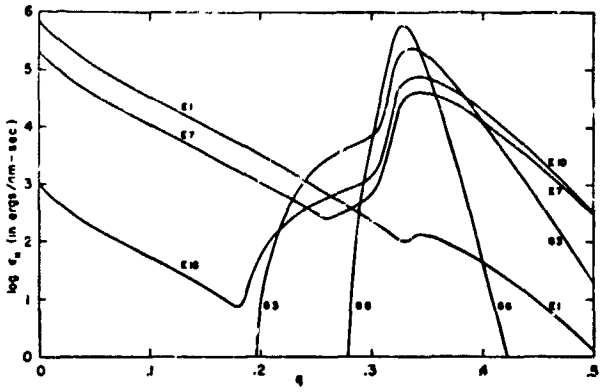


FIGURE 2.—Logarithm of the energy generation due to hydrogen burning as a function of mass fraction. Curves are labeled with the model numbers.

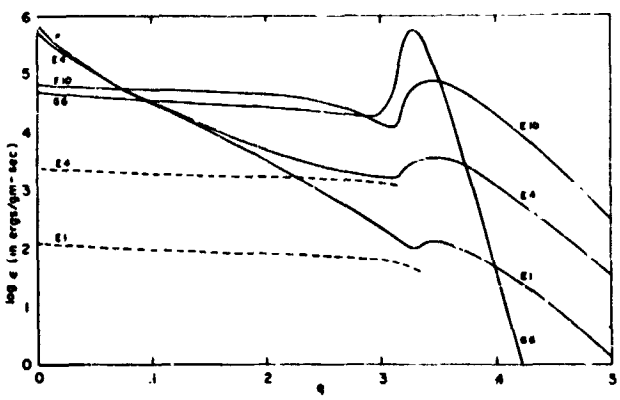


FIGURE 3.—Logarithm of the energy generation as a function of mass fraction. Solid curves refer to the contributions from both hydrogen burning and gravitational contraction. The dashed curve refers only to gravitational contraction. Curves are labeled with the model numbers.

Shrinkage of the convective core is at first slow because β_c remains almost constant. Hence the evolution of the center approximates Lane's law, $\rho_c \sim T_c^3$. When the shell source becomes important for $X_c < 10^{-4}$, ρ_c increases at a slightly greater rate than T_c^3, the only concession to isothermality that the core makes. At this point (Model E7), $L_\varrho/L \approx 0.5$ and $L_{H,\,shell}/L \approx 0.1$. It seems that as soon as L_ϱ/L grows much larger, contraction of the core begins to be offset by an expansion of the envelope (cf. Hayashi and Cameron 1962). This expansion is due to an increase of $L_{H,\,shell}$, compensating for the reduction of $L_{H,\,core}$. Exactly why the envelope expands as $L_{H,\,shell}$ increases will be made evident in the next section. The result on the H—R diagram is that the star takes another turn away from the main sequence.

When the hydrogen content at the center of the star vanishes after Model E10, the convective core does not disappear. This contrasts with the results on stars of lower mass (Hoyle 1960; Polak 1962; Hayashi and Cameron 1962), where apparently it is just the nuclear-energy generation that maintains the steep temperature gradient. However, in the case of $30M_\odot$, β is still so low that convection persists near the center.

The time scale of this phase is short enough (8.8×10^4 years) that the shell source, which remains peaked at $q_s = 0.342$, undergoes very little hydrogen depletion. Because of the struggle of the various energy sources to maintain the luminosity, L increases but slightly. Thus the structure of the envelope remains nearly constant.

V. GRAVITATIONAL CONTRACTION PHASE

This phase is taken to start when $L_{H,\,core}/L < 0.001$. The hydrogen content of the convective core is negligible, and the sole nuclear-energy source is the hydrogen-burning shell. In Model E9 the rise of the shell source had terminated the upward growth of L_ϱ/L, which reached a maximum value of 0.75. This fraction becomes all the more impressive when we compare it with 0.56 for a star of $15.6M_\odot$ (Hayashi and Cameron 1962). Table 5 shows that even at the onset of helium burning (Model G6) the contribution of gravitational contraction to the luminosity is still a sizable fraction. This suggests that for stars more massive than $30M_\odot$ hydrogen burning in a shell may not become dominant.

During the shell-burning phase, the computational difficulty of fitting models increases. The reason lies in the shift of the $U-V$ curve to the left, as seen in Figure 4. This is caused by the developing chemical inhomogeneity, which began after the initial main-sequence state, and by the shell source, which produces the leftward loop on the $U-V$ plane. Because of the short time scale, the amount of hydrogen depletion is very small

TABLE 3.—*Evolutionary Models of a Star of 30 M⊙ during Gravitational Contraction (G) Phase*

	Models					
	1	2	3	4	5	6
log C	−3.231	−3.216	−3.195	−3.178	−3.165	−3.152
q_1	+0.733	+0.739	+0.749	+0.756	+0.762	+0.769
$q_1=q_3$	+0.530	+0.522	+0.508	+0.498	+0.489	+0.481
λ_2	+0.620	+0.606	+0.582	+0.560	+0.541	+0.522
X_3	+0.542	+0.523	+0.490	+0.467	+0.445	+0.425
q_s	+0.342	+0.342	+0.338	+0.328	+0.328	+0.328
β_s	+0.589	+0.576	+0.556	+0.553	+0.553	+0.515
log T_s	+7.596	+7.607	+7.626	+7.654	+7.668	+7.683
log $(r_s/R\odot)$	+0.073	+0.048	+0.001	−0.058	−0.100	−0.142
q_4	+0.155	+0.123	+0.093	+0.078	+0.069	+0.064
β_c	+0.587	+0.604	+0.631	+0.651	+0.665	+0.674
log T_c	+7.873	+7.907	+7.971	+8.035	+8.101	+8.168
log ρ_c	+1.386	+1.520	+1.760	+1.992	+2.217	+2.433
$L_{H,\,shell}/L$	+0.307	+0.391	+0.485	+0.536	+0.561	+0.575
Lg/L	+0.693	+0.609	+0.515	+0.464	+0.439	+0.425
log $(L/L\odot)$	+5.480	+5.495	+5.516	+5.533	+5.546	+5.560
log $(R/R\odot)$	+1.098	+1.144	+1.239	+1.353	+1.492	+1.680
log T_s	+4.581	+4.562	+4.520	+4.467	+4.401	+4.310
τ (10^4 years)	0.00	+0.10	+0.30	+0.50	+0.70	+0.90

and the shell source remains peaked at roughly the same mass fraction. Since the shell provides the driving mechanism for the expanding envelope while the core is contracting, r at the outer boundary of the shell remains roughly constant. Thus it is here that the change of log r with q is greatest (see Fig. 5), that is, U is a minimum, isolating the energy-producing regions from the envelope. Unless small increments in q are taken in the numerical integrations through the shell, the solutions diverge. Since the region behaves like an outer boundary ($U_s \to 0$), V tends to blow up for small changes of the input parameters, β_c and T_c.

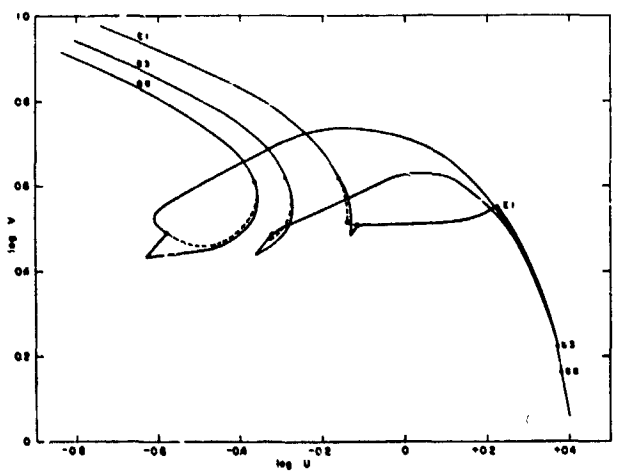


FIGURE 4.—Evolution of a star of 30 M⊙ in the U–V plane during the hydrogen-exhaustion (E) and gravitational-contraction (G) phases. Curves are labeled with the model numbers. Dots and jumps represent fitting points. The dashed lines in the solutions represent the assumed radiative zone (see Paper I).

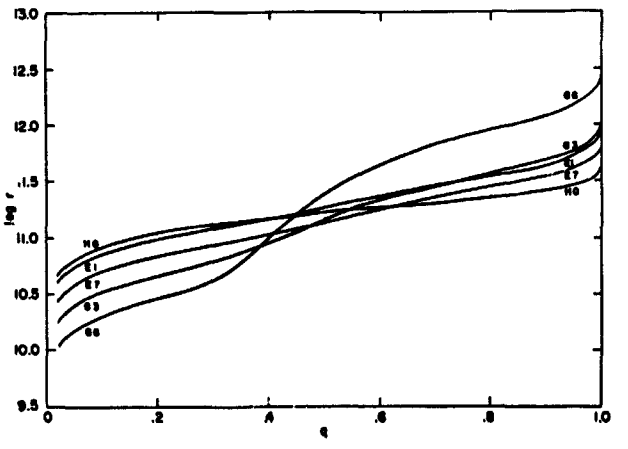


FIGURE 5.—Logarithm of the radius as a function of mass fraction. Curves are labeled with the model numbers.

Hayashi, Hoshi, and Sugimoto (1962) have interpreted the envelope expansion in terms of the $U-V$ plane. For values of the polytropic index $3 \leq N > 1$, where $P=K\rho^{1+1/N}$ and K and N are constants, the centrally condensed solutions of a polytropic envelope form a leftward loop on the $U-V$ plane and converge to a point $U=0$ and $V=N+1$ as $r \to 0$. For an electron-scattering envelope, N is 3 near the surface, decreases inside, and then approaches 3 again as $r \to 0$. Hence such an envelope approximates the polytropic case $N=3$, where $V \to 4$ as $U \to 0$ (see Fig. 4). Interior to q_s, $L(r) \ll L_{\mathrm{H,\ shell}}$ so that the radiative index becomes large, since $(n+1)_{\mathrm{rad}} \sim (1-\beta)/\kappa L(r)$. Hence the core approximates roughly the isothermal case $N=\infty$, which loops outward in the $U-V$ plane much as in Figure 4.

It remains to see how the shell source actually expands the envelope. We have already noted that, in order to tap additional nuclear fuel, the temperature in the core rises and starts shell burning. Consequent to hydrogen exhaustion in the core, the temperature in the shell must continue to rise so that $L_{\mathrm{H,\ shell}}$ can assume more of the total luminosity burden. As $L_{\mathrm{H,\ shell}}$ increases during the evolution, $(n+1)_{\mathrm{rad}}$ decreases, steepening the temperature gradient beyond q_s (see Fig. 6). But actually $n+1$ can only decrease until the adiabatic value is reached. Therefore, to compensate for the increasing luminosity, β decreases (see Fig. 7). This relative increase in radiation pressure expands the envelope.

To make this point more definite, we look at the structure of the envelope, which may be considered a polytrope of some characteristic index.

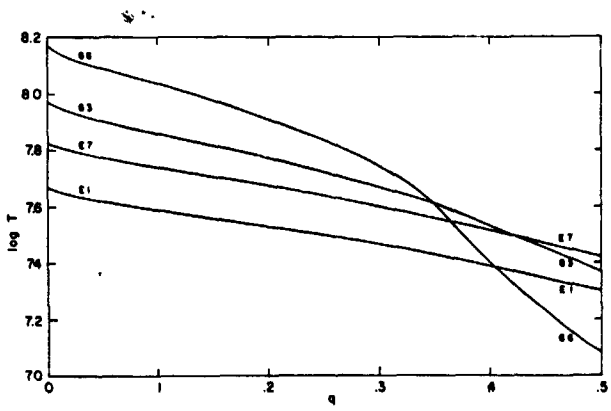


FIGURE 6.—Logarithm of the temperature as a function of mass fraction. Curves are labeled with the model numbers.

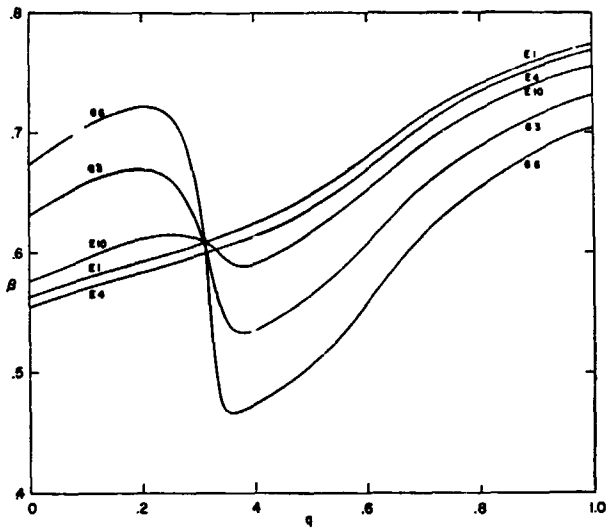


FIGURE 7.—Ratio of gas pressure to total pressure as a function of mass fraction. Curves are labeled with the model numbers.

The effective termination point of the envelope must be taken just outside the shell peak, where U is a minimum on the $U-V$ plane. Now it may be shown generally for a polytrope of any index and constant $\beta=\beta_c$ that $R \sim (T_c/\beta_c\rho_c)^{1/2}$. Eliminating ρ, we have $R \sim (1-\beta_c)^{1/2}/\beta_c T_c$. Thus the expansion may be considered as fundamentally due to the relative increase in radiation pressure, as stated above. Further, the same result can be obtained by examining equation (8). A decrease in β causes expansion by making ϵ_g negative, and an increase causes contraction by making ϵ_g positive. The consequences may be seen by comparing Figures 5–7. During phases of roughly constant β, the change of T will determine the change in R.

If we take the stellar core to be a polytrope, then we may similarly explain the changes of shell radius, r_s (see Tables 1 and 3). By extension, a star with m shell sources may be subdivided into $m+1$ zones whose structure is assumed to be polytropic. Then, approximately, the changes of β and T at a given shell should determine the changes of r at the next outer shell, unless the mass fraction of the shell changes by a drastic amount. Qualitative agreement is found between this picture of evolution and the accurate published results on the evolution of stars of $4M_\odot$ and $15.6M_\odot$ from the initial main sequence through carbon burning (Hayashi, Nishida, and

Sugimoto 1962; Hayashi and Cameron 1962; Hayashi, Hoshi, and Sugimoto 1962).

In this regard we note that the last model before helium burning in a star of $1.3M_\odot$ considered by Schwarzschild and Selberg (1962) had $\beta_s = 0.87$. Since they neglected radiation pressure and used $\beta = 1$ throughout, it might be expected that the radii they derive are too small. Indeed the track of their theoretical models on the H–R diagram deviates leftward of the observed red-giant branch in globular clusters. One reason may be the neglect of varying β in their models; apart from other factors, for a closer agreement with observations β should be explicitly included.

Since the envelope evolves independently of the core (each with a fixed mass), the expansion of the envelope must lower the local density. Since β is also decreasing in the envelope, the temperature near q_s actually does not change very much (see Fig. 6). Beyond the shell, however, it [illegible] considerably, because the temperature gradient is approaching the adiabatic value. This increase of the temperature gradient has two major effects: (1) The semiconvective zone moves steadily inward; by Model G6 it has reached $q = 0.48$. The hydrogen discontinuity grows to $X_2 - X_3 = 0.10$, although $\mu_2/\mu_3 = 0.91$ because the actual hydrogen content is lower at small q. (2). The shell source becomes narrower, on account of the decreasing temperature outside (and the decreasing hydrogen content inside). A curious result, not occurring in stars of lower mass, is that q_s moves slightly inward at first because of the rising temperature inside and the rather broad shell distribution of hydrogen, but thereafter remains stationary as it encounters less and less hydrogen. The total amount of hydrogen depletion (in the shell) is 0.0004, and the maximum local change in X is 0.011 at $q = 0.329$.

In the core, although β_c steadily rises, the central convection persists and recedes only to a roughly constant $q_4 = 0.06$ at the onset of helium burning. Hence convection probably never disappears completely from the center of very massive stars.

Model G6 corresponds to the onset of helium burning. At this point, $L_{He}/L \approx 0.004$ and T_c is 150 million degrees. This value of T_c is almost the same as the corresponding value for a star of $15.6M_\odot$ because of the steep temperature dependence of the energy-generation formula. However, ρ_c is only 270 gm/cm^3, lower by a factor of 10 than in the star of $15.6M_\odot$. But in the heavier star the core mass is four times as large.

The gravitational contraction phase is very short (9×10^3 years), although large changes in the stellar radius take place.

VI. FINAL REMARKS

Although the computed values of the *fractional* luminosities should not be considered exact (since we have neglected in the computations to take account of the radiative flux that is absorbed in the expansion of the envelope), nevertheless Figure 8 shows up a brief peak in the luminosity due to gravitational contraction, whereas the nuclear-energy sources are spread over longer time intervals. Since the total energy available to the star as a result of gravitational contraction is small compared with the nuclear-energy store, it is expected that the time scale of contraction phases will be very short. Indeed, the post–main-sequence age at the onset of helium burning is 4.9 million years, of which 4.8 million are spent burning almost half of the initial hydrogen content.

During this time the interior structure of the star undergoes drastic changes, especially in the

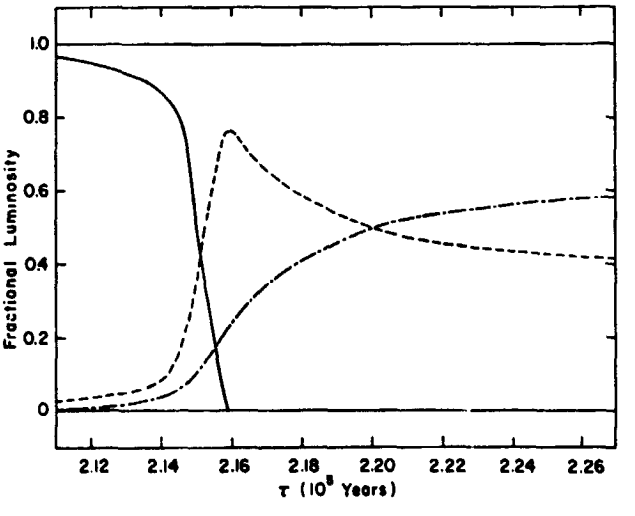


FIGURE 8.—The fraction of the total luminosity contributed by hydrogen burning in the core *(solid curve)*, gravitational contraction in the core *(dashed curve)*, and hydrogen burning in the shell *(dash-dot-curve)* is plotted against the time elapsed since the last model of the hydrogen-burning phase (Model H4 of Paper I).

short (0.1×10^6 years) contraction phase, as shown in Figure 9. One complication not occurring in stars of lower mass may arise from the existence of the semiconvective zone. That is, the inward-moving semiconvection brings hydrogen-enriched material toward the shell source, which is simultaneously moving outward. A hydrogen flare-up might be expected to occur. However, the results of Table 3 show that ΔX is only about 0.1 at the radiative-semiconvective interface. Moreover, from the results on a star of $15.6 M_\odot$ (Hayashi and Cameron 1962; Hayashi, Hoshi, and Sugimoto 1962), the total luminosity is expected to increase little during helium burning, so that the structure of the envelope (and hence of the semiconvective zone) will change very slowly; the outward motion of q_s is also rather slow. Thus the star could probably accommodate quietly the sharp increase in X, despite the narrowness of the shell source. However, it seems unlikely that the inner edge of the semiconvective zone will ever penetrate the shell before the whole envelope becomes convective during the rightward swing of the evolutionary track in the H–R diagram toward red carbon-burning models. We need only worry about possible hydrogen flaring in stars of mass greater than $30 M_\odot$.

The evolution on the theoretical H–R diagram is shown in Figure 10 for the phases considered in this paper. The ages are given in units of 10^4 years. At $\tau = 9.07$ the spectral class becomes B0, and at $\tau = 9.67$ it is B3. The star is expected never to return to the region of O stars during its active life.

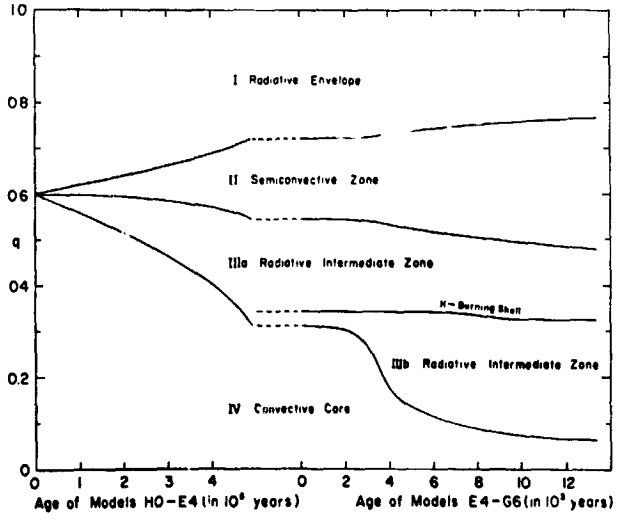


FIGURE 9.—Evolution of the structural zones from the initial main sequence to the onset of helium burning.

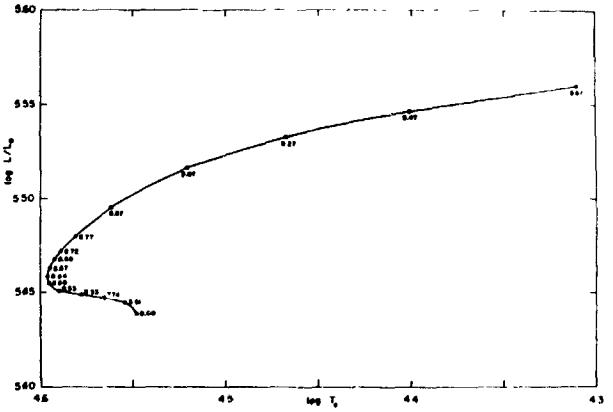


FIGURE 10.—Evolution of a star of 30 M⊙ on a plot of luminosity versus effective temperature, during the phases of hydrogen exhaustion and gravitational contraction. Numbers attached to the models *(filled circles)* represent the age in units of 10^4 years.

REFERENCES

HASELGROVE, C. B., and HOYLE, F. 1956, *M.N.*, **116**, 515.

HAYASHI, C., and CAMERON, R. C. 1962, *Ap. J.*, **136**, 166.

HAYASHI, C., HOSHI, R., and SUGIMOTO, D. 1962, *Prog. Theoret. Phys. Suppl.* (Kyoto), No. 22.

HAYASHI, C., NISHIDA, M., and SUGIMOTO, D. 1962, *Prog. Theoret. Phys.* (Kyoto), **27**, 1233.

HOFMEISTER, E., KIPPENHAHN, R. and WEIGERT, A. 1963, Stellar Evolution Conference (Goddard Institute for Space Studies, New York).

HOYLE, F. 1960, *M.N.*, **120**, 22.

KUSHWAHA, R. S. 1957, *Ap. J.*, **125**, 242.

POLAK, E. J. 1962, *Ap. J.*, **136**, 465.

REEVES, H. 1962, *Stellar Energy Sources* (Goddard Institute for Space Studies, NASA Report).

REIZ, A. 1963, Stellar Evolution Conference (Goddard Institute for Space Studies, New York).

SAKASHITA, S., ÔNO, Y., and HAYASHI, C. 1959, *Prog. Theoret. Phys.* (Kyoto), **21**, 315.

SCHWARZSCHILD, M. 1958, *Structure and Evolution of the Stars* (Princeton, N.J.: Princeton University Press).

SCHWARZSCHILD, M., and HARM, R. 1958, *Ap. J.*, **128**, 348.

SCHWARZSCHILD, M. and SELBERG, H. 1962, *Ap. J.*, **136**, 150.

STOTHERS, R. 1963, *Ap. J.*, **138**, 1074 ("Paper I").

N64-33451

THE RV TAURI POPULATION*

RICHARD STOTHERS

Our understanding of the RV Tauri stars has come mainly from photometric and spectrographic studies based on several tabulations of these stars since 1921.[1–9] Although members of this group are too distant for accurate parallax derterminations, their absolute magnitudes are known through the discovery and identification of some in globular clusters.[5,10] It is the purpose of this paper to analyze the spatial distribution of the RV Tauri stars in the field and to compare properties of these stars with properties of similar stars found in globular clusters.

The data are taken mainly from the *General Catalogue of Variable Stars*[8] (*GCVS*), which lists 92 stars of this class based on criteria that are essentially those of Payne-Gaposchkin, Brenton, and Gaposchkin.[4] RV Tauri stars are supergiants with (1) comparatively stable periodicity of light variation, whose amplitude may reach 3 magnitudes, (2) alternating shallow and deep minima, which occasionally interchange, (3) double-period of 30–150 days, and (4) F to K spectral class, which is earliest near maximum light. From the *GCVS* list we disregard those variables designated RV? and include only stars with periods in the range 39–110 days, since this is roughly the range of RV Tauri stars in globular clusters. Although some types of red variables in the same period range have light curves resembling those of RV Tauri stars, we assume that those variables without known spectra are yellow. In some cases photographic magnitudes were obtained from other sources.[4,6] Since the magnitude of CN Centauri is uncertain at minimum, it has been omitted from the distance determinations. Preston, Krzeminski, Smak, and Williams have shown that UU Herculis and V453 Ophiuchi are RV Tauri stars.[9] Apart from these two variables, their list of RV Tauri stars does not otherwise conflict with assignments given by the *GCVS*, in the period range 39-110 days. We have used their magnitudes and period for V 453 Ophiuchi. The total number of variables whose distances we may determine is 45, of which 25 have known spectra.

Preston *et al.*[9] have defined three groups of RV Tauri stars on spectroscopic criteria: A.—All spectral features indicate type G or K, except that TiO bands may occur during deep light minima. B.—The type based on the Ca II lines differs from that based on the hydrogen lines; strong CN bands occur at and around light minima. C.—The spectrum resembles that group B, except that CN bands never appear. Moreover, the stars in group A are redder than those in groups B and C (which cannot be distinguished photometrically). The RV Tauri stars may be distinguished from the yellow semiregulars on the basis of U,B,V colors and strength of hydrogen emission. At the end of this paper we shall comment on these subdivisions.

The distribution of RV Tauri stars in longitude is far from scraggly, contrary to the results indicated by the smaller numer of stars used in earlier investigations.[7] Indeed, Figure 1 shows a strong concentration of stars in the direction of the galactic center. Their spatial concentration depends on a knowledge of distances, which we may obtain from globular clusters.

In general, RV Tauri stars in globular clusters are specified by the same criteria as those in the field. Small deviations from these criteria have been discussed by Rosino[5] and Joy.[6] Other differences will be pointed out below. Using the mean photographic magnitudes of the nine certain RV Tauri stars in clusters (from the writer's previous paper,[10] hereafter called Paper I), we

*Published in the *Publications of the Astronomical Society of the Pacific*, 76 (449): 98–105, April 1964.

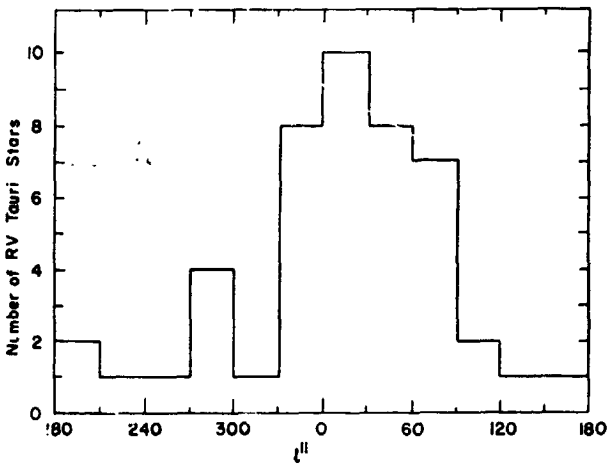


FIGURE 1.—Frequency distribution of the field RV Tauri stars in longitude.

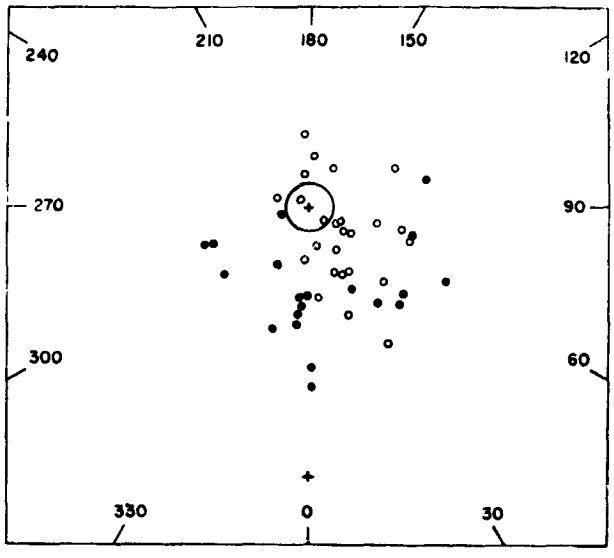


FIGURE 2.—Field RV Tauri stars with known spectra (open circles) and unknown spectra (filled circles) are projected on the galactic plane. The positions of the sun and galactic center are denoted by + signs. The circle around the sun has a radius of 1 kpc.

may obtain a period-luminosity relation, if the distance moduli of the clusters are known. We take the moduli from Hogg's list,[11] where she assumed that $M_{pg}=0.0$ for the RR Lyrae stars. For M 2, M 5, and M 13, we have used the values of M_v of the RR Lyrae stars corrected for reddening listed by Arp[12] and have added C.I. $=+0.1$. For M 22, which closely resembles M 13,[13,14] and for ω Centauri, which has a main sequence turnoff at the same magnitude as M 3[15] (also listed by Arp), we assume analogous values of M_{pg}. For M 56 and NGC 6712 we adopt the mean of all the known values, $M_{pg}=+0.3$. Applying these corrections to Hogg's distance moduli, we obtain absolute magnitudes of the cluster RV Tauri stars. The period-luminosity relation is then $<M_{pg}>=-4.0+0.026P\pm0.36$, where P is in days, in reasonable agreement with previous determinations which depended on indirect or statistical methods.[5,6,7,17,18,19,20,21]

To calculate the distance of the RV Tauri stars in the field, we use

$$<M_{pg}>=<m_{pg}>-10-5\log r-a$$

where r is in kpc and $a=0.7r$ if $|b^{II}|<20°$ or $0.25\csc|b^{II}|$ if $|b^{II}|>20°$. The first formula for extinction is used by Payne-Gaposchkin[16] and also happens to be the mean of extinctions derived for five RV Tauri stars by Kameny.[7]

Figure 2 shows the RV Tauri stars projected on the galactic plane. Variables with known spectra are indicated by open circles, those with unknown spectra by filled circles; the crosses denote the positions of the sun and the galactic center. It is difficult to corroborate the previously suspected[7] groupings near Ophiuchus (30°), Aquila (40°), and Gemini (180°), on the meager data available and the uncertain extinction. Other clumpings at 60° and 75° are also suspect, although the Sagittarius group (0°) is probably real. There is no apparent correlation of period with position in the Galaxy.

Figure 3 shows a vertical cross-section of the Galaxy, where the sun is assumed to lie at 10 kpc from the galactic center. As indicated also by Figures 1 and 2, the number of variables increases in the direction of the galactic center. Their apparent absence at low galactic heights is due to interstellar extinction. The distribution strongly resembles that of the long-period variables with $P<250$ days, that is, intermediate between the distribution of long-period variables with $P>250$ days (Population I) and the distribution of the halo RR Lyrae stars (Population II).[18] The vertical density gradient y, where $\log N=x-y$ (r sin b) and N is the relative number of stars, is 1.1, and the median height $\overline{r \sin b}$ is 0.45 kpc. Both these quantities are similar to those for the long-period variables with $P<250$ days.

Perepelkina suspected that the RV Tauri stars form an intermediate subsystem.[17] Their relatively low radial velocities and latitudes indicate

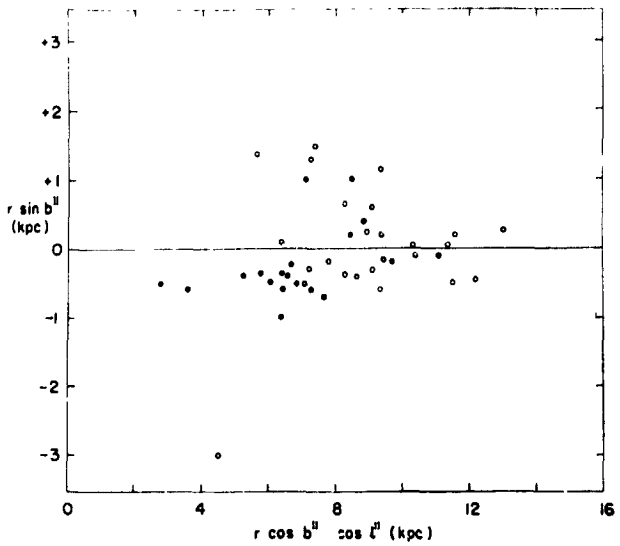


FIGURE 3.—Vertical cross-section of the Galaxy, showing the distribution of field RV Tauri stars on the assumption that the sun lies at 10 kpc from the galactic center. Same symbols as in Figure 2.

a weak Population I, but their excessive blueness (like the W Virginis stars) and discontinuous velocity curves indicate Population II.[7] The yellow semiregulars, however, seem to be real Population II objects.

The general similarity of the RV Tauri stars in globular clusters to those in the field and the approximately equal ratios of number of RV Tauri stars to number of RR Lyrae stars[7] indicate that field and cluster RV Tauri stars belong to the same family. At maximum light the RV Tauri stars are usually the brightest cluster members and are found close to the cluster center (compared with the long-period variables).[10] Therefore, because of the long relaxation times of globular clusters, it seems unlikely that the RV Tauri stars in the field have evaporated from clusters, even apart from the observation that they do not form a halo population.

Besides their spatial distribution, some other differences between the field and cluster variables should be mentioned. While Rosino found no period-spectrum relation for the variables in the field,[8] Paper I showed such a relation for the cluster variables, although the stars seem to be separated chiefly into a 60- and a 90-day group. Kameny[7] found that the short-period RV Tauri stars in the clusters were slightly bluer than those in the field. No very long-period RV Tauri stars have yet been identified in globular clusters. Moreover, in the clusters there is a marked absence of RV Tauri stari stars in the period range 68–87 days. We note in Table I an actual anticorrelation of period frequency with the corresponding field variables. Lastly, we mention that

TABLE I.—*Period Frequencies of the RV Tauri Stars*

P (days)	Globular Clusters		Field	
	RV	All Variables	RV (with Sp.)	RV (all)
28–47	0	1	3	5
48–67	5	7	6	15
68–87	0	2	15	23
88–107	4	13	3	5
>107	0	16	5	24

no RV Tauri stars in globular clusters have light amplitudes exceeding 2 magnitudes, whereas about one-third of the field stars in the same period range have greater amplitudes. All these differences may be due to different ages and chemical compositions.

The results of work by Preston *et al.*[9] indicate that, at least spectroscopically, the RV Tauri stars do not comprise a homogeneous class. Their group A of these stars shows TiO at minimum and resembles kinematically an intermediate (disk) population.[6] Group B is probably also related kinematically to the disk, and we note that it, too, shows the presence of metals in the strong CN bands. The radial velocities of group C stars are very high and suggest a halo population; the absence of CN is also suggestive. We note that the two variables from globular clusters that Preston *et al.* studied probably belong to the "halo" group C; it is interesting that, whereas groups A and B show a wide range in period, the three field variables from group C have periods lying in the "forbidden" range of the clusters.

If the RV Tauri stars plotted on the vertical cross-section of the Galaxy (Fig. 3) are distinguished according to groups, then it appears that groups B and C adhere very well to the above galactic assignments, whereas group A populates

both the disk and the halo. However, it is clear that RV Tauri stars do not really range far enough to separate groups into halo and disk populations. Therefore the suggestion by Preston *et al.* that groups B and C belong to a larger family raises the further speculation that all the groups form really one family. At least this has been indicated by previous work.

In this connection, we note that there is no apparent relation between spectroscopic criteria and shape of light curve or period. Moreover, the colors appear to form a continuous sequence, despite the spectroscopic differences. Although Preston *et al.* could not obtain reliable luminosity classes at the dispersion they used, Rosino's data[5] suggest that the luminosity class falls with increasing period, in agreement with the results on variables in globular clusters.[10] Comparison of the general period-luminosity relations for the variables in globular clusters and those in the field suggests that any difference in the luminosity between the cluster and field RV Tauri stars does not exceed 1 mag. (Paper I, Fig. 11). We believe that the statistical identification of luminosities in the clusters and in the field is sufficient to outline the general spatial distribution of RV Tauri stars, as considered in this paper.

In conclusion, the RV Tauri stars seems to form an intermediate subsystem between Population I and II. Except for the underabundance of metals and the period-frequency anomaly, the RV Tauri stars in globular clusters resemble their counterparts in the field. Perhaps long-period variables in clusters with late integrated spectra should be investigated for RV Tauri behavior; in this connection we note the 105-day variable in NGC 6712 (GA). Certainly the known semiregular and irregular variables with undetermined periods should be checked for periods in the range 68–87 days.

REFERENCES

[1] H. Ludendorff, *A.N.*, **214**, 217, 1921.

[2] H. Ludendorff, in *Handbuch der Astrophysik*, Vol. 6, G. Eberhard, A. Kohlschütter, and H. Ludendorff, eds. (Berlin: Springer 1928), p. 174.

[3] B. P. Gerasimovic, *Harvard Obs. Circ.* No. 341, 1929.

[4] C. H. Payne-Gaposchkin, V. K. Brenton, and S. Gaposchkin, *Harvard Obs. Ann.*, **113**, 1; 1943.

[5] L. Rosino, *Ap. J.*, **113**, 60, 1951.

[6] A. H. Joy, *Ap. J.*, **115**, 25, 1952.

[7] F. E. Kameny, Dissertation, Harvard University, 1956.

[8] B. V. Kukarkin, P. P. Parenago, Yu. I. Efremov, and P. N. Kholopov, *General Catalog of Variable Stars* (Moscow: Academy of Sciences, U.S.S.R. Press, 1958).

[9] G. W. Preston, W. Krzeminski, J. Smak, and J. A. Williams, *Ap. J.*, **137**, 401, 1963.

[10] R. Stothers, *A. J.*, **68**, 242, 1963.

[11] H. S. Hogg, in *Handbuch der Physik*, Vol. 53, S. Flügge, ed. (Berlin: Springer 1959), p. 129.

[12] H. Arp, *Ap. J.* **135**, 971, 1962.

[13] H. C. Arp and W. G. Melbourne, *A. J.*, **64**, 28, 1959.

[14] A. Sandage and G. Wallerstein, *Ap. J.*, **131**, 598, 1960.

[15] E. P. Belserene, *A. J.*, **64**, 58, 1959.

[16] C. H. Payne-Gaposchkin, *Variable Stars and Galactic Structure* (London: Athlone Press, 1954).

[17] E. D. Perepelkina, *Perem. Zvezdy* (*Variable Stars*), **7**, 230, 1950.

[18] P. P. Parenago, *Astr. Zhurnal U.S.S.R.*, **11**, 95, 1934.

[19] A. H. Joy, *Pub. A. S. P.*, **46**, 51, 1934.

[20] W. S. Adams, A. H. Joy, M. L. Humason, and A. M. Brayton, *Ap. J.*, **81**, 187, 1935.

[21] D. McLaughlin, *Pub. Michigan Obs.*, **7**, 57, 1938.

II. CELESTIAL MECHANICS AND GEODESY

LONG-PERIOD CONTRIBUTIONS TO THE DISTURBING FUNCTIONS OF THE EARTH FROM THE SEVENTH, NINTH, AND ELEVENTH ZONAL HARMONICS*

T. L. FELSENTREGER AND W. J. WICKLESS

INTRODUCTION

It is the purpose of this paper to present explicit formulas for the long-period terms due to the seventh, ninth, and eleventh zonal harmonics in the disturbing function of the earth in the case of an artificial earth satellite. The formulas are given for terms of the satellite's orbital elements and the Delaunay variables. G. Giacaglia (1) has given general expressions for the long-period terms due to any of the zonal harmonics, which can be expressed in terms of the orbital elements. The apparent differences between the results of this paper and those in Giacaglia's have been verified as due to the errors in the latter as it appears in the A.J.

The contributions of the long-period terms to the mean motion of the argument of perigee are also given.

THE DISTURBING FUNCTION

The earth's gravitational potential at a distance r from the center of the earth is

$$U=\frac{\mu}{r}\left[1-\sum_{n=2}^{\infty}\left(\frac{R}{r}\right)^{n}J_nP_n\ (\sin\phi)\right],$$

where

$\mu = GM$

G = gravitational constant

M = mass of the earth

R = radius of the earth

J_n = zonal harmonic coefficients ($n+2$, 3,...)

P_n = Legendre polynomials ($n=2$, 3,...)

ϕ = geocentric latitude.

Here, the earth's radius is adopted as the unit of length. The seventh, ninth, and eleventh Legendre polynomials are

$$P_7\ (\sin\phi)=\frac{1}{16}(429\sin^7\phi-693\sin^5\phi+315\sin^3\phi-35\sin\phi)$$

$$P_9\ (\sin\phi)=\frac{1}{128}(12155\sin^9\phi-25740\sin^7\phi+18018\sin^5\phi-4620\sin^3\phi+315\sin\phi)$$

$$P_{11}\ (\sin\phi)=\frac{1}{256}(88179\sin^{11}\phi-230945\sin^9\phi+218790\sin^7\phi-90090\sin^5\phi+15015\sin^3\phi-693\sin\phi).$$

*Published as *Goddard Space Flight Center Document X-547-64-333*, August 1964.

Let

a = semi-major axis of satellite's orbit

e = eccentricity of orbit

i = inclination of orbital plane to equatorial plane

ℓ = mean anomaly

f = true anomaly

g = argument of perigee.

The Delaunay variables L, G, and H are

$$L=\sqrt{\mu a}$$

$$G=L\sqrt{1-e^2}$$

$$H=G\cos i.$$

Use is also made of the relations

$$\sin\phi=\sin i\sin(f+g)$$

$$\frac{a}{r}=\frac{L^2}{G^2}(1+e\cos f).$$

The long-period terms in the expansion of U as a Fourier series in ℓ and g are given by

$$\frac{1}{2\pi}\int_0^{2\pi} U\,d\ell, \qquad \text{(see reference 2)}$$

making use of the relation

$$d\ell=\frac{L}{G}\,\frac{r^2}{a^2}df.$$

Denoting the long-period parts of U_7, U_9, and U_{11} by Δ_7F_{2p}, Δ_9F_{2p}, and $\Delta_{11}F_{2p}$, respectively, we have

$$\Delta_7F_{2p}=-\frac{21\mu^9J_7e\sin i}{16384L^3G^{13}}\left[10\left(5-135\frac{H^2}{G^2}+495\frac{H^4}{G^4}-429\frac{H^6}{G^6}\right)\left(33-30\frac{G^2}{L^2}+5\frac{G^4}{L^4}\right)\sin g\right.$$

$$-15\left(3-69\frac{H^2}{G^2}+209\frac{H^4}{G^4}-143\frac{H^6}{G^6}\right)\left(11-14\frac{G^2}{L^2}+3\frac{G^4}{L^4}\right)\sin 3g$$

$$\left.+33\left(1-15\frac{H^2}{G^2}+27\frac{H^4}{G^4}-13\frac{H^6}{G^6}\right)\left(1-2\frac{G^2}{L^2}+\frac{G^4}{L^4}\right)\sin 5g\right]$$

$$\Delta_9F_{2p}=-\frac{3\mu^{11}J_9e\sin i}{524288L^3G^{17}}\left[210\left(7-308\frac{H^2}{G^2}+2002\frac{H^4}{G^4}-4004\frac{H^6}{G^6}+2431\frac{H^8}{G^8}\right)\times\right.$$

$$\times\left(7-1001\frac{G^2}{L^2}+385\frac{G^4}{L^4}-35\frac{G^6}{L^6}\right)\sin g-10780\left(1-40\frac{H^2}{G^2}+234\frac{H^4}{G^4}-416\frac{H^6}{G^6}+221\frac{H^8}{G^8}\right)\times$$

$$\times\left(39-65\frac{G^2}{L^2}+29\frac{G^4}{L^4}-3\frac{G^6}{L^6}\right)\sin 3g+12012\left(1-32\frac{H^2}{G^2}+146\frac{H^4}{G^4}-200\frac{H^6}{G^6}+85\frac{H^8}{G^8}\right)\times$$

$$\times\left(5-11\frac{G^2}{L^2}+7\frac{G^4}{L^4}-\frac{G^6}{L^6}\right)\sin 5g-2145\left(1-20\frac{H^2}{G^2}+54\frac{H^4}{G^4}-52\frac{H^6}{G^6}+17\frac{H^8}{G^8}\right)\times$$

$$\left.\times\left(1-3\frac{G^2}{L^2}+3\frac{G^4}{L^4}-\frac{G^6}{L^6}\right)\sin 7g\right]$$

$$\Delta_{11}F_{2p}=\frac{11\mu^{13}J_{11}e\sin i}{67108864L^3G^{21}}\Big[126\Big(21-1365\frac{H^2}{G^2}+13650\frac{H^4}{G^4}-46410\frac{H^6}{G^6}+29393\frac{H^8}{G^8}-23933\frac{H^{10}}{G^{10}}\Big)\times$$

$$\times\Big(41423-77292\frac{G^2}{L^2}+45738\frac{G^4}{L^4}-8652\frac{G^6}{L^6}+63\frac{G^8}{L^8}\Big)\sin g$$

$$-163800\Big(1-61\frac{H^2}{G^2}+570\frac{H^4}{G^4}-1802\frac{H^6}{G^6}+2261\frac{H^8}{G^8}-969\frac{H^{10}}{G^{10}}\Big)\times\Big(323-680\frac{G^2}{L^2}+462\frac{G^4}{L^4}-112\frac{G^6}{L^6}+7\frac{G^8}{L^8}\Big)\sin 3g$$

$$+7020\Big(5-265\frac{H^2}{G^2}+2130\frac{H^4}{G^4}-5746\frac{H^6}{G^6}+6137\frac{H^8}{G^8}-2261\frac{H^{10}}{G^{10}}\Big)\times\Big(323-816\frac{G^2}{L^2}+678\frac{G^4}{L^4}-200\frac{G^6}{L^6}+15\frac{G^8}{L^8}\Big)\sin 5g$$

$$-49725\Big(1-41\frac{H^2}{G^2}+250\frac{H^4}{G^4}-514\frac{H^6}{G^6}+437\frac{H^8}{G^8}-133\frac{H^{10}}{G^{10}}\Big)\times\Big(19+60\frac{G^2}{L^2}+66\frac{G^4}{L^4}-28\frac{G^6}{L^6}-3\frac{G^8}{L^8}\Big)\sin 7g$$

$$+20995\Big(1-25\frac{H^2}{G^2}+90\frac{H^4}{G^4}-130\frac{H^6}{G^6}+85\frac{H^8}{G^8}-21\frac{H^{10}}{G^{10}}\Big)\times\Big(1-4\frac{G^2}{L^2}+6\frac{G^4}{L^4}-4\frac{G^6}{L^6}+\frac{G^8}{L^8}\Big)\sin 9g\Big]$$

CONTRIBUTIONS TO dg/dt

Since the Delaunay set of variables is canonical with respect to the Hamiltonian F, which includes Δ_7F_{2p}, Δ_9F_{2p}, and $\Delta_{11}F_{2p}$, we have

$$\frac{dg}{dt}=-\frac{\partial F}{\partial G}\quad\text{(see reference 2).}$$

Therefore, a computation of $\partial(\Delta_iF_{2p})/\partial g(i=7,9,11)$ provides the long-period terms in dg/dt due to the seventh, ninth, and eleventh zonal harmonics.
The results are

$$\frac{\partial(\Delta_7F_{2p})}{\partial G}=\frac{21\mu^9J_7}{16384L^3G^{14}e\sin i}\Big\{10\Big[(\sin^2 i-e^2)\Big(5-135\frac{H^2}{G^2}+495\frac{H^4}{G^4}-429\frac{H^6}{G^6}\Big)\Big(33-30\frac{G^2}{L^2}+5\frac{G^4}{L^4}\Big)$$

$$+e^2\sin^2 i\Big(65-2025\frac{H^2}{G^2}+8415\frac{H^4}{G^4}-8151\frac{H^6}{G^6}\Big)\Big(33-30\frac{G^2}{L^2}+5\frac{G^4}{L^4}\Big)$$

$$+20e^2\sin^2 i\Big(5-135\frac{H^2}{G^2}+495\frac{H^4}{G^4}-429\frac{H^6}{G^6}\Big)\frac{G^2}{L^2}\Big(3-\frac{G^2}{L^2}\Big)\Big]\sin g$$

$$-15\Big[(\sin^2 i-e^2)\Big(3-69\frac{H^2}{G^2}+209\frac{H^4}{G^4}-143\frac{H^6}{G^6}\Big)\Big(11-14\frac{G^2}{L^2}+3\frac{G^4}{L^4}\Big)+e^2\sin^2 i\Big(39-1035\frac{H^2}{G^2}+3553\frac{H^4}{G^4}$$

$$-2717\frac{H^6}{G^6}\Big)\Big(11-14\frac{G^2}{L^2}+3\frac{G^4}{L^4}\Big)+4e^2\sin^2 i\Big(3-69\frac{H^2}{G^2}+209\frac{H^4}{G^4}-143\frac{H^6}{G^6}\Big)\frac{G^2}{L^2}\Big(7-3\frac{G^2}{L^2}\Big)\Big]\sin 3g$$

$$+33\Big[(\sin^2 i-e^2)\Big(1-15\frac{H^2}{G^2}+27\frac{H^4}{G^4}-13\frac{H^6}{G^6}\Big)\Big(1-2\frac{G^2}{L^2}+\frac{G^4}{L^4}\Big)+e^2\sin^2 i\Big(13-225\frac{H^2}{G^2}+459\frac{H^4}{G^4}$$

$$-247\frac{H^6}{G^6}\Big)\Big(1-2\frac{G^2}{L^2}+\frac{G^4}{L^4}\Big)+4e^2\sin^2 i\Big(1-15\frac{H^2}{G^2}+27\frac{H^4}{G^4}-13\frac{H^6}{G^6}\Big)\frac{G^2}{L^2}\Big(1-\frac{G^2}{L^2}\Big)\Big]\sin 5g\Big\}$$

$$\frac{\partial(\Delta_9 F_{2p})}{\partial G}=\frac{3\mu^{11}J_9}{524288L^3G^{18}e\sin i}\Bigg\{210\Bigg[(\sin^2 i-e^2)\left(7-308\frac{H^2}{G^2}+2002\frac{H^4}{G^4}-4004\frac{H^6}{G^6}+2431\frac{H^8}{G^8}\right)$$

$$\times\left(715-1001\frac{G^2}{L^2}+385\frac{G^4}{L^4}-35\frac{G^6}{L^6}\right)+e^2\sin^2 i\left(119-5852\frac{H^2}{G^2}+42042\frac{H^4}{G^4}-92092\frac{H^6}{G^6}+60775\frac{H^8}{G^8}\right)\left(715\right.$$

$$\left.-1001\frac{G^2}{L^2}+385\frac{G^4}{L^4}-35\frac{G^6}{L^6}\right)+14e^2\sin^2 i\left(7-308\frac{H^2}{G^2}+2002\frac{H^4}{G^4}-4004\frac{H^6}{G^6}\right.$$

$$\left.+2431\frac{H^8}{G^8}\right)\frac{G^2}{L^2}\left(143-110\frac{G^2}{L^2}+15\frac{G^4}{L^4}\right)\Bigg]\sin g$$

$$-10780\Bigg[(\sin^2 i-e^2)\left(1-40\frac{H^2}{G^2}+234\frac{H^4}{G^4}-416\frac{H^6}{G^6}+221\frac{H^8}{G^8}\right)\left(39-65\frac{G^2}{L^2}+29\frac{G^4}{L^4}-3\frac{G^6}{L^6}\right)$$

$$+e^2\sin^2 i\left(17-760\frac{H^2}{G^2}+4914\frac{H^4}{G^4}-9568\frac{H^6}{G^6}+5525\frac{H^8}{G^8}\right)\left(39-65\frac{G^2}{L^2}+29\frac{G^4}{L^4}-3\frac{G^6}{L^6}\right)$$

$$+2e^2\sin^2 i\left(1-40\frac{H^2}{G^2}+234\frac{H^4}{G^4}-416\frac{H^6}{G^6}+221\frac{H^8}{G^8}\right)\frac{G^2}{L^2}\left(65-58\frac{G^2}{L^2}+9\frac{G^4}{L^4}\right)\Bigg]\sin 3g$$

$$+12012\Bigg[(\sin^2 i-e^2)\left(1-32\frac{H^2}{G^2}+146\frac{H^4}{G^4}-200\frac{H^6}{G^6}+85\frac{H^8}{G^8}\right)\left(5-11\frac{G^2}{L^2}+7\frac{G^4}{L^4}-\frac{G^6}{L^6}\right)+e^2\sin^2 i\left(17-608\frac{H^2}{G^2}\right.$$

$$\left.+3066\frac{H^4}{G^4}-4600\frac{H^6}{G^6}+2125\frac{H^8}{G^8}\right)\left(5-11\frac{G^2}{L^2}+7\frac{G^4}{L^4}-\frac{G^6}{L^6}\right)+2e^2\sin^2 i\left(1-32\frac{H^2}{G^2}+146\frac{H^4}{G^4}-200\frac{H^6}{G^6}\right.$$

$$\left.+85\frac{H^8}{G^8}\right)\frac{G^2}{L^2}\left(11-14\frac{G^2}{L^2}+3\frac{G^4}{L^4}\right)\Bigg]\sin 5g$$

$$-2145\Bigg[(\sin^2 i-e^2)\left(1-20\frac{H^2}{G^2}+54\frac{H^4}{G^4}-52\frac{H^6}{G^6}+17\frac{H^8}{G^8}\right)\left(1-3\frac{G^2}{L^2}+3\frac{G^4}{L^4}-\frac{G^6}{L^6}\right)+e^2\sin^2 i\left(17-380\frac{H^2}{G^2}\right.$$

$$\left.+1134\frac{H^4}{G^4}-1196\frac{H^6}{G^6}+425\frac{H^8}{G^8}\right)\left(1-3\frac{G^2}{L^2}+3\frac{G^4}{L^4}-\frac{G^6}{L^6}\right)+6e^2\sin^2 i\left(1-20\frac{H^2}{G^2}+54\frac{H^4}{G^4}-52\frac{H^6}{G^6}\right.$$

$$\left.+17\frac{H^8}{G^8}\right)\frac{G^2}{L^2}\left(1-2\frac{G^2}{L^2}+\frac{G^4}{L^4}\right)\Bigg]\sin 7g\Bigg\}$$

$$\frac{\partial(\Delta_{11}F_{2p})}{\partial G}=\frac{11\mu^{13}J_{11}}{67108864L^3G^{22}e\sin i}\Bigg\{126\Bigg[(\sin^2 i-e^2)\left(21-1365\frac{H^2}{G^2}+13650\frac{H^4}{G^4}-46410\frac{H^6}{G^6}+62985\frac{H^8}{G^8}\right.$$

$$\left.-29393\frac{H^{10}}{G^{10}}\right)\left(41423-77292\frac{G^2}{L^2}+45738\frac{G^4}{L^4}-8652\frac{G^6}{L^6}+63\frac{G^8}{L^8}\right)+e^2\sin^2 i\left(441-31395\frac{H^2}{G^2}+341250\frac{H^4}{G^4}\right.$$

$$\left.-1253070\frac{H^6}{G^6}+1826565\frac{H^8}{G^8}-911183\frac{H^{10}}{G^{10}}\right)\times\left(41423-77292\frac{G^2}{L^2}+45738\frac{G^4}{L^4}-8652\frac{G^6}{L^6}+63\frac{G^8}{L^8}\right)$$

$$+72e^2\sin^2 i\left(21-1365\frac{H^2}{G^2}+13650\frac{H^4}{G^4}-46410\frac{H^6}{G^6}+62985\frac{H^8}{G^8}-29393\frac{H^{10}}{G^{10}}\right)\times\frac{G^2}{L^2}\left(2147\right.$$

$$\left.-2541\frac{G^2}{L^2}+721\frac{G^4}{L^4}-7\frac{G^6}{L^6}\right)\Bigg]\sin g$$

$$-163800\left[(\sin^2 i-e^2)\left(1-61\frac{H^2}{G^2}+570\frac{H^4}{G^4}-1802\frac{H^6}{G^6}+2261\frac{H^8}{G^8}-969\frac{H^{10}}{G^{10}}\right)\times\left(323-680\frac{G^2}{L^2}+462\frac{G^4}{L^4}\right.\right.$$
$$\left.-112\frac{G^6}{L^6}+7\frac{G^8}{L^8}\right)+e^2\sin^2 i\left(21-1403\frac{H^2}{G^2}+14250\frac{H^4}{G^4}-48654\frac{H^6}{G^6}+65569\frac{H^8}{G^8}-30039\frac{H^{10}}{G^{10}}\right)$$
$$\times\left(323-680\frac{G^2}{L^2}+462\frac{G^4}{L^4}-112\frac{G^6}{L^6}+7\frac{G^8}{L^8}\right)+8e^2\sin^2 i\left(1-61\frac{H^2}{G^2}+570\frac{H^4}{G^4}-1802\frac{H^6}{G^6}+2261\frac{H^8}{G^8}\right.$$
$$\left.\left.-969\frac{H^{10}}{G^{10}}\right)\times\frac{G^2}{L^2}\left(170-231\frac{G^2}{L^2}+84\frac{G^4}{L^4}-7\frac{G^6}{L^6}\right)\right]\sin 3g$$
$$+7020\left[(\sin^2 i-e^2)\left(5-265\frac{H_2}{G^2}+2130\frac{H^4}{G^4}-5746\frac{H^6}{G^6}+6137\frac{H^8}{G^8}-2261\frac{H^{10}}{G^{10}}\right)\times\left(323-816\frac{G^2}{L^2}+678\frac{G^4}{L^4}\right.\right.$$
$$\left.-200\frac{G^6}{L^6}+15\frac{G^8}{L^8}\right)+e^2\sin^2 i\left(105-6095\frac{H^2}{G^2}+53250\frac{H^4}{G^4}-155142\frac{H^6}{G^6}+177973\frac{H^8}{G^8}-70091\frac{H^{10}}{G^{10}}\right)$$
$$\times\left(323-816\frac{G^2}{L^2}+678\frac{G^4}{L^4}-200\frac{G^6}{L^6}+15\frac{G^8}{L^8}\right)+24e^2\sin^2 i\left(5-265\frac{H^2}{G^2}+2130\frac{H^4}{G^4}-5746\frac{H^6}{G^6}+6137\frac{H^8}{G^8}\right.$$
$$\left.\left.-2261\frac{H^{10}}{G^{10}}\right)\times\frac{G^2}{L^2}\left(68-113\frac{G^2}{L^2}+50\frac{G^4}{L^4}-5\frac{G^6}{L^6}\right)\right]\sin 5g$$
$$-49725\left[(\sin^2 i-e^2)\left(1-41\frac{H^2}{G^2}+250\frac{H^4}{G^4}-514\frac{H^6}{G^6}+437\frac{H^8}{G^8}-133\frac{H^{10}}{G^{10}}\right)\times\left(19-60\frac{G^2}{L^2}+66\frac{G^4}{L^4}-28\frac{G^6}{L^6}+3\frac{G^8}{L^8}\right)\right.$$
$$+e^2\sin^2 i\left(21-943\frac{H^2}{G^2}+6250\frac{H^4}{G^4}-13878\frac{H^6}{G^6}+12673\frac{H^8}{G^8}-4123\frac{H^{10}}{G^{10}}\right)\times\left(19-60\frac{G^2}{L^2}+66\frac{G^4}{L^4}-28\frac{G^6}{L^6}+3\frac{G^8}{L^8}\right)$$
$$\left.+24e^2\sin^2 i\left(1-41\frac{H^2}{G^2}+250\frac{H^4}{G^4}-514\frac{H^6}{G^6}+437\frac{H^8}{G^8}-133\frac{H^{10}}{G^{10}}\right)\times\frac{G^2}{L^2}\left(5-11\frac{G^2}{L^2}+7\frac{G^4}{L^4}-\frac{G^6}{L^6}\right)\right]\sin 7g$$
$$+20995\left[(\sin^2 i-e^2)\left(1-25\frac{H^2}{G^2}+90\frac{H^4}{G^4}-130\frac{H^6}{G^6}+85\frac{H^8}{G^8}-21\frac{H^{10}}{G^{10}}\right)\times\left(1-4\frac{G^2}{L^2}+6\frac{G^4}{L^4}-4\frac{G^6}{L^6}+\frac{G^8}{L^8}\right)\right.$$
$$+e^2\sin^2 i\left(21-575\frac{H^2}{G^2}+2250\frac{H^4}{G^4}-3510\frac{H^6}{G^6}+2465\frac{H^8}{G^8}-651\frac{H^{10}}{G^{10}}\right)\times\left(1-4\frac{G^2}{L^2}+6\frac{G^4}{L^4}-4\frac{G^6}{L^6}+\frac{G^8}{L^8}\right)$$
$$\left.+8e^2\sin^2 i\left(1-25\frac{H^2}{G^2}+90\frac{H^4}{G^4}-130\frac{H^6}{G^6}+85\frac{H^8}{G^8}-21\frac{H^{10}}{G^{10}}\right)\times\frac{G^2}{L^2}\left(1-3\frac{G^2}{L^2}+3\frac{G^4}{L^4}-\frac{G^6}{L^6}\right)\right]\sin 9g\Bigg\}.$$

REFERENCES

1. Giacaglia, Giorgio E. O., "The Influence of High-Order Zonal Harmonics on the Motion of an Artificial Satellite without Drag," Astronom. J. 69(4):303–308, May 1964.
2. Brouwer, D., "Solution of the Problem of Artificial Satellite Theory without Drag," Astronom. J. 64(9): 378–397, November 1959.

NOTES ON VON ZEIPEL'S METHOD*

GIORGIO E. O. GIACAGLIA

1. INTRODUCTION

Since the rediscovery of von Zeipel's method by D. Brouwer (1959) and its successful application to the problem of artificial satellites, many other problems have been solved by that same method, thus proving its great applicability. It is the purpose of these notes to present the general equations of von Zeipel's method and discuss briefly their applicability.

The reduction of the order of a differential canonical system can, in theory, be performed by obtaining, one way or another, integrals of the system. One of them is the Hamiltonian itself when it is time independent. Actually, this integral of the system (physically its "energy"), can describe completely the geometry of the solutions in a phase space of $2n$ dimensions where $2n$ is the order of the system. When this order is 2, then the solution is completely specified and the use of the Hamiltonian reduces it to a first order differential equation which can be integrated by quadrature. The introduction of p integrals in a system of n degrees of freedom ($2n^{th}$ order), reduces it to one of $n-p$ degrees of freedom which can be integrated immediately when $n-p\leq 1$ (where, of course, p cannot be greater than n).

A few comments can be made with respect to the more famous methods of reduction to show their eventual connection with von Zeipel's method.

2. FROM HAMILTON TO VON ZEIPEL

In the discussion that follows only methods that have been used in connection with differential systems describing the motion of a physical system are considered. The presentation does not necessarily follow a chronological order.

Consider then a system of n degrees of freedom given by $2n$ first order canonical equation

$$\dot{x}_j=\frac{\partial H}{\partial y_j} \qquad (j=1, 2, \cdots, n) \qquad (1)$$
$$\dot{y}_j=-\frac{\partial H}{\partial x_j}$$

where the Hamiltonian $H=H(x_1,\ldots,x_n,y_2,\ldots y_n)$ is presumed to be time independent. If this is not the case, the introduction of time as a new canonical coordinate x_{n+1} (the associated momentum being $-H$) always reduces the latter to the former case. The degree of freedom will however increase by one.

A canonical transformation of the variables (x,y) to new variables (x',y') will be, *in this exposition*, equivalent to the problem of finding a generating function $S=S(x',y,t)$ such that

$$y'_j=\frac{\partial S}{\partial x'_j} \qquad (j=1, 2, \cdots, n) \qquad (2)$$
$$x_j=\frac{\partial S}{\partial y_j}.$$

It is easily seen that this is a sufficient condition to satisfy the Jacobi-Poincare relation

$$\sum_{j=1}^{n}(x_j\,dy_j-x'_j\,dy'_j)=dW. \qquad (3)$$

which is valid whether or not S is an explicit function of time. The Hamiltonian of the new system will be equal to that of the old one inasmuch as one is obtained from the other by introducing the transformation of variables expressed by Equations (2) when $\partial S/\partial t=0$.

a. *HAMILTON-JACOBI*—The method introduced by Hamilton and Jacobi consists in obtaining a canonical transformation such that the new

*Published as *Goddard Space Flight Center Document X-547-64-161*, June 1964.

Hamiltonian is identically zero. In such a case, the new variables are all constants.

b. *LINDSTEDT'S METHOD*—Lindstedt's method is a particular application of the Hamilton-Jacobi method for cases where the Hamiltonian is expanded in terms of small parameters. In this particular case the solution gives the coordinates as linear functions of time and the momenta as constants (usually called action angle variables). The comparison with the Hamilton-Jacobi method is purely formal since the method devised by Lindstedt is quite original. Actually, the real difference between von Zeipel's and this method is that Lindstedt does not make use of a generating function.

c. *WHITTAKER'S METHOD* (solution by series). This method obtains n integrals of the system by reducing the Hamiltonian to a function of the products $p_j=x_jy_j(j=1,2,\ldots,n)$. In this case, since

$$\dot{x}_j=\frac{\partial H}{\partial y_j}=\frac{\partial H}{\partial p_j}x_j$$

$$\dot{y}_j=-\frac{\partial H}{\partial x_j}=-\frac{\partial H}{\partial p_j}y_j,$$

it follows that

$$\dot{x}_j\dot{y}_j+x_jy_j=0 \text{ or } p_j=\text{const } (j+1,2,\ldots,n).$$

d. *DELAUNAY'S METHOD*—This method, as Lindstedt's, can be applied only when the Hamiltonian consists of a "zero order" part (the corresponding system having a known solution) and a "disturbing function" that has a small numerical factor. The basic approach of the von Zeipel's method is the same as that of Delaunay's method; however, the latter one makes no use of a generating function and breaks the disturbing function into parts which are treated separately. The Hamiltonian must be constructed after the transformation is performed for each particular part.

A few more techniques could be mentioned but one deserves more attention than all the others. The concept of adiabatic invariants in Quantum Mechanics is quite analogous to the concept of "mean variables" in von Zeipel's method, or to a certain extent to what Whittaker calls Adelphic Integrals.

3. THE VON ZEIPEL'S METHOD (1916)

It has been quite common, after Delaunay, to use the negative of the Hamiltonian. Thus, if $F=-H$ and if $\ell_j(j=1,2,\ldots,n)$ and $L_j(j=1,2,\ldots,n)$ are the coordinates and momenta respectively, then

$$\dot{\ell}_j=-\frac{\partial F}{\partial L_j}$$

$$(j=1, 2, \cdots, n) \qquad (4)$$

$$\dot{L}_j=\frac{\partial F}{\partial \ell_j}.$$

Suppose

$$F=F(\ell,L;\epsilon) \qquad (5)$$

where ϵ is a "small parameter" and ℓ and L indicate the sets $(\ell_1,\ldots,\ell n)$ and $(L_1,\ldots,L_n)$. A canonical transformation involving the parameter ϵ will be given by a generating function

$$S=S(\ell,L^*;\epsilon)$$

such that

$$L_j=\frac{\partial S}{\partial \ell_j}$$

$$(j=1, 2, \cdots, n)$$

$$\ell_j^*=\frac{\partial S}{\partial L_j^*} \qquad (6)$$

where (ℓ^*,L^*) are the new coordinates and momenta. If F^* is the negative of the new Hamiltonian, then we assume

$$F^*(\ell^*,L^*;\epsilon)=F(\ell(\ell^*,L^*;\epsilon),L(\ell^*,L^*;\epsilon);\epsilon) \qquad (7)$$

or, from Equations (6),

$$F^*\left(\frac{\partial S}{\partial L^*},\ L^*;\ \epsilon\right)=F\left(\ell,\frac{\partial S}{\partial L};\ \epsilon\right). \qquad (8)$$

In a more restrictive sense it is assumed that the series

$$\bar{F}=\sum_{k=0}^{N}\epsilon^kF_k(\ell,\ L) \qquad (9)$$

represents the negative of the Hamiltonian to the required degree of precision and converges to $F(\ell,L;\epsilon)$ as $N\to\infty$. From this point $\bar{F}$ is written

as F without danger of confusion. Furthermore, it is assumed that

$$S(\ell, L^*; \epsilon)$$

$$F_k(\ell, L)$$

$$F^*(\ell^*, L^*; \epsilon)$$

are developable in Taylor's series in the neighborhood of $\epsilon=0$, so that the series

$$S=\sum_{k=0}^{\infty}\epsilon^k S_k(\ell, L^*)_{\epsilon=0}$$

$$F_p\left(\ell, \frac{\partial S}{\partial \ell}\right)=\sum_{k=0}^{\infty}\frac{\epsilon^k}{k!}\left(\frac{d^k F_p}{d\epsilon^k}\right)_{\epsilon=0} \tag{10}$$

are convergent for sufficiently small ϵ.

By the conservation property

$$\sum_{k=0}^{\infty}\epsilon^k F_k\left(\ell, \frac{\partial S}{\partial \ell}\right)=\sum_{k=0}^{\infty}\frac{\epsilon^k}{k!}\left(\frac{d^k F^*}{d\epsilon^k}\right)_{\epsilon=0} \tag{11}$$

where it is important to note that $\partial S/\partial \ell$ contains ϵ through Equation (6). Equating the coefficients of like powers in ϵ in both sides of Equation (11) gives a system of partial differential equations in S and F^*. The next step is obtaining this system.

4. DIFFERENTIAL EQUATIONS OF THE VON ZEIPEL'S METHOD

The m^{th} derivative of F_k with respect to ϵ at the point $\epsilon=0$ is obtained as follows.

Consider

$$\frac{dF_k}{d\epsilon}=\sum_{i=1}^{n}\frac{\partial F_k}{\partial L_i}\frac{dL_i}{d\epsilon}=\sum_{i=1}^{n}\frac{\partial F_k}{\partial L_i}\frac{d}{d\epsilon}\left(\frac{\partial S}{\partial \ell_i}\right).$$

Using Equation (10) it follows that

$$\left(\frac{dF_k}{d\epsilon}\right)=\sum_{i=1}^{n}\frac{\partial F_k}{\partial L_i}\frac{d}{d\epsilon}\left\{\sum_{j=0}^{\infty}\epsilon^j\frac{\partial S_j(\ell, L^*)}{\partial \ell_i}\right\}=\sum_{i=1}^{n}\sum_{j=1}^{\infty}j\,\epsilon^{j-1}\frac{\partial F_k}{\partial L_i}\left(\frac{\partial S_j}{\partial \ell_i}\right)_{\epsilon=0}. \tag{12}$$

Let us now compute

$$\frac{d^{m-1}}{d\epsilon^{m-1}}\left(\epsilon^{j-1}\frac{\partial F_k}{\partial L_i}\right)$$

where

$$F_k=F_k\left(\ell, \frac{\partial S}{\partial \ell}\right).$$

Applying Leibniz' formula, this becomes

$$\frac{d^{m-1}}{d\epsilon^{m-1}}\left(\epsilon^{j-1}\frac{\partial F_k}{\partial L_i}\right)=\sum_{\nu=0}^{\min(m-1,\, j-1)}\binom{m-1}{\nu}\frac{d^\nu \epsilon^{j-1}}{d\epsilon^\nu}\frac{d^{m-1-\nu}}{d\epsilon^{m-1-\nu}}\left(\frac{\partial F_k}{\partial L_i}\right).$$

For $\epsilon=0$ the only possible choice is $j<m$. Then

$$\left\{\frac{d^{m-1}}{d\epsilon^{m-1}}\left(\epsilon^{j-1}\frac{\partial F_k}{\partial L_i}\right)\right\}_{\epsilon=0}=\binom{m-1}{j-1}(j-1)!\left\{\frac{d^{m-j}}{d\epsilon^{m-j}}\left(\frac{\partial F_k}{\partial L_i}\right)\right\}_{\epsilon=0}.$$

Now using Equation (12)

$$\left(\frac{d^m F_k}{d\epsilon^m}\right)_{\epsilon=0}=\sum_{i=1}^{n}\sum_{j=1}^{\infty}j\binom{m-1}{j-1}(j-1)!\left\{\frac{d^{m-j}}{d\epsilon^{m-j}}\left(\frac{\partial F_k}{\partial L_i}\right)\right\}_{\epsilon=0}\left(\frac{\partial S_j}{\partial \ell_i}\right)_{\epsilon=0}.$$

It is now desirable to rewrite the above equation as

$$\left(\frac{d^m F_k}{d\epsilon^m}\right)_{\epsilon=0}=\sum_{i_1=1}^{n}\sum_{j_1=1}^{\infty}j_1\frac{(m-1)!}{(m-j_1)!}\left(\frac{\partial S_{j_1}}{\partial \ell_{i_1}}\right)_{\epsilon=0}\left[\frac{d^{m-j_1}}{d\epsilon^{m-j}}\left(\frac{\partial^1 F_k}{\partial L_{i_1}}\right)\right]_{\epsilon=0}. \tag{13}$$

Equation (13) is now applied to find

$$\left[\frac{d^{m-j_1}}{d\epsilon^{m-j_1}}\left(\frac{\partial^1 F_k}{\partial L_{i_1}}\right)\right]_{\epsilon=0}^{\cdot}.$$

The result is

$$\left[\frac{d^{m-j_1}}{d\epsilon^{m-j_1}}\left(\frac{\partial^1 F_k}{\partial L_{i_1}}\right)\right]_{\epsilon=0}^{\cdot}=\sum_{i_2=1}^{n}\sum_{j_2=1}^{\infty}j_2\frac{(m-j_1-1)!}{(m-j_1-j_2)!}\left(\frac{\partial S_{j_2}}{\partial \ell_{i_2}}\right)_{\epsilon=0}\times\left[\frac{d^{m-j_1-j_2}}{d\epsilon^{m-j_1-j_2}}\left(\frac{\partial^2 F_k}{\partial L_{i_1}\partial L_{i_2}}\right)\right]_{\epsilon=0}^{\cdot}$$

The process is repeated up to the point where

$$m-j_1-j_2-\cdots-j_N=0, \tag{14}$$

so that

$$\left[\frac{d^{m-j_1-j_2-\cdots-j_N}}{d\epsilon^{m-j_1-j_2-\cdots-j_N}}\left(\frac{\partial^N F_k}{\partial L_{i_1}\partial L_{i_2}\cdots\partial L_{i_N}}\right)\right]_{\epsilon=0}^{\cdot}=\left(\frac{\partial^N F_k}{\partial L_{i_1}\cdots\partial L_{i_N}}\right)_{\epsilon=0}^{\cdot}.$$

Substituting these successive derivatives into Equation (13), it follows that

$$\left(\frac{d^m F_k}{d\epsilon^m}\right)_{\epsilon=0}^{\cdot}=\sum_{i_1=1}^{n}\sum_{j_1=1}^{\infty}j_1\frac{(m-1)!}{(m-j_1)!}\left(\frac{\partial S_{j_1}}{\partial \ell_{i_1}}\right)_{\epsilon=0}\times\sum_{i_2=1}^{n}\sum_{j_2=1}^{\infty}j_2\frac{(m-j_1-1)!}{(m-j_1-j_2)!}\left(\frac{\partial S_{j_2}}{\partial \ell_{i_2}}\right)_{\epsilon=0}$$

$$\times\sum_{i_3=1}^{n}\sum_{j_3=1}^{\infty}j_3\frac{(m-j_1-j_2-1)!}{(m-j_1-j_2-j_3)!}\left(\frac{\partial S_{j_3}}{\partial \ell_{i_3}}\right)_{\epsilon=0}\times\cdots\times\sum_{i_N=1}^{n}\sum_{j_N=1}^{\infty}j_N\frac{(m-j_1-j_2-\cdots-j_{N-1}-1)!}{(m-j_1-j_2-\cdots-j_N)!}\left(\frac{\partial S_{j_N}}{\partial \ell_{i_N}}\right)_{\epsilon=0}$$

$$\times\left(\frac{\partial^N F_k}{\partial L_{i_1}\partial L_{i_2}\cdots\partial L_{i_N}}\right)_{\epsilon=0}^{\cdot}.$$

The numerical factors are reduced to

$$j_1 j_2\cdots j_N\frac{m!}{m(m-j_1)(m-j_1-j_2)\cdots(m-j_1-j_2-\cdots-j_{N-1})}\equiv m!\,C(m;j_1,j_2,\cdots,j_N)$$

and the second summation does not run in general up to infinity but to a limit given by condition (14). Thus the above relation becomes

$$\left(\frac{d^m F_k}{d\epsilon^m}\right)_{\epsilon=0}^{\cdot}=\sum_{(m)}\prod_{p=1}^{N}\left(\sum_{i_p=1}^{n}\right)m!\,C(m;j_1,\cdots,j_N)\prod_{p=1}^{N}\left(\frac{\partial S_{j_p}}{\partial \ell_{i_p}}\right)_{\epsilon=0}\left(\frac{\partial^N F_k}{\partial L_{i_1}\cdots\partial L_{i_N}}\right)_{\epsilon=0}^{\cdot} \tag{15}$$

where $\sum_{(m)}$ stands for summation over all possible positive integers j_p whose sum is m (according to Equation (14)), and the first product $\prod_{p=1}^{N}$ refers to the summation signs $\sum_{i_p=1}^{n}$. There are n of these integers. Equation (15) will be valid even for $m=0$ with the definitions

$$\frac{d^0 F_k}{d\epsilon^0}\equiv F_k$$

and

$$C(0;-)\equiv 1.$$

From Equations (10) and (11) it follows that

$$F=\sum_{k=0}^{\infty}\epsilon^k F_k(\ell, L)=\sum_{k=0}^{\infty}\epsilon^k\sum_{m=0}^{\infty}\frac{\epsilon^m}{m!}\left(\frac{d^m F_k}{d\epsilon^m}\right)_{\epsilon=0}=\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\frac{\epsilon^{m+k}}{m!}\left(\frac{d^m F_k}{d\epsilon^m}\right)_{\epsilon=0}=\sum_{\nu=0}^{\infty}\sum_{m=0}^{\nu}\frac{\epsilon_\nu}{m!}\left(\frac{d^m F_{\nu-m}}{d\epsilon^m}\right)_{\epsilon=0}.$$

The substitution of these results into Equation (14) leads to

$$F=\sum_{\nu=0}^{\infty}\sum_{m=0}^{\nu}\sum_{(m)}\prod_{p=1}^{N}\left(\sum_{i_p=1}^{n}\right)C(m; j_1, \cdots, j_N)\prod_{p=1}^{N}\left(\frac{\partial S_{j_p}}{\partial \ell_{i_p}}\right)_{\epsilon=0}\left(\frac{\partial^N F_{\nu-m}}{\partial L_{i_1}\cdots\partial L_{i_N}}\right)_{\epsilon=0}\epsilon^\nu. \tag{16}$$

In a complete similar way, if

$$F^*=\sum_{k=0}^{\infty}\epsilon^k F_k^*\left(\frac{\partial S}{\partial L^*}, L^*\right),$$

then

$$F^*=\sum_{\nu=0}^{\infty}\sum_{m=0}^{\nu}\sum_{(m)}\prod_{p=1}^{N}\left(\sum_{i_p=1}^{n}\right)C(m; j_1, \cdots, j_N)\prod_{p=1}^{N}\left(\frac{\partial S_{i_p}}{\partial L_{i_p}^*}\right)_{\epsilon=0}\left(\frac{\partial^N F_{\nu-m}^*}{\partial \ell_{i_1}^*\cdots\partial \ell_{i_N}^*}\right)_{\epsilon=0}\epsilon^\nu. \tag{17}$$

It is important to note that in Equation (16), $\epsilon=0$ is equivalent to $L_r=\partial S_0/\partial \ell_r (r=1,2,\cdots,n)$, and in Equation (17) $\epsilon=0$ is equivalent to $\ell_r^*=\partial S_0/\partial L_r^* (r=1,2,\cdots,n)$, according to Equation (10). The equality of factors of the same power of ϵ in Equations (16) and (17) gives the partial differential equations for the von Zeipel's method

$$\sum_{m=0}^{\nu}\sum_{(m)}\prod_{p=1}^{N}\left(\sum_{i_p=1}^{n}\right)C(m; j_1, \cdots, j_N)\prod_{p=1}^{N}\left\{\frac{\partial S_{j_p}}{\partial \ell_{i_p}}\left(\frac{\partial^N F_{\nu-m}}{\partial L_{i_p}\cdots\partial L_{i_N}}\right)+-\frac{S\partial_{j_p}}{\partial L_{i_p}^*}\left(\frac{\partial^N F_{\nu-m}}{\partial \ell_{i_1}^*\cdots\partial \ell_{i_N}^*}\right)\right\}_{\epsilon=0}=0 \tag{18}$$

for $\nu=0, 1, 2, \cdots$.

For instance, Equation (18) gives:

$\underline{\nu=0}$

$$F_0\left(\ell, \frac{\partial S_0}{\partial \ell}\right)=F_0^*\left(\frac{\partial S_0}{\partial L^*}, L^*\right) \tag{19}$$

$\underline{\nu=1}$

$$F_1\left(\ell, \frac{\partial S_0}{\partial \ell}\right)+\sum_{i=1}^{n}\left(\frac{\partial S_1}{\partial \ell_i}\frac{\partial F_0}{\partial L_i}\right)_{L_i=\frac{\partial S_0}{\partial \ell_i}}=F_1^*\left(\frac{\partial S_0}{\partial L^*}, L^*\right)+\sum_{i=1}^{n}\left(\frac{\partial S_1}{\partial L_i^*}\frac{\partial F_0^*}{\partial \ell_i^*}\right)_{\ell_i^*=\frac{\partial S_0^*}{\partial L_i^*}} \tag{20}$$

$\underline{\nu=2}$

$$F_2\left(\ell, \frac{\partial S_0}{\partial \ell}\right)+\sum_{i=1}^{n}\left(\frac{\partial S_1}{\partial \ell_i}\frac{\partial F_1}{\partial L_i}\right)_{L_1=\frac{\partial S_0}{\partial \ell_i}}+\sum_{i=1}^{n}\left(\frac{\partial S_2}{\partial \ell_i}\frac{\partial F_0}{\partial L_i}\right)_{L_i=\frac{\partial S_0}{\partial \ell_i}}+\frac{1}{2}\sum_{i,\,j=1}^{n}\left(\frac{\partial S_1}{\partial \ell_i}\frac{\partial S_1}{\partial \ell_j}\frac{\partial^2 F_0}{\partial L_i \partial L_j}\right)_{L_k=\frac{\partial S_0}{\partial \ell_k}}$$

$$=F_2^*\left(\frac{\partial S_0}{\partial L^*}, L^*\right)+\sum_{i=1}^{n}\left(\frac{\partial S_1}{\partial L_i^*}\frac{\partial F_1^*}{\partial \ell_i^*}\right)_{\ell_i^*=\frac{\partial S_0}{\partial L_i^*}}+\sum_{i=1}^{n}\left(\frac{\partial S_2}{\partial L_i^*}\frac{\partial F_0^*}{\partial \ell_i^*}\right)_{\ell_i^*=\frac{\partial S_0}{\partial L_i^*}}+\frac{1}{2}\sum_{i,\,j=1}^{n}\left(\frac{\partial S_1}{\partial L_i^*}\frac{\partial S_1}{\partial L_j^*}\frac{\partial^2 F_0^*}{\partial \ell_i^* \partial \ell_j^*}\right)_{\ell_k^*=\frac{\partial S_0}{\partial L_k^*}} \tag{21}$$

$\nu=3$

$$F_3\left(\ell,\frac{\partial S_0}{\partial \ell}\right)+\sum_{i=1}^{n}\left(\frac{\partial S_1}{\partial \ell_i}\frac{\partial F_2}{\partial L_i}\right)_{L_i=\frac{\partial S_0}{\partial \ell_i}}+\sum_{i=1}^{n}\left(\frac{\partial S_2}{\partial \ell_i}\frac{\partial F_1}{\partial L_i}\right)_{L_i=\frac{\partial S_0}{\partial \ell_i}}+\sum_{i,\,j=1}^{n}\frac{1}{2}\left(\frac{\partial S_1}{\partial \ell_i}\frac{\partial S_1}{\partial \ell_j}\frac{\partial^2 F_1}{\partial L_i \partial L_j}\right)_{L_k=\frac{\partial S_0}{\partial \ell_k}}+$$

$$+\sum_{i,\,j,\,k=1}^{n}\frac{1}{6}\left(\frac{\partial S_1}{\partial \ell_i}\frac{\partial S_1}{\partial \ell_j}\frac{\partial S_1}{\partial \ell_k}\frac{\partial^3 F_0}{\partial L_i \partial L_j \partial L_k}\right)_{L_p=\frac{\partial S_0}{\partial \ell_p}}+\sum_{i,\,j=1}^{n}\left(\frac{\partial S_1}{\partial \ell_i}\frac{\partial S_2}{\partial \ell_j}\frac{\partial^2 F_0}{\partial L_i \partial L_j}\right)_{L_k=\frac{\partial S_0}{\partial \ell_k}}+\sum_{i=1}^{n}\left(\frac{\partial S_3}{\partial \ell_i}\frac{\partial F_0}{\partial L_i}\right)_{L_i=\frac{\partial S_0}{\partial \ell_i}}$$

$$=F_3^*\left(\frac{\partial S_0}{\partial L^*},L^*\right)+\sum_{i=1}^{n}\left(\frac{\partial S_1}{\partial L_i^*}\frac{\partial F_2^*}{\partial \ell_i^*}\right)_{\ell_i^*=\frac{\partial S_0}{\partial L_i^*}}+\sum_{i=1}^{n}\left(\frac{\partial S_2}{\partial L_i^*}\frac{\partial F_1^*}{\partial \ell_i^*}\right)_{\ell_i^*=\frac{\partial S_0}{\partial L_i^*}}+\sum_{i,\,j=1}^{n}\frac{1}{2}\left(\frac{\partial S_1}{\partial L_i^*}\frac{\partial S_1}{\partial L_j^*}\frac{\partial^2 F_1^*}{\partial \ell_i^* \partial \ell_j^*}\right)_{\ell_k^*=\frac{\partial S_0}{\partial L_k^*}}$$

$$+\sum_{i,\,j,\,k=1}^{n}\frac{1}{6}\left(\frac{\partial S_1}{\partial L_i^*}\frac{\partial S_1}{\partial L_j^*}\frac{\partial S_1}{\partial L_k^*}\frac{\partial^3 F_0^*}{\partial \ell_i^* \partial \ell_j^* \partial \ell_k^*}\right)_{\ell_p^*=\frac{\partial S_0}{\partial L_p^*}}+\sum_{i,\,j=1}^{n}\left(\frac{\partial S_1}{\partial L_i^*}\frac{\partial S_2}{\partial L_j^*}\frac{\partial F_0^*}{\partial \ell_i^* \partial \ell_j^*}\right)_{\ell_k^*=\frac{\partial S_0}{\partial L_k^*}}$$

$$+\sum_{i=1}^{n}\left(\frac{\partial S_2}{\partial L_i^*}\frac{\partial F_0^*}{\partial \ell_i^*}\right)_{\ell_i^*=\frac{\partial S_0}{\partial L_i^*}} \tag{22}$$

where use has been made of the coefficients

$C(1;\,1)=1$
$C(2;\,2)=1 \quad C(2;\,1,\,1)=\frac{1}{2}$
$C(3;\,3)=1 \quad C(3;\,2,\,1)=\frac{2}{3} \quad C(3;\,1,\,2)=\frac{1}{3} \quad C(3;\,1,\,1,\,1)=\frac{1}{6}.$

5. ELIMINATION OF VARIABLES

Since the solution of the system is known where F is reduced to F_0, the problem is to eliminate variables which are not present in F_0. Suppose a canonical transformation is found in such a way that p of the n coordinates ($p\leq n$) have been eliminated from the Hamiltonian, that is

$$F^*=F^*(\ell_{p+1}^*,\cdots,\ell_n,L_1^*,L_2^*,\cdots,L_n^*). \tag{23}$$

The equations of motion then yield

$$L_k^*=C_k(\text{const.}) \qquad (k=1,2,\cdots,p). \tag{24}$$

If these constants are replaced in F^*, then

$$F^*=F^*(\ell_{p+1}^*,\cdots,\ell_n^*,C_1,C_2,\cdots,C_p,L_{p+1}^*,\cdots,L_n^*)$$

and the problem is reduced to one of $n-p$ degrees of freedom.

a. If $p=n$, the problem is completely solved, since

$$L_k^*=C_k \qquad (k=1,2,\cdots,n)$$

and

$$\ell_k^*=\omega_k(C_1,\cdots,C_n)t+\ell_k^*(0). \qquad (k=1,2,\cdots,n)$$

b. If $p=n-1$, the problem is integrable by quadrature. In fact,

$$L_k^* = C_k \qquad (k=1,2,\cdots,n-1)$$

$$\dot{L}_n^* = \frac{\partial F^*}{\partial \ell_n^*} = \Lambda(C_1, C_2, \cdots, C_{n-1}; L_n^*, \ell_n^*)$$

$$\dot{\ell}_n^* = -\frac{\partial F^*}{\partial L_n^*} = \lambda_n'(C_1, C_2, \cdots, C_{n-1}; L_n^*, \ell_n^*).$$

Since

$$F^*(C_1,\cdots,C_{n-1}; \ell_n^*, L_n^*) = C = \text{const.},$$

then

$$L_n^* = L_n^*(C, C_1, \cdots, C_{n-1}; \ell_n^*)$$

and therefore

$$\dot{\ell}_n^* = \lambda_n(C, C_1, \cdots, C_{n-1}; \ell_n^*)$$

and

$$t - t_0 = \int_{\ell_{n(0)}^*}^{\ell_n^*} \frac{d\zeta}{\lambda_n(C, C_1, \cdots, C_{n-1}; \zeta)}$$

The coordinate ℓ_n^* becomes a known function of time as well as L_n^*.

Therefore, the equations

$$\dot{\ell}_k^* = -\frac{\partial F^*}{\partial L_k^*} = \lambda_k\Big(C_1, C_2, \cdots, C_{n-1}; L_n^*(t), \ell_n^*(t)\Big) (k=1, 2, \cdots, n-1)$$

can be integrated by quadrature.

The von Zeipel's method consists in the elimination of some of the coordinates (angular variables) and the reduction of the problem to case (b) and possibly (a). The adaptability of this method is based on a set of hypotheses which are listed below in Roman numerals.

(I) The new and old corresponding variables differ by a quantity at least of the first order, i.e.

$$\ell_i' - \ell_i = 0(\epsilon)$$

$$(i=1, 2, \cdots, n)$$

$$L_i^* - L_i = 0(\epsilon).$$

This automatically fixes S_0 to correspond to the identity transformation since for $\epsilon=0$, the above conditions give

$$\ell_i^* \equiv \ell_i$$

$$(i=1, 2, \cdots, n)$$

$$L_i^* \equiv L_i.$$

Therefore

$$S_0 = \sum_{i=1}^{n} \ell_i L_i^*. \tag{25}$$

If expression (25) is substituted into Equations (19), (20), (21) and (22), then

$\nu=0$

$$F_0(\ell, L^*)=F_0^*(\ell, L^*) \tag{26}$$

$\nu=1$

$$F_1(\ell, L^*)+\sum_{i=1}^{n}\left(\frac{\partial S_1}{\partial \ell_i}\right)_{L_i=L_i^*}\frac{\partial F_0}{\partial L_i^*}=F_1^*(\ell, L^*)+\sum_{i=1}^{n}\left(\frac{\partial S_1}{\partial L_i^*}\right)_{\ell_i^*=\ell_i}\frac{\partial F_0^*}{\partial \ell_i} \tag{27}$$

$\nu=2$

$$F_2(\ell, L^*)+\sum_{i=1}^{n}\left(\frac{\partial S_i}{\partial \ell_i}\right)_{L_i=L_i^*}\frac{\partial F_i}{\partial L_i^*}+\sum_{i=1}^{n}\left(\frac{\partial S_2}{\partial \ell_i}\right)_{L_i=L_i^*}\frac{\partial F_0}{\partial L_i^*}+\frac{1}{2}\sum_{i,\,j=1}^{n}\left(\frac{\partial S_1}{\partial \ell_i}\frac{\partial S_1}{\partial \ell_j}\right)_{L_k=L_k^*}\frac{\partial^2 F_0}{\partial L_i^*\partial L_j^*}=F_2^*(\ell, L^*)$$

$$+\sum_{i=1}^{n}\left(\frac{\partial S_1}{\partial L_i^*}\right)_{\ell_i^*=\ell_i}\frac{\partial F_1^*}{\partial \ell_i^*}+\sum_{i=1}^{n}\left(\frac{\partial S_2}{\partial L_i^*}\right)_{\ell_i^*=\ell_i}\frac{\partial F_0^*}{\partial \ell_i}+\frac{1}{2}\sum_{i,\,j=1}^{n}\left(\frac{\partial S_1}{\partial L_i^*}\frac{\partial S_1}{\partial L_j^*}\right)_{\ell_k^*=\ell_k}\frac{\partial^2 F_0^*}{\partial \ell_i\,\partial \ell_j}. \tag{28}$$

and similarly for Equation (22).

It is seen that $S_k(L^*, \ell)$ and $F_k^*(L^*, \ell)$ are unknown functions. In order to perform a particular solution toward the elimination of certain angular variables in F^* we impose conditions (which are usually suitable in Celestial Mechanics) on the functions S_k and F_k^*. They are

II) $F_k^*(L^*, \ell)$ does not depend on $\ell_i(i=1, 2,\cdots,p\leq n)$ for any $k\geq 0$.

III) $S_k(L^*, \ell)$ only depends on the $\ell_i(i=1,2,\cdots,n)$ through trigonometric functions, for any $k>0$. This avoids "secular perturbations" in the momenta L_j, or in other words differences

$$L_j-L_j^*=\frac{\partial(S-S_0)}{\partial \ell_j}$$

are periodic functions of the $\ell_k(k=1, 2,\cdots,n)$.

The application of these conditions, together with the obvious fact that F_0 does not depend on angular variables $\ell_i(i=1,2,\cdots,p\leq n)$ which are to be eliminated, yields the relations

$$F_0(\ell_{p+1},\cdots, \ell_n, L_1^*,\cdots,L_n^*)=F_0^*(\ell_{p+1},\cdots\ell_n,L_1^*\cdots,L_n^*) \tag{29}$$

$$F_{1s}=F_1^*$$

$$F_{1p}+\sum_{i=1}^{n}\left(\frac{\partial S_1}{\partial \ell_i}\right)_{L_i=L_i^*}\frac{\partial F_0}{\partial L_i}=\sum_{i=p+1}^{n}\left(\frac{\partial S_1}{\partial L_i^*}\right)_{\ell_i^*=\ell_i}\frac{\partial F_0^*}{\partial \ell_i} \tag{30}$$

$$F_{2s}+P_{2s}=F_2^*+P_{2s}^*$$

$$F_{2p}+P_{2p}+\sum_{i=1}^{n}\left(\frac{\partial S_2}{\partial \ell_i}\right)_{L_i=L_i^*}\frac{\partial F_0}{\partial L_i^*}=P_{2p}^*+\sum_{i=p+1}^{n}\left(\frac{\partial S_2}{\partial L_i}\right)_{\ell_i^*=\ell_i}\frac{\partial F_0^*}{\partial \ell_i} \tag{31}$$

and so forth. The functions F_{1s} and F_{1p}, F_{2s} and F_{2p}, P_{2s} and P_{2p}, P_{2s}^* and P_{2p}^* are the portions of F_1, F_2, P_2 and P_2^* which are respectively independent of and dependent on the $\ell_i(i=1,2,\cdots,p)$, and where

$$P_2=\sum_{i=1}^{n}\left(\frac{\partial S_1}{\partial \ell_i}\right)_{L_i=L_i^*}\frac{\partial F_1}{\partial L^*}+\frac{1}{2}\sum_{i,\,j=1}^{n}\left(\frac{\partial S_1}{\partial \ell_i}\frac{\partial S_1}{\partial \ell_j}\right)_{L_k=L_k^*}\frac{\partial^2 F_0}{\partial L_i^*\partial L_j^*}$$

$$P_2^*=\sum_{i=p+1}^{n}\left(\frac{\partial S_1}{\partial L_i^*}\right)_{\ell_i^*=\ell_i}\frac{\partial F_1^*}{\partial \ell_i^*}+\frac{1}{2}\sum_{i,\,j=p+1}^{n}\left(\frac{\partial S_1}{\partial L_i^*}\frac{\partial S_1}{\partial L_j^*}\right)_{\ell_k^*=\ell_k}\frac{\partial^2 F_0^*}{\partial \ell_i\partial \ell_j}. \tag{32}$$

In the usual problems of Celestial Mechanics F_0 does not depend on any angular variable so that the Equations (30), (31), (32) and the corresponding equations for higher order are much simplified. Thus, the additional hypotheses will be considered.

IV) F_0 and thus F_0^* depend only on the momenta L_i^*

V) The angular variables ℓ_i $(i=1,2,\cdots,m)$ corresponding to momenta $L_i(i=1,2,\cdots,m)$ that are present in F_0 have been eliminated to the k^{th} order.

The next problem is the possibility of elimination of angular variables whose conjugate momenta are not present in F_0. At this stage the Hamiltonian of the system is

$$F^*=F_0^*(L_1^*,\cdots,L_m^*)+F_1^*(\ell_{m+1},\cdots,\ell_n,L_1^*,\cdots,L_n^*)+\cdots+F_k^*(\ell_{m+1},\cdots,\ell_n,\cdots,L_n^*) \tag{33}$$

where

$$L_j^*=C_j=\text{const}(j=1,2,\cdots,m),$$

and the old and new variables are related by

$$\begin{aligned} L_j-L_j^*&=\frac{\partial S_1}{\partial \ell_j}+\frac{\partial S_2}{\partial \ell_j}+\cdots+\frac{\partial S_K}{\partial \ell_j} \\ \ell_j^*-\ell_j&=\frac{\partial S_1}{\partial L_j^*}+\frac{\partial S_2}{\partial L_j^*}+\cdots+\frac{\partial S_K}{\partial L_j^*}. \end{aligned} \qquad (j=1,2,\cdots,n) \tag{34}$$

Assume a new canonical transformation from the variables $(\ell_{m+1}^*,\cdots,\ell_n^*,L_{m+1}^*,\cdots,L_n^*)$ to the variables $(\ell_{m+1}^{**},\cdots,\ell_n^{**},L_{m+1}^{**},\cdots,L_n^{**})$ and let

$$S^*=S^*(\ell_{m+1}^*,\cdots,\ell_n^*,L_{m+1}^{**},\cdots,L_n^{**}) \tag{35}$$

be its generating function. Then, since $L_j^*=C_j=\text{const}$ $(j=1,2,\cdots,m)$,

$$F_0^*(L_1^{**},\cdots,L_m^{**})=F_0^{**}(L_1^{**},\cdots,L_m^{**})=\text{const } L_k^{**}=L_k^*=C_k=\text{const } (k=1,2,\cdots,m) \tag{36}$$

$$F_1^*(\ell_{m+1}^*,\cdots,\ell_n^*;C_1,C_2,\cdots,C_m,L_{m+1}^{**},\cdots,L_n^{**})=F_1^{**}(\ell_{m+1}^*,\cdots,\ell_n^*;L_{m+1}^{**},\ldots,L_n^{**}). \tag{37}$$

The last equation implies that the elimination of further variables is possible if and only if F_1^* does not depend on them. For in this case

$$F_1^*(\ell_{m+p+1}^*,\cdots,\ell_n^*,C_1,C_2,\cdots,C_n,L_{m+1}^{**},\cdots,L_n^{**})=F_1^{**}(\ell_{m+p+1}^*,\cdots,\ell_n^*;L_{m+1}^{**},\cdots,L_n^{**})$$

and

$$\begin{aligned} F_2^*(\ell_{m+1}^*,\cdots,\ell_n^*;L_{m+1}^{**},\cdots,L_n^{**})+\sum_{i=m+1}^{n}\frac{\partial S_1^*}{\partial \ell_i^*}\frac{\partial F_1^*}{\partial L_i^{**}} \\ =F_2^{**}(\ell_{m+p+1}^*,\cdots,\ell_n^*;L_{m+1}^{**},\cdots,L_n^{**})+\sum_{i=m+p+1}^{n}\frac{\partial S_1^*}{\partial L_i^{**}}\frac{\partial F_1^{**}}{\partial \ell_i^*}. \end{aligned} \tag{38}$$

which defines S_1^*. It is important to note that in such a case S_1^* will be defined by an equation involving *2nd order* terms; these terms are therefore necessary to obtain *first order* "perturbations." This fact is exactly what happens in Brouwer's theory on artificial satellites (1959), where

a) The elimination of g^* is possible because F_1^* is independent of this variable.

b) The development for long period perturbations (those of argument g^*) needs the evaluation of 2nd order terms.

This type of reasoning can be carried on up to any order in exactly the same way. It may then happen that the elimination of a certain angular variable by obtaining S_1^* requires the evaluation of terms of the kth order.

However if F_1^* depends on the angular variables to be eliminated the problem cannot be solved unless it happens that the remaining system has one degree of freedom. For example, this is the case of the perturbations on the motion of an artificial satellite by the moon.

6. SMALL DIVISORS

The case of critical inclination for the theory of artificial satellites of an oblate planet for which P_2 is the dominant zonal harmonic and $J_4 \neq -J_2^2$, is a well known example of the problem of small divisors. Here, only a particular aspect of the question is dealt with. Consider the solution of Equation (30) in the usual case where F_0^* does not depend on the ℓ_i. The characteristic associated system is

$$\frac{d\ell_1}{\dfrac{\partial F_0}{\partial L_1^*}} = \frac{d\ell_2}{\dfrac{\partial F_0}{\partial L_2^*}} = \cdots \equiv \frac{d\ell_p}{\dfrac{\partial F_0}{\partial L_p^*}} = \frac{dS_1}{F_{1p}}. \tag{39}$$

Should one of the partials $\partial F_0/\partial L_i^*$ happen to be zero, the general solution would certainly be discontinuous since a "small divisor" is present. However this divisor is not exactly zero because the quantity $\partial F_0/\partial L_i^*$ is evaluated to first order only.

In the case of critical inclination it is necessary to take

$$S = S_0 + \epsilon^{1/2} S_{1/2} + \epsilon S_1 + \epsilon^{3/2} S_{3/2} + \cdots.$$

However, in doing so the separation of "long periodic" and "secular" perturbations is lost. The integration leads, in most cases, to elliptic integrals (Hori, 1960).

The question of small divisors usually arises whenever the problem presents cases of libration as particular solutions.

Another case to be mentioned is the resonance for an artificial satellite whose period is commensurable with the period of rotation of the Earth when tesseral harmonics are included. Again, expansion in powers of $\epsilon^{1/2}$ can be used to solve the problem (Morando, 1962).

Finally it is important to note that singularities in the Equations (39) reflect singular points in the hypersurface defined by the Hamiltonian of the system in a phase-space of $2(n-p)$ dimensions if p variables have already been eliminated.

7. SUMMARY

The general differential equations of the von Zeipel's method have been given to any order. It is hoped that this will avoid tedious Taylor expansions if one needs to go to order higher than the second.

At the same time, the brief discussion on the applicability and a few pathological cases of the method, may give some guidance toward the solution of new problems.

REFERENCES

1. Brouwer, D., Astron. J., **64**, 378–397, 1959.
2. Brouwer, D., Clemence, G., "Methods of Celestial Mechanics," Acad. Press, 1961.
3. Hori, G., Astron. J., **65**, 291, 1960.
4. Mello, S. F., Seminars of the Bureau des Longitudes, Paris, 1963.
5. Morando, B., Séance du 15 janvier 1962, Bureau des Longitudes, Paris.
6. Poincaré, H., "Les Méthodes Nouvelles de la Mécanique Céleste," Vol. I, Dover Publ., 1959.
7. Siegel, W., "Vorlesungen über Himmelsmechanik," Springer Verl., 1956.
8. Von Zeipel, H., Arkiv Mat., Astron., Physik, **11**, No. 1, 1916.

N63-21025

A SERIES SOLUTION FOR SOME PERIODIC ORBITS IN THE RESTRICTED THREE-BODY PROBLEM ACCORDING TO THE PERTURBATION METHOD*

N66 37952

SU-SHU HUANG†

A series is obtained for those periodic orbits surrounding the more massive of the two finite bodies in the restricted three-body problem. The expansion is in terms of the mass of the less massive finite body. The initial conditions predicted by the series for several periodic orbits are compared with those obtained by purely numerical processes. They are in good agreement for the case corresponding to the earth-moon system. Also, a simple theory of nearly periodic orbits in the neighborhood of a periodic orbit is developed and numerically verified by examples. Finally, it is suggested that, if asteroids avoid places where their orbits would be commeasurable with the period of Jupiter, then artificial satellites and dust particles may avoid certain areas around the earth as a result of the presence of the moon.

INTRODUCTION

A periodic solution may be regarded as a solution of the differential equations of motion that satisfies, in addition to the initial conditions, the condition that after a lapse of one period, P, both coordinates and velocities return to their initial values. Thus, the problem of finding a periodic orbit in celestial mechanics resembles the problem of finding the eigen function for an eigen value in quantum mechanics, since the eigen value is determined by boundary conditions. Indeed, it is this basic concept that led to the derivation of a series solution for those periodic orbits in the restricted three-body problem that are revolving around the more massive of the two finite mass points. The recent papers by Szebehely give a detailed discussion of the restricted three-body problem (References 1 and 2).

The mathematical method used here follows the standard technique employed in classical mechanics under these circumstances. Indeed, it is very similar to the method Hill used in his lunar theory (References 3 and 4; also see Reference 5). Actually, the present analysis is somewhat parallel to the analysis that led Hill to his variation orbit, for both his and the present method depend upon some series expansions. However, in his lunar theory Hill expanded the solution in terms of the ratio of the mean motion of the sun to that of the moon in the rotating coordinate system. In this present theory for artificial satellites orbiting the earth in the earth-moon system the solution is expanded in terms of the mass of the moon. The mathematical simplicity in terms of the lunar mass is obvious; however, the result cannot be applied to the study of the motion of the moon because the moon is revolving around a relatively much less massive body (i.e., the earth) in the earth-sun system; consequently, it is senseless to expand the solution in terms of the solar mass.

EQUATIONS OF MOTION

The total mass of the two finite bodies is considered the unit of mass, and their separation the unit of length. The unit of time is such that the gravitational constant is unity. If we adopt the line joining the two finite mass points as the x-axis and the location of the greater of these masses as the origin, the equations of motion of the third,

*Published as *NASA Technical Note* D-2488, September 1964.

†*Goddard Space Flight Center* and Catholic University, Washington, D.C.

infinitesimal body in this rotating system of reference become:

$$\frac{d^2x}{dt^2} - 2\frac{dy}{dt} = x - \mu - (1-\mu)\frac{x}{(r_1')^3} - \mu\frac{x-1}{(r_2')^3}, \quad (1)$$

$$\frac{d^2y}{dt^2} + 2\frac{dx}{dt} = y - (1-\mu)\frac{y}{(r_1')^3} - \mu\frac{y}{(r_2')^3}, \quad (2)$$

where it is assumed that the mass $1-\mu$ is at the origin. Consequently the mass μ is at point (1, 0). Also, r_1' and r_2' are the distances of the third body from the $1-\mu$ and μ components, respectively. It is obvious that r_1' is also the distance of the third body from the origin; so

$$r_1' = r. \quad (3)$$

We are interested in the periodic orbits revolving around the $1-\mu$ component. Therefore, we make the transformations

$$x = r\cos\theta, \quad (4)$$

$$y = r\sin\theta. \quad (5)$$

This brings Equations 1 and 2 into the forms

$$\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 - 2r\frac{d\theta}{dt} = r - (1-\mu)\frac{1}{r^2} - \mu\frac{r}{(r_2')^3} + \mu\left[\frac{1}{(r_2')^3} - 1\right]\cos\theta, \quad (6)$$

$$r\frac{d^2\theta}{dt^2} + 2\frac{dr}{dt}\frac{d\theta}{dt} + 2\frac{dr}{dt} = -\mu\left[\frac{1}{(r_2')^3} - 1\right]\sin\theta. \quad (7)$$

EQUATIONS OF PERTURBATION

When μ equals zero all circles with their center at the origin are the periodic solutions of the problem. We can now argue that if μ is small, the periodic orbit deviates only slightly from a circular one. The deviation obviously depends upon μ. Therefore, the periodic solutions may be written as

$$r = r_0 + \mu r_1(t) + \mu^2 r_2(t) + \cdots, \quad (8)$$

$$\theta = \lambda t + \mu\theta_1(t) + \mu^2\theta_2(t) + \cdots, \quad (9)$$

where r_0 and λ are to be determined and are independent of time. Substituting Equations 8 and 9 into Equations 6 and 7 gives equations of different orders of approximation after a series of long but straightforward calculations.

The zeroth order (μ^0) approximation gives

$$(\lambda+1)^2 = \frac{1-\mu}{r_0^3}, \quad (10)$$

which is simply Kepler's third law in the problem of two bodies, the $1-\mu$ component and the third, infinitesimal body. The term is $\lambda+1$ instead of λ because the equations are expressed in the rotating coordinate system. Also, λ may have positive or negative values, corresponding respectively to the direct and retrograde motion of the third body.

The first order (μ^1) approximation yields

$$\frac{d^2r_1}{dt^2} - 2(\lambda+1)r_0\frac{d\theta_1}{dt} - 3(\lambda+1)^2 r_1 = \frac{1}{2}r_0 + \frac{3}{2}r_0\cos 2\lambda t + \frac{1}{8}r_0^2(9\cos\lambda t + 15\cos 3\lambda t), \quad (11)$$

$$r_0\frac{d^2\theta_1}{dt^2} + 2(\lambda+1)\frac{dr_1}{dt} = -\frac{3}{2}r_0\sin 2\lambda t - \frac{3}{8}r_0^2(\sin\lambda t + 5\sin 3\lambda t), \quad (12)$$

if the terms involving third and higher orders of r_0 are neglected.

With the same degree of approximation in regard to the series in r_0, the second order (μ^2)

equations are:

$$\frac{d^2 r_2}{dt^2} - 2(\lambda+1) r_0 \frac{d\theta_2}{dt} - 3(\lambda+1)^2 r_2 = r_0 \left(\frac{d\theta_1}{dt}\right)^2 + 2(\lambda+1) r_1 \frac{d\theta_1}{dt} - 3(\lambda+1)^2 \frac{r_1^2}{r_0}$$

$$+ \frac{1}{2} r_1 (1 + 3\cos 2\lambda t) + \frac{1}{4} r_0 r_1 (9\cos \lambda t + 15 \cos 3\lambda t)$$

$$- 3 r_0 \theta_1 \sin 2\lambda t - \frac{r_0^2 \theta_1}{8} (45 \sin 3\lambda t + 9 \sin \lambda t) , \qquad (13)$$

and

$$r_0 \frac{d^2 \theta_2}{dt^2} + 2(\lambda+1) \frac{dr_2}{dt}$$

$$= - 2 \frac{dr_1}{dt} \frac{d\theta_1}{dt} - r_1 \frac{d^2 \theta_1}{dt^2} - \frac{3}{2} r_1 \sin 2\lambda t - 3 r_0 \theta_1 \cos 2\lambda t$$

$$- \frac{1}{4} r_0 r_1 (3 \sin \lambda t + 15 \sin 3\lambda t) - \frac{1}{8} r_0^2 \theta_1 (3 \cos \lambda t + 45 \cos 3\lambda t) . \qquad (14)$$

In a similar way equations of higher orders can be derived in terms of the solutions of the equations of lower orders.

THE FIRST ORDER EQUATIONS

The solutions of Equations 11 and 12 can be easily found:

$$r_1 = - \frac{r_0}{6(\lambda+1)^2} + A_1' \cos(\lambda+1)t + A_2' \sin(\lambda+1)t - \frac{2B_1'}{3(\lambda+1)} + k_1 \cos \lambda t + k_2 \cos 2\lambda t + k_3 \cos 3\lambda t , \qquad (15)$$

$$r_0 \theta_1 = l_1 \sin \lambda t + l_2 \sin 2\lambda t + l_3 \sin 3\lambda t + B_1' t + B_2' - 2A_1' \sin(\lambda+1)t + 2A_2' \cos(\lambda+1)t . \qquad (16)$$

A_1', A_2', B_1', and B_2' are arbitrary constants which make Equations 15 and 16 the most general solution, k_n and l_n are defined by

$$k_1 = \frac{15+6}{8\lambda(2\lambda+1)} r_0^2 , \qquad (17)$$

$$k_2 = - \frac{3(2\lambda+1)}{2\lambda(3\lambda^2 - 2\lambda - 1)} r_0 , \qquad (18)$$

$$k_3 = - \frac{5(5\lambda+2)}{8\lambda(8\lambda^2 - 2\lambda - 1)} r_0^2 , \qquad (19)$$

$$l_1 = - \frac{3(10\lambda^2 + 12\lambda + 3)}{8\lambda^2(2\lambda+1)} r_0^2 , \qquad (20)$$

$$l_2 = \frac{3(11\lambda^2 + 10\lambda + 3)}{8\lambda^2(3\lambda^2 - 2\lambda - 1)} r_0 , \qquad (21)$$

$$l_3 = \frac{5(6\lambda^2 + 4\lambda + 1)}{8\lambda^2(8\lambda^2 - 2\lambda - 1)} r_0^2 . \qquad (22)$$

It should be observed that λ (or, equivalently, r_0), which appears in Equation 10, is not an integration constant in the perturbed case, although it is in the unperturbed case. Indeed, in the unperturbed case λ is the only integration constant that does not vanish under the assumed initial conditions. The integration constants in the perturbed case are A_1', A_2', B_1', and B_2'. In general, if higher orders of μ are considered, A_1, A_2, B_1, and B_2 are defined by

$$A_1 = A_1' \mu + A_1'' \mu^2 + \cdots , \quad \text{etc.},$$

because, as we shall see later, the complementary

functions in different orders of approximation are of the same form. Consequently, the integration constants in different orders of μ can be combined into four arbitrary constants A_1, A_2, B_1, and B_2 to agree with the expected number of four that occur in the solution of two second order differential equations (Equations 6 and 7).

The integration constants A_1, A_2, and B_2 are not particularly relevant, but it is important to note the dynamical meaning of B_1. If λ' is the mean angular velocity in the perturbed case, we have, from Equations 9 and 16,

$$\lambda' = \lambda + \frac{B_1' \mu}{r_0}$$

or in general

$$\lambda' = \lambda + \frac{B_1}{r_0} .$$

Therefore, $B_1'\mu/r_0$ (or B_1/r_0) is the change in the mean angular velocity from the unperturbed case to the perturbed case. Hence the four integration constants for the perturbed case may be taken as A_1, A_2, λ', and B_2 (or A_1', A_2', λ', and B_2' in the first approximation), and λ, or equivalently r_0, acts in the perturbed case as a standard for comparison with the perturbed orbits and is consequently a parameter.

We can set $\lambda' = \lambda$ for the sake of simplicity because both the perturbed and unperturbed periodic orbits form a continuous family. This leaves $B_1' = 0$.

Even with $B_1' = 0$, the general solution does not give the periodic orbits because there are two fundamental periods $2\pi/(\lambda+1)$ and $2\pi/\lambda$, with several harmonics of the latter. However, since the existence of some periodic solutions has been proved (see for example Reference 1), these periodic orbits must correspond to the particular integral obtained from Equations 15 and 16 by setting $A_1' = A_2' = B_1' = 0$. Then the solutions contain only terms with period $2\pi/\lambda$ of the fundamental oscillation, and shorter periods corresponding to its harmonics. Thus the periodic orbits around the $1-\mu$ component may be given, to the first order of μ and the second order of r_0, by

$$r = r_0 + \mu r_1 , \tag{23}$$

$$\theta = \lambda t + \mu \theta_1 , \tag{24}$$

where

$$r_1 = -\frac{r_0}{6(\lambda+1)^2} + k_1 \cos \lambda t + k_2 \cos 2\lambda t + k_3 \cos 3\lambda t , \tag{25}$$

$$r_0 \theta_1 = l_1 \sin \lambda t + l_2 \sin 2\lambda t + l_3 \sin 3\lambda t . \tag{26}$$

It follows from the solution given by Equations 15 and 16 that, in general, a periodic solution can be obtained for any given value of λ only by setting A_1, A_2, B_1, and B_2 equal to zero. Thus, one periodic orbit is associated with one value of the period. However, if λ is a ratio of two integers, periodic orbits may exist for arbitrary (small) values of these constants. If they do, a large number of periodic orbits would be found for some particular values of the period.

THE SECOND ORDER EQUATIONS

When the solutions of r_1 and $r_0\theta_1$ given by Equations 25 and 26 are substituted into the second order equations (13 and 14) and the resulting equations are simplified, we obtain:

$$\frac{d^2 r_2}{dt^2} - 2(\lambda+1) r_0 \frac{d\theta_2}{dt} - 3(\lambda+1)^2 r_2 = \beta_0 + \sum_{n=1} \beta_n \cos n\lambda t , \tag{27}$$

$$r_0 \frac{d^2 \theta_2}{dt^2} + 2(\lambda+1) \frac{dr_2}{dt} = \sum_{n=1} \rho_n \sin n\lambda t . \tag{28}$$

These equations have the same form as Equations 11 and 12, except for more terms on the right-hand side. Therefore, the solution can be derived in the same manner as in the case of the first order equations, although finding the explicit expressions of the solutions is much more tedious because of the lengthy equations that define β_n and ρ_n. The author has evaluated only β_0. From this the average radius of the periodic orbit may be derived:

$$\langle r \rangle = r_0 \left\{ 1 - \frac{\mu}{6(\lambda+1)^2} + \mu^2 \left[\frac{1}{18(\lambda+1)^4} + \frac{3(19\lambda^2+14\lambda+3)}{32\lambda^2(\lambda+1)^2(3\lambda^2-2\lambda-1)} + \frac{9(2\lambda+1)^2}{8\lambda^2(3\lambda^2-2\lambda-1)^2} \right] \right\}, \quad (29)$$

correct to the second order both in μ and r_0.

It follows from Equation 29 that for a given value of λ the average radius of the periodic orbit around the earth in the presence of the moon is slightly less than that given by Equation 10, which is for the case of the absence of the moon. This fact may be easily understood because the overall long-range effect of the moon on a satellite that is revolving around the earth is to reduce the central attractive force of the earth on the satellite. The argument becomes physically apparent if we imagine the moon and its orbit to be replaced by an annular ring of the same mass as the moon, with this mass uniformly distributed. From this reasoning the presence of the sun may be expected to further reduce, by a small amount, the average value of the radius of the satellite's orbit. The prediction has been verified by actual calculations.

The similarity between the present calculation and the perturbation theory in quantum mechanics is apparent. In neither case is the convergence of the series solution proven. But, it will be shown in the following sections that the periodic orbits derived in this way agree perfectly with those obtained by the trial and error method, just as the effectiveness of the perturbation method in quantum mechanics is based on its ability to predict empirical results.

On the other hand, the present perturbation method differs in many ways from that in quantum mechanics. For example, the main purpose of the perturbation theory in quantum mechanics is to find the new eigen value as a result of perturbation, whereas here we are interested in the variation in the orbital nature for a given value of λ.

Although the approach parallels Hill's determination of the variation orbit, differences do exist. In the first place, Hill started with a set of differential equations already approximated by the neglecting of terms involving the ratio of the mean distance of the moon to that of the sun. The present investigation uses the equations in the restricted three-body problem. Secondly, Hill was concerned only in obtaining a particular solution but we are interested in the general solution that involves four aribtrary constants, A_1, A_2, B_1, and B_2, which determine, as shall be seen later, those nearly periodic orbits in the neighborhood of the exact periodic solution. Consequently, the present solution gives an entire family of orbits in the neighborhood of any periodic solution that we can determine. Needless to say, for orbits that depart greatly from the periodic one, the approximation employed in this analysis breaks down. Consequently, those orbits can no longer be represented by the equations derived.

The differences between Hill's analysis and the present one clearly reflect divergent problems faced in different times. Indeed, the present solution of periodic and nearly periodic orbits around a more massive component, expanded in terms of the mass of the less massive component, would have had little practical significance in Hill's time.

NUMERICAL APPLICATIONS

In a previous paper (Reference 6), by a numerical method (Reference 7) a synchronous orbit around the earth was derived, under the idealization of the restricted three-body problem with $\mu = 0.012149$. Now we are able to derive it from Equations 23–26.

In the XY coordinate system with the origin at the mass point $1-\mu$, the initial conditions of

the synchronous orbit, derived by successive approximations, are

$$\left.\begin{aligned} x_0 &= 0.1095908\upsilon\ , \\ y_0 &= 0\ , \\ \dot{x}_0 &= 0\ , \\ \dot{y}_0 &= 2.8927303\ , \end{aligned}\right\} \qquad (30)$$

which yield

$$P=0.23802754 \qquad (31)$$

for the period.

We can now calculate for this case the values of k_n and l_n from Equations 17–22, with

$$\lambda=\frac{2\pi}{P}=26.396884, \qquad (32)$$

which gives $r_0=0.10958800$ from Equation 10.

Table 1

Values of k_n and l_n for $\mu = 0.012149$ and $\lambda = 26.396884$.

n	k_n	l_n
1	4.249394×10^{-4}	8.756117×10^{-4}
2	-1.644862×10^{-4}	2.296898×10^{-4}
3	-6.901187×10^{-6}	8.365787×10^{-6}

The computed values are listed in Table 1. Substituting these values of k_n and l_n in the solution given by Equations 23–26 and transforming variables from r and θ to x and y, in accordance with Equations 4 and 5, shows that the values $x(t)$,$y(t)$,$\dot{x}(t)$, and $\dot{y}(t)$ derived from the present formulas agree to seven significant figures with those obtained from direct integration under the initial conditions given by Equations 30. In particular, at $t=0$ our formulas predict the initial conditions

$$\left.\begin{aligned} x_0 &= 0.10959080\ , \\ y_0 &= 0\ , \\ \dot{x}_0 &= 0\ , \\ \dot{y}_0 &= 2.8927300\ , \end{aligned}\right\} \qquad (33)$$

for the periodic orbit with P given by Equation 31. The agreement between Equations 30 and 33 must be regarded as satisfactory.

As would be expected, the prediction of periodic orbits by Equations 23 and 24 becomes less and less accurate as the period increases. A few cases are given in Table 2 to show the progressive worsening of the prediction. However, it should be noted that even at $P=1.6$, the two sets of calculations—one from the numerical approach and the other from the present formulas—still give results that agree to the fourth significant figure.

ORBITS IN THE NEIGHBORHOODS OF THE PERIODIC ONES

It is obvious that for orbits which are not exactly periodic, the general solution given by

Table 2

Initial Conditions Derived by the Series Solution Compared with those Obtained by the Method of Successive Approximation ($\mu = 0.012149$, $y_0 = \dot{x}_0 = 0$).

P	Series Solution		Successive Approximation		
	x_0	$\dot{y}_0$	x_0	$\dot{y}_0$	Jacobian Constant
0.23802754	0.10959080	2.8927300	0.10959080	2.8927303	9.6968861
0.39999890	0.15239388	2.3936373	0.15239410	2.3936354	7.2832706
0.59999666	0.19580805	2.0503747	0.19580870	2.0503734	5.9498741
0.79999290	0.23271671	1.8277779	0.23271790	1.8277824	5.2292161
0.99999026	0.26506832	1.6657674	0.26506980	1.6657871	4.7756998
1.1999940	0.29394940	1.5398118	0.29395020	1.5398618	4.4638648
1.4000154	0.32005211	1.4376335	0.32005020	1.4377350	4.2365564
1.6000716	0.34385494	1.3522620	0.34384680	1.3524439	4.0638285

Equations 23 and 24 together with Equations 15 and 16 should be applied instead of the particular integral given by Equations 23–26. Let us now consider the behavior of these orbits when we make the initial values of x_0 and $\dot{y}_0$ only slightly different from those corresponding to the periodic orbit, while maintaining $y_0 = x_0 = 0$.

The initial values of r, θ, $\dot{r}$, $\dot{\theta}$ will be denoted by r_i, θ_i, $\dot{r}_i$, $\dot{\theta}_i$. Thus, by setting $t=0$ in Equations 15 and 23

$$r_i = r_0 + \mu\left[-\frac{r_0}{6(\lambda+1)^2} - \frac{2B_1}{3(\lambda+1)} + k_1 + k_2 + k_3 + A_1\right]. \tag{34}$$

By assuming $\theta_i = 0$ as usual

$$B_2 = -2A_2 \tag{35}$$

from Equations 16 and 24. We have assumed $\dot{x}_0 = 0$ which is equivalent to $\dot{r}_i = 0$; hence

$$A_2 = B_2 = 0. \tag{36}$$

Finally, it is easy to obtain

$$\theta_i = \lambda + \frac{\mu}{r_0}\left[l_1\lambda + 2l_2\lambda + 3l_3\lambda - 2A_1(\lambda+1) + B_1\right]. \tag{37}$$

The initial values of r and $\dot{\theta}$ for the true periodic orbit will be denoted by $r_{i,p}$ and $\dot{\theta}_{i,p}$. They are given by

$$r_{i,p} = r_0 + \mu\left[-\frac{r_0}{6(\lambda+1)^2} + k_1 + k_2 + k_3\right]. \tag{38}$$

$$\dot{\theta}_{i,p} = \lambda + \frac{\mu\lambda}{r_0}\left(l_1 + 2l_2 + 3l_3\right). \tag{39}$$

From

$$\Delta r_i = r_i - r_{i,p}. \tag{40}$$

$$\Delta\dot{\theta}_i = \dot{\theta}_i - \dot{\theta}_{i,p}, \tag{41}$$

we derive

$$A_1 = -\frac{\Delta r_i}{\mu} - \frac{2r_0\Delta\dot{\theta}_i}{(\lambda+1)\mu}, \tag{42}$$

$$B_1 = -6(\lambda+1)\frac{\Delta r_i}{\mu} - \frac{3r_0\Delta\dot{\theta}_i}{\mu}. \tag{43}$$

We can now examine the behavior of orbits close to the periodic ones. The explicit expression for θ is, from Equations 16 and 24,

$$\vartheta = \left(\lambda + \frac{\mu}{r_0}B_1\right)t + \frac{\mu}{r_0}\left[\sum_{i=1}^{3} l_i \sin i\lambda t - 2A_1 \sin(\lambda+1)t\right]. \tag{44}$$

Let us define time as $t=t_n$ when $\theta = 2\pi n$, and $t=t_{n-1}$ when $\theta = 2\pi(n-1)$, where n is an integer. Therefore P_n, defined by,

$$P_n = t_n - t_{n-1}. \tag{45}$$

is the period of the n^{th} cycle of a nearly periodic orbit. It follows from Equation 44 that

$$P_n = \frac{2\pi r_0}{\lambda r_0 + \mu B_1} - \frac{\mu}{\lambda r_0 + \mu B_1}\left[\sum_{i=1}^{3} 2l_i \sin\frac{i\lambda P_n}{2}\cos\frac{i\lambda\left(t_n + t_{n-1}\right)}{2} - 4A_1 \sin\frac{(\lambda+1)P_n}{2}\cos\frac{(\lambda+1)\left(t_n + t_{n-1}\right)}{2}\right]. \tag{46}$$

from which we note immediately that the mean period of the nearly periodic orbit is

$$\langle P_n\rangle = \frac{2\pi r_0}{\lambda r_0 + \mu B_1}. \tag{47}$$

So the first term on the right-hand side of Equation 46 is associated with the mean period, but the rest denotes small oscillations of the P_n value around its mean value for various values of n.

Since the amplitudes of various oscillating terms in Equation 46 are small, λP_n may be set equal to 2π in the argument of the sine function. All terms involving l_i vanish as a result of this approximation and Equation 46 can be reduced to:

$$P_n = \langle P_n \rangle + \frac{4\mu A_1}{\lambda r_0 + \mu B_1} \sin\frac{\pi}{\lambda} \cos\left[(\lambda+1)\, t_n - \frac{\pi}{\lambda}\right]. \tag{48}$$

where A_1 and B_1 are given by Equations 42 and 43, respectively, and t_n may be considered to be

$$t_n = n\langle P_n \rangle. \tag{49}$$

Since there is a periodic orbit for each value of λ in the range of interest, we can compare a nearly periodic orbit with any periodic orbit in the former's neighborhood. In our calculation we have fixed a single value of λ for both the periodic and nearly periodic orbit in order to derive A_1 and B_1 from Equations 42 and 43. Obviously a slightly different choice of the value of λ will give different values to these two constants for the same nearly periodic orbit.

As a simple example the two orbits may be compared by starting with the same r_i, i.e., $\Delta r_i = 0$. Then A_1 and B_1 are functions of $\Delta\theta_i$ alone and $\Delta\theta_i$ is related to the difference Δy_0, between the initial value, y_0, of the nearly periodic orbit and that of the exactly periodic orbit. If, furthermore,

$$n \ll \frac{\lambda}{20\Delta\dot{\theta}_i}, \tag{50}$$

Equation 48 reduces to

$$P_n = \frac{2\pi}{\lambda - 3\Delta\dot{\theta}_i} - \frac{8\Delta\dot{\theta}_i}{\lambda(\lambda+1)} \sin\frac{\pi}{\lambda} \cos\frac{(2n-1)\pi}{\lambda}. \tag{51}$$

From Equation 51 values of P_n have been computed up to $n = 28$ for two cases, with λ given by Equation 32. The results are given in the second and fourth columns of Table 3. In both cases $\lambda/\Delta\theta_i$ is of the order of several thousand. Therefore, the use of Equation 51 is justified. For comparison P_n has also been computed by direct integration. This was done by computing successive times, t_n, when the orbit crossed the positive axis. A four point interpolation has been used for obtaining t_n from the integrated tables. In this way, from the initial conditions corresponding respectively to two values of y_0 in Table 3, t_n, and consequently P_n, have been computed and the latter has been tabulated in the third and fifth columns of Table 3. The agreement between the second and third columns as well as that between the fourth and fifth columns may be regarded as satisfactory.

Table 3

Time Intervals for Successive Cycles of Nearly Periodic Orbits ($\mu = 0.012149$, $\lambda = 26.396884$).*

n	Time for $\dot{y}_0 = 2.8920000$		Time for $\dot{y}_0 = 2.8930000$	
	Equation 51*	Direct Integration	Equation 51*	Direct Integration
1	0.2378562	0.2378561	0.2380910	0.2380909
2	0.2378557	0.2378556	0.2380912	0.2380911
3	0.2378547	0.2378546	0.2380915	0.2380915
4	0.2378534	0.2378533	0.2380920	0.2380920
5	0.2378517	0.2378516	0.2380927	0.2380926
6	0.2378497	0.2378497	0.2380934	0.2380933
7	0.2378477	0.2378476	0.2380941	0.2380941
8	0.2378456	0.2378456	0.2380949	0.2380948
9	0.2378437	0.2378435	0.2380956	0.2380956
10	0.2378419	0.2378420	0.2380963	0.2380962
11	0.2378405	0.2378404	0.2380968	0.2380967
12	0.2378394	0.2378394	0.2380972	0.2380971
13	0.2378389	0.2378388	0.2380974	0.2380973
14	0.2378388	0.2378388	0.2380974	0.2380974
15	0.2378392	0.2378391	0.2380973	0.2380973
16	0.2378400	0.2378399	0.2380970	0.2380968
17	0.2378413	0.2378411	0.2380965	0.2380965
18	0.2378429	0.2378428	0.2380959	0.2380958
19	0.2378448	0.2378447	0.2380952	0.2380951
20	0.2378469	0.2378467	0.2380944	0.2380944
21	0.2378489	0.2378488	0.2380937	0.2380937
22	0.2378509	0.2378506	0.2380929	0.2380928
23	0.2378527	0.2378528	0.2380923	0.2380923
24	0.2378542	0.2378540	0.2380917	0.2380916
25	0.2378554	0.2378553	0.2380913	0.2380912
26	0.2378560	0.2378558	0.2380910	0.2380908
27	0.2378562	0.2378562	0.2380910	0.2380911
28	0.2378559	0.2378559	0.2380911	0.2380910

* In computing $\Delta\dot{\theta}_i$ the initial condition given by Equation 33 is considered to correspond to the exact periodic orbit used as the reference for comparison.

SOME REMARKS ON RELATED PROBLEMS

Although the present calculation was performed in order to understand the general effect of the

moon on the motion of the earth's satellites, it will serve equally well in understanding the general behavior of the motion of inner planets and asteroids as a result of the perturbation by the major planets, especially Jupiter. Indeed, the smallness of μ in such cases ($\mu = 9.539 \times 10^{-4}$ for the sun-Jupiter system) will make the result derived with the first order equations a good approximation to the problem. However, because of the relatively large sizes of the orbits of the inner planets, it may be necessary to include in the solution terms involving the third and perhaps higher powers of r_0.

The coefficients k_n and l_n contain in their denominators a factor $(n\lambda)^2-(\lambda+1)^2$ where n is an integer. Thus if

$$n\lambda = \lambda + 1$$

the coefficients k_n and l_n diverge. Therefore, periodic orbits cannot be obtained in this way.

However, the Kirkwood gaps in the asteroid belt concide with positions where asteroids, if present, would have periods commeasurable with the period of Jupiter's orbit. Recently, this problem was investigated theoretically by Brouwer (Reference 8).

The presence of the Kirkwood gaps are obviously due to the perturbation by Jupiter. Since perturbation increases with μ, it follows that stronger gaps would be present in systems of two revolving bodies of increasing values of μ. Hence, we may immediately predict stronger gaps around the earth in the earth-moon system than those around the sun in the asteroid belt in the sun-Jupiter system. Such gaps around the earth can be computed easily from the condition of commeasurability of the moon's period and the period of any satellite in such a zone. Of course, a result of this argument is that these zones should be avoided in launching satellites that are intended to stay in orbit for a long time. Being short, the period of the synchronous orbit is not near any strongly commeasurable gap.

Another interesting consequence of the commeasurable gaps around the earth may be found in the distribution of dust particles in the earth-moon system. It is not now known whether dust particles all move at random or partially revolve around the earth. If the latter should be the case, the presence of zones around the earth void of dust particles like the Kirkwood gaps in the asteroid belt around the sun, would be inevitable. Consequently, their detection may be an interesting subject of investigation in the field of space research.

ACKNOWLEDGMENTS

It is the author's pleasure to express his sincere thanks to Dr. D. Brouwer, Dr. G. Hori, Dr. V. Szebehely, and Dr. T. Y. Wu for their valuable discussions.

REFERENCES

1. Szebehely, V. G., "The Restricted Problem of Three Bodies," paper given at the Summer Institute in Dynamical Astronomy, Yale Univ., 1961.
2. Szebehely, V. G., "Zero Velocity Curves and Orbits in the Restricted Problem of Three Bodies," *Astronom. J.* 68(3):147–151, April 1963.
3. Hill, G. W., "Researches in the Lunar Theory, Chapter II," *Amer. J. Math.* 1:5,129–147, and 245–260, 1878.
4. Hill, G. W., "On the Part of the Motion of the Lunar Perigee Which is a Function of the Mean Motions of the Sun and Moon," *Acta Math.* 8:1–36, 1886.
5. Brouwer, D., and Clemence, G. M., "Methods of Celestial Mechanics," New York: Academic Press, 1961, Chapter 12.
6. Huang, S.-S., "The Hypothetical Four-Body Problem," *Astronom. J.* to be published, 1964.
7. Huang, S.-S., and Wade, C., Jr., "Preliminary Study of Periodic Orbits of Interest for Moon Probes. II," *Astronom. J.* 68(6):388–391, August 1963.
8. Brouwer, D., "The Problem of the Kirkwood Gaps in the Asteroid Belt," *Astronom. J.* 68(3):152–159, April 1963.

ON THE APPLICATION OF PFAFF'S METHOD IN THE THEORY OF VARIATION OF ASTRONOMICAL CONSTANTS*

PETER MUSEN

Cartan's integral invariant is taken as a foundation of the theory of variation of astronomical parameters. The differential equations for the general perturbations are obtained as the first system of Pfaffian equations associated with the linear differential form appearing in the intergral invariant.

The equations for general perturbations of the Gibbsian unit vectors, of the Gibbsian rotation vector and of Euler's parameters are derived. The utilization of the Gibbsian rotation vector represents an extension of Strömgren's and of the author's ideas on special perturbations to the problems of general perturbations. Euler's parameters find their application in Hansen's lunar theory.

The case of redundant elements is treated by introducing constraints and Lagrangian multipliers. Cartan's integral invariant expressed in terms of vectorial kinematic elements represents a powerful tool in the search for new sets of elements and it leads to differential equations having a compact form, which is convenient for the programming and use of electronic machines.

NOTATIONS

m_i = the mass of the i-th point in the system of points
$\mathbf{r}_i$ = the position vector of the i-th point
$\mathbf{v}_i$ = the velocity of the i-th point in the system of points
T = the kinetic energy
U = the force function
F = U − T
Ω = the disturbing function
$[\Omega]$ = the disturbing function averaged over the orbit of the disturbed body
$\text{grad}_{\mathbf{s}}\ \phi$—the gradient of ϕ with respect $\mathbf{s}$; $\text{grad}_i\ \phi$—the gradient of ϕ with respect to $\mathbf{s}_i$
k = the Gaussian constant
M = the mass of the sun
m = the mass of the disturbed body
$\mu = k^2(M+m)$
$\mathbf{r}$ = the position vector of the disturbed body
$r = |\mathbf{r}|$
$\mathbf{v}$ = the heliocentric velocity of the disturbed body
$\ell, \omega, i, \Omega, e, a, n$ = the standard elliptic elements of the disturbed body
u = the eccentric anomaly of the disturbed body
$\mathbf{c} = \mathbf{r} \times \mathbf{v}$ = the vector of the angular momentum
$\mathbf{P}, \mathbf{Q}, \mathbf{R}$ = the Gibbsian vectorial elements of the disturbed body
$\mathbf{e} = \mu e \mathbf{P}$ = the Laplacian vector

*Published as *NASA Technical Note* D-2301, April 1964

$\mathbf{g}$ = the Gibbsian rotation vector
$\ell', \omega', \Omega', \cdots, \mathbf{P}', \mathbf{Q}', \mathbf{R}', \cdots$
$\ell'', \omega'', \Omega'', \cdots, \mathbf{P}'', \mathbf{Q}'', \mathbf{R}'', \cdots$
the elements of disturbing bodies
$\mathbf{r}'^0$ = the unit vector in the direction from the sun to the disturbing body
$p = a(1-e^2)$

INTRODUCTION

In treating the problem of general perturbations of planets and satellites it is felt that a general and flexible method is needed which would permit an easy transformation from one system to another and would also facilitate a proper choice of elements under the different circumstances.

The standard way to develop the equations for the general perturbations in elements is based on the computation of the matrix either of Lagrangian or of Poissonian brackets. This computation is relatively simple for the classical elliptic elements, but it can become cumbersome if some algebraic combinations of the classical elements are taken as a new set of elements.

An additional difficulty may arise if, for the sake of the symmetry of the development of the disturbing function, some redundant elements are introduced. As a result, additional constraints appear in the problem. It is necessary to point out, that the presence of these constraints does not necessarily mean that the problem of the determination of the constants of integration will become more complicated. Sometimes the presence of constraints makes this determination easier. Hill's method of general perturbations is a well known example in celestial mechanics where a redundant constant of integration is present.

However, the difficulties associated with the presence of a redundant constant in this particular case should not be generalized to all problems in celestial mechanics. One should not be hesitant to make use of redundant elements if the problem of the determination of constants of integration for all orders of perturbations can be made symmetrical and if the programming can be made more efficient.

In order to have a more general view and to facilitate the search for new types of elements, an application of Pfaff's method is made in this article to the problems of celestial mechanics. This method permits the formation of the equations for the variation of elements in a straightforward manner if these elements are obtained from the classical elements by any type of functional transformation.

Surprisingly, such a direct method has not found wider applications in the theory of variation of astronomical constants. The pioneering work of Bilimovich (Reference 1) stands rather apart from the main stream. Bilimovich succeeded in deducing Milankovich's equations (Reference 2) for the variation of the Laplacian and momentum vectors in a rather simple way. He also deduced the classical equations for the variation of the elliptic elements in a very simple manner.

The existence of Cartan's (Reference 3) integral invariant for dynamical systems is the foundation of the Pfaffian method. Using a hydrodynamical analogy we can say that the theorem of the circulation for an ideal fluid in the phase space is taken as a basis for the theory of perturbations presented herein. The equations for the variation of the elements represent the first system of Pfaffian equations associated with the differential form appearing under the integral sign in the expression for circulation.

In this work we suggest the use of a slightly more general form of Pfaffian equations than in Bilimovich's work by permitting the constraints to be present and by forming the equations with Lagrangian multipliers.

We shall develop the equations for the general perturbations of Gibbsian vectors $\mathbf{P}$ and $\mathbf{Q}$ and also of the elements the author has suggested in his article on Strömgren's (Reference 4) perturbations (Reference 5) and in the articles on Hansen's (Reference 6) lunar theory (References 7 and 16).

In recent times electronic equipment is being used more and more for the purpose of developing general perturbations. We might expect that in the near future the utilization of machines in this domain will become even wider. This fact will have an increasingly greater impact on the theoretical thought and thus the search for different types of new elements is in order now.

PFAFFIAN EQUATIONS OF MOTION WITH REDUNDANT COORDINATES

Let us consider the material system of N points having the position vectors $\mathbf{r}_i$ and the velocity vectors $\mathbf{v}_i$ $(i=1, 2, \cdots, N)$. The Pfaffian linear form associated with this system is

$$\phi = \sum_{i=1}^{N} m_i \mathbf{v}_i \cdot d\mathbf{r}_i - (T - U)\, dt \,. \quad (1)$$

A functional transformation,

$$\mathbf{v}_i = \mathbf{v}_i(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n; q_1, q_2, \cdots, q_m) \,,$$

$$\mathbf{r}_i = \mathbf{r}_i(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n; q_1, q_2, \cdots, q_m) \,,$$

with the imposed conditions

$$f_p(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n; q_1, q_2, \cdots, q_m) = a_p$$

brings Equation 1 to the form

$$\phi \quad \sum_{i=1}^{n} \mathbf{X}_i(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n; q_1, q_2, \cdots, q_m) \cdot d\mathbf{x}_i$$

$$+ \sum_{j=1}^{m} Q_j(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n; q_1, q_2, \cdots, q_m) dq_j + F dt \,, \quad (2)$$

where $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_n$ are vectors in k-dimensional Euclidean space and $q_1, q_2, \cdots, q_m$ are scalars.

The condition

$$kn + m - s = 6N$$

must be satisfied. In the process of transformation from Equation 1 to Equation 2 any additive total differential can be neglected

For present purposes we can assume that all constraints are scleronomic. We write the first system of Pfaffian equations, using the notations suggested by Bilimovich, in the form:

$$\sum_{i=1}^{n} (\mathrm{grad}_i \phi - d\mathbf{X}_i) \cdot \delta\mathbf{x}_i + \sum_{j=1}^{m} \left(\frac{\partial \phi}{\partial q_j} - dQ_j\right) \delta q_j + \left(\frac{\partial \phi}{\partial t} - dF\right) \delta t = 0$$

with the additional conditions

$$\sum_{i=1}^{n} \mathrm{grad}_i f_p \cdot \delta\mathbf{x}_i + \sum_{j=1}^{m} \frac{\partial f_p}{\partial q_j} \delta q_j = 0 \quad (i = 1, 2, \cdots, n;\ j = 1, 2, \cdots, m;\ p = 1, 2, \cdots s)$$

imposed on the variations by the constraints. Introducing Lagrangian multipliers, we have

$$\mathrm{grad}_i \phi - d\mathbf{X}_i + \sum_{p=1}^{s} \lambda_p\, dt\, \mathrm{grad}_i f_p = 0 \quad (3)$$

$$\frac{\partial \phi}{\partial q_j} - dQ_j + \sum_{p=1}^{s} \lambda_p\, dt\, \frac{\partial f_p}{\partial q_j} = 0 \,, \quad (4)$$

$$\frac{\partial F}{\partial t} dt - dF = 0 \,. \quad (5)$$

The Lagrangian multipliers $\lambda_p (p=1, 2, \cdots, s)$ can be determined with the help of the equations

$$\sum_{i=1}^{n} \mathrm{grad}_i f_p \cdot d\mathbf{x}_i + \sum_{j=1}^{m} \frac{\partial f_p}{\partial q_j} dq_j = 0, \ (p = 1, 2, \cdots, s). \quad (6)$$

The existence of a Cartan integral invariant on the hypersurface

$$f_p = a_p \quad (p = 1, 2, \cdots, s)$$

defined in (kn+m) -dimensional space serves as a basis for deducing the Pfaffian equations.

Sometimes, for the sake of the symmetry, it might be advantageous to increase beyond six the number of the osculating elements. This is done, for example, by Milankovich (Reference 2), by Herrick (Reference 8) and by the author (Reference 9). Then the relations containing the redundant elements can be understood as the scleronomic constraints and Equations 3–6 can be applied.

PFAFFIAN EXPRESSION FOR PLANETARY MOTION

In this article we deduce Bilimovich's expression for ϕ for disturbed planetary motion using the Eckert-Brouwer (Reference 10) expression for the variation of the position vector of a planet.

Originally the Eckert-Brouwer formula was designed for the purpose of orbit correction. However, experience has shown that this elegant formula can also be used in the vectorial theory of general perturbations as in the present article and in the article published recently by Musen and Carpenter (Reference 11).

We also suggest herein some new sets of elements, canonical and uncanonical, which follow directly from the form of ϕ.

The differential form (Equation 1) in the case of disturbed planetary motion is

$$\phi = v \cdot dr - \left(\frac{v^2}{2} - \frac{\mu}{r} - \Omega\right) dt \ .$$

Making use of the expression for dr as obtained by Eckert and Brouwer,

$$dr = d\psi \times r + \frac{v}{n} dl + \frac{r}{a} da + \left(\frac{r + p - 2a}{ep} r + \frac{r + p}{ep} \cdot \frac{r \cdot v}{a^2 n^2} v\right) de \ ,$$

we deduce

$$v \cdot dr = r \times v \cdot d\psi + \frac{v^2}{n} dl + r \cdot v \frac{da}{a} + \frac{r \cdot v}{ep}\left[(r + p)\left(1 + \frac{v^2}{a^2 n^2}\right) - 2a\right] de \ . \qquad (7)$$

Substituting

$$\begin{aligned} r \cdot v &= \sqrt{\mu a}\, e \sin u \ , \\ r \times v &= R\sqrt{\mu a\left(1 - e^2\right)} \ , \end{aligned} \qquad (8)$$

$$\begin{aligned} v^2 &= \mu\left(\frac{2}{r} - \frac{1}{a}\right) \ , \\ n &= \frac{\sqrt{\mu}}{a^{3/2}} \ , \end{aligned} \qquad (9)$$

into Equation 7 we obtain

$$v \cdot dr = \frac{\mu}{n}\left(\frac{2}{r} - \frac{1}{a}\right) dl + \left(\sqrt{\frac{\mu}{a}}\, e\, da + \sqrt{\mu a}\, 2\frac{a}{r} de\right) \sin u + \sqrt{\mu a\left(1 - e^2\right)}\, R \cdot d\Psi \ . \qquad (10)$$

Differentiating Equation 8 and taking

$$\frac{r}{a} = 1 - e \cos u$$

into account, we have

$$2d(r \cdot v) = \left(e\sqrt{\frac{\mu}{a}} da + 2\sqrt{\mu a}\, de\right) \sin u + 2\sqrt{\mu a}\left(1 - \frac{r}{a}\right) du \ . \qquad (11)$$

Differentiating Kepler's equation we have

$$du = \frac{a}{r} \sin u \, de + \frac{a}{r} \, dl \ . \tag{12}$$

Eliminating du from Equation 11 by means of Equation 12 we obtain

$$2d(r \cdot v) = \left(e \sqrt{\frac{\mu}{a}} \, da + 2 \sqrt{\mu a} \, \frac{a}{r} \, de \right) \sin u + 2 \sqrt{\mu a} \left(\frac{a}{r} - 1 \right) dl$$

Taking Equation 9 into account we deduce

$$2d(r \cdot v) = \left(\frac{v^2}{n} - \sqrt{\mu a} \right) dl + \left(e \sqrt{\frac{\mu}{a}} \, da + 2 \frac{a}{r} \sqrt{\mu a} \, de \right) \sin u \ .$$

Comparing the last equation with Equation 10 we conclude that

$$v \cdot dr = \sqrt{\mu a} \, dl + \sqrt{\mu a (1 - e^2)} \, R \cdot d\psi + 2d(r \cdot v) \ . \tag{13}$$

Decomposing the infinitesimal rotation $d\Psi$ along the axes **P**, **Q**, and **R** we have

$$d\psi = dP \cdot QR + dQ \cdot RP + dR \cdot PQ$$

and Equation 13 becomes

$$v \cdot dr = \sqrt{\mu a} \, dl + \sqrt{\mu a (1 - e^2)} \, Q \cdot dP + 2d(r \cdot v) \ . \tag{14}$$

Neglecting the total differential 2d (**r**·**v**) we have

$$\phi = \sqrt{\mu a} \, dl + \sqrt{\mu a (1 - e^2)} \, Q \cdot dP + F \, dt \tag{15}$$

with $F = \mu/2a + \Omega$. Equation 15 was deduced by Bilimovich on the basis of a different consideration.

EQUATIONS FOR VARIATION OF VECTORIAL ELEMENTS

Milankovich has deduced the vectorial equations for the determination of general perturbations of vectors **c** and **e**.

The vectors **P** and **Q** constitute a better choice for the development of the disturbing function and the computation of the position and velocity vectors.

The elements we choose are

$$l, \ a, \ e, \ P \text{ and } Q$$

For reasons of symmetry we take the expression for ϕ in the form

$$\phi = \sqrt{\mu a} \, dl + \frac{1}{2} \sqrt{\mu a (1 - e^2)} \, (Q \cdot dP - P \cdot dQ) + r \, dt$$

with the three additional conditions

$$P \cdot P = 1, \quad Q \cdot Q = 1, \quad P \cdot Q = 0 \ . \tag{16}$$

The Pfaffian equations (Equations 3–6) for the variation of constants in this particular case become:

$$\text{grad}_P \, \phi - d \left[\frac{1}{2} \sqrt{\mu a (1 - e^2)} \, Q \right] + (uP + wQ) \, dt = 0 \ , \tag{17}$$

$$\mathrm{grad}_{\mathbf{Q}}\,\phi + d\left[\frac{1}{2}\ \sqrt{\mu a\left(1-e^2\right)}\ \mathbf{P}\right] + (v\mathbf{Q} + w\mathbf{P})\,dt = 0\,, \tag{18}$$

$$\frac{\partial \phi}{\partial a} = 0\,, \tag{19}$$

$$\frac{\partial \phi}{\partial e} = 0\,, \tag{20}$$

$$\frac{\partial \phi}{\partial l} - d\sqrt{\mu a} = 0\,, \tag{21}$$

where u, v and w here are the Lagrangian multipliers associated with the constraints (Equation 16).

After some easy transformations we deduce from Equations 17–21:

$$na^2\sqrt{1-e^2}\,\frac{d\mathbf{P}}{dt} + (A + w)\,\mathbf{P} + v\mathbf{Q} + \mathrm{grad}_{\mathbf{Q}}\,F = 0\,, \tag{22}$$

$$na^2\sqrt{1-e^2}\,\frac{d\mathbf{Q}}{dt} + (A - w)\,\mathbf{Q} - u\mathbf{P} - \mathrm{grad}_{\mathbf{P}}\,F = 0\,, \tag{23}$$

$$\frac{1}{2}\,na\,\frac{dl}{dt} + \frac{1}{2}\,na\,\sqrt{1-e^2}\,\mathbf{Q}\cdot\frac{d\mathbf{P}}{dt} + \frac{\partial F}{\partial a} = 0\,, \tag{24}$$

$$\mathbf{Q}\cdot\frac{d\mathbf{P}}{dt} = -\,\mathbf{P}\cdot\frac{d\mathbf{Q}}{dt} = +\,\frac{\sqrt{1-e^2}}{na^2\,e}\,\frac{\partial F}{\partial e}\,, \tag{25}$$

$$\frac{\partial F}{\partial l} - \frac{1}{2}\,na\,\frac{da}{dt} = 0\,, \tag{26}$$

where

$$A = \frac{1}{4}\left(na\sqrt{1-e^2}\,\frac{da}{dt} - \frac{2na^2\,e}{\sqrt{1-e^2}}\,\frac{de}{dt}\right). \tag{27}$$

By multiplying Equation 22 by $\mathbf{P}$ and Equation 23 by $\mathbf{Q}$ we deduce:

$$A + w = -\mathbf{P}\cdot\mathrm{grad}_{\mathbf{Q}}\,F\,,$$

$$A - w = +\mathbf{Q}\cdot\mathrm{grad}_{\mathbf{P}}\,F\,,$$

$$A = +\frac{1}{2}\,\mathbf{Q}\cdot\mathrm{grad}_{\mathbf{P}}\,F - \frac{1}{2}\,\mathbf{P}\cdot\mathrm{grad}_{\mathbf{Q}}\,F\,, \tag{28}$$

$$w = -\frac{1}{2}\,\mathbf{Q}\cdot\mathrm{grad}_{\mathbf{P}}\,F - \frac{1}{2}\,\mathbf{P}\cdot\mathrm{grad}_{\mathbf{Q}}\,F\,. \tag{29}$$

Multiplying Equation 22 by $\mathbf{Q}$ and Equation 23 by $\mathbf{P}$ and taking Equation 25 into account, we have

$$v = -\mathbf{Q}\cdot\mathrm{grad}_{\mathbf{Q}}\,F - \frac{1-e^2}{e}\,\frac{\partial F}{\partial e}\,, \tag{30}$$

$$u = -\mathbf{P}\cdot\mathrm{grad}_{\mathbf{P}}\,F - \frac{1-e^2}{e}\,\frac{\partial F}{\partial e}\,. \tag{31}$$

Substituting Equations 29–31 into Equations 22 and 23 and taking the identity

$$\mathbf{I} = \mathbf{PP} + \mathbf{QQ} + \mathbf{RR}$$

into account, we obtain

$$\frac{d\mathbf{P}}{dt} = -\frac{1}{na^2\sqrt{1-e^2}}\,\mathbf{RR}\cdot\operatorname{grad}_{\mathbf{Q}} F + \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial F}{\partial e}\,\mathbf{Q}\,, \tag{32}$$

$$\frac{d\mathbf{Q}}{dt} = +\frac{1}{na^2\sqrt{1-e^2}}\,\mathbf{RR}\cdot\operatorname{grad}_{\mathbf{P}} F - \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial F}{\partial e}\,\mathbf{P}\,. \tag{33}$$

These two last equations are new. Equations 24 and 26 can be written in the classical form:

$$\frac{da}{dt} = \frac{2}{na}\frac{\partial F}{\partial l}\,, \tag{34}$$

$$\frac{dl}{dt} = -\frac{1-e^2}{na^2 e}\frac{\partial F}{\partial e} - \frac{2}{na}\frac{\partial F}{\partial a}\,. \tag{35}$$

From Equations 27, 28, and 34 we deduce

$$\frac{de}{dt} = +\frac{1-e^2}{na^2 e}\frac{\partial F}{\partial l} + \frac{\sqrt{1-e^2}}{na^2 e}\left(\mathbf{P}\cdot\operatorname{grad}_{\mathbf{Q}} F - \mathbf{Q}\cdot\operatorname{grad}_{\mathbf{P}} F\right). \tag{36}$$

The system Equations 32–36 forms a complete set to be used in the development of perturbations.

From Equations 32 and 33 we see that the constraints (Equation 16) are satisfied. Because of the existence of these constraints the determination of the additive constants of integration associated with Equations 32 and 33 becomes extremely simple. This simplicity justifies in part the choice of $\mathbf{P}$ and $\mathbf{Q}$ as the basic elements. Changing the notation, we shall designate any disturbed element E as a series

$$E + E_1 + E_2 + E_3 + \cdots\,,$$

where E now means the undisturbed value and E_k are the perturbations of k-th order. We deduce from Equation 16:

$$\begin{aligned}
&\mathbf{P}\cdot\mathbf{P}_1 = 0\,,\\
&\mathbf{P}\cdot\mathbf{P}_2 + \tfrac{1}{2}\mathbf{P}_1^2 = 0\,, \qquad (37)\\
&\mathbf{P}\cdot\mathbf{P}_3 + \mathbf{P}_1\cdot\mathbf{P}_2 = 0\,,\\
&\qquad\cdots\cdots\cdots\cdots\cdots
\end{aligned}$$

$$\begin{aligned}
&\mathbf{Q}\cdot\mathbf{Q}_1 = 0\,,\\
&\mathbf{Q}\cdot\mathbf{Q}_2 + \tfrac{1}{2}\mathbf{Q}_1^2 = 0\,, \qquad (38)\\
&\mathbf{Q}\cdot\mathbf{Q}_3 + \mathbf{Q}_1\cdot\mathbf{Q}_2 = 0\,,\\
&\qquad\cdots\cdots\cdots\cdots\cdots
\end{aligned}$$

$$\begin{aligned}
&\mathbf{Q}\cdot\mathbf{P}_1 + \mathbf{P}\cdot\mathbf{Q}_1 = 0\,,\\
&\mathbf{Q}\cdot\mathbf{P}_2 + \mathbf{P}\cdot\mathbf{Q}_2 + \mathbf{P}_1\cdot\mathbf{Q}_1 = 0\,, \qquad (39)\\
&\mathbf{Q}\cdot\mathbf{P}_3 + \mathbf{P}\cdot\mathbf{Q}_3 + \mathbf{P}_1\cdot\mathbf{Q}_2 + \mathbf{P}_2\cdot\mathbf{Q}_1 = 0\,.\\
&\qquad\cdots\cdots\cdots\cdots\cdots
\end{aligned}$$

For $\mathbf{P}_k$, $\mathbf{Q}_k$ ($k=1, 2, \cdots$) we obtain from Equations 32 and 33 equations of the form:

$$\frac{d\mathbf{P}_k}{dt} = f_k\left(\mathbf{P}, \mathbf{Q}, a, l, e; \mathbf{P}_1, \mathbf{Q}_1, a_1, l_1, e_1; \cdots; \mathbf{P}_{k-1}, \mathbf{Q}_{k-1}, a_{k-1}, l_{k-1}, e_{k-1}\right),$$

$$\frac{d\mathbf{Q}_k}{dt} = \phi_k\left(\mathbf{P}, \mathbf{Q}, a, l, e; \mathbf{P}_1, \mathbf{Q}_1, a_1, l_1, e_1; \cdots; \mathbf{P}_{k-1}, \mathbf{Q}_{k-1}, a_{k-1}, l_{k-1}, e_{k-1}\right).$$

Let

$$(\mathbf{P}_k) = \int \mathbf{f}_k \, dt \, ,$$

$$(\mathbf{Q}_k) = \int \boldsymbol{\phi}_k \, dt \, ,$$

where the integration is performed in a formal manner. We can write

$$\mathbf{P}_k = (\mathbf{P}_k) + p_{k1}\,\mathbf{P} + p_{k2}\,\mathbf{Q} + p_{k3}\,\mathbf{R} \, ,$$

$$\mathbf{Q}_k = (\mathbf{Q}_k) + q_{k1}\,\mathbf{P} + q_{k2}\,\mathbf{Q} + q_{k3}\,\mathbf{R} \, ,$$

where p_{k1}, q_{k1} are the constants of integration. The expressions $(\mathbf{P}_k)$, $(\mathbf{Q}_k)$ contain the secular, the mixed and the purely periodic terms, but there are no constant terms.

Let [f] designate the constant part in the multiple Fourier series of some function $f(l, l', l'', \cdots)$. Taking into account

$$[(\mathbf{P}_k)] = [(\mathbf{Q}_k)] = 0 \, ,$$

we deduce from Equations 37–39:

$$\begin{aligned} p_{11} &= 0 \, , \\ p_{21} &= -\tfrac{1}{2}[\mathbf{P}_1^2] \, , \\ p_{31} &= -[\mathbf{P}_1 \cdot \mathbf{P}_2] \, , \\ &\cdots\cdots\cdots\cdots \end{aligned} \tag{40}$$

$$\begin{aligned} q_{12} &= 0 \, , \\ q_{22} &= -\tfrac{1}{2}[\mathbf{Q}_1^2] \, , \\ q_{32} &= -[\mathbf{Q}_1 \cdot \mathbf{Q}_2] \, , \\ &\cdots\cdots\cdots\cdots \end{aligned} \tag{41}$$

$$\begin{aligned} q_{11} &= -p_{12} \\ q_{21} &= -p_{22} - [\mathbf{P}_1 \cdot \mathbf{Q}_1] \, , \\ q_{31} &= -p_{32} - [\mathbf{P}_1 \cdot \mathbf{Q}_2] - [\mathbf{P}_2 \cdot \mathbf{Q}_1] \, . \\ &\cdots\cdots\cdots\cdots \end{aligned} \tag{42}$$

We see from Equations 40–42 that in the process of the determination of the perturbations of k-th order in $\mathbf{P}$ and $\mathbf{Q}$ the three constants p_{k2}, p_{k3}, q_{k3} can be considered as independent. The constants q_{k1} are determined in terms of p_{k2} and of the constants of the lower orders. The constants p_{k1} and q_{k2} are also functions of the constants of the lower orders. Although there are redundant elements there are no redundant constants which cannot be determined easily. For the disturbed position vector $\mathbf{r}$ we have the standard expression

$$r = Pa(\cos u - e) + Qa\sqrt{1 - e^2}\sin u \, ,$$

$$u - e\sin u = l \, ,$$

where all symbols now designate the disturbed elements.

In order to verify the correctness of Equations 32, 33, and 36 we shall deduce them from the corresponding classical equations. From

$$\mathbf{P} = \begin{Bmatrix} +\cos\omega\cos\Omega - \sin\omega\sin\Omega\cos i \\ +\cos\omega\sin\Omega + \sin\omega\cos\Omega\cos i \\ +\sin\omega\sin i \end{Bmatrix} ,$$

$$\mathbf{Q} = \begin{Bmatrix} -\sin\omega\cos\Omega - \cos\omega\sin\Omega\cos i \\ -\sin\omega\sin\Omega + \cos\omega\cos\Omega\cos i \\ +\cos\omega\sin i \end{Bmatrix} ,$$

$$\mathbf{R} = \begin{Bmatrix} +\sin\Omega\sin i \\ -\cos\Omega\sin i \\ +\cos i \end{Bmatrix} ,$$

we have

$$\frac{dP}{dt} = Q\frac{d\omega}{dt} + R\sin\omega\frac{di}{dt} + k \times P\frac{d\Omega}{dt} , \quad (43)$$

$$\frac{dQ}{dt} = -P\frac{d\omega}{dt} + R\cos\omega\frac{di}{dt} + k \times Q\frac{d\Omega}{dt} , \quad (44)$$

and as a consequence of the two last equations:

$$\sin i\frac{d\Omega}{dt} = R \cdot \left(-\frac{dP}{dt}\cos\omega + \frac{dQ}{dt}\sin\omega\right) , \quad (45)$$

$$\frac{di}{dt} = R \cdot \left(+\frac{dP}{dt}\sin\omega + \frac{dQ}{dt}\cos\omega\right) , \quad (46)$$

$$\frac{d\omega}{dt} + \cos i\frac{d\Omega}{dt} = Q \cdot \frac{dP}{dt} = -P \cdot \frac{dQ}{dt} . \quad (47)$$

From

$$\frac{\partial F}{\partial E} = \frac{\partial P}{\partial E} \cdot \mathrm{grad}_P F + \frac{\partial Q}{\partial E} \cdot \mathrm{grad}_Q F ,$$

where

$$E = \omega, \Omega, i ,$$

and Equations 43 and 44, we deduce

$$\frac{\partial F}{\partial \omega} = Q \cdot \mathrm{grad}_P F - P \cdot \mathrm{grad}_Q F , \quad (48)$$

$$\frac{\partial F}{\partial i} = \sin\omega\left(R \cdot \mathrm{grad}_P F\right) + \cos\omega\left(R \cdot \mathrm{grad}_Q F\right) , \quad (49)$$

$$\frac{\partial F}{\partial \Omega} = k \times P \cdot \mathrm{grad}_P F + k \times Q \cdot \mathrm{grad}_Q F . \quad (50)$$

Substituting

$$k = (P\sin\omega + Q\cos\omega)\sin i + R\cos i$$

into Equation 50, we obtain

$$-\sin\omega\left(R \cdot \mathrm{grad}_Q F\right) + \cos\omega\left(R \cdot \mathrm{grad}_P F\right) = -\frac{1}{\sin i}\frac{\partial F}{\partial \Omega} + \mathrm{ctg}\, i\frac{\partial F}{\partial \omega} .$$

From Equations 36, 45–47 and Equations 32, 33, 48–50 we obtain the classical equations:

$$\frac{d\Omega}{dt} = \frac{1}{na^2\sqrt{1-e^2}\sin i}\frac{\partial F}{\partial i} ,$$

$$\frac{di}{dt} = -\frac{1}{na^2\sqrt{1-e^2}\sin i}\frac{\partial F}{\partial \Omega} + \frac{\cos i}{na^2\sqrt{1-e^2}\sin i}\frac{\partial F}{\partial \omega} ,$$

$$\frac{d\omega}{dt} = +\frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial F}{\partial e} - \cos i\frac{d\Omega}{dt} ,$$

$$\frac{de}{dt} = +\frac{1-e^2}{na^2 e}\frac{\partial F}{\partial l} - \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial F}{\partial \omega} .$$

Let the matrix of direction-cosines between (**P**, **Q**, **R**) and (**P′**, **Q′**, **R′**) be:

$$\mathbf{P}\cdot\mathbf{P}' = \alpha_1 , \quad \mathbf{P}\cdot\mathbf{Q}' = \alpha_2 , \quad \mathbf{P}\cdot\mathbf{R}' = \alpha_3 ,$$

$$\mathbf{Q}\cdot\mathbf{P}' = \beta_1 , \quad \mathbf{Q}\cdot\mathbf{Q}' = \beta_2 , \quad \mathbf{Q}\cdot\mathbf{R}' = \beta_3 ,$$

$$\mathbf{R}\cdot\mathbf{P}' = \gamma_1 , \quad \mathbf{R}\cdot\mathbf{Q}' = \gamma_2 , \quad \mathbf{R}\cdot\mathbf{R}' = \gamma_3 .$$

The disturbing function depends upon α_1, α_2, β_1 and β_2. Taking into account

$$\mathrm{grad}_{\mathbf{P}}\,\alpha_1 = \mathrm{grad}_{\mathbf{Q}}\,\beta_1 = \mathbf{P}' ,$$

$$\mathrm{grad}_{\mathbf{Q}}\,\beta_2 = \mathrm{grad}_{\mathbf{P}}\,\alpha_2 = \mathbf{Q}' ,$$

$$\mathrm{grad}_{\mathbf{Q}}\,\alpha_1 = \mathrm{grad}_{\mathbf{Q}}\,\alpha_2 = \mathrm{grad}_{\mathbf{P}}\,\beta_1 = \mathrm{grad}_{\mathbf{P}}\,\beta_2 = 0 ,$$

we obtain:

$$\mathrm{grad}_{\mathbf{P}}\,F = \frac{\partial F}{\partial \alpha_1}\,\mathbf{P}' + \frac{\partial F}{\partial \alpha_2}\,\mathbf{Q}' , \qquad (51)$$

$$\mathrm{grad}_{\mathbf{Q}}\,F = \frac{\partial F}{\partial \beta_1}\,\mathbf{P}' + \frac{\partial F}{\partial \beta_2}\,\mathbf{Q}' , \qquad (52)$$

$$\mathbf{R}\cdot\mathrm{grad}_{\mathbf{P}}\,F = \frac{\partial F}{\partial \alpha_1}\,\gamma_1 + \frac{\partial F}{\partial \alpha_2}\,\gamma_2 ,$$

$$\mathbf{R}\cdot\mathrm{grad}_{\mathbf{Q}}\,F = \frac{\partial F}{\partial \beta_1}\,\gamma_1 + \frac{\partial F}{\partial \beta_2}\,\gamma_2 ;$$

and the Equations 32 and 33 become

$$\frac{d\mathbf{P}}{dt} = -\frac{1}{na^2\sqrt{1-e^2}}\left(\frac{\partial F}{\partial \beta_1}\,\gamma_1 + \frac{\partial F}{\partial \beta_2}\,\gamma_2\right)\mathbf{R} + \frac{\sqrt{1-e^2}}{na^2\,e}\,\frac{\partial F}{\partial e}\,\mathbf{Q} ,$$

$$\frac{d\mathbf{Q}}{dt} = +\frac{1}{na^2\sqrt{1-e^2}}\left(\frac{\partial F}{\partial \alpha_1}\,\gamma_1 + \frac{\partial F}{\partial \alpha_2}\,\gamma_2\right)\mathbf{R} - \frac{\sqrt{1-e^2}}{na^2\,e}\,\frac{\partial F}{\partial e}\,\mathbf{P} .$$

If the frame (**P′**, **Q′**, **R′**) is a basic reference frame, then the last two equations become:

$$\frac{d\mathbf{P}}{dt} = -\frac{1}{na^2\sqrt{1-e^2}}\left(\frac{\partial F}{\partial \beta_1}\,\gamma_1 + \frac{\partial F}{\partial \beta_2}\,\gamma_2\right)\mathbf{R} + \frac{\sqrt{1-e^2}}{na^2\,e}\,\frac{\partial F}{\partial e}\,\mathbf{Q} + \mathbf{P}\times\boldsymbol{\psi}' , \qquad (53)$$

$$\frac{d\mathbf{Q}}{dt} = +\frac{1}{na^2\sqrt{1-e^2}}\left(\frac{\partial F}{\partial \alpha_1}\,\gamma_1 + \frac{\partial F}{\partial \alpha_2}\,\gamma_2\right)\mathbf{R} - \frac{\sqrt{1-e^2}}{na^2\,e}\,\frac{\partial F}{\partial e}\,\mathbf{P} + \mathbf{Q}\times\boldsymbol{\psi}' , \qquad (54)$$

where

$$\boldsymbol{\psi}' = \psi_1'\,\mathbf{P}' + \psi_2'\,\mathbf{Q}' + \psi_3'\,\mathbf{R}'$$

is the absolute angular velocity of rotation of the frame (**P′**, **Q′**, **R′**), and

$$\psi_1' = +\sin\omega'\sin i'\,\frac{d\Omega'}{dt} + \cos\omega'\,\frac{di'}{dt} ,$$

$$\psi_2' = +\cos\omega'\sin i'\,\frac{d\Omega'}{dt} - \sin\omega'\,\frac{di'}{dt} ,$$

$$\psi_3' = \cos i'\,\frac{d\Omega'}{dt} + \frac{d\omega'}{dt} .$$

From Equations 53 and 54 we deduce the following system of scalar equations for the variation of

the direction-cosines:

$$\frac{d\alpha_1}{dt} = -\frac{1}{na^2\sqrt{1-e^2}}\gamma_1\left(\frac{\partial F}{\partial \beta_1}\gamma_1 + \frac{\partial F}{\partial \beta_2}\gamma_2\right) + \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial F}{\partial e}\beta_1 + (\alpha_2\psi_3' - \alpha_3\psi_2') ,$$

$$\frac{d\alpha_2}{dt} = -\frac{1}{na^2\sqrt{1-e^2}}\gamma_2\left(\frac{\partial F}{\partial \beta_1}\gamma_1 + \frac{\partial F}{\partial \beta_2}\gamma_2\right) + \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial F}{\partial e}\beta_2 + (\alpha_3\psi_1' - \alpha_1\psi_3') ,$$

$$\frac{d\alpha_3}{dt} = -\frac{1}{na^2\sqrt{1-e^2}}\gamma_3\left(\frac{\partial F}{\partial \beta_1}\gamma_1 + \frac{\partial F}{\partial \beta_2}\gamma_2\right) + \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial F}{\partial e}\beta_3 + (\alpha_1\psi_2' - \alpha_2\psi_1') ,$$

and

$$\frac{d\beta_1}{dt} = +\frac{1}{na^2\sqrt{1-e^2}}\gamma_1\left(\frac{\partial F}{\partial \alpha_1}\gamma_1 + \frac{\partial F}{\partial \alpha_2}\gamma_2\right) - \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial F}{\partial e}\alpha_1 + (\beta_2\psi_3' - \beta_3\psi_2') ,$$

$$\frac{d\beta_2}{dt} = +\frac{1}{na^2\sqrt{1-e^2}}\gamma_2\left(\frac{\partial F}{\partial \alpha_1}\gamma_1 + \frac{\partial F}{\partial \alpha_2}\gamma_2\right) - \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial F}{\partial e}\alpha_2 + (\beta_3\psi_1' - \beta_1\psi_3') ,$$

$$\frac{d\beta_3}{dt} = +\frac{1}{na^2\sqrt{1-e^2}}\gamma_3\left(\frac{\partial F}{\partial \alpha_1}\gamma_1 + \frac{\partial F}{\partial \alpha_2}\gamma_2\right) - \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial F}{\partial e}\alpha_3 + (\beta_1\psi_2' - \beta_2\psi_1') .$$

Substituting Equations 51 and 52 into Equation 36 we obtain

$$\frac{de}{dt} = +\frac{1-e^2}{na^2 e}\frac{\partial F}{\partial l} + \frac{\sqrt{1-e^2}}{na^2 e}\left(\alpha_1\frac{\partial F}{\partial \beta_1} - \beta_1\frac{\partial F}{\partial \alpha_1} + \alpha_2\frac{\partial F}{\partial \beta_2} - \beta_2\frac{\partial F}{\partial \alpha_2}\right) .$$

Equations 32, 33, and 36 can be applied to the determination of the lunar and solar long period effects in the motion of close satellites. In the case of close satellites the short period terms, depending upon the mean anomaly of the satellite, can be neglected and the remaining portion $[\Omega]$ of the disturbing function can be developed (References 12 and 13) into the series of polynomials in e and in

$$\alpha = \alpha_1 \cos f' + \alpha_2 \sin f' ,$$

$$\beta = \beta_1 \cos f' + \beta_2 \sin f' ,$$

$$\gamma = \gamma_1 \cos f' + \gamma_2 \sin f' .$$

Taking into account

$$\mathrm{grad}_P\,\alpha = \mathrm{grad}_Q\,\beta = 1 ,$$

$$\mathrm{grad}_Q\,\alpha = \mathrm{grad}_P\,\beta = 0 ,$$

we obtain from Equations 32, 33 and 36

$$\frac{dP}{dt} = -\frac{\gamma R}{na^2\sqrt{1-e^2}}\frac{\partial[\Omega]}{\partial \beta} + \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial[\Omega]}{\partial e}Q ,$$

$$\frac{dQ}{dt} = +\frac{\gamma R}{na^2\sqrt{1-e^2}}\frac{\partial[\Omega]}{\partial \alpha} - \frac{\sqrt{1-e^2}}{na^2 e}\frac{\partial[\Omega]}{\partial e}P ,$$

and

$$\frac{de}{dt} = + \frac{\sqrt{1 - e^2}}{na^2 e} \left(a \frac{\partial [\Omega]}{\partial \beta} - \beta \frac{\partial [\Omega]}{\partial i} \right) .$$

A very elegant and convenient development of the disturbing function for close satellites was given by Kaula (Reference 14). The use of the Gibbsian rotation vector $\mathbf{g}$ (Reference 15) was suggested in an article (Reference 5) on the improvement of Strömgren's theory (Reference 4).

The basic vectors $\mathbf{P}$ and $\mathbf{Q}$ can be represented as relatively simple functions of $\mathbf{g}$ and the introduction of $\mathbf{g}$ removes the necessity of keeping constraints (Equation 16) as the additional conditions. The components of $\mathbf{g}$ in the inertial system are

$$g_1 = + \frac{\cos \frac{1}{2} (\omega - \Omega)}{\cos \frac{1}{2} (\omega + \Omega)} \operatorname{tg} \frac{i}{2} ,$$

$$g_2 = - \frac{\sin \frac{1}{2} (\omega - \Omega)}{\cos \frac{1}{2} (\omega + \Omega)} \operatorname{tg} \frac{i}{2} ,$$

$$g_3 = + \operatorname{tg} \frac{1}{2} (\omega + \Omega) .$$

Equation 47 shows that the Pfaffian (Equation 15) can be written as

$$\phi = \sqrt{\mu a}\, dl + \sqrt{\mu a (1 - e^2)}\, (d\omega + \cos i\, d\Omega) + F\, dt . \qquad (55)$$

This form was originally used by Bilimovich. Taking

$$\omega - \Omega = - 2 \arctan \frac{g_2}{g_1} ,$$

$$\omega + \Omega = 2 \arctan g_3 ,$$

and

$$\tan^2 \frac{i}{2} = \frac{g_1^2 + g_2^2}{1 + g_3^2}$$

into account, we deduce from Equation 55

$$\phi = \sqrt{\mu a}\, dl + \frac{2 \sqrt{\mu a (1 - e^2)}}{1 + g^2} s \cdot dg + F\, dt , \qquad (56)$$

where we put

$$s = k + g \times k . \qquad (57)$$

The Pfaffian equation corresponding to $\mathbf{g}$ is

$$\operatorname{grad}_g \left\{ \frac{2 \sqrt{\mu a (1 - e^2)}}{1 + g^2} s \cdot dg \right\} - d \left\{ \frac{2 \sqrt{\mu a (1 - e^2)}}{1 + g^2} s \right\} + \operatorname{grad}_g F\, dt = 0$$

or, after the differentiation is performed,

$$\frac{4na^2 \sqrt{1 - e^2}}{1 + g^2} \left(k \times \frac{dg}{dt} + \frac{sg \cdot \frac{dg}{dt} - gs \cdot \frac{dg}{dt}}{1 + g^2} \right)$$

$$- \left(na \sqrt{1 - e^2} \frac{da}{dt} - \frac{2na^2 e}{\sqrt{1 - e^2}} \frac{de}{dt} \right) \frac{s}{1 + g^2} + \operatorname{grad}_g F = 0 . \qquad (58)$$

Taking into account

$$\mathbf{s}\mathbf{g} \cdot \frac{d\mathbf{g}}{dt} - \mathbf{g}\mathbf{s} \cdot \frac{d\mathbf{g}}{dt} = (\mathbf{g} \times \mathbf{s}) \times \frac{d\mathbf{g}}{dt} ,$$

we can write Equation 58 in the form

$$\frac{4na^2\sqrt{1-e^2}}{1+g^2}\,\mathbf{h} \times \frac{d\mathbf{g}}{dt} - \left(na\sqrt{1-e^2}\,\frac{da}{dt} - \frac{2na^2 e}{\sqrt{1-e^2}}\,\frac{de}{dt}\right)\frac{\mathbf{s}}{1+g^2} + \operatorname{grad}_{\mathbf{g}} F = 0 , \quad (59)$$

where

$$\mathbf{h} = \mathbf{k} + \frac{\mathbf{g} \times \mathbf{s}}{1+g^2} = \frac{\mathbf{k} + \mathbf{g} \times \mathbf{k} + \mathbf{g}\mathbf{g} \cdot \mathbf{k}}{1+g^2} . \quad (60)$$

We have from Equation 59:

$$\frac{4na^2\sqrt{1-e^2}}{1+g^2}\,\mathbf{s} \times \left(\mathbf{h} \times \frac{d\mathbf{g}}{dt}\right) + \mathbf{s} \times \operatorname{grad}_{\mathbf{g}} F = 0 ,$$

or, after developing the triple product,

$$\mathbf{h} \cdot \mathbf{s}\,\frac{d\mathbf{g}}{dt} = \mathbf{h}\mathbf{s} \cdot \frac{d\mathbf{g}}{dt} + \frac{1+g^2}{4na^2\sqrt{1-e^2}}\,\mathbf{s} \times \operatorname{grad}_{\mathbf{g}} F . \quad (61)$$

From Equations 57 and 60 we have

$$\mathbf{h} \cdot \mathbf{s} = 1 , \quad (62)$$

and Equation 61 becomes

$$\frac{d\mathbf{g}}{dt} = \mathbf{h}\mathbf{s} \cdot \frac{d\mathbf{g}}{dt} + \frac{1+g^2}{4na^2\sqrt{1-e^2}}\,\mathbf{s} \times \operatorname{grad}_{\mathbf{g}} F . \quad (63)$$

The Pfaffian equation associated with e is simply

$$\frac{\partial \phi}{\partial e} = 0$$

or

$$-\frac{2na^2 e}{\sqrt{1-e^2}}\,\frac{\mathbf{s} \cdot d\mathbf{g}}{1+g^2} + \frac{\partial F}{\partial e}\,dt = 0 . \quad (64)$$

From Equations 63 and 64 we obtain

$$\frac{d\mathbf{g}}{dt} = \mathbf{h}\,\frac{(1+g^2)\sqrt{1-e^2}}{2na^2 e}\,\frac{\partial F}{\partial e} + \frac{1+g^2}{4na^2\sqrt{1-e^2}}\,\mathbf{s} \times \operatorname{grad}_{\mathbf{g}} F . \quad (65)$$

Multiplying Equation 59 by $\mathbf{h}$ and taking Equations 62 and 34 into account, we deduce

$$\frac{de}{dt} = + \frac{1-e^2}{na^2 e}\,\frac{\partial F}{\partial l} - \frac{(1+g^2)\sqrt{1-e^2}}{2na^2 e}\,\mathbf{h} \cdot \operatorname{grad}_{\mathbf{g}} F . \quad (66)$$

The use of the Gibbsian rotation vector represents an application and an extension of Strömgren's and of the author's ideas from the domain of special perturbations to the domain of general perturbations. The use of the vector $\mathbf{g}$ becomes especially convenient if the basic reference plane

nearly coincides with the osculating orbit plane at the initial epoch. Then the components of $\mathbf{g}$ become small—of the order of perturbations.

Now let a, e, l be the disturbed and **P**, **Q** the un-disturbed elements and put

$$r_0 = Pa(\cos u - e) + Qa\sqrt{1 - e^2}\,\sin u .$$

$$u - e\sin u = l$$

Writing the matrix of rotation Γ, which transforms $\mathbf{r}_0$ into the disturbed position vector $\mathbf{r}$, in the Gibbsian form

$$\Gamma = I + \frac{2}{1 + g^2}\left[g \times I + g \times (g \times I)\right] , \quad (67)$$

we have

$$r = \Gamma \cdot r_0 ,$$

or, in the developed form,

$$r = r_0 + \frac{2}{1 + g^2}\left[g \times r_0 + g \times (g \times r_0)\right] ,$$

or

$$r = \frac{1 - g^2}{1 + g^2} r_0 + \frac{2}{1 + g^2}\left[g \times r_0 + gg \cdot r_0\right] .$$

EQUATIONS FOR VARIATION OF EULER'S PARAMETERS

It is also of interest to deduce the equations for the variation of elements in terms of Euler parameters:

$$\begin{aligned} \lambda_1 &= \sin\tfrac{1}{2} i \cos\tfrac{1}{2}(\omega - \Omega) , \\ \lambda_2 &= \sin\tfrac{1}{2} i \sin\tfrac{1}{2}(\omega - \Omega) , \\ \lambda_3 &= \cos\tfrac{1}{2} i \sin\tfrac{1}{2}(\omega + \Omega) , \\ \lambda_4 &= \cos\tfrac{1}{2} i \cos\tfrac{1}{2}(\omega + \Omega) . \end{aligned} \quad (68)$$

The additional constraint is

$$\lambda_1^2 + \lambda_2^2 + \lambda_3^2 + \lambda_4^2 = 1 . \quad (69)$$

Two of these parameters, namely λ_1 and λ_2, were used by Hansen (Reference 6) in his lunar theory. The author has suggested the use of all four in the theory of artificial satellites (Reference 7) and in a modification of Hansen's lunar theory (Reference 16). Writing the Pfaffian (Equation 55) in the form

$$\phi = \sqrt{\mu a}\, dl + \sqrt{\mu a(1 - e^2)}\,\frac{d(\omega + \Omega) + \tan^2\frac{i}{2}\, d(\omega - \Omega)}{1 + \tan^2\frac{i}{2}} + F\, dt$$

and taking Equation 68 into account, we deduce

$$\phi = \sqrt{\mu a}\, dl + 2\sqrt{\mu a(1 - e^2)}\left(\lambda_1 d\lambda_2 - \lambda_2 d\lambda_1 + \lambda_4 d\lambda_3 - \lambda_3 d\lambda_4\right) + F\, dt . \quad (70)$$

The Pfaffian equations associated with the λ-parameters take the form:

$$\frac{\partial\phi}{\partial\lambda_1} + 2d\left[\sqrt{\mu a(1 - e^2)}\,\lambda_2\right] - w\lambda_1\, dt = 0 \quad (71)$$

$$\frac{\partial\phi}{\partial\lambda_2} - 2d\left[\sqrt{\mu a(1 - e^2)}\,\lambda_1\right] - w\lambda_2\, dt = 0 \quad (72)$$

$$\frac{\partial\phi}{\partial\lambda_3} - 2d\left[\sqrt{\mu a(1 - e^2)}\,\lambda_4\right] - w\lambda_3\, dt = 0 \quad (73)$$

$$\frac{\partial\phi}{\partial\lambda_4} + 2d\left[\sqrt{\mu a(1 - e^2)}\,\lambda_3\right] - w\lambda_4\, dt = 0 , \quad (74)$$

where w is the Lagrangian multiplier associated with the constraint (Equation 69). Putting

$$A = na\sqrt{1-e^2}\,\frac{da}{dt} - \frac{2na^2 e}{\sqrt{1-e^2}}\,\frac{de}{dt}, \tag{75}$$

we can write Equations 71–74 in the form:

$$-A\lambda_2 - 4na^2\sqrt{1-e^2}\,\frac{d\lambda_2}{dt} - \frac{\partial F}{\partial \lambda_1} + w\lambda_1 = 0, \tag{76}$$

$$+A\lambda_1 + 4na^2\sqrt{1-e^2}\,\frac{d\lambda_1}{dt} - \frac{\partial F}{\partial \lambda_2} + w\lambda_2 = 0, \tag{77}$$

$$+A\lambda_4 + 4na^2\sqrt{1-e^2}\,\frac{d\lambda_4}{dt} - \frac{\partial F}{\partial \lambda_3} + w\lambda_3 = 0, \tag{78}$$

$$-A\lambda_3 - 4na^2\sqrt{1-e^2}\,\frac{d\lambda_3}{dt} - \frac{\partial F}{\partial \lambda_4} + w\lambda_4 = 0. \tag{79}$$

From this last set of equations and Equation 69 we have

$$A = -\lambda_2\frac{\partial F}{\partial \lambda_1} + \lambda_1\frac{\partial F}{\partial \lambda_2} - \lambda_3\frac{\partial F}{\partial \lambda_4} + \lambda_4\frac{\partial F}{\partial \lambda_3}, \tag{80}$$

and

$$w = B - 4na^2\sqrt{1-e^2}\left(-\lambda_1\frac{d\lambda_2}{dt} + \lambda_2\frac{d\lambda_1}{dt} + \lambda_3\frac{d\lambda_4}{dt} - \lambda_4\frac{d\lambda_3}{dt}\right), \tag{81}$$

where, for the sake of brevity, we put

$$B = \lambda_1\frac{\partial F}{\partial \lambda_1} + \lambda_2\frac{\partial F}{\partial \lambda_2} + \lambda_3\frac{\partial F}{\partial \lambda_3} + \lambda_4\frac{\partial F}{\partial \lambda_4}. \tag{82}$$

From the Pfaffian equation

$$\frac{\partial \phi}{\partial e} = 0,$$

we deduce, after some transformations,

$$-\lambda_1\frac{d\lambda_2}{dt} + \lambda_2\frac{d\lambda_1}{dt} + \lambda_3\frac{d\lambda_4}{dt} - \lambda_4\frac{d\lambda_3}{dt} = -\frac{\sqrt{1-e^2}}{2na^2 e}\frac{\partial F}{\partial e}, \tag{83}$$

and Equation 81 becomes

$$w = B + \frac{2(1-e^2)}{e}\frac{\partial F}{\partial e}.$$

Substituting this value of w into Equations 76–79 we obtain:

$$\frac{d\lambda_1}{dt} = + \frac{1}{4na^2\sqrt{1-e^2}}\left(\frac{\partial F}{\partial \lambda_2} - \lambda_1 A - \lambda_2 B\right) - \frac{\sqrt{1-e^2}}{2na^2 e}\lambda_2 \frac{\partial F}{\partial e} ,$$

$$\frac{d\lambda_2}{dt} = - \frac{1}{4na^2\sqrt{1-e^2}}\left(\frac{\partial F}{\partial \lambda_1} + \lambda_2 A - \lambda_1 B\right) + \frac{\sqrt{1-e^2}}{2na^2 e}\lambda_1 \frac{\partial F}{\partial e} ,$$

$$\frac{d\lambda_3}{dt} = - \frac{1}{4na^2\sqrt{1-e^2}}\left(\frac{\partial F}{\partial \lambda_4} + \lambda_3 A - \lambda_4 B\right) + \frac{\sqrt{1-e^2}}{2na^2 e}\lambda_4 \frac{\partial F}{\partial e} ,$$

$$\frac{d\lambda_4}{dt} = + \frac{1}{4na^2\sqrt{1-e^2}}\left(\frac{\partial F}{\partial \lambda_3} - \lambda_4 A - \lambda_3 B\right) - \frac{\sqrt{1-e^2}}{2na^2 e}\lambda_3 \frac{\partial F}{\partial e} .$$

Taking Equations 80 and 82 into account we deduce from the last set of equations:

$$\frac{d\lambda_1}{dt} = \frac{1}{4na^2\sqrt{1-e^2}}\left[+\left(\lambda_4^2 + \lambda_3^2\right)\frac{\partial F}{\partial \lambda_2} - \left(\lambda_1\lambda_4 + \lambda_2\lambda_3\right)\frac{\partial F}{\partial \lambda_3} - \left(\lambda_2\lambda_4 - \lambda_1\lambda_3\right)\frac{\partial F}{\partial \lambda_4}\right] - \frac{\sqrt{1-e^2}}{2na^2 e}\lambda_2\frac{\partial F}{\partial e} ,$$

$$\frac{d\lambda_2}{dt} = \frac{1}{4na^2\sqrt{1-e^2}}\left[-\left(\lambda_3^2 + \lambda_4^2\right)\frac{\partial F}{\partial \lambda_1} - \left(\lambda_2\lambda_4 - \lambda_1\lambda_3\right)\frac{\partial F}{\partial \lambda_3} + \left(\lambda_1\lambda_4 + \lambda_2\lambda_3\right)\frac{\partial F}{\partial \lambda_4}\right] + \frac{\sqrt{1-e^2}}{2na^2 e}\lambda_1\frac{\partial F}{\partial e} ,$$

$$\frac{d\lambda_3}{dt} = \frac{1}{4na^2\sqrt{1-e^2}}\left[+\left(\lambda_1\lambda_4 + \lambda_2\lambda_3\right)\frac{\partial F}{\partial \lambda_1} + \left(\lambda_2\lambda_4 - \lambda_1\lambda_3\right)\frac{\partial F}{\partial \lambda_2} - \left(\lambda_1^2 + \lambda_2^2\right)\frac{\partial F}{\partial \lambda_4}\right] + \frac{\sqrt{1-e^2}}{2na^2 e}\lambda_4\frac{\partial F}{\partial e} ,$$

$$\frac{d\lambda_4}{dt} = \frac{1}{4na^2\sqrt{1-e^2}}\left[+\left(\lambda_2\lambda_4 - \lambda_1\lambda_3\right)\frac{\partial F}{\partial \lambda_1} - \left(\lambda_1\lambda_4 + \lambda_2\lambda_3\right)\frac{\partial F}{\partial \lambda_2} + \left(\lambda_1^2 + \lambda_2^2\right)\frac{\partial F}{\partial \lambda_3}\right] - \frac{\sqrt{1-e^2}}{2na^2 e}\lambda_3\frac{\partial F}{\partial e} .$$

A similar set of equations was recently deduced by the author in less direct manner (Reference 16). From Equations 34, 75 and 80 we obtain

$$\frac{de}{dt} = + \frac{1 - e^2}{na^2 e} \frac{\partial F}{\partial l} + \frac{\sqrt{1 - e^2}}{2na^2 e} \left(\lambda_2 \frac{\partial F}{\partial \lambda_1} - \lambda_1 \frac{\partial F}{\partial \lambda_2} + \lambda_3 \frac{\partial F}{\partial \lambda_4} - \lambda_4 \frac{\partial F}{\partial \lambda_3} \right)$$

The disturbed position vector $\mathbf{r}$ is given by the formula:

$$r = \Gamma \cdot \begin{bmatrix} a(\cos u - e) \\ a\sqrt{1 - e^2} \sin u \\ 0 \end{bmatrix},$$

where

$$u - e \sin u = l ,$$

where a, e, l are the disturbed elements. The expression for the matrix Γ in terms of λ_1, λ_2, λ_3, λ_4 is given in Reference 7. We repeat it here for the sake of completeness:

$$\Gamma = [\lambda_{ij}] ,$$

$$\lambda_{11} = + \lambda_1^2 - \lambda_2^2 - \lambda_3^2 + \lambda_4^2 ,$$

$$\lambda_{12} = - 2(\lambda_1 \lambda_2 + \lambda_3 \lambda_4) ,$$

$$\lambda_{13} = + 2(\lambda_1 \lambda_3 - \lambda_2 \lambda_4) ,$$

$$\lambda_{21} = + 2(\lambda_3 \lambda_4 - \lambda_1 \lambda_2) ,$$

$$\lambda_{31} = + 2(\lambda_1 \lambda_3 + \lambda_2 \lambda_4) ,$$

$$\lambda_{22} = - \lambda_1^2 + \lambda_2^2 - \lambda_3^2 + \lambda_4^2 ,$$

$$\lambda_{32} = + 2(\lambda_1 \lambda_4 - \lambda_2 \lambda_3) ,$$

$$\lambda_{23} = - 2(\lambda_1 \lambda_4 + \lambda_2 \lambda_3) ,$$

$$\lambda_{33} = - \lambda_1^2 - \lambda_2^2 + \lambda_3^2 + \lambda_4^2 .$$

We can also form different sets of purely vectorial elements by taking Equation 15 as a basis. Writing Equation 15 in an equivalent form

$$\phi = - \sqrt{\mu a}\, dl + \sqrt{\mu a(1 - e^2)}\, \mathbf{P} \cdot d\mathbf{Q} - F\, dt \tag{84}$$

and putting

$$\mathbf{p} = \sqrt{\mu a}\, \mathbf{P} ,$$

$$\mathbf{q} = \sqrt{1 - e^2}\, \mathbf{Q} - l\mathbf{P} ,$$

we deduce from Equation 84

$$\phi = \mathbf{p} \cdot d\mathbf{q} - F\, dt .$$

Thus $\mathbf{p}$ and $\mathbf{q}$ represent a purely vectorial canonical set of elements. We have

$$\frac{d\mathbf{q}}{dt} = + \operatorname{grad}_{\mathbf{p}} F ,$$

$$\frac{d\mathbf{p}}{dt} = - \operatorname{grad}_{\mathbf{q}} F .$$

CONCLUSION

The Pfaffian method is suggested for use in the search for new sets of elements and in the formation of equations for their variations. The application of vector analysis permits the formation of Pfaffian equations in a very compact and elegant form. A method of integration similar to von Zeipel's method (Reference 17) can be based on the expansion of the Pfaffian differential form in powers of a small parameter.

REFERENCES

1. Bilimovich, A., "Über die Anwendungen der Pfaffschen Methode in der Störungstheorie," *Astr. Nachr.* 273:161–178, 1943.
2. Milankovich, M., "On Application of Vectorial Elements in the Computation of the Planetary Perturbations," *Bull. Acad. Math. Natur.* (A), Belgrade, No. 6, 1939 (In Serbian).
3. Cartan, E., "Leçons sur les Invariants Intégraux," Paris; A. Hermann et Fils, 1922.

4. STRÖMGREN, B., *Pub. Med. Kobenhavns Obs.* No. 65, p. 5, 1929.
5. MUSEN, P., "Om Strömgren's Method of Special Perturbations," *J. Astronaut. Sci.* 8:48–51, Summer 1961.
6. HANSEN, P. A., "Fundamenta," Gotha: Carolum Glasser, p. 331, 1838.
7. MUSEN, P., "Application of Hansen's Theory to the Motion of an Artificial Satellite in the Gravitational Field of the Earth," *Journ. Geophys. Res.* 64(12): 2271–2279, December 1959.
8. Herrick, S., "A Modification of the Variation of Constants Method for Special Perturbations," Astron. Soc. Pacific Pub. 60:321–323, October 1948.
9. MUSEN, P., "Special Perturbations of the Vectorial Elements," *Astron. J.* 59(7):262–267, August 1954.
10. ECKERT, W. J., and BROUWER, D., "The Use of Rectangular Coordinates in the Differential Corrections of Orbits," *Astron. J.* 46(13):125–132, August 16, 1937.
11. MUSEN, P., and CARPENTER, L., "On the General Planetary Perturbations in Rectangular Coordinates," *J. Geophys. Res.* 68(9):2727–2734, May 1, 1963.
12. MUSEN, P., "On the Long-Period Lunar and Solar Effects in the Motion of an Artificial Satellite, 2," *J. Geophys. Res.* 66(9):2797–2805, September 1961.
13. COOK, G. E., "Luni-Solar Perturbations of the Orbit of an Earth Satellite," Royal Aeronautical Establishment Report, Farnborough, 1961.
14. KAULA, W. M., "Development of the Lunar and Solar Disturbing Functions for a Close Satellite," *Astron. J.* 67(5):300–303, June 1962; also NASA Technical Note D-1126, January 1962.
15. GIBBS, J. W., "Vector Analysis," New York: Charles Scribner's Sons, 1901, pp. 343-345.
16. MUSEN, P., "On a Modification of Hansen's Lunar Theory," *J. Geophys. Res.* 68(5):1439–1456, March 1, 1963; also NASA Technical Note D-1745, June 1963.
17. VON ZEIPEL, H., "Recherches sur le Mouvement des Petites Planètes," *Arkiv Matem. Astron. o Fysik*, 11(1):1–58, 1916–1917.

ON THE SIMULTANEOUS COMPUTATION OF THE SECULAR AND RESONANCE EFFECTS IN THE MOTION OF CELESTIAL BODIES*

PETER MUSEN

The stability of the orbit of a celestial body depends predominantly upon the long period perturbations in its elements. The long period effects produced by the sun and the moon can prolong or shorten the lifetime of a satellite considerably. In every attempt to solve the problem of stability of the asteroidal ring the long period effects must be taken into account. Unfortunately, no convenient expansion of the disturbing function exists for large values of the eccentricity, large values of the inclinations, or large values of the ratio of the semimajor axes of the disturbed and disturbing body. Thus numerical integration must be used.

In the problem of determining orbital stability the disturbing function cannot be used in its standard form, because it contains the combined effects of the long and short period terms. If the short period terms are not removed, then the integration step will be too short, and the accumulation of the errors from rounding off will be too great over a long time interval. Thus, before integration all short period effects were removed from the disturbing function by a numerical process which avoids the development into series. In a previous article by the author the suggestion was made to use Halphen's method to compute the effects caused by the secular terms.

However, the author feels that the results of computation cannot be considered complete if they do not include the near resonance effects. As a result, Halphen's method is extended by inclusion of the near resonance effects caused by the commensurability of the mean motions of the disturbed and disturbing bodies. Liouville's method reduces the problem to a single harmonic analysis. The introduction of a nonsingular set of elements permits the extension of the method to the cause of near circular orbits and to the case of low inclinations of the orbital planes.

LIST OF SYMBOLS

k = the Gaussian constant

ϵ = the angle between the equatorial and the ecliptical planes 1950.0

$\mathbf{r}$ = the position vector of the disturbed body

u = the eccentric anomaly of the disturbed body

f = the true anomaly of the disturbed body

$\ell, \omega, \Omega, \pi, i, e, a,$ = the standard elliptic elements of the disturbed body

$\mathbf{P}, \mathbf{Q},$ = Gibbs' vectorial elements of the disturbed body

σ = the angular distance of the departure point in the orbit plane of the disturbed body from the ascending node

χ = the orbital true longitude of the perigee

$$\xi = e \cos \chi$$

$$\eta = e \sin \chi$$

*Published as *NASA Technical Note* D-2152, June 1964.

L = the mean longitude of the disturbed body
$L_1 = \ell + \chi$ = the orbital mean longitude of the disturbing body

$$X = r \cos(f + \chi)$$

$$Y = r \sin(f + \chi)$$

$$p = \frac{\cos \frac{\Omega + \sigma}{2}}{\cos \frac{\Omega - \sigma}{2}} \tan \frac{i}{2}$$

$$q = \frac{\sin \frac{\Omega + \sigma}{2}}{\cos \frac{\Omega - \sigma}{2}} \tan \frac{i}{2}$$

$$s = \tan \frac{\Omega - \sigma}{2}$$

$\mathbf{r}'$ = the position vector of the disturbing body
u' = the eccentric anomaly of the disturbing body
ℓ', ω', Ω', π', i', e', a', n' = the standard elliptic elements of the disturbing body
$\mathbf{P}'$, $\mathbf{Q}'$, $\mathbf{R}'$ = Gibbs' vectorial elements of the disturbing body
ρ = the distance between the disturbed and the disturbing body
S = the component of the disturbing force in the direction of the radius vector of the disturbed body
T = the component of the disturbing force lying in the orbital plane and normal to the radius vector
W = the component of the disturbing force normal to the orbital plane.

INTRODUCTION

The problem of determining the long period effects in the motions of celestial bodies is a central problem in the determination of orbital stability over a long time interval. Two types of long period perturbations are of primary importance in treating the problem of stability:

1. The effects produced by the secular terms in the disturbing function. The arguments of the secular terms contain neither the mean anomaly of the disturbing body nor the mean anomaly of the disturbed body.

2. The effects produced by the commensurability of mean motions. If the ratio of the mean motions of the disturbed and the disturbing bodies can be approximated by a rational number, then a small divisor will appear when the integration is performed analytically. The terms producing such small divisors are named "the critical terms."

The long period effects as well as their interactions can be determined analytically, but only under the restrictions that the eccentricities or inclinations are small and that only one critical term is present.

The secular effects have been treated numerically with Halphen-Goriachev theory (References 1–4). The only restriction imposed on the elements by this method is that a very close approach between the disturbed and the disturbing bodies cannot be considered. The secular effects of the first and second order were computed by means of numerical integration over an interval of 10–20 years for artificial satellites, and over an interval of many thousands of years for minor planets.

Several unexpected features and large variations of the elements affecting their stability have been discovered (References 4 and 5).

However, the scheme previously used becomes incomplete if there are critical terms in the development of the disturbing function. The effects produced by the critical terms in the elements might become comparable with those produced by the secular terms, and in addition, the problem of the mutual influence of terms of both types arises.

COMPUTATION OF THE SYSTEM OF FORMULAS FOR SECULAR AND RESONANCE EFFECTS

The system of formulas given in this article is based on the application of the method of Liouville (Reference 6) and on Halphen's method. It takes into account the secular as well as the critical effects. The equations for variations of the elements are also given in the form which includes the cases of small eccentricity and small inclination. Let F(u, u′) be a function of the form

$$F(u, u') = \sum_{j=0}^{+\infty} \sum_{j'=-\infty}^{+\infty} \left[A_{j,j'} \cos(jl - j'l') + B_{j,j'} \sin(jl - j'l') \right]. \quad (1)$$

where

$$l = u - e \sin u ,$$

$$l' = u' - e' \sin u' ,$$

and the coefficients $A_{j,j'}$, $B_{j,j'}$ are functions of all the remaining elements. If it is assumed that

$$\frac{n'}{n} \approx \frac{i}{i'}$$

where i and i′ are relative primes, then the critica arguments have the form

$$k(il - i'l') .$$

By using the Liouville substitution (Reference 6),

$$l = i'\sigma' ,$$

$$l' = i\sigma' - \frac{\theta}{i'} ,$$

Equation 1 can be written in the form:

$$F(u, u') = \sum_{j=0}^{+\infty} \sum_{j'=-\infty}^{+\infty} \left\{ A_{j,j'} \cos\left[(ji' - j'i)\sigma' + \frac{j'}{i'}\theta \right] + B_{j,j'} \sin\left[(ji' - j'i)\sigma' + \frac{j'}{i'}\theta \right] \right\}. \quad (2)$$

The angle σ' is not contained in the argument if

$$ji' - j'i = 0 ,$$

or, consequently, if

$$\left.\begin{aligned} j &= pi , \\ j' &= pi' , \end{aligned}\right\} \quad (3)$$

where p=0, 1, 2,⋯. Let us designate the combination of all terms in F(u, u′) which do not contain σ' by $\Psi(\theta)$:

$$\Psi(\theta) = \sum_{p=0}^{+\infty} \left(A_{pi,pi'} \cos p\theta + B_{pi,pi'} \sin p\theta \right) .$$

By taking Equation 3 into account, it may be deduced from Equation 2 that

$$\Psi(\theta) = \frac{1}{2\pi} \int_0^{2\pi} F(u, u')\, d\sigma'$$

$$= \sum_{p=0}^{+\infty} \left(A_{pi,pi'} \cos p\theta + B_{pi,pi'} \sin p\theta \right) . \quad (4)$$

Putting $\theta = 0, \pi, \pi/2, 3\pi/2, \pi/4$, in Equation 4 yields:

$$\Psi(0) = A_{0,0} + A_{i,i'} + A_{2i,2i'} + A_{3i,3i'} + \cdots ,$$

$$\Psi(\pi) = A_{0,0} - A_{i,i'} + A_{2i,2i'} - A_{3i,3i'} + \cdots ,$$

$$\Psi\left(\frac{\pi}{2}\right) = A_{0,0} + B_{i,i'} - A_{2i,2i'} - B_{3i,3i'} + \cdots ,$$

$$\Psi\left(\frac{3\pi}{2}\right) = A_{0,0} - B_{i,i'} - A_{2i,2i'} + B_{3i,3i'} + \cdots ,$$

$$\Psi\left(\frac{\pi}{4}\right) = A_{0,0} + \frac{\sqrt{2}}{2}\left(A_{i,i'} + B_{i,i'}\right) + B_{2i,2i'} + \cdots .$$

It is rarely necessary to go beyond the critical term with the argument $2il - 2i'l'$. Thus, the coefficients $A_{2i,2i'}$, $B_{2i,2i'}$, etc. can be neglected and a set of simple formulas results:

$$A_{i,i'} = \frac{\Psi(0) - \Psi(\pi)}{2} , \quad (5)$$

$$B_{i,i'} = \frac{\Psi\left(\frac{\pi}{2}\right) - \Psi\left(\frac{3\pi}{2}\right)}{2} , \quad (6)$$

$$A_{2i,2i'} = \frac{\Psi(0) - \Psi\left(\frac{\pi}{2}\right) + \Psi(\pi) - \Psi\left(\frac{3\pi}{2}\right)}{4}, \tag{7}$$

$$B_{2i,2i'} = \Psi\left(\frac{\pi}{4}\right) - \frac{\sqrt{2}+1}{4}\left[\Psi(0) + \Psi\left(\frac{\pi}{2}\right)\right] + \frac{\sqrt{2}-1}{4}\left[\Psi(\pi) + \Psi\left(\frac{3\pi}{2}\right)\right]. \tag{8}$$

In the case of a large or moderate eccentricity set

$$u = i'\phi ;$$

then

$$l' = u' - e'\sin u'$$

$$= i\phi - \frac{ie}{i'}\sin i'\phi - \frac{\theta}{i'}, \tag{9}$$

and Equation 4 takes a form more convenient for numerical computations:

$$\Psi(\theta) = \frac{1}{2\pi}\int_0^{2\pi} F(u, u')\,\frac{r}{a}\,d\phi . \tag{10}$$

This last integral is computed as a simple arithmetical mean of the integrands of equidistant values of the angle ϕ. Now a collection of formulas can be established for the actual computation of the effect of resonance in the elements on the basis of the Liouville equations (Equations 5–8). Setting

$$A_1(\alpha) = \begin{bmatrix} +1 & 0 & 0 \\ 0 & +\cos\alpha & -\sin\alpha \\ 0 & +\sin\alpha & +\cos\alpha \end{bmatrix},$$

$$A_3(\alpha) = \begin{bmatrix} +\cos\alpha & -\sin\alpha & 0 \\ +\sin\alpha & +\cos\alpha & 0 \\ 0 & 0 & +1 \end{bmatrix},$$

we can deduce the Gibbs vectorial elements **P**, **Q**, **R** and **P′**, **Q′**, **R′** by means of the formulas:

$$[\mathbf{P}, \mathbf{Q}, \mathbf{R}] = A_1(\epsilon) \cdot A_3(\Omega) \cdot A_1(i) \cdot A_3(\omega) ,$$

$$[\mathbf{P}', \mathbf{Q}', \mathbf{R}'] = A_1(\epsilon) \cdot A_3(\Omega') \cdot A_1(i') \cdot A_3(\omega') .$$

In the case of an artificial satellite the elements ω and Ω are referred to the equator and

$$[\mathbf{P}, \mathbf{Q}, \mathbf{R}] = A_3(u) \cdot A_1(i) \cdot A_3(\omega) .$$

The coordinates of the disturbed body are

$$\left.\begin{aligned} x &= a\mathbf{P}_x(\cos u - e) + a\sqrt{1-e^2}\,\mathbf{Q}_x \sin u , \\ y &= a\mathbf{P}_y(\cos u - e) + a\sqrt{1-e^2}\,\mathbf{Q}_y \sin u , \\ z &= a\mathbf{P}_z(\cos u - e) + a\sqrt{1-e^2}\,\mathbf{Q}_z \sin u , \end{aligned}\right\} \tag{11}$$

and the coordinates of the disturbing body are

$$\left.\begin{aligned} x' &= a'P_x'(\cos u' - e') + a'\sqrt{1-e'^2}\,Q_x'\sin u' \ , \\ y' &= a'P_y'(\cos u' - e') + a'\sqrt{1-e'^2}\,Q_y'\sin u' \ , \\ z' &= a'P_z'(\cos u' - e') + a'\sqrt{1-e'^2}\,Q_z'\sin u' \ . \end{aligned}\right\} \tag{12}$$

The square of the mutual distance is

$$\rho^2 = r^2 + r'^2 - 2\mathbf{r}\cdot\mathbf{r}' \ . \tag{13}$$

Let the critical argument be

$$\theta = il - i'l' \ .$$

Set

$$u = i'\phi \ ,$$

$$u' - e'\sin u' = i\phi - \frac{\theta}{i'} - \frac{ie}{i'}\sin i'\phi \ .$$

The radial component S, the orthogonal component T, and the normal component W of the disturbing force can be considered functions of the angle ϕ and the critical argument θ. Thus

$$rS = km'\left(\frac{1}{\rho^3} - \frac{1}{r'^3}\right)\mathbf{r}\cdot\mathbf{r}' - km'\frac{r^2}{\rho^3} = rS(\phi, \theta) \ , \tag{14}$$

$$rT = km'\left(\frac{1}{\rho^3} - \frac{1}{r'^3}\right)(\mathbf{R}\cdot\mathbf{r}\times\mathbf{r}') = rT(\phi, \theta) \ , \tag{15}$$

$$W = km'\left(\frac{1}{\rho^3} - \frac{1}{r'^3}\right)\mathbf{R}\cdot\mathbf{r}' = W(\phi, \theta) \ . \tag{16}$$

From the following standard equations for variations of the elements

$$\frac{dn}{dt} = -\frac{3k}{a^2}\left(S\frac{ae\sin f}{\sqrt{1-e^2}} + T\frac{a^2}{r}\sqrt{1-e^2}\right) \ ,$$

$$\frac{de}{dt} = \frac{na(1-e^2)}{ke}\left(S\frac{ae\sin f}{\sqrt{1-e^2}} + T\frac{a^2}{r}\sqrt{1-e^2}\right) - \frac{na\sqrt{1-e^2}}{ke}Tr \ ,$$

$$\frac{d\pi}{dt} = \frac{na\sqrt{1-e^2}}{ke}\left[-S\cos f + T\left(1 + \frac{1}{1-e^2}\frac{r}{a}\right)\sin f\right] + 2\sin^2\frac{i}{2}\frac{d\Omega}{dt} \ ,$$

$$\frac{dL}{dt} = -\frac{2nar}{k}S + \left(1 - \sqrt{1-e^2}\right)\frac{d\pi}{dt} + 2\sqrt{1-e^2}\sin^2\frac{i}{2}\frac{d\Omega}{dt} \ ,$$

$$\sin i\frac{d\Omega}{dt} = \frac{nar}{k\sqrt{1-e^2}}W\sin(f+\omega) \ ,$$

$$\frac{di}{dt} = \frac{nar}{k\sqrt{1-e^2}}W\cos(f+\omega) \ .$$

the short period and the secular terms may be eliminated by using the process described above, and the transformed equations will contain only the effect of the critical terms. By designating

$$A_{1,1}' = K_1 , \qquad B_{1,1}' = K_2 , \qquad A_{2i,2i}' = K_3 , \qquad B_{2i,2i}' = K_4 ,$$

the derivative of an element E becomes

$$\frac{dE}{dt} = K_1 \cos\theta + K_2 \sin\theta + K_3 \cos 2\theta + K_4 \sin 2\theta . \tag{17}$$

Let us introduce the following notations:

$$\frac{f(\phi, 0) - f(\phi, \pi)}{2} = f_1(\phi) ,$$

$$\frac{f\left(\phi, \frac{\pi}{2}\right) - f\left(\phi, \frac{3\pi}{2}\right)}{2} = f_2(\phi) ,$$

$$\frac{f(\phi, 0) - f\left(\phi, \frac{\pi}{2}\right) + f(\phi, \pi) - f\left(\phi, \frac{3\pi}{2}\right)}{4} = f_3(\phi) ,$$

$$f\left(\phi, \frac{\pi}{4}\right) - \frac{\sqrt{2}+1}{4}\left[f(\phi, 0) + f\left(\phi, \frac{\pi}{2}\right)\right] + \frac{\sqrt{2}-1}{4}\left[f(\phi, \pi) + f\left(\phi, \frac{3\pi}{2}\right)\right] = f_4(\phi) .$$

These values of the coefficients K_j (j = 1, 2, 3, 4) are deduced:

in dn/dt

$$K_j^{(n)} = -\frac{3k}{2\pi a}\int_0^{2\pi}\left(S_j\, e \sin u + T_j \sqrt{1-e^2}\right) d\phi ; \tag{18}$$

in de/dt

$$K_j^{(e)} = \frac{\sqrt{a(1-e^2)}}{2\pi}\int_0^{2\pi}\left[\sqrt{1-e^2}\ S_j \sin u + T_j\left(-\frac{3}{2}\, e + 2\cos u - \frac{1}{2}\, e\cos 2u\right)\right] d\phi ; \tag{19}$$

in dπ/dt

$$K_j^{(\pi)} = \frac{\sqrt{a(1-e^2)}}{2\pi e}\int_0^{2\pi}\left[-S_j(\cos u - e) + T_j\left(1 + \frac{r}{a}\,\frac{1}{1-e^2}\right)\sqrt{1-e^2}\ \sin u\right] d\phi + 2\sin^2\frac{i}{2}\, K_j^{(\Omega)} ; \tag{20}$$

in dL/dt

$$K_j^{(L)} = \frac{e^2}{1+\sqrt{1-e^2}}\, K_j^{(\pi)} + \sqrt{1-e^2}\, K_j^{(\Omega)}\, 2\sin^2\frac{i}{2} - \frac{2\sqrt{a}}{2\pi}\int_0^{2\pi} S_j\,\frac{r^2}{a^2}\, d\phi ; \tag{21}$$

in dΩ/dt

$$K_j^{(\Omega)} = \frac{1}{2\pi \sin i}\sqrt{\frac{a}{1-e^2}}\int_0^{2\pi} W_j\,\frac{r}{a}\left[(\cos u - e)\sin\omega + \sqrt{1-e^2}\ \sin u\cos\omega\right] d\phi ; \tag{22}$$

in di/dt

$$K_j^{(i)} = \frac{1}{2\pi}\sqrt{\frac{a}{1-e^2}}\int_0^{2\pi} W_j \frac{r}{a}\left[(\cos u - e)\cos\omega - \sqrt{1-e^2}\,\sin u \sin\omega\right] d\phi \,. \tag{23}$$

It is interesting to note that Equations 18–23 are similar in form to the equations for the computation of secular effects. The cases of small inclination and small eccentricity deserve special attention. Equations 20 and 22 contain a "small divisor" if i or e is very small. Instead of the standard orbital elements, a nonsingular set of elements can be introduced. The set of elements

$$\left.\begin{aligned} p &= \frac{\cos\frac{\Omega+\sigma}{2}}{\cos\frac{\Omega-\sigma}{2}}\tan\frac{i}{2}\,, \\ q &= \frac{\sin\frac{\Omega+\sigma}{2}}{\cos\frac{\Omega-\sigma}{2}}\tan\frac{i}{2}\,, \\ s &= \tan\frac{\Omega-\sigma}{2}\,, \end{aligned}\right\} \tag{24}$$

$$\left.\begin{aligned} \xi &= e\cos\chi\,, \\ \eta &= e\sin\chi\,, \end{aligned}\right\} \tag{25}$$

determines the position of the osculating ellipse without the introduction of a singularity for $i=0$ or $e=0$. The use of Equations 25 for the computation of perturbations was suggested by the author in his article on Strömgren's perturbations (Reference 7). In fact, p, q, and s are the components of Gibbs' rotation vector (Reference 8). From the equations

$$\sin i\,\frac{d\Omega}{dt} = \frac{rW}{\sqrt{a(1-e^2)}}\sin(f+\omega)\,,$$

$$\frac{di}{dt} = \frac{rW}{\sqrt{a(1-e^2)}}\cos(f+\omega)\,,$$

$$\frac{d\sigma}{dt} = \cos i\,\frac{d\Omega}{dt}\,,$$

it is deduced that

$$\left.\begin{aligned} \frac{dp}{dt} &= \frac{W}{2\sqrt{a(1-e^2)}}\left[(1+p^2)X + (pq-s)Y\right], \\ \frac{dq}{dt} &= \frac{W}{2\sqrt{a(1-e^2)}}\left[(pq+s)X + (1+q^2)Y\right], \\ \frac{ds}{dt} &= \frac{W}{2\sqrt{a(1-e^2)}}\left[(ps-q)X + (qs+p)Y\right]. \end{aligned}\right\} \tag{26}$$

It can be seen that no small divisor is present, even if the inclination is small. X and Y are the coordinates of the disturbing body in an ideal system of coordinates with the X- and Y-axes in the orbital plane. They can be written in the form:

$$\begin{aligned} X &= r\cos(f+\chi) \\ &= a\left[\left(\frac{1+\sqrt{1-e^2}}{2} + \frac{\xi^2-\eta^2}{2}\,\frac{1}{1+\sqrt{1-e^2}}\right)\cos\lambda + \frac{\xi\eta}{1+\sqrt{1-e^2}}\sin\lambda - \xi\right], \end{aligned} \tag{27}$$

$$\begin{aligned} Y &= r\sin(f+\chi) \\ &= a\left[\left(\frac{1+\sqrt{1-e^2}}{2} - \frac{\xi^2-\eta^2}{2}\,\frac{1}{1+\sqrt{1-e^2}}\right)\sin\lambda + \frac{\xi\eta}{1+\sqrt{1-e^2}}\cos\lambda - \eta\right], \end{aligned} \tag{28}$$

where λ is the eccentric orbital longitude,

$$\lambda = u + \chi \,. \tag{29}$$

Also (Reference 9)

$$\frac{d\xi}{dt} = \sqrt{a(1-e^2)}\left[S\sin(f+\chi) + T\left(1+\frac{r}{a}\frac{1}{1-e^2}\right)\cos(f+\chi) + \frac{r}{a}\frac{1}{1-e^2}\,T\xi\right], \tag{30}$$

$$\frac{d\eta}{dt} = \sqrt{a(1-e^2)}\left[-S\cos(f+\chi) + T\left(1+\frac{r}{a}\frac{1}{1-e^2}\right)\sin(f+\chi) + \frac{r}{a}\frac{1}{1-e^2}\,T\eta\right]. \tag{31}$$

We write the critical argument in the form:

$$\alpha = iL_1 - i'l' \,.$$

Then the Liouville substitution becomes

$$L_1 = i'\sigma' \,,$$

$$l' = i\sigma' - \frac{\alpha}{i'} \,.$$

Considering the disturbing force as a function of σ' and α, set

$$\frac{f(\sigma', 0) - f(\sigma', \pi)}{2} = f_1(\sigma'), \text{ etc.}$$

From Equations 26, 30, and 31 the following expressions are deduced for the coefficients of the critical terms:

$$K_j^{(p)} = \frac{1}{4\pi\sqrt{a(1-e^2)}}\int_0^{2\pi} W_j\left[(1+p^2)\,X + (pq-s)\,Y\right]d\sigma' \,, \tag{32}$$

$$K_j^{(q)} = \frac{1}{4\pi\sqrt{a(1-e^2)}}\int_0^{2\pi} W_j\left[(pq+s)\,X + (1+q^2)\,Y\right]d\sigma' \,, \tag{33}$$

$$K_j^{(s)} = \frac{1}{4\pi\sqrt{a(1-e^2)}}\int_0^{2\pi} W_j\left[(ps-q)\,X + (qs+p)\,Y\right]d\sigma' \,, \tag{34}$$

$$K_j^{(\xi)} = \frac{1}{2\pi}\sqrt{\frac{1-e^2}{a}}\int_0^{2\pi}\left[+S_jY + T_j\left(1+\frac{r}{a}\frac{1}{1-e^2}\right)X + \frac{r^2}{a^2}\frac{1}{1-e^2}\,T_j\xi\right]\frac{a}{r}\,d\sigma' \,, \tag{35}$$

$$K_j^{(\eta)} = \frac{1}{2\pi}\sqrt{\frac{1-e^2}{a}}\int_0^{2\pi}\left[-S_jX + T_j\left(1+\frac{r}{a}\frac{1}{1-e^2}\right)Y + \frac{r^2}{a^2}\frac{1}{1-e^2}\,T_j\xi\right]\frac{a}{r}\,d\sigma' \,. \tag{36}$$

The expression for the position vector of the disturbed body to be used in the computations of the integrands is:

$$r = \frac{A_3(\epsilon)\cdot\Gamma}{1+p^2+q^2+s^2}\begin{bmatrix}X\\Y\\0\end{bmatrix}, \tag{37}$$

where the elements γ_{ij} (i is the row ind· x and j is the column index) of the matrix Γ are determined from the formulas (Reference 10):

$$\left.\begin{aligned}
\gamma_{11} &= +p^2 - q^2 - s^2 + 1 , & \gamma_{21} &= +2(s+pq) , & \gamma_{31} &= +2(sp-q) ,\\
\gamma_{12} &= -2(s-pq) , & \gamma_{22} &= -p^2 + q^2 - s^2 + 1 , & \gamma_{32} &= +2(p+sq) ,\\
\gamma_{13} &= +2(q+ps) , & \gamma_{23} &= -2(p-sq) , & \gamma_{33} &= -p^2 - q^2 + s^2 - 1 .
\end{aligned}\right\} \quad (38)$$

X and Y are given by Equations 27–29. Kepler's equation is replaced by

$$\lambda - \xi \sin\lambda + \eta \cos\lambda = L_1 \qquad (39)$$

in a way similar to Herget's method of writing Kepler's equation for small eccentricities (Reference 11). Similar systems of equations can be established for numerical integration of the secular effects for the case of a small eccentricity or a small inclination.

The effects of the secular and critical terms must be integrated together. Again Halphen's method may be used for the computations of the secular effects. A complete collection of formulas is given in the author's previous article (Reference 3). However, if i or e, or both, are small, the system requires some modification.

For the values, say, $\lambda = 0°, 10°, \cdots, 350°$, compute:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{1+p^2+q^2+s^2} \begin{bmatrix} P_x' & P_y' & P_z' \\ Q_x' & Q_y' & Q_z' \\ R_x' & R_y' & R_z' \end{bmatrix} \cdot \Gamma \cdot \begin{bmatrix} X \\ Y \\ 0 \end{bmatrix} . \qquad (40)$$

X and Y are defined by Equations 27 and 28. Kepler's equation is taken in the form of Equation 39. Then, as in the previous article, put

$$\alpha = x' + e'a ,$$

$$\beta = y ,$$

$$\gamma = z .$$

Equations 4–9 of Reference 3 remain unchanged, but Equation 10, p. 121, for the variations of the elements must be replaced by:

$$\frac{dn}{dt} = -\frac{3k}{a}\frac{1}{2\pi}\int_0^{2\pi}\left[\frac{S_0(Y\xi - X\eta)}{\sqrt{1-e^2}} + T_0\sqrt{1-e^2}\right]d\lambda , \qquad (41)$$

$$\frac{d\xi}{dt} = \sqrt{\frac{1-e^2}{a}}\frac{1}{2\pi}\int_0^{2\pi}\left[S_0 Y + T_0\left(1+\frac{r}{a}\frac{1}{1-e^2}\right)X + \left(\frac{r}{a}\right)^2\frac{a}{1-e^2}T_0\xi\right]d\lambda , \qquad (42)$$

$$\frac{d\eta}{dt} = \sqrt{\frac{1-e^2}{a}}\frac{1}{2\pi}\int_0^{2\pi}\left[-S_0 X + T_0\left(1+\frac{r}{a}\frac{1}{1-e^2}\right)Y + \left(\frac{r}{a}\right)^2\frac{a}{1-e^2}T_0\eta\right]d\lambda , \qquad (43)$$

$$\frac{dp}{dt} = \frac{1}{2\sqrt{a\left(1-e^2\right)}} \frac{1}{2\pi} \int_0^{2\pi} W_0 \frac{r}{a} \left[\left(1+p^2\right) X + (pq - s) Y\right] d\lambda \ , \tag{44}$$

$$\frac{dq}{dt} = \frac{1}{2\sqrt{a\left(1-e^2\right)}} \frac{1}{2\pi} \int_0^{2\pi} W_0 \frac{r}{a} \left[(pq + s) X + \left(1+q^2\right) Y\right] d\lambda \ , \tag{45}$$

$$\frac{ds}{dt} = \frac{1}{2\sqrt{a\left(1-e^2\right)}} \frac{1}{2\pi} \int_0^{2\pi} W_0 \frac{r}{a} \left[(ps - q) X + (qs + p) Y\right] d\lambda \ , \tag{46}$$

$$\frac{dL_1}{dt} = -\frac{2}{\sqrt{a}} \frac{1}{2\pi} \int_0^{2\pi} \left(\frac{r}{a}\right)^2 S_0 \, d\lambda + \frac{\xi \frac{d\eta}{dt} - \eta \frac{d\xi}{dt}}{1 + \sqrt{1-e^2}} \ . \tag{47}$$

CONCLUSIONS

The combined effects of the secular and critical terms have a deciding influence in determining the orbital stability over a long time interval. Some approximate schemes based on the process of averaging lead to conclusions concerning the stability of the ring of minor planets which definitely require a further check. The proposed scheme will at least give a more accurate answer over an interval of some thousands of years. Our knowledge about the stability of the orbits of artificial satellites is still incomplete. The described method will be programmed in order to investigate these problems. The application of the method of secular effects leads to some conclusions which were not anticipated. It remains to be seen what interesting conclusions will follow from the superposition of all long period terms and from the computations of their direct actions, as well as their interactions, in the problems of orbital stability of the minor planets and artificial satellites.

REFERENCES

1. Halphen, G. H., "Traité des Fonctions Elliptiques et de Leurs Applications," Paris: Gauthier-Villars et Fils, 1888, V.2.
2. Goriachev, N. N., "On the Method of Halphen of the Computation of Secular Perturbations," University of Tomsk, 1937, pp. 1–115.
3. Musen, P., "A Discussion of Halphen's Method, Part 1," *Rev. Geophys* 1 (1):85–122, 1963.
4. Smith, A. J., Jr., "A Discussion of Halphen's Method for Secular Perturbations and its Application to the Determination of Long Range Effects in the Motions of Celestial Bodies. Part II.," NASA Technical Report (in press).
5. Shute, B. E., "A Cislunar Orbit (IMP)" *Astronom. J.* 67(5):283, June 1962 (abstract).
6. Liouville, J., "Note Sur le Calcul des Inégalités Périodiques de Mouvement des Planétes," *J. Math. Pures Appliquées* 1:197–210, 1836.
7. Musen, P., "On Stromgrens Method of Special Perturbations," *J. Astronaut. Sci.* 8(2):48, 1961.
8. Gibbs, J. W., "Vector Analysis," (E. B. Wilson, ed.): New York: Charles Scribner's Sons, 1901, pp. 343–345.
9. Tisserand, F., "Traité de la Mécanique céleste," Paris: Gauthier-Villars et Fils, 1896, V. 4, p. 328.
10. Musen, P. "Application of Hansen's Theory to the Motion of an Artificial Satellite in the Gravitational Field of the Earth," *J. Geophys, Res.* 64(12): 2271–2279, December 1959.

N64-23737

DETERMINATION OF THE TRIAXIALITY OF THE EARTH FROM OBSERVATIONS ON THE DRIFT OF THE SYNCOM II SATELLITE*

C. A. WAGNER

SUMMARY

The near 24-hour Syncom II satellite (with an almost circular orbit) has been under continuous observation by range and range-rate radar and minitrack stations for 7 months since mid-August 1963, when the orbit was relocated, placing its mean longitude at about 55 degrees west of Greenwich. During the first 4 months of this period, the satellite was allowed to drift free in the gravity fields of the earth, sun, and moon. In this first free-drift period, the satellite experienced a mean daily drift acceleration of its ascending node (with respect to Greenwich) of

$$-(1.27\pm.02)\times10^{-3} \text{ degrees/day}^2. \quad (1)$$

The average growth of the semimajor axis for this period was

$$(.0993\pm.0042) \text{ km/day}. \quad (2)$$

These values, checked by a simulated particle trajectory run on the Goddard ITEM program, confirm a significant longitude-dependent earth-gravity potential. The existence of a "triaxial earth" has been a speculation of geodesists since the early years of this century.

During the last 3 months of this 7-month drift period, starting at the end of November 1963, Syncom II was relocated at about 60 degrees west of Greenwich. In this period, the mean daily drift acceleration of the ascending node was

$$-(1.32\pm.02)\times10^{-3} \text{ degrees/day}^2. \quad (3)$$

The average growth of the semimajor axis for this period was

$$(.0994\pm.0080) \text{ km/day}. \quad (4)$$

Combining the results of (1) through (4) above for the two separate drift periods, it is estimated (on the basis of a triaxial geoid only) that the absolute magnitude of the longitude dependent-gravity coefficient J_{22} is

$$J_{22}=-(1.7\pm.05)\times10^{-6}.$$

This value corresponds to a difference in major and minor earth-equatorial radii of

$$(a_e-b_e)=213\pm6 \text{ feet}$$

The best present estimate of the position of the earth's major equatorial axis is

$$\lambda_{22}=19\pm6 \text{ degrees west of Greenwich.}$$

In view of previous estimates of the higher order tesseral harmonics of the earth's field, the true value of J_{22}, separated from the small influence of gravity anomalies of third and higher order on the reduction for a triaxial earth only at "synchronous" altitudes, will probably be somewhat higher than the -1.7×10^{-6} reported herein. The true value of J_{22}, however, is not likely to be greater than -2.2×10^{-6} or smaller than -1.6×10^{-6}. The true location of the earth equator's major axis is not expected to differ significantly from that reported herein, when all higher tesseral harmonics are accounted for. (See Appendix B.)

*Published as *Goddard Space Flight Center Document* X-621-64-90, April 1964.

The reported value of $J_{22} = -1.7 \times 10^{-6}$ implies that a maximum tangential velocity correction of

$$\Delta V_T = 5.36 \text{ ft./sec./year}$$

is required to keep a satellite with a 24-hour circular equatorial orbit continuously "on station" at a longitude midway between the longitudes of the equatorial major and minor axes of the earth. The original "conservative" Syncom I design requirement of $\Delta V_T = 17$ ft./sec./year correction capability was based on the longitude-dependent earth field of Izsak (January 1961), which is now outdated. (See Appendix B and Reference 4.)

INTRODUCTION

SUMMARY OF PREVIOUS INVESTIGATIONS AND DISCUSSION OF RESULTS

The question of the existence and extent of the longitude-dependence of the earth's external gravity field has concerned geodesists since the early years of this century. (See Reference 1 for example.) The existence of a longitude-dependent field implies the existence of inhomogeneities and states of stress within the earth which are of fundamental importance to all dynamical theories of the earth's interior.

In Appendix B, I have summarized 23 reductions of gravimetric, astro-geodetic, and satellite gravity data reporting longitude-dependent terms in the earth's external gravity field represented as a series of spherical harmonics. Ten of these reductions are based on worldwide surface-gravity measurements only. Although surface measurements have the advantage of providing an excellent sampling of the field in latitude and longitude, they are seriously affected by even small uncertainties in station position with respect to the mean earth geoid, as these are of the same order-of-magnitude as the *reported* geoid deviations caused by longitude-dependent gravity.

Of the 12 satellite reduced tesseral gravity fields, those of Kozai (1962), Izsak (July 1963), Kaula (September 1963), and Guier (1963) show good general agreement in the eastern hemisphere when the constants are used as a set. All of these observers are aware, nevertheless, of the high degree of uncertainty in the reported values of the individual coefficients themselves, this uncertainty being due mainly to unresolved station-datum errors and to the limited sampling of the field from observations on a small number of medium-altitude, medium-inclination satellites.

As late as July 1963, Izsak stated (Reference 2): "It might be some time before one can arrive at definite conclusions regarding the longitude dependence of the earth's gravity field." The presence of Syncom II, high over the earth with a nearly stationary, narrow figure-8 ground track centered close to the equator, dramatically alters this gloomy outlook. The 24-hour satellite is high enough to be unaffected by the earth's atmosphere, yet is close enough to the earth to be protected from the solar wind by the "magnetosphere" of the earth, and to remain essentially unaffected by sun or moon gravity.

In theory, as the ground track is nearly stationary, any small earth-gravity anomaly in longitude will in time, cause significant drift of the ground-track configuration. In theory too, only observations of the longitude of the satellite from a single ground station are necessary to reveal this effect of the "tesseral" gravity field over an extended period of time. Long-term observations of the longitude drift of one or more 24-hour satellites should reveal the exact nature of the tesseral-gravity field to at least the third order without essential difficulty. The great height of the 24-hour satellite tends to cancel out the individual contributions to the longitudinal drift of anomalies higher than about the fourth order. It is fortunate that the initial slow westward drift of the ground track of Syncom II (August 1963 to March 1964) appears to have occurred relatively close to a point midway between the triaxial earth's major and minor equatorial axes where the perturbation of the second tesseral-gravity anomaly, for which the reduction was made, is greatest. A weighted average of the longitude-perturbation fields of Kozai (1962), Zhongolovitch (1957), Kaula (September 1963), and Izsak (July 1963), at the altitude and longitude of Syncom II during this period, shows that the second tesseral should be contributing about 80 to 85 percent of the perturbing force. If further observations of this and other 24-hour satellites confirm this estimate, then the magnitude of the dominant J_{22} potential term, separated from higher order effects, will increase by 25 percent at most from the value reported here, which is based on the assumption

that only the triaxial earth-gravity field is being measured by the drift observations. The reported location of the major equatorial axis of the triaxial earth ellipsoid is not expected to change significantly with this later refinement.

In summary, long-term observations on the drift of Syncom II have already established to a high degree of certainty that:

1. The earth must be considered to be a "triaxial ellipsoid" (for example, having a sea-level surface of this form) for the purposes of 24-hour satellite design. (For broader geodetic purposes; a significant longitude dependent gravity field exists, defining:

$$2.2\times10^{-6}<|J_{22}|<1.6\times10^{-6};$$
$$-25^\circ<\lambda_{22}<-13^\circ.)$$

2. The difference between the major and minor equatorial radii of that ellipsoid is not less than 200 feet nor greater than 225 feet.

3. The location of the major equatorial axis of the "triaxial geoid" is between 13 degrees and 25 degrees west of Greenwich.

It may be added that the study of simulated 24-hour satellite drift in a triaxial earth field, influenced also by sun and moon gravity and by sun-radiation pressure perturbations, shows that the theory of longitude drift presented in this report is substantially unaffected by all perturbations except that due to the earth's elliptical equator and possible higher order longitude-dependent earth-gravity anomalies.

1. Basic Theory of this Reduction

(Determination of the Longitude Drift and Orbit Expansion for a 24-Hour Satellite With a Near-Circular Orbit Affected by a Small But Persistent Tangential Perturbing Force.)

The dominant perturbations of a 24-hour equatorial satellite in a higher order earth-gravity field have been derived many times in the literature (References 3, 4, 5, and 6). In these references, the perturbations were found by directly linearizing the equations of motion themselves and displaying the perturbed motion in appropriate geographic coordinates; no attempt was made to treat the drift of the inclined 24-hour satellite.

In this report, I will depart from this rather involved and difficult-to-visualize procedure of linearization-of-the-equations-of-motion. Instead, I will show how simple it is to derive the dominant drift and orbit-expansion equations for the 24-hour satellite by dealing with what can be called "the perturbation of the 2-body energy" of the geographically stationary satellite by the small but persistent longitude-dependent earth-gravity force. This paper will not discuss in detail the limits of validity of the expressions derived. Instead, to assess the accuracy with which these expressions predict the satellite's behavior, simulated trajectories with typical Syncom II orbit elements have been run on Goddard's particle program "ITEM." These trajectories (Appendix A) confirm the validity of the derived drift equations to an accuracy well within the "noise levels" in the orbital elements reported for Syncom II. The equations are essentially the same as those which P. Musen has derived by a more general but more complex "energy perturbation" method (Reference 7).

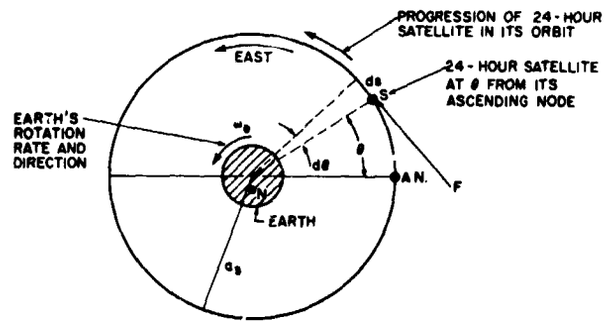


FIGURE 1.—Orbit plane of a 24-hour satellite, looking southerly.

In Figure 1, F is a small earth-gravity perturbation force acting tangentially to an initial circular 24-hour satellite orbit; ds is a small arc length of the satellite's path around the earth. At the beginning of the dynamics, the total energy (the sum of potential and kinetic) of the satellite in a spherical earth-gravity field (Reference 8) is

$$E=-\frac{\mu_E}{2a_s} \qquad (1)$$

where μ_E is the earth's gaussian gravity constant (3.986×10^5 km^3/sec^2). The energy added to the satellite by F per day is

$$\Delta E = \oint F ds = 2\pi a_s \bar{F}, \tag{2}$$

where $\bar{F} = (1/2\pi)\oint F d\theta \cdot F$, in units of force per unit mass, is the orbit averaged energy perturbing force. If the orbit is purely circular, only a tangential perturbation force can cause a change in the total energy. The ITEM simulated trajectories in Appendix A and the real Syncom II orbits both maintain eccentricities of the order of 0.0001 for periods up to 100 days. Equation (2) assumes the eccentricity is zero for the 24-hour satellite of semimajor axis a_s.

From (1), the change in energy of a 24-hour satellite is accompanied by a change in semi-major axis expressed by

$$\Delta E = \frac{\mu_E a_s}{2(a_s)^2}, \quad \text{or}$$

$$\Delta a_s = \frac{2(a_s)^2 \Delta E}{\mu_E}. \tag{3}$$

Substituting (2) into (3), the change in semimajor axis of the 24-hour near-circular orbit, per day, is approximately given by

$$\Delta a_s = \frac{4\pi (a_s)^3 \bar{F}}{\mu_E}. \tag{4}$$

From Kepler's third law, the period of a 24-hour orbit as a function of its semimajor axis is

$$T_s = \frac{2\pi (a_s)^{3/2}}{(\mu_E)^{1/2}}. \tag{5}$$

Thus, if the semimajor axis changes by Δa_s, the period change is given by

$$\Delta T_s = \frac{3\pi (a_s)^{1/2} \nabla a_s}{(\mu_E)^{1/2}}. \tag{6}$$

Substituting (4) into (6), the change in period, per day, of a 24-hour circular orbit is given by

$$\Delta T_s = \frac{12\pi^2 (a_s)^{7/2} \bar{F}}{(\mu_E)^{1/2}}. \tag{7}$$

The apparent net longitudinal drift rate of the 24-hour satellite's ground track with respect to the surface of the earth after the first sidereal day is

$$\dot{\lambda}(t = 1 \text{ sidereal day}) = -\frac{(\Delta T_s) 2\pi}{T_s} (\text{radians/sidereal day}) \tag{8}$$

(See Reference 9 for example). The minus sign is taken in (8) because a gain in period is accompanied by a decrease in net geographic longitude for the initially 24-hour satellite (for example, for the daily geographic position of the ascending node). Combining Equations (7) and (5) in (8) gives

$$\dot{\lambda}(t = 1 \text{ sid. day}) = -\frac{12\pi^2 \bar{F}}{\mu_E/(a_s)^2} \text{ (rad./sid. day)}. \tag{9}$$

As the gain in semimajor axis is small over one day (and, in fact, small compared to a_s, for the entire libration period of the satellite in the triaxial earth field), the drift rate will continue to build up linearly with time initially, adding increments of (9) each day. Thus, the net longitudinal drift acceleration of an initially 24-hour satellite is

$$\ddot{\lambda} = -\frac{12\pi^2 \bar{F}}{\mu_E/(a_s)^2} \text{ (rad./sid. day}^2\text{)} \tag{10}$$

Rewriting (4) as

$$\dot{a} = \frac{4\pi (a_s)^3 F}{\mu_E}, \text{ (length units/sid. day)} \tag{11}$$

gives the expansion rate of the initially 24-hour near-circular satellite orbit due to a small but persistently acting orbit-averaged tangential perturbing force $\bar{F}$.

2. Evaluation of the Perturbing Force

Figure 2 shows the position of the 24-hour satellite with respect to the earth and the celestial sphere. F_r, F_ϕ and F_λ, earth-gravity perturbing forces in the radial, latitude, and longitude directions, are assumed to be acting on the satellite at s. Considering only the earth-gravity perturbation forces arising from the ellipticity of the earth's equator (Reference 10), Appendix B gives these forces as

$$F = \frac{\mu_E (R_0/a_s)^2}{(a_s)^2} \left\{ 9 J_{22} \cos^2 \phi \cos 2(\lambda - \lambda_{2^{2,0}}) \right\} \tag{12}$$

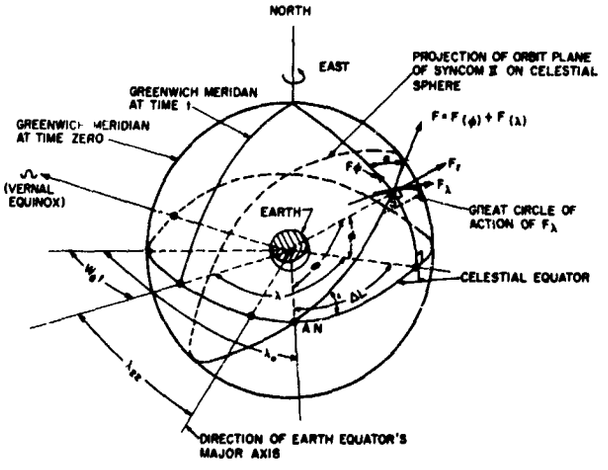


FIGURE 2.—Position of a 24-hour satellite with a near-circular orbit with respect to the earth and the celestial sphere.

$$F_\phi = \frac{\mu_E (R_0/a_s)^2}{(a_s)^2}\{6J_{22}\cos\phi\sin\phi\cos 2(\lambda-\lambda_{22})\}, \tag{12A}$$

$$F_\lambda = \frac{\mu_E (R_0/a_s)^2}{(a_s)^2}\{6J_{22}\cos\phi\sin 2(\lambda-\lambda_{22})\}. \tag{13}$$

As long as the orbit is nearly circular, F_r will have negligible contribution to F. The contribution to F from F_ϕ is:

$$F_{(\phi)} = F_\phi\cos\alpha = K\sin\phi\cos\phi\cos\alpha\cos 2(\lambda-\lambda_{22}) \tag{14}$$

K is a constant for a single orbit.

$$K = 6J_{22}\left(\frac{\mu_E R_0^2}{(a_s)^4}\right). \tag{14A}$$

In the right spherical triangle AN, s, L, note the following trigonometric relations:

$$\cos(i) = \frac{\tan\Delta L}{\tan\theta}, \tag{15A}$$

$$\cos\alpha = \frac{\tan\phi}{\tan\theta}, \tag{15B}$$

$$\sin\phi = \sin(i)\sin\theta, \tag{15C}$$

$$\sin\alpha = \frac{\cos(i)}{\cos\phi}. \tag{15D}$$

From (15A),

$$\Delta L = \tan^{-1}[\tan\theta\cos(i)]. \tag{16}$$

Let the geographic longitude of the satellite at the ascending node (AN) be λ_0. Counting time from this orbital position, the geographic longitude of the 24-hour satellite at s in its near-circular orbit is

$$\lambda = \lambda_0 + \Delta L - \omega_{et} \text{ (Figure 2)}$$

or, using (16),

$$\lambda = \lambda_0 + \tan^{-1}[\tan\theta\cos(i)] - \omega_{et} \tag{17}$$

w_e is the earth's sidereal rotation rate. For the 24-hour satellite (starting the dynamics with S. at A.N., for convenience), $\theta \doteq w_{er}$, so that (17) becomes

$$\lambda = \lambda_0 + \tan^{-1}[\tan\theta\cos(i)] - \theta, \tag{18}$$

approximately. The function $\tan^{-1}[\tan\theta\cos(i)] - \theta$ is even about $\theta=0$ and $\theta=\pi/2$, with a period of π, and behaves like a somewhat distorted sine function (Figure 3 and Appendix C). Call this function $\Delta\lambda$ and note that for $i<33$ degrees, $\Delta\lambda$ is always less than 5°. Thus, using (18) and assuming i is sufficiently small ($i<45°$ proves to be a sufficient restriction on the inclination), $\cos 2(\lambda-\lambda_{22})$ for the 24-hour satellite can be approximated by

$$\cos 2(\lambda-\lambda_{22}) \doteq \cos 2(\lambda_0-\lambda_{22}) - 2\Delta\lambda\sin 2(\lambda_0-\lambda_{22}) = -\cos 2\gamma_0 + 2\Delta\lambda\sin 2\gamma_0. \tag{18A}$$

Similarly, $\sin 2(\lambda-\lambda_{22})$ can be approximated by

$$\sin 2(\lambda_0-\lambda_{22}) + 2\Delta\lambda\cos 2(\lambda_0-\lambda_{22}) = -\sin 2\gamma_0 - 2\Delta\lambda\cos 2\gamma_0 \doteq \sin 2(\lambda-\lambda_{22}) \tag{18B}$$

In Figure 3, γ_0 is the geographic longitude of the node of the 24-hour nearly circular satellite orbit with respect to the minor equatorial axis. With these expansions [(18A) and (18B)], and using (15B), (14) becomes

$$F_{(\phi)} = K\frac{\sin\phi\cos\phi\tan\phi}{\tan\theta}\{-\cos 2\lambda_0 + 2\Delta\lambda\sin 2\lambda_0\}.$$

Using (15C) in the above expression, the contribution to the perturbing force F due to F_φ becomes

$$F_{(\phi)} = K\frac{\sin^2(i)\sin 2\nu}{2}\{-\cos 2\lambda_0 + 2\Delta\lambda\sin 2\lambda_0\}. \tag{19}$$

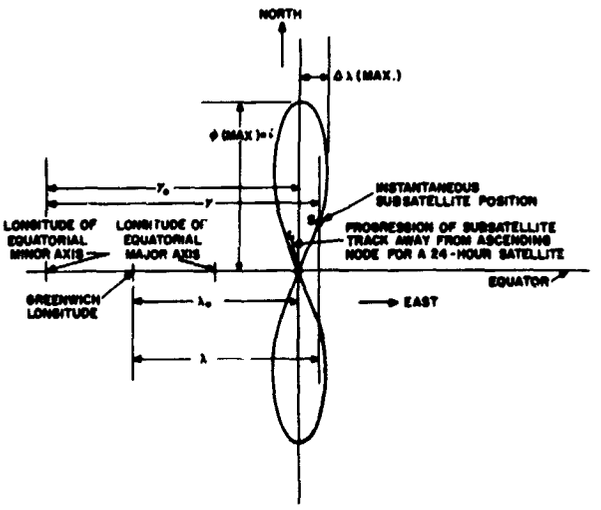


FIGURE 3.—Geographic subsatellite track of 24-hour satellite in a near-circular orbit.

Writing $\Delta\lambda \doteq \Delta\lambda\ (\max) \sin 2\theta$, (19) becomes

$$F_{(\phi)} \doteq K \sin^2 (i)\ [-\cos 2\gamma_0]\ \frac{\sin 2\theta}{2}$$

$$-K \sin^2 (i)\ \Delta\lambda\ (\max)\ [\sin 2\gamma_0 \sin^2 2\theta] \quad (20)$$

Averaging $F_{(\phi)}$ over $0 \leq \theta \leq 2\pi$, (20) gives

$$\bar{F}_{(\phi)} = \frac{1}{2\pi}\int_0^{2\pi} F_{(\phi)} d\theta$$

$$= -\frac{\sin^2 i \Delta\lambda\ (\max)}{2}[K \sin 2\gamma_0] \quad (21)$$

The contribution to F from F_λ is

$$F\omega = F_\lambda \cos (90-\alpha) = F_\lambda \sin \alpha$$
$$= K \sin \alpha \cos \phi \{-\sin 2\gamma_0 - 2\Delta\lambda \cos 2\gamma_0\}, \quad (22)$$

from (13) and (18B). Using (15D) in (22), and noting that $\Delta\lambda \doteq \Delta\lambda\ (\max) \sin 2\theta$, as before, gives the contribution to F from F_λ as

$$F\omega = K \cos (i) \{-\sin 2\gamma_0 - 2\Delta\lambda (\max) \sin 2\theta \cos 2\gamma_0\}. \quad (23)$$

Averaging $F_{(\lambda)}$ over $0 \leq \theta \leq 2\pi$, (23) gives

$$\bar{F}\omega = -K \cos (i) \sin 2\gamma_0. \quad (24)$$

Thus, combining the contributions of the latitude and longitude perturbations to the average perturbation force over a single 24-hour orbit, (21) and (24) sum to produce

$$\bar{F}\ (\text{total}) = -\bar{F}_{(\phi)} + \bar{F}_{(\gamma)}$$
$$= -K \sin 2\gamma_0 \left\{\cos (i) + \frac{\sin^2 (i)\ \Delta\lambda(\max)}{2}\right\}. \quad (25)$$

3. Completion of the Derivation of the Drift Equations

Appendix C shows that

$$\Delta\lambda(\max) = \tan^{-1}[\sec (i)] - 45°.$$

It is also shown there that, to a high degree of accuracy for $i < 50°$,

$$\cos (i) + \frac{\sin^2 \Delta\lambda\ (\max)}{2} \doteq \frac{1+\cos^2 (i)}{2}.$$

Numerically integrated orbits have shown that the drift theory for a 24-hour satellite stemming from (25) is in error by more than 2 percent for $i > 45°$. With this restriction on orbit inclination, using the above approximation for the inclination factor, we can rewrite the longitude drift and orbit expansion equations (10) and (11), evaluating $\bar{F}$ by (25), giving

$$\dot{\gamma} = \frac{12\pi^2 K}{\mu_E/(a_s)^3}\left[\frac{\cos^2 (i)+1}{2}\right] \sin 2\gamma_0,$$
(rad./sid. day²) (26)

$$\dot{a} = -\frac{4\pi (a_s)^2 K}{\mu_E}\left[\frac{\cos^2 (i)+1}{2}\right] \sin 2\gamma_0,$$
(length units/sid. day). (27)

Substituting (14A) into (26) and (27) reduces these expressions to

$$\dot{\gamma} = 72\pi^2 J_{22}(R_0/a_s)^2\left[\frac{\cos^2(i)+1}{2}\right] \sin 2\gamma_0,$$
(rad./sid. day²) (28)

$$\dot{a} = -24\pi J_{22}(R_0/a_s) R_0\left[\frac{\cos^2(i)+1}{2}\right] \sin 2\gamma_0,$$
(length unit/sids. day). (29)

Define a nondimensional change of semimajor axis from a_s during the drift as

$$a_1 = \frac{a - a_s}{a_s} = \frac{\Delta a}{a_s} \text{ so that, } \dot{a}_1 = \frac{\dot{a}}{a_s}. \quad (29A)$$

With (29A), (29) becomes

$$\dot{a}_1 = -24\pi J_{22}(R_0/a_s)^2\left\{\frac{\cos^2(i)+1}{2}\right\}\sin 2\gamma_0, \quad (1/\text{sid. day}) \quad (30)$$

Define:

$$A_{22} = -72\pi^2 J_{22}(R_0/a_s)^2\left\{\frac{\cos^2(i)+1}{2}\right\}, \quad (\text{rad./sid. day}^2). \quad (30A)$$

With (30A), (28) and (30) become;

$$\ddot{\gamma} + A_{22}\sin 2\gamma_0 = 0, \ (\text{rad./sid. day}^2) \quad (31)$$

$$\dot{a}_1 - \frac{A_{22}\sin 2\gamma_0}{3\pi} = 0, \ (1/\text{sid. day}). \quad (32)$$

Note that π in (32) has dimensions of rad./sid. day. It must be understood that (31) describes the *net* daily geographic acceleration of the initially 24-hour satellite with respect to the earth's minor equatorial axis. Stated another way, (31) describes the geographic drift of the entire originally stationary, figure-8 ground track (Figure 3). Similarly, (32) describes the net daily orbit-expansion rate of the 24-hour satellite. In particular, it it convenient to treat the motion of the ascending node of the orbit in geographic longitude as a reference for the entire configuration. In what follows, therefore, γ will refer always to the geographic longitude of the ascending node east of the equatorial minor axis; γ_0 will refer to the initial geographic longitude of the A.N. east of the minor axis, at the start of the dynamics under consideration. (31) and (32) can thus be rewritten in terms of the general nodal longitude position γ, to give the relevant partially uncoupled long-term drift and orbit-expansion differential equations for the near-24-hour near-circular orbit satellite:

$$\ddot{\gamma} + A_{22}\sin 2\gamma = 0, \ (\text{rad./sid. day}^2) \quad (33)$$

$$\dot{a}_1 - \frac{A_{22}\sin 2\gamma}{3\pi} = 0, \ (1/\text{sid. day}). \quad (34)$$

4. General Considerations of the Solutions of the Drift Equation

Equation (33) can be integrated directly for the geographic drift rate by noting that

$$\ddot{\gamma} = \frac{d\dot{\gamma}}{dt} = \frac{d(\dot{\gamma})^2}{2\dot{\gamma}dt} = \frac{d(\dot{\gamma})^2}{2d\gamma}.$$

Thus (33) can be separated to

$$d(\dot{\gamma})^2 = -2A_{22}\sin 2\gamma d\gamma. \quad (35)$$

Since the variables $(\dot{\gamma})^2$ and γ are separated in (35), (35) integrates to

$$(\dot{\gamma})^2 = A_{22}\cos 2\gamma + C_1. \quad (36)$$

With the initial condition that $\dot{\gamma} = \dot{\gamma}_0$ at $\gamma = \gamma_0$, (36) becomes

$$\dot{\gamma} = [(\dot{\gamma}_0)^2 + A_{22}(\cos 2\gamma - \cos 2\gamma_0)]^{1/2}, \quad (37)$$

giving the drift rate of the 24-hour satellite as a function of the initial drift rate $\dot{\gamma}_0$, the earth gravity constant A_{22}, the initial longitude east of the minor axis γ_0, and the instantaneous longitude γ. Returning to the semicoupled system of equations (33) and (34), the explicit dependence of the equations on the location from the minor axis and the magnitude of the equatorial ellipticity may be eliminated by multiplying (33) by $\frac{1}{3}\pi$ and adding the resulting equation to (34), giving

$$\ddot{\gamma} + 3\pi\dot{a}_1 = 0. \quad (38)$$

(38) can be rewritten as

$$\frac{d(\dot{\gamma})}{dt} + 3\pi\dot{a}_1 = 0 = d(\dot{\gamma}) + 3\pi\dot{a}_1 dt = d(\dot{\gamma}) + 3\pi d(a_1) \quad (39)$$

Separation of the variables $\dot{\gamma}$ and a_1 is thus achieved in (39). (39) integrates directly to

$$3\pi a_1 + \dot{\gamma} = C_2. \quad (40)$$

With the initial conditions: $a_1 = 0$, when $\dot{\gamma} = 0$ (the satellite is in the momentarily stationary ground-track configuration); $C_2 = 0$. If γ_0 in (37) is also the longitude of this initially stationary orbit, $(\dot{\gamma}_0)^2 = 0$ there, and (37) in (40) yields for a_1, the semimajor axis change from "synchronism" in the drift motion,

$$a_1 = \pm\frac{(A_{22})^{1/2}(\cos 2\gamma - \cos 2\gamma_0)^{1/2}}{3\pi}, \quad (41)$$

(41) shows explicitly that the semimajor axis is bounded in long-term drift from a stationary orbit. From (33), since $A_{22} > 0$, if $0 < \gamma_0 < 90°$, drift proceeds *towards* the nearest longitude of the earth's equatorial minor axis (in $a-\gamma$ direction). If $-90° < \gamma_0 < 0°$, (33) shows that drift

again proceeds *toward* the nearest minor axis longitude (in $a+\gamma$ direction). Thus, in all cases of drift from a stationary geographic configuration, $\cos 2\gamma - \cos 2\gamma_0$ is a positive function which has a maximum when $\gamma=0$ (when the satellite has drifted over the longitude of the minor axis). Thus (41) gives (for the librations of a 24-hour satellite)

$$a_1 \text{ (max)} = \left| \frac{(A_{22})^{1/2}(1-\cos 2\gamma_0)^{1/2}}{3\pi} \right|. \qquad (42)$$

Again it is noted that π in (42) has units of rad/sid. day. An absolute maximum semimajor axis change in the drift occurs when the "synchronous" condition is established near the longitude of the major equatorial axis. Here, $\gamma_0 = -90°$, $\cos 2\gamma_0 = -1$, and

$$a_1 \text{ (absolute maximum for a librating 24-hour satellite)} = \frac{(2A_{22})^{1/2}}{3\pi} \qquad (43)$$

For the constant $J_{22} = -1.7\times10^{-6}$ (derived in this study from long-term observations on the drift of the Syncom II satellite) and using the additional constants from this study: $i=33°$, $a_s \doteq 42166$ km, $R_0 = 6368.388$ km; (30A) gives

$$A_{22} = 23.2\times10^{-6} \text{ (rad./sid. day}^2\text{)}.$$

(43) then gives

a_1 (absolute max.) $= .72\times10^{-3}$, from which, by (29A),

Δa (absolute max. from a "synchronous" condition near the equatorial major axis, for a satellite of $i=33°$) $= 30.7$ km.

Thus the assumption made in (10) and (11), to approximate the slightly varying semimajor axis by a_s (a constant) throughout the drift motion, appears amply justified.

Figure 4 is a graph of (41) for a_1 vs. γ (the longitude with respect to the nearest minor axis location) as a function of γ_0, the longitude in the initially stationary configuration.

Note that (41) allows equal $\pm$values for a_1 for each γ. Suppose the satellite is initially at $+\gamma_0$ (position 1 in Figure 4) from the nearest location

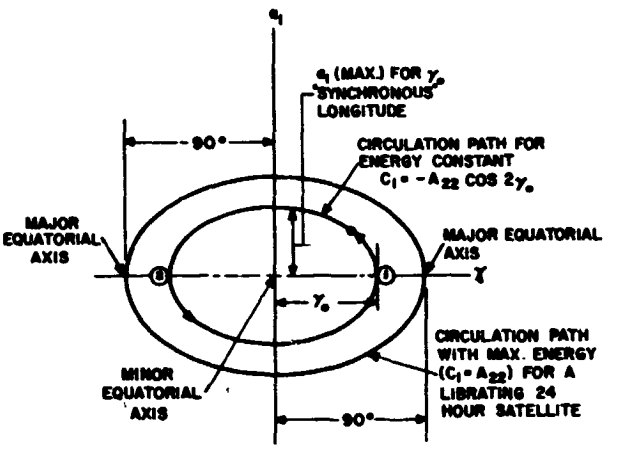


FIGURE 4.—Libration with longitude of the semimajor axis of a 24-hour satellite as a function of the longitude of the initially stationary configuration.

of the minor axis: From (33), $\sin 2\gamma_0$ being positive, the satellite begins to drift west (attaining a negative drift rate) towards the minor axis. But, from (40), since $C_2=0$, $a_1 = -(\gamma)/3\pi > 0$; the drift therefore proceeds counterclockwise in Figure 4, around the central point of the minor axis and $a_1=0$, along the upper portion of the two-valued arc determined from (41).

The same situation holds for the motion beginning or stemming away from the "synchronous" longitude at $-\gamma_0$, position 2 in Figure 4. Here $\sin 2\gamma_0$ is negative, and the drift proceeds at a positive rate to the east. Again from (40), as soon as the satellite leaves position 2, $a_1 = -(\gamma)/3\pi < 0$, and the circulation continues in a counterclockwise direction. Every trajectory in the phase plane $a_1 \leftrightarrow \gamma$ may be conveniently defined by the constant C_1 of the "energy integral" of the drift motion (36). Since (33) is the equation-of-motion defining the large-angle oscillations of a mathematical pendulum (in the case of the 24-hour-orbit satellite, the point of symmetry is the minor axis where $2\gamma=0$), it can be expected that the general solutions in that theory apply to the long-term librations of the "synchronous satellite" (Appendix D). For example, in (36), with a momentarily "synchronous" condition at γ_0 being given by $\gamma_0=0$, the "energy constant" is evaluated as

$$C_1 = -A_{22} \cos 2\gamma_0.$$

With this evaluation, (36) becomes

$$(\dot{\gamma})^2 - A_{22}\cos 2\gamma = -A_{22}\cos 2\gamma_0. \qquad (44)$$

Solving for the initially "synchronous" longitude as a function of any longitude in the drift and the corresponding longitude rate, (44) gives

$$\gamma_0 = \frac{1}{2}\cos^{-1}\left[\cos 2\gamma - \frac{(\dot{\gamma})^2}{A_{22}}\right]. \qquad (45)$$

Since $(\dot{\gamma})^2/A_{22} \geq 0$, the argument of $\cos^{-1}$ in (45) is always less than or equal to 1. Thus, as long as $\cos 2\gamma - (\dot{\gamma})^2/A_{22} \geq -1$, (45) will give a real solution for the momentarily "synchronous" longitude with respect to the minor axis. But, if $\cos 2\gamma - (\dot{\gamma})^2/A_{22} < -1$, there will be no real momentarily "synchronous" configuration for the near-24-hour satellite. With this energy, the world-circulation regime commences, corresponding to the over-the-top, high-energy regime of the mathematical pendulum (Reference 3). The above inequality implies that, for the commencement of "world circulation" for the near-24-hour satellite,

$$(\dot{\gamma})^2 \geq A_{22}(1+\cos 2\gamma), \text{ or}$$
$$(\dot{\gamma})^2 \geq 2A_{22}\cos^2\gamma. \qquad (46)$$

When $2\gamma = 0$, or the satellite is over the minor axis, (46) allows the maximum possible drift rate for a librating 24-hour satellite:

$$\dot{\gamma}(\max)_{\text{for libration}} = (2A_{22})^{1/2}, \text{ (rad./sid. day)}. \qquad (47)$$

For example, using the reported value of $A_{22} = 23.2\times10^{-6}$ rad./day^2 for the inclination of the Spncom II satellite, (47) gives

$$\dot{\gamma}(\max)_{\text{for libration with } J_{22}=-1.7\times10^{-6},\ i=33^\circ} = (46.4\times10^{-6})^{1/2} = .39 \text{ degrees/day}. \qquad (48)$$

5. Approximations to the Exact Drift Solutions for Periods Very Close to Synchronous

Expanding the drift from the "synchronous" longitude ($\gamma = \gamma_0$ in this section) in a Taylor series, with respect to increments of time Δ_t from the momentarily stationary condition,

$$\gamma(\Delta t) = \gamma_0 + \dot{\gamma}_0\Delta t + \ddot{\gamma}_0\frac{(\Delta t)^2}{2} + \dddot{\gamma}_0\frac{(\Delta t)^3}{6} + \gamma_0^{[4]}\frac{(\Delta t)^4}{24} + \gamma_0^{[5]}\frac{(\Delta t)^5}{120} + \gamma_0^{[6]}\frac{(\Delta t)^6}{720} + \gamma_0^{[7]}\frac{(\Delta t)^7}{5040} + \gamma_0^{[8]}\frac{(\Delta t)^8}{40320} + \cdots \qquad (50)$$

Differentiating (33) six times with respect to time, it is clear that all derivatives in (50) can be written as functions of A_{22}, γ_0 and $\dot{\gamma}_0$. Noting that $\gamma(\Delta_t) - \gamma_0 = \Delta\lambda$ (the geographic longitude with respect to the "synchronous" configuration) and $\dot{\gamma}_0 = 0$, (50) can be shown to reduce to the expansion

$$\Delta\lambda = (-A_{22}\sin 2\gamma_0)\frac{(\Delta t)^2}{2} + [(A_{22})^2\sin 4\gamma_0]\frac{(\Delta t)^4}{24} + [(A_{22})^3\sin 2\gamma_0\,(4\sin^2 2\gamma_0 - 1)]\frac{(\Delta t)^6}{180} - [(A_{22})^4\sin 4\gamma_0(34\sin^2 2\gamma_0 - 1)]\frac{(\Delta t)^8}{10080} + \ldots \qquad (51)$$

It is apparent that, as $\Delta_t \to 0$, the higher order terms of (51) become increasingly more insignificant to the total drift, in comparison to the terms of lower order.

In Appendix D, the exact "elliptic integral" of motion from (33) is calculated from a synchronous longitude of 60° east of the minor axis. This calculation demonstrates that the simple term-inclusion-time criterion below gives an adequately converging series to the "exact" drift. In the actual reduction, all higher order terms in (51) which are less in magnitude than the root mean square (rms) error of the observed Syncom II longitudes, are ignored. Section 7 of this report shows that this rms error of longitude determination for the ascending equator crossings of Syncom II from August 1963 to March 1964 has been of the order of $\pm.025$ degrees. Thus, 0.025° is used below in forming the minimum-time-term-inclusion criterion for each term of (51).

A_{22} is assumed to be 23.2×10^{-6} rad./sid. day^2.

A) For inclusion of the $(\Delta t)^4$ term:

$|\sin 4\gamma_0|$ is maximum when $\gamma_0 = \pm 22.5^\circ$ and $\pm 67.5^\circ$.

Therefore,

$$|\Delta\gamma_{\max} \text{ (from the fourth-order term)}| = (A_{22})^2\frac{(\Delta t)^4}{24}. \qquad (52A)$$

Solving (52A) for Δt, when $|\Delta\lambda_{max}$ (4th order)$|$ $=.025°$,

Δt (min. fourth-order term inclusion)
$=(.025\times24/57.3\times[23.2\times10^{-6}]^2)^{1/4}$
$=66.5$ sid. days from "synchronism".

B) For inclusion of the $(\Delta t)^6$ term:

$|\sin 2\gamma_0\ (4\sin^2 2\gamma_0-1)|$ is maximum when γ_0 $=\pm45°$.

Therefore,

$|\Delta\gamma_{max}$ (from the sixth-order term)$|$
$$=(A_{22})^3\frac{(\Delta t)^6}{60}. \qquad (52B)$$

Solving (52B) for Δt, when $|\Delta\lambda_{max}$ (sixth order)$|$ $=.025°$,

Δt (min. for sixth-order term inclusion)
$=(.025\times60/57.3\times[23.2\times10^{-6}]^3)^{1/6}$
$=113.$ sid. days from "synchronism".

C) For inclusion of the $(\Delta t)^8$ term:

$|\sin^4\gamma_0\ (34\sin^2 2\gamma_0-1)|$ is maximum wh $=\pm5914°$.

Therefore,
$|\Delta\lambda_{max}$(from the eighth-order term)$|$
$$=21.2(A_{22})^4\frac{(\Delta t)^8}{10080} \qquad (52C)$$

Solving (52C) for Δt, when $|\Delta\lambda_{max}$ (eight. order)$|$ $=.025°$,

(min. for eight-order term inclusion)
$=(10080\times.025/21.2\times57.3\times[23.2\times10^{-6}]^4)^{1/8}$
$=171.$ sid. days from "synchronism".

Similarly, expanding $a_1\ (t)$ in a Taylor series about the time of "synchronism", $(t_0,\ \gamma_0)$:

$$a_1(t_0+\Delta t)=(a_1)_0+(\dot a_1)_0(\Delta t)+(\ddot a_1)_0\frac{(\Delta t)^2}{2}+(\dddot a_1)_0\frac{(\Delta t)^3}{6}+\cdots. \qquad (53)$$

But, from (34),

$$(\dot a_1)_0=\frac{A_{22}\sin 2\gamma_0}{3\pi}. \qquad (53A)$$

Differentiating (53A) with respect to time,

$$(\ddot a_1)_0=\frac{2\dot\gamma_0A_{22}\cos 2\gamma_0}{3\pi}=0, \qquad (53B)$$

since $\gamma_0=0$. Differentiating (53B) with respect to time,

$$(\dddot a_1)_0=\frac{-4(\dot\gamma_0)^2A_{22}\sin 2\gamma_0}{3\pi}+\frac{2\ddot\gamma_0A_{22}\cos 2\gamma_0}{3\pi}=\frac{-(A_{22})^2\sin 4\gamma_0}{3\pi}, \qquad (54)$$

using equation (33). From the conventional definition of a_1, $(a_1)_0=0$. (53) then becomes

$$a_1\text{ (at }\Delta t\text{ from "synchronism")}=\frac{(A_{22}\sin 2\gamma_0)\Delta t}{3\pi}-\frac{(A_{22})^2\sin 4\gamma_0(\Delta t)^3}{18\pi}+\cdots, \qquad (54A)$$

with the results of (53A), (53B), and (54) in (53).

Section 7 shows that the rms error of semimajor axis determination for Syncom II (including sun and moon "noise") is of the order of ±0.5 km. Therefore, the rms error to be expected in a_1 is of the order of $.5/42166.=1.185\times10^{-5}$. Following the procedure for the longitude drift, 1.185×10^{-5} is used below to determine the minimum time for the inclusion of the terms beyond the first on the righthand side of (54A), to ensure adequate convergence of the infinite series for $a_1(\Delta t)$.

A) For inclusion of the $(\Delta t)^3$ term:

$|\sin 4\gamma_0|$ is maximum when γ_0 $=\pm22.5°$ and $\pm67.5°$.

Therefore,

$|a_{1(max)}$, [from the third-order term of (54A)]$|$
$$=\frac{(A_{22})^2(\Delta t)^3}{18\pi}. \qquad (55)$$

Solving (55) for Δt, when $|a_1$ (max)$|$ $=1.185\times10^{-5}$,

Δt [min. for the third-order term inclusion in (54A)] $=\left[\frac{1.185\times10^{-5}\times18\pi}{23.2\times10^{-12}}\right]^{1/3}$
$=$ 108 sid days from synchronism.

From a "synchronous" configuration at 54.8° west of Greenwich, on or about 6 September 1963, Syncom II drifted to 59.2° west of Greenwich on 28 November 1963, where it was "stopped" by the tangential firing of on-board cold-gas jets. A second free-drift period followed from a "synchronous" configuration at 59 2° west on about 29 November 1963, to 66.3° west on 18 March 1964, where the on-board tangential jets were fired to speed up the westward drift. Of the 34 separate orbits calculated by the Goddard Data and Tracking Systems Directorate for these free-drift periods, only 7 fell outside the minimum 66-day period around a condition of "synchronism",

for which the inclusion of higher order terms in (54A) would be necessary in reducing the drift data according to that theory. The data reduction of Section 7 includes only those orbits falling within the minimum 36-day period around "synchronism". Further refinement of this reduction to include the 7 outside-of-synchronous orbits (according to the criterion of this chapter), will be made in the near future. This refinement is not expected to materially affect the results of this report.

6. Determination of Earth-Equatorial Ellipticity from Two Observations of Drift Acceleration At a Given Longitude Separation

Given two independent near-synchronous drifts (in the sense discussed previously), whose momentarily synchronous longitudes $(\gamma_0)_1$ and $(\gamma_0)_2$ are separated by $\nabla\lambda$. Let the two drift accelerations at these two "synchronous" configurations be $(\ddot{\gamma}_0)_1$ and $(\ddot{\gamma}_0)_2$. The drift accelerations may be determined from drift-data reduction according to the theory of (51).

From (33),

$$(\ddot{\gamma}_0)_1 = -(A_{22})_1 \sin 2(\gamma_0)_1 \tag{56}$$

$$(\ddot{\gamma}_0)_2 = -(A_{22})_2 \sin 2[(\gamma_0)_1 + \nabla\lambda], \tag{57}$$

since $(\gamma_0)_2 - (\gamma_0)_1 = \Delta\lambda = (\lambda_0)_2 - (\lambda_0)_1$. Expanding (57) and dividing by (56) gives

$$\left[\frac{(\ddot{\gamma}_0)_2}{(\ddot{\gamma}_0)_1}\right]\left[\frac{(A_{22})_1}{(A_{22})_2}\right] = \cos 2\nabla\lambda + \sin 2\nabla\lambda \cot 2(\gamma_0)_1 \tag{58}$$

Solving (58) for $(\gamma_0)_1$,

$$(\gamma_0)_1 = \frac{1}{2}\tan^{-1}\left\{\frac{\sin(\nabla\lambda)}{\frac{(\ddot{\gamma}_0)_2(A_{22})_1}{(\ddot{\gamma}_0)_1(A_{22})_2} - \cos 2(\nabla\lambda)}\right\}. \tag{59}$$

The quadrant of $(\gamma_0)_1$ is either the first or the fourth, because drift acceleration is always in the direction of the *nearest* longitude extension of the earth's minor equatorial axis. Once the minor axis is located by (59), the absolute value of J_{22} in the earth's triaxial gravity field can be determined through (56) and (30A), for example, as

$$J_{22} = \frac{(A_{22})_1}{72\pi^2[R_0/(a_s)_1]^2\left[\frac{\cos^2(i)_1+1}{2}\right]} = \frac{(\ddot{\gamma}_0)_1}{72\pi^2 \sin 2(\gamma_0)_1[R_0/(a_s)_1]^2\left[\frac{\cos^2(i)_1+1}{2}\right]} \tag{60}$$

Note that the units of $(\ddot{\gamma}_0)_1$ in (60) must be those of radians/sidereal day^2 so that J_{22} will be dimensionless. Note also that in (59), using the result of (30A),

$$\frac{(A_{22})_1}{(A_{22})_2} \simeq \left[\frac{(a_s)_2}{(a_s)_1}\right]^2\left[\frac{\cos^2(i)_1+1}{\cos^2(i)_2+1}\right]. \tag{61}$$

Using (59), since $(\lambda_0)_1$ is known from the data reduction (the geographic longitude of the "synchronous" configuration), the geographic longitude of the nearest minor axis location can be calculated as

$$\gamma_{22} = (\lambda_0)_1 - (\gamma_0)_1. \tag{61A}$$

Similarly, the geographic longitude of the nearest major equatorial axis location can be calculated from

$$\lambda_{22} = (\lambda_0)_1 - (\gamma_0)_1 + 90^\circ \tag{61B}$$

(See figure 3).

Following the theory of Reference 10, the difference in major and minor equatorial radii of the earth's triaxial geoid (a_0-b_0), is related to the gravity constant J_{22} by

$$a_0-b_0=-6R_0J_{22}. \tag{62}$$

7. Reduction of 27 Syncom II Orbits to Determine the Earth's Equatorial Ellipticity

Appendix A tabulates the 27 Syncom II orbits from which the reduction below was made. Table 1 gives the estimated ascending equator crossings nearest to the epoch of these orbits. These were calculated by hand, and therefore are listed only to 0.01 degrees and 0.01 days. The technique used was to locate from the Nautical Almanac, the geographic longitude of the ascending node at epoch through the reported right ascension of the ascending node for the orbit, and the hour-angle of the vernal equinox calculated at epoch. The geographic longitude of the ascending equator crossing was then estimated by turning the earth back through the orbit angle from the ascending node to the satellite at epoch. This latter quantity was estimated as $\omega-M$ for the near-circular orbit of Syncom II. A correction factor to this orbit angle- -the ratio of the satellite's period to the earth's sidereal period—was applied for orbits whose period was sufficiently different from the earth's. The nodal longitude at epoch, minus this reduced nodal excursion angle, is the estimated "ascending equator crossing nearest to epoch" reported in Table 1. (See Appendix E for an example of this calculation.)

Table 2 gives the Goddard-reported semimajor axes for these 27 orbits. Truncating equations (51) and (54A) at their first righthand terms:

$$\Delta\lambda \text{ (longitude drift from "synchronism")} \doteq -(A_{22}\sin 2\gamma_0)\frac{(\Delta t)^2}{2} \tag{63}$$

$$a_1 \text{ (semimajor axis change from "synchronism")} \doteq A_{22}\frac{\sin 2\gamma_0(\Delta t)}{3\pi}. \tag{64}$$

Let the drift time be given from a certain arbitrary base time by T. Let T_0 be the time of "synchronism" from the base time. Let the drift be given from a certain arbitrary geographic longitude by λ. Let λ_0 be the geographic longitude from this base longitude, of the "synchronous" configuration. Then:

$$\Delta_t = T-T_0, \text{ and}$$
$$\Delta\lambda=\lambda-\lambda_0.$$

With these changes, (63) and (64) become (noting that $a_1=\frac{a-a_s}{a_s}$):

$$\lambda=\lambda_0-\frac{(A_{22})(\sin 2\gamma_0)}{2}-(T^2-2TT_0T_0^2), \text{ or}$$

$$\lambda=\left\{\lambda_0-\frac{T_0^2(A_{22}\sin 2\lambda_0)}{2}\right\}+T\left\{T_0A_{22}\sin 2\gamma_0\right\}+T^2\left\{-\frac{A_{22}\sin 2\gamma_0}{2}\right\}. \tag{65}$$

(Note that from (65), $\dot{\lambda}=-A_{22}\sin 2\gamma_0=\gamma_0$ from (33). This result is valid only for orbits sufficiently close to "synchronous," as discussed previously.)

$$a=a_s+a_s\left[\frac{A_{22}\sin 2\gamma_0}{3\pi}\right][T-T_0], \text{ or}$$

$$a=a_s\left(1-\frac{T_0A_{22}\sin 2\gamma_0}{3\pi}\right)+T\left(\frac{a_sA_{22}\sin 2\gamma_0}{3\pi}\right). \tag{66}$$

TABLE 1.—*Estimated Ascending Equator Crossings Nearest the Epoch of 27 Syncom II Orbits.*

First Drift Orbit #	Time from 20.0 Aug. 1963 (days)	Ascending Equator Crossing in Degrees West of 50.0° West	Second Drift Orbit #	Time from 26.0 Nov. 1963 (days)	Ascending Equator Crossing in Degrees West of 50.0° West
1-1	2.12	4.89	2-1	1.86	9.17
1-2	7.11	4.83	2-2	7.84	9.17
1-3	11.09	4.78	2-3	13.83	9.22
1-4	16.08	4.74	2-4	20.81	9.38
1-5	20.07	4.77	2-5	41.75	10.15
1-6	23.06	4.78	2-6	44.74	10.36
1-7	28.05	4.85	2-7	55.71	11.02
1-8	31.04	4.90	2-8	64.69	11.76
1-9	38.02	5.06	2-9	71.67	12.32
1-10	42.01	5.09	2-10	76.66	12.81
1-11	48.09	5.45	2-11	83.64	13.49
1-12	54.97	5.68			
1-13	62.95	6.09			
1-14	70.93	6.60			
1-15	77.91	7.14			
1-16	83.90	7.61			
	~100.0	First free drift period ends at an ascending equator crossing of ~9.15° west of 50.0° west.			

TABLE 2.—*Goddard-Reported Semimajor Axes for 27 Syncom II Orbits.*

First Drift Orbit#	Time from 20.0 Aug. 1963 (days)	Semimajor Axis: (42160.0+Data; km)	Second Drift Orbit#	Time from 26.0 Nov. 1963 (days)	Semimajor Axis: (42160.0+Data; km)
1-1	2.27	4.58	2-1	2.04	5.89
1-2	6.71	4.52	2-2	8.00	7.20
1-3	11.00	6.02	2-3	14.00	7.18
1-4	16.00	6.39	2-4	20.71	8.17
1-5	20.00	6.35	2-5	41.71	8.01
1-6	23.08	6.55	2-6	44.25	9.90
1-7	28.08	7.70	2-7	55.88	11.43
1-8	31.08	7.42	2-8	64.83	11.91
1-9	38.08	7.51	2-9	71.67	12.89
1-10	42.08	8.88	2-10	76.79	13.31
1-11	49.08	9.14	2-11	83.71	14.89
1-12	55.08	9.78			
1-13	63.08	11.51			
1-14	71.00	11.09			
1-15	78.00	12.15			
1-16	84.21	12.51			
	~100.0	First free drift period ends with a semimajor axis of ~42174.5 km.			

(66) and (65) may be written with determinable coefficients as

$$\lambda = d_0 + d_1 T + d_2 T^2, \tag{67}$$

$$a = e_0 + e_1 T, \tag{68}$$

where:

$$d_0 = \lambda_0 - \frac{T^2 A_{22} \sin 2\gamma_0}{2},$$

$$d_1 = (+A_{22} T_0 \sin 2\gamma_0),$$

$$d_2 = -\frac{A_{22} \sin 2\gamma_0}{2}, \tag{69}$$

$$e_0 = a_s\left(1 - \frac{T_0 A_{22} \sin 2\gamma_0}{3\pi}\right),$$

$$e_1 = a_s A_{22} \frac{\sin 2\gamma_0}{3\pi}$$

From (69),

$$T_0 = -d_1/2d_2, \tag{70}$$

$$\lambda_0 = d_0 - \frac{(d_1)^2}{4d_2}, \tag{71}$$

$$\dot{\gamma}_0 = -A_{22} \sin 2\gamma_0 = 2d_2. \tag{72}$$

Alternately, and as an internal check on the theory of the coupling of the drift and orbit expansion,

$$\dot{\gamma}_0 = -A_{22} \sin 2\gamma_0 = \frac{-3\pi e_1}{a_s}, \text{ implying}$$

$$d_2 = 3\pi e_1 / 2a_s. \tag{73}$$

In equation (73), the units of d_2 must be radians/sidereal day^2, and the units of e_1 must be length/sidereal day so that the equation will be dimensionally correct. The semimajor axis at the "synchronous" configuration is calculated from (68) for $T = T_0$:

$$a_s = e_0 + e_1 T_0. \tag{74}$$

For the first drift period (orbits 1–1 through 1–16), the best estimates (in the "least squares" sense) of the coefficients $(d)_1$ and $(e)_1$, obtained by fitting (67) and (68) to the data in Tables 1 and 2, have been found to be:

$$(d_0)_1 = (4.941 \pm .018) \text{ degrees}$$
$$(d_1)_1 = -(.0216 \pm .0010) \text{ degrees/solar day}$$
$$(d_2)_1 = (6.37 \pm .11) \times 10^{-4} \text{ degrees/solar day}^2$$
$$= (6.33 \pm .11) \times 10^{-4} \text{ degrees/sid. day}^2$$

$(e_0)_1 = (4.35 \pm .19)$ km
$(e_1)_1 = (.0993 \pm .0042)$ km/solar day
$(.0990 \pm .0042)$ km/sid. day.

The mean value of the inclination during this period was

$$(i)_1 = 33.018 \pm .005 \text{ degrees.}$$

From (70),

$$(T_0)_1 = \left(16.95 \begin{matrix} +1.09 \\ -1.05 \end{matrix}\right) \text{ days from 29.0 August 1963.}$$

From (71),

$$(\lambda^0)_1 = (4.76 \pm .03) \text{ degrees west of 50.0 degrees west longitude.}$$

From (72),

$$(\dot{\gamma}_0)_1 = -(1.27 \pm .02) \times 10^{-3} \text{ degrees/solar day}^2 = -(2.20 \pm .04) \times 10^{-6} \text{ rad./sid. day}^2.$$

From (74) and the above value of $(T_0)_1$,

$$(a_e)_1 = (42166.0 \pm .2) \text{ km.}$$

For the second drift period (orbits 2-1 through 2-11), the best estimates (in the "least squares" sense) of the coefficients $(d)_2$ and $(e)_2$, obtained by fitting (67) and (68) to the data in Tables 1 and 2, have been found to be:

$(d_0)_2 = (9.156 \pm .017)$ degrees
$(d_1)_2 = -(.0030 \pm .0010)$ degrees/solar day
$(d_2)_2 = (6.59 \pm .11) \times 10^{-4}$ degrees/solar day^2
$= (6.55 \pm .11) \times 10^{-4}$ degrees/sid. day^2

$(e_0)_2 = (5.70 \pm .42)$ km
$(e_1)_2 = (.0994 \pm .0080)$ km/solar day
$= (.0990 \pm .0080)$ km/sid. day.

The mean value of the inclination during this period was

$$(i)_2 = 32.851 \pm .010 \text{ degrees.}$$

From (70),

$$(T_0)_2 = 2.3 \pm .8 \text{ days from 26.0 November 1963.}$$

From (71),

$$(\lambda_0)_2 = 9.15 \pm .02 \text{ degrees west of 50.0 degrees west longitude.}$$

From (72),

$$(\dot{\gamma}_0)_2 = -(1.32 + .02) \times 10^{-3} \text{ degrees/solar day}^2 = -(2.29 \pm .04) \times 10^{-6} \text{ rad./sid. day}^2.$$

From (74) and the above value of $(T_0)_2$,

$$(a_e)_2 = (42165.9 \pm .4) \text{ km.}$$

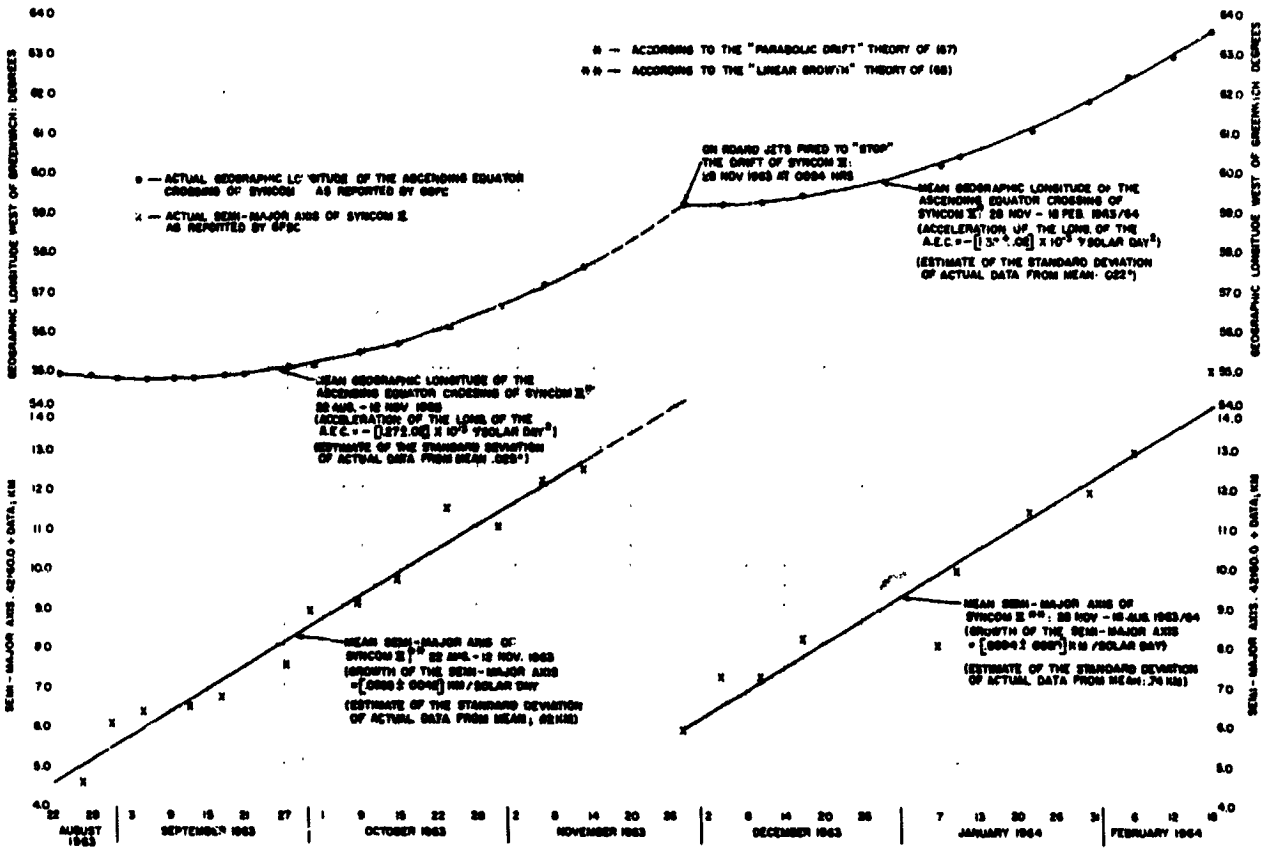


FIGURE 5.—Drift of the ascending equator crossing and growth of the semimajor axis of Syncom II.

(See Figure 5 for a graph of this orbit data and reduction for the two drift periods.) Combining the above results of the two free-drift periods, from (61),

$$\frac{(A_{22})_1}{(A_{22})_2} = (42165.9 \pm .4/42166.0 \pm .2)^2 \frac{[\cos^2 (33.018 \pm .005) + 1]}{[\cos^2 (32.851 \pm .010) + 1]}$$
$$= .99845 \pm .00014.$$

The longitude separation between the two drift periods is given by

$$\nabla\lambda = (\lambda_0)_2 - (\lambda_0)_1 = [-(59.15 \pm .02)] - [-(54.76 \pm .03)] \text{ degrees}$$
$$= -(4.39 \pm .05) \text{ degrees geographic longitude.}$$

Thus

$$2\nabla\lambda = -(8.78 \pm .10) \text{ degrees geographic longitude.}$$

Therefore, from (59), the location of the minor equatorial axis with respect to the "synchronous" longitude during the first free-drift period (54.76 ± .03 degrees west of Greenwich) is

$$(\gamma_0)_1 = \frac{1}{2} \tan^{-1} \left\{ \frac{\sin [-(8.78 \pm .10)]}{\frac{(1.32 \pm .02)}{(1.27 \pm .02)} (.99845 \pm .00014 - \cos [-(8.78 \pm .10)])} \right\}$$

$$= 54 \begin{matrix} +4 \\ -6 \end{matrix} \text{ degrees east of the minor equatorial axis.}$$

From (61B), the best estimate of the location of the major equatorial axis is

$$\lambda_{22} = -55 - \left(54^{+4}_{-6}\right) + 90 = -\left(19^{+4}_{-6}\right) \text{ degrees geographic longitude}$$

From (60), the best estimate of the triaxial gravity coefficient J_{22} is

$$J_{22} = \frac{-(2.20 \pm .04) \times 10^{-6}}{72\pi^2 \left[\sin 2\left(54^{+4^\circ}_{-6^\circ}\right)\right](6378.2/42166.0 \pm .3)^2 \left[\frac{\cos^2(33.018 \pm .005) + 1}{2}\right]}$$

$$= -\left(1.67^{+.07}_{-.03}\right) \times 10^{-6}.$$

The mean equatorial radius, taken as $R_0 = 6378.2$ km, is a compromise between a number of currently used values. It is stated above without error. The likely error in (a_e) has been increased arbitrarily by 0.1 km. to account for the likely uncertainty in R_0.

Using the above estimate of J_{22} from observations on Syncom II drift, the difference between the major and minor equatorial radii of the triaxial geoid is, by (62),

$$a_0 - b_0 = 64^{+3}_{-1} \text{ meters} = 210^{+10}_{-4} \text{ feet.}$$

Comparing the deviation due to earth ellipticity with other higher order earth-gravity deviations (Appendix B and Reference 10), we note that the above figure implies a maximum deviation from the mean earth sphere, due to the ellipticity of the equator, of

$$\Delta R_0 = 105^{+5}_{-2} \text{ feet.}$$

REFERENCES

1. Heiskanen and Meinesz: 1958, *The Earth and Its Gravity Field*, (McGraw-Hill).
2. Izsak: July, 1963, "Tesseral Harmonics In The Geopotential," *Nature* July 13, 1963, pp. 137–139.
3. Blitzer, Boughton, Kang and Page: 1962, "Effect of Ellipticity of the Equator On 24-Hour Nearly Circular Satellite Orbits," *Journal of Geophysica Research*, Vol. 67, pp. 329–335.
4. Frick and Garber: 1962, private communication (Rand Corp. Memo. RM-2296).
5. Barrett: 1962, "The Perturbations of a Synchronous Satellite Resulting From The Gravitational Field of a Triaxial Earth," GSFC Document X-623-62-160, Sept. 10, 1962, NASA.
6. Wagner: 1964, "The Drift of a 24-Hour Equatorial Satellite Due To An Earth Gravity Field Through 4th Order," NASA TN-D-2103, Feb. 1964.
7. Musen: 1962, "On The Motion Of A 24-Hour Satellite," *Journal of Geophysical Research*, Vol. 67 #3, pp. 1123–1132.
8. Thomson: 1961, *Introduction To Space Dynamics*, p. 72 (John Wiley and Sons).
9. Isley: 1962, "A Sommary of Constants Associated With Orbital Analysis of Earth Satellites, Including the Influence Of Their Uncertainties Upon Gravitational Measurements For Synchronous Satellites," GSFC Document X-623-62-169.
10. Wagner: 1962, "The Gravitational Potential of a Triaxial Earth," GSFC Document X-623-62-206, Oct. 62, NASA.
11. Salvadori and Schwarz: 1954, *Differential Equations in Engineering Problems*, Chapter 9 (Prentice Hall).
12. Jahnke and Emde: 1945 ed., *Table Of Functions*, Chapter V (Dover Publications).
13. Kozai: 1961, "The Earth's Gravitational Potential Derived from Motions of Three Satellites," *Astronomical Journal*, Vol. 66, pp. 8–10.
14. O'Keefe, Eckels and Squires: 1959, "The Gravitational Field of the Earth," *Astronomical Journal*, vol. 64, pp. 245–253.

LIST OF SYMBOLS

J_{nm}, λ_{nm} = Spherical harmonic constants (order n, power m) of the earth's gravity potential

F = A gravity force per unit mass acting on a 24-hour satellite

θ = (except in Appendix F) The argument from the ascending node to the satellite position for the 24-hour orbit

a, a_s = The instantaneous semimajor axis, and the "momentarily synchronous" semimajor axis of the orbit of the 24-hour earth satellite. (a_s, estimated to within 2 km, is 42166 km.)

ds = A small arc length of a space trajectory

μ_E = The earth's gaussian gravity constant (3.986×10^5 km^3/sec.^{2})

T_p, T_s = The orbital period for a satellite, and the "momentarily synchronous" period of a 24-hour satellite (i.e., the earth's sidereal rotation period)

λ, r, ϕ = Geographic longitude, geocentric radius, and geocentric latitude of the 24-hour satellites position

$(\dot{}), (\ddot{}), ()^n = \frac{d()}{dt}, \frac{d^2()}{dt^2}, \frac{d^n()}{dt^n}$: time differentials

R_0 = The mean equatorial radius of the earth (6378.2 km)

i = The inclination of the orbit of the 24-hour satellite

λ_0 = The "initial" geographic longitude of the satellite, or the ascending node of the 24-hour satellite's orbit at the start of the dynamics under consideration

ω_e = The earth's sidereal rotation rate ($.7292115 \times 10^{-4}$ rad./sec.)

t = Real time

$\Delta( )$ = A small argument ( )

γ_0 = The geographic longitude (positive to the east) of the ascending node of the 24-hour satellite's orbit with respect to the earth's minor equatorial axis' longitude location, at the start of the dynamics under consideration

γ = The geographic longitude (positive to the east) of the 24-hour satellite, or the ascending node of the satellite's orbit with respect to the longitude of the earth's minor equatorial axis

$a_1 = (a - a_s)/a_s$; a nondimensional semimajor axis change for the 24-hour satellite's orbit, with respect to the "momentarily synchronous" semimajor axis

A_{22} = The driving function causing drift and orbit expansion of a 24-hour satellite in a "triaxial" earth-gravity field; a constant for a given 24-hour orbit inclination

$( )_0$ = The argument ( ) at the start of the dynamics under consideration

$( )_n$ = The argument ( ) at a specified location n (except in Appendix A: ( )s; the argument for the simulated trajectory)

$\nabla\lambda$ = The geographic longitude difference between two "momentarily synchronous" 24-hour satellite configurations

a_0, b_0 The major and minor equatorial radii of the "triaxial" earth $\left(R_0 = \frac{a_0 + b_0}{2}\right.$, according to the definition in Reference 10$\left.\right)$

w = The argument of perigee in a satellite orbit: The orbit angle (from the center of the earth) from the ascending node to perigee

M = The mean anomaly of the satellite in its orbit: The orbit angle (from the center of the earth) from perigee to a point M in the orbit, where $M = \frac{2\pi t}{T_p}$, t being the real time since perigee passage and T_p the period of the satellite's orbit

(e_n), (d_n) = Determinable coefficients in the drift and orbit-expansion equations (67) and (68)

T_0 = The time of "synchronism" from an arbitrary base time of reckoning T

$F(i)$ = The inclination factor in the triaxial driving function A_{22}

V_E = The gravity potential of the earth

g_0, g_s = The radial acceleration of the earth's gravity field at the earth's surface, and at the altitude of the "synchronous" satellite

m = A test mass

$F(k, \phi)$ = The elliptic integral of the first kind with argument (or amplitude) ϕ and modulus k

Λ = The longitude location of the vernal equinox

APPENDIX A

REDUCTION OF SIMULATED PARTICLE TRAJECTORIES FOR EARTH EQUATORIAL ELLIPTICITY

Tables A–1 and A–2 present data taken from a numerically integrated particle trajectory of a triaxial earth in the presence of the sun and moon's gravity field. Only perturbed equations of motion from a periodically rectified Keplerian reference orbit are actually integrated by the digital computer program (called ITEM at the Goddard Space Flight Center). For the 3 months' real orbit time of these trajectories, the accumulated truncation and roundoff error in the numerical integration is believed to be negligible for the purposes of this reduction. The initial position and velocity conditions for these simulated trajectories were the same as those reported for the "actual" Syncom II orbits 1–2 (for the trajectory of Table A–1) and 2–3 (for the trajectory of Table A–2). The program used the earth gaussian-gravity constant

$$\mu_E = 3.9862677 \times 10^5 \text{ km}^3/\text{sec}^2,$$

which is the gravity constant used by the GSFC Data and Tracking Systems Directorate in computing the elements of satellite orbits from radar and Minitrack observations. The best estimates (in the "least squares" sense) of the coefficients $(d)_{s,1}$ and $(e)_{s,1}$, obtained by fitting the drift and orbit expansion equations (67) and (68) to the data in Table A–1, have been found to be

$$(d_0)_{s,1} = 4.841 \pm .004 \text{ degrees}$$

$$(d_1)_{s,1} = -(1.22 \pm .03) \times 10^{-2} \text{ degrees/solar day}$$

$$(d_2)_{s,1} = (6.303 \pm .038) \times 10^{-4} \text{ degrees/solar day}^2$$
$$= (6.268 \pm .038) \times 10^{-4} \text{ degrees/sid. day}^2$$

$$(e_0)_{s,1} = 5.45 \pm 41 \text{ km}$$

$$(e_1)_{s,1} = (.091 \pm .010) \text{ km/solar day}$$
$$= (.091 \pm .010) \text{ km/sid. day.}$$

The mean value of the inclination during this first simulated trajectory period was

$$(i)_{s,1} = 33.005 \pm .003 \text{ degrees.}$$

From (70),

$$(T_0)_{s,1} = 9.68 \pm .30 \text{ days from 26.709 August 1963}$$

From (71),

$$(\lambda_0)_{s,1} = 4.782 \begin{matrix} +.008 \\ -.007 \end{matrix} \text{ degrees west of 50.0 degrees west longitude.}$$

From (72),

$$(\dot{\gamma}_0)_{s,1} = -(1.261 \pm .008) \times 10^{-3} \text{ degrees/solar day}^2$$
$$= -(2.188 \pm .013) \times 10^{-6} \text{ rad./sid. day}^2.$$

TABLE A-1.—*Data from Simulated Trajectory Beginning with the Elements of Syncom II Orbit 1-2.* ($J_{22} = -1.68 \times 10^{-6}$, $R_0 = 6378.388$ km, $\gamma_{22} = -108.0$: *Input into Trajectory Program*)

Time from 26.709 Aug. 1963: (Solar Days)	Ascending Equator Crossing: (Degrees West of 50.0° West Geog. Long.)	Semimajor Axis (42160.0+Data; km)	Inclination (32.0+Data: Degrees)
2.390	4.816	5.27	1.089
8.374	4.783	7.09	1.072
14.358	4.792	6.01	1.056
20.341	4.861	8.12	1.043
26.324	4.954	7.13	1.025
32.308	5.101	8.98	1.019
38.292	5.291	8.31	.997
44.276	5.537	9.67	.991
47.268	5.678	10.38	.983
50.260	5.821	10.03	.972
53.253	5.975	9.42	.967
56.245	6.144	9.74	.966
59.237	6.326	11.09	.960
62.229	6.522	11.94	.957

TABLE A-2.—*Data from Simulated Trajectory Beginning with the Elements of Syncom II Orbit 2-3.* ($J_{22} = -1.68 \times 10^{-6}$, $R_0 = 6378.388$ km, $\gamma_{22} = -108.0$: *Imput into Trajectory Program*)

Time from 10.0 Dec. 1963: (days)	Ascending Equator Crossing: (Degrees West of 50.0° West Geog. Long.)	Semimajor Axis (42160.0+Data; km)	Inclination (32.0+Data; Degrees)
0.823	9.243	6.88	.881
5.809	9.351	7.30	.881
10.796	9.495	9.41	.877
15.783	9.666	8.15	.864
20.769	9.885	9.85	.864
25.756	10.134	10.60	.850
30.743	10.401	9.95	.842
35.730	10.708	11.81	.841
40.717	11.044	12.18	.825
45.704	11.412	11.58	.816
50.692	11.830	13.93	.808
55.679	12.259	13.11	.790
58.672	12.534	13.07	.785
60.667	12.724	13.69	.784

From (74) and the above value of $(T_0)_{s,2}$,

$$(a_s)_{s,2} = 42166.3 \pm .4 \text{ km.}$$

The best estimates (in the "least squares" sense) of the coefficients $(d)_{s,2}$ and $(e)_{s,2}$, obtained by fitting the drift and orbit expansion equations (67) and (68) to the data in Table A-2, have been found to be

$$(d_0)_{s,2} = 9.224 \pm .004 \text{ degrees}$$

$$(d_1)_{s,2} = (1.830 \pm .028) \times 10^{-2} \text{ degrees/solar day}$$

$$(d_2)_{s,2} = (6.501 \pm .042) \times 10^{-4} \text{ degrees/solar day}^2$$

$$= (6.465 \pm .042) \times 10^{-4} \text{ degrees/sid. day}^2$$

$$(e_0)_{s,2} = 7.19 \pm .37 \text{ km}$$

$$(e_1)_{s,2} = (.111 \pm .010) \text{ km/solar day}$$

$$= (.111 \pm .010) \text{ km/sid. day.}$$

The mean value of the inclination during the second simulated trajectory period is

$$(i)_{s,2} = 32.836 \pm .003 \text{ degrees.}$$

From (70),

$$(T_0)_{s,2} = -(14.07 \pm .30) \text{ days from 10.0 December 1963.}$$

From (71),

$$(\lambda_0)_{s,2} = 9.095 \begin{matrix} +.009 \\ -.009 \end{matrix} \text{ degrees west of 50.0 degrees west longitude.}$$

From (72),

$$(\dot{\gamma}_0)_{s,2} = -(1.300 \pm .008) \times 10^{-3} \text{ degrees/solar day}^2$$

$$= -(2.257 \pm .015) \times 10^{-6} \text{ rad./sid. day}^2.$$

From (74) and the above value of $(T_0)_{s,2}$,

$$(a_s)_{s,2} = 42165.6 \pm .5 \text{ km.}$$

A graph of these trajectory simulations is seen in Figure A-1.

Combining the above results for the two simulated trajectories: from (61),

$$\frac{(A_{22})_{s,1}}{(A_{22})_{s,2}} = [(42165.6 \pm .5/42166.3 \pm .4)^2] \left[\frac{\cos^2 (33.005 \pm .003) + 1}{\cos^2 (32.836 \pm .003) + 1} \right]$$

$$= .99840 \pm .00007$$

$$\nabla\lambda = (\lambda_0)_{s,2} - (\lambda_0)_{s,1} = [-(59.095 \pm .009)] - [-(54.782 \pm .008)],$$

$$\therefore 2\nabla\lambda = -(8.626 \pm .034) \text{ degrees geographic longitude.}$$

Therefore, from (59), the location of the minor equatorial axis with respect to the "synchronous longitude" during the first simulated trajectory (54.782 ± .008 degrees west of Greenwich) is

$$(\gamma_0)_{s,1} = \frac{1}{2} \tan^{-1} \left\{ \frac{\sin [-(8.626 \pm .034)]}{\frac{(1.300 \pm .008}{(1.261 \pm .008} (.99840 \pm .00007) - \cos [-(8.626 \pm .034)]} \right.$$

$$= 52.5 \pm 2.5 \text{ degrees east of the minor equatorial axis.}$$

From (61A), the best estimate of the geographic location of the nearest extension of the equatorial minor axis from the simulated trajectory data is

$$(\gamma_{22})_s = -54.8 - (52.5 \pm 2.5) = -(107.3 \pm 2.5) \text{ degrees geographic longitude.}$$

This compares well with the input value of $(\gamma_{22})_s = -108.0°$ used to compute the simulated trajectories. From (60), the best estimate of the triaxial gravity coefficient J_{22} from the simulated data (according

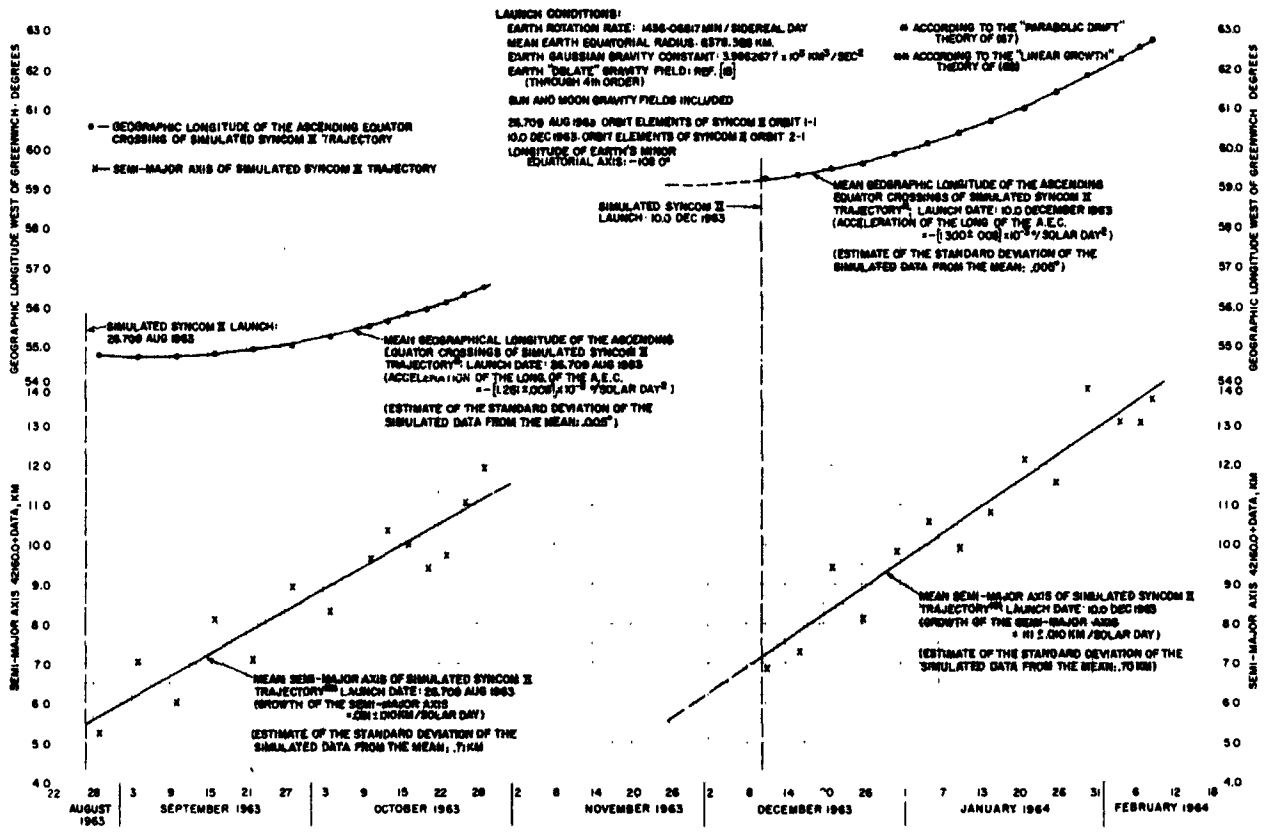


FIGURE A-1.—Drift of the ascending equator crossing and growth of the semimajor axis of simulated Syncom II trajectories.

to the theory of this report) is

$$(J_{22})_s = \left\{ \frac{-(2.188 \pm .013) \times 10^{-6}}{72\pi^2[\sin 2(52.5 \pm 2.5)](6378.388/42166.3 \pm .4)^2 \left[\frac{\cos^2 (33.005 \pm .003) + 1}{2}\right]} \right\}$$

$$= -(1.64 \pm .03) \times 10^{-6}.$$

The mean equatorial radius used in the simulation is $R_0 = 6378.388$ km, the same used to compute the "actual" Syncom II orbits from the radar and Minitrack observations.

The above value of $(J_{22})_s$ compares reasonably well with the input value of $(J_{22})_s = -(1.68) \times 10^{-6}$ used to compute the simulated trajectories.

The model error implicit in the difference between the reduced and inputed geodetic coefficients for the simulated trajectories warrants an adjustment of the J_{22} and λ_{22} reported in Section 7 from the reduction of the "actual" Syncom II orbits. The values below appear sufficient to cover all the *known* uncertainties of this reduction for a triaxial earth:

$$J_{22} \text{ (actual-adjusted)} = -(1.70 \pm .05) \times 10^{-6}$$

$$\lambda_{22} \text{ (actual-adjusted)} = -(19 \pm 6) \text{ degrees geographic longitude.}$$

As Appendix B will show, the principal *unknown* uncertainty of the reduction is the possible influence of higher order earth tesseral anomalies on the drift of Syncom II. When all relevant higher order anomalies in the earth's gravity potential are evaluated, the adjusted values above will probably remain

representative for an "average" triaxial potential field sufficient to consider for the future design of synchronous satellites. As a guess, the author would increase the upper limit of J_{22} to about $-(1.80)\times10^{-6}$ (based on some of the recent gravity potentials in Appendix B) for design purposes, based on an "average" triaxial geoid. A lower limit of $J_{22}=1.60\times10^{-6}$ for this purpose appears justifiable. The variance in the location of the major equatorial axis for the "average" triaxial geoid is not likely to change appreciably from the value quoted for the adjusted figure. The author is presently studying these higher order earth-gravity effects. The accumulated influence on synchronous satellites of all higher order earth anomalies, is believed to be small compared to the 2nd order anomaly. But it appears that close and continuing observations on the drift of these satellites will be rewarded in time by revelation of many of these "tesseral" anomalies to about 4th order with an absolute precision almost as good as that reported here for the 2nd order effect.

APPENDIX B

THE EARTH GRAVITY POTENTIAL AND FORCE FIELD USED IN THIS REPORT: COMPARISON WITH PREVIOUS INVESTIGATIONS

The gravity potential used as the basis for the data reduction in this study is the exterior potential of the earth derived in reference 10 for geocentric spherical coordinates referenced to the earth's spin axis and its center of mass. The infinite series of spherical harmonics is truncated after J_{44}. The nontesseral harmonic constants J_{20}, J_{30} and J_{40} are derived from reference 13.

The earth radius R_0 used in this study is:

$$R_0 = 6378.388 \text{ km.}$$

The earth's gaussian gravity constant used is:

$$\mu_E = 3.9862677 \times 10^5 \text{ km}^3/\text{sec}^2.$$

Neither of these values, taken from reference 14, nor the "zonal geoid" of Reference 13, is felt to be the most accurate known to date. They are the values used by the GSFC Tracking and Data Systems Directorate to calculate the orbit elements of Syncom II from radar and Minitrack observations. They were chosen to insure consistency between the data of this study and these published orbits, inasmuch as the "triaxial" reduction for which this study has been undertaken is not significantly affected by the probable errors in these values. The second-order tesseral harmonic constants used in the simulation studies were

$$J_{22} = -1.68 \times 10^{-6}$$

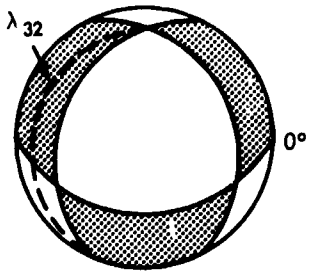
$$\lambda_{22} = -18°.$$

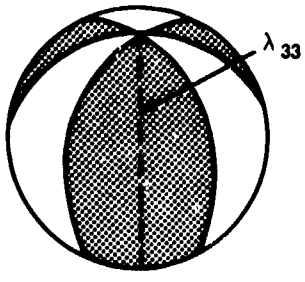
These are the values shown on the "tesseral geoid" below (for the J_{22} harmonic). At a later point in the analysis, the slightly different values reported in the abstract were estimated. The most accurate "zonal geoid" is probably that of Kozai (1962) [See Reference 6], with the following earth constants;

$$R_0 = 6.378.2 \text{ km}$$

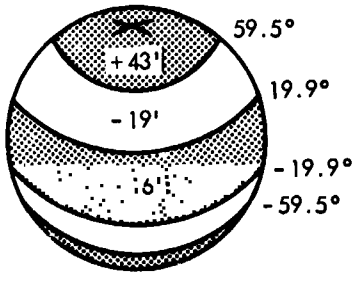
$$\mu_E = 3.98603 \times 10^5 \text{ km}^3/\text{sec}^2.$$

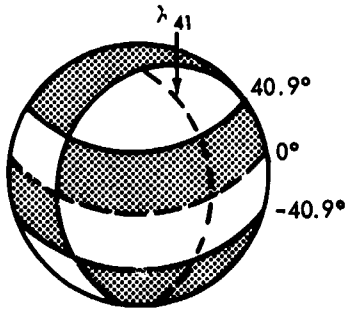
The earth's gravity potential (to fourth order, probably sufficient to account for all significant perturbations on a 24-hour satellite) may be illustrated as follows (following Reference 6, Appendix B):



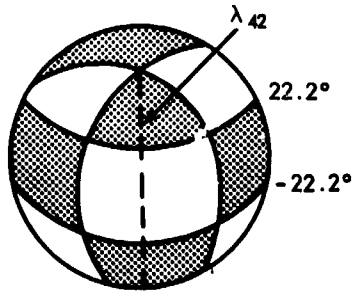


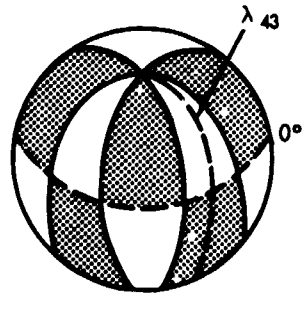
$$-15J_{32}\frac{R_0^3}{r^3}\cos^2\phi\sin\phi\cos 2(\lambda-\lambda_{32})-15J_{33}\frac{R_0^3}{r^3}\cos^3\phi\cos 3(\lambda-\lambda_{33})$$



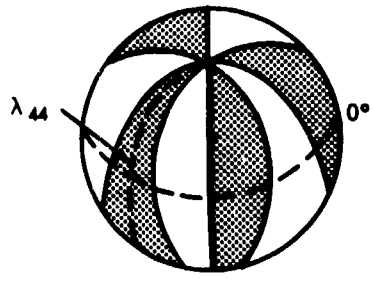


$$-\frac{J_{40}R_0^4}{8r^4}(35\sin^4\phi-30\sin^2\phi+3)-\frac{J_{41}R_0^4}{8r^4}[140\sin^3\phi-60\sin\phi]\cos\phi\cos(\lambda-\lambda_{41})$$

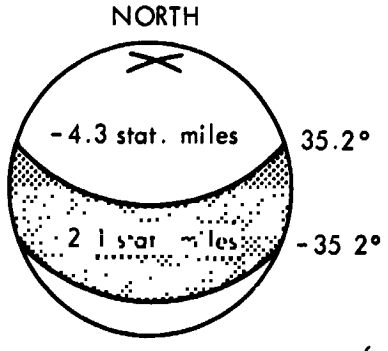


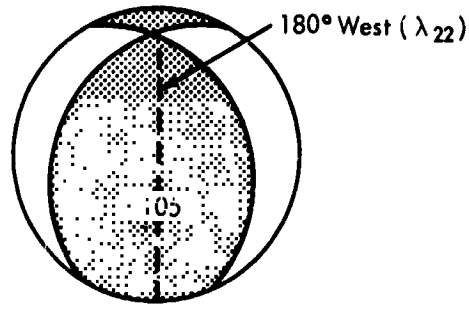


$$-\frac{J_{42}R_0^4}{8r^4}[420\sin^2\phi-60]\cos^2\phi\cos 2(\lambda-\lambda_{42})-\frac{J_{43}R_0^4}{8r^4}840\sin\phi\cos^3\phi\cos 3(\lambda-\lambda_{43})$$

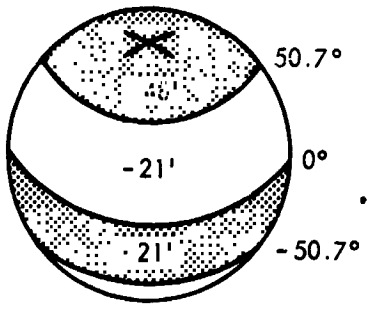


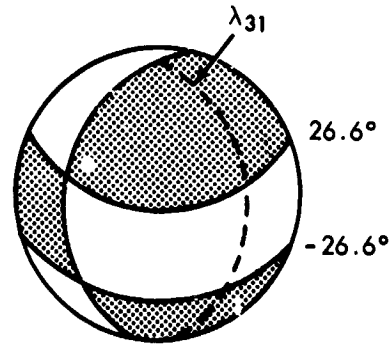
$$-\frac{J_{44}R_0^4}{8r^4}840\cos^4\phi\cos 4(\lambda-\lambda_{44})$$





$$V_E=\frac{\mu_E}{r}\left\{1-\frac{\mu_{20}R_0^2}{2r^2}(3\sin^2\phi-1)-3J_{22}\frac{R_0^2}{r^2}\cos^2\phi\cos 2(\lambda-\lambda_{22})\right.$$





$$\left.-\frac{J_{30}R_0^3}{2r^3}(5\sin^3\phi-3\sin\phi)-\frac{J_{31}R_0^3}{2r^3}\cos\phi\cos(\lambda-\lambda_{31})(15\sin^2\phi-3)\right\} \qquad \text{(B-1)}$$

The earth-gravity field (per unit test mass) whose potential is (B-1) is given as the gradient of (B-1), or

$$\bar{F}=rF_r+\lambda F_\lambda+\phi F_\phi\pm\Delta V_E=r\frac{\partial V_E}{\partial r}+\frac{\lambda}{r\cos\phi}\frac{\partial V_E}{\partial\lambda}+\frac{\phi}{r}\frac{\partial V_E}{\partial\phi}, \qquad \text{(B-2)}$$

or

$$\begin{aligned}Fr=\frac{\mu_E}{r^2}\Big\{&-1+(R_0/)^2[3/2J_{20}(3\sin^2\phi-1)+9J_{22}\cos^2\phi\cos 2(\lambda-\lambda_{22})\\&+2(R_0/r)J_{30}(5\sin^2\phi-3)(\sin\phi)+6(R_0/r)J_{31}(5\sin^2\phi-1)\cos\phi\cos(\lambda-\lambda_{31})\\&+60(R_0/r)J_{32}\cos^2\phi\sin\phi\cos 2(\lambda-\lambda_{32})+60(R_0/r)J_{33}\cos^3\phi\cos 3(\lambda-\lambda_{33})\\&+5/8(R_0/r^2J_{40}(35\sin^4\phi-30\sin^2\phi+3)\\&+25/2(R_0/r)^2J_{41}(7\sin^2\phi-3)\cos\phi\sin\phi\cos(\lambda-\lambda_{41})\\&+75/2(R_0/r)^2J_{42}(7\sin^2-1)\cos^2\phi\cos 2(\lambda-\lambda_{42})\\&+525(R_0/r)^2J_{43}\cos^3\phi\sin\phi\cos 3(\lambda-\lambda_{43})+525(R_0/r)^2J_{44}\cos^4\phi\cos 4(\lambda-\lambda_{44})]\Big\}. \qquad \text{(B-3)}\end{aligned}$$

$$\begin{aligned}F_\lambda=\frac{\mu_E}{r^2}(R_0/r)^2\{&6J_{22}\cos\phi\sin 2(\lambda-\lambda_{22})+3/2(R_0/r)J_{31}[5\sin^2\phi-1]\sin(\lambda-\lambda_{31})\\&+30(R_0/r)J_{32}\cos\phi\sin\phi\sin 2(\lambda-\lambda_{32})+45(R_0/r)J_{33}\cos^2\phi\sin 3(\lambda-\lambda_{33})\\&+5/2(R_0/r)^2J_{41}[7\sin^2\phi-3]\sin\phi\sin(\lambda-\lambda_{41})+15(R_0/r)^2J_{42}(7\sin^2\phi-1)\cos\phi\sin 2(\lambda-\lambda_{42})\\&+315(R_0/r)^2J_{43}\cos^2\phi\sin\phi\sin 3(\lambda-\lambda_{43})\\&+420(R_0/r)^2J_{44}\cos^3\phi\sin 4(\lambda-\lambda_{44})\}. \qquad \text{(B-4)}\end{aligned}$$

Table B-1.—*Tesseral Coefficients in the Earth's Gravity Potential, To Fourth Order, As Reported 1916–1963.**

	J_{22}	λ_{22}	J_{21}	λ_{21}	J_{32}	λ_{32}	J_{31}	λ_{31}	J_{42}	λ_{42}	J_{41}	λ_{41}	J_{33}	λ_{33}	J_{43}	λ_{43}	J_{44}	λ_{44}
(1) Guier (Nov. 1963)	-1.80×10^{-6}	−10 4°	-1.77×10^{-6}	6.3°	$-.226\times10^{-6}$	− 2.6°	$-.304\times10^{-6}$	24.1°	$-.73\times10^{-6}$	219°	$-.273\times10^{-6}$	38.6°	$-.0791\times10^{-6}$	− .7°	$-.0102\times10^{-6}$	35.0°
(2) Kaula (Sept. 1963)	−1.51	−12.1	−1.65	+ 5.3	−.144	46.4	− .145	15.8	− .471	132	− 078	44.2	−.0265	22.6	−.0038	23.3
(3) Izsak (July 1963)	−1.05	−11.2	−1.1	+ 3.2	−.20	−21.8	− .14	20.0	− .43	−132.1	−.13	37.0	−.026	11.5	−.019	14.8
(4) Kaula (May 1963)	−1.4	−21.5	−1.6	− 1.9	−.15	35.8	− .12	18.5	− .53	+126.3	−.12	44.5	−.019	10.7	−.0038	23.3
(5) Cohen (May 1963)	−2.08	−14.1							− .775	201	−.288	34.6	− 162	−4.3		
(6) Kaula (Jan. 1963)	−1.62	−21.4	−1.81	− 3.57	−.145	6.6	− .112	37.6	− .479	114.5	−.072	47.7	−.0088	5.9	−.0132	28.4
(7) Kozai (Oct. 1962)	−1.2	−26.4	−1.9	+ 4.6	−.14	−16.8	− .10	42.6	− .52	−122.5	−.082	65.2	−.035	0.5	−.031	14.9
(8) Newton (April 1962)	−2.15	−10.9							−2.53	+170.9						
(9) Newton (Jan. 1962)	−6.16	−11.0														
(10) Kozai (June 1961)	−2.32	−27.5	−3.95	+22.0	−.41	+31.0	−1.91	51.3	− .264	+163.5	−.167	54.0	−.046	−13.0	−.056	50.3
(11) Kaula (June 1961)	− .55	−13.3	−2.30	20.6	−.06	− .9	− .46	22.6	−1.85	194	−.42	21.1	−.091	− .5	−.024	26.4
(12) Izsak (Jan. 1961)	−5.55	−33.15														
(13) Kaula (1961)	−1.68	−38.5							−1.15	−13.0						
(14) Kranowski (1961?)	−5.53	+15.0														
†(15) Kaula (1960)	− .62	−20 9	− 98	55 4	−	13.5	− .19	14.3	− .46	−132.3	−.081	48.6	−.01	−30.0	−.02	22.5
†(16) Jeffreys (1960)	−4.17	0														
†(17) Uotila (1957)	−3.5	− 6.0														
†(18) Zhongolovitch (1957)	−5.95	− 7.7	−2.21	−25.7	−.6	−26.4	− .54	13.0	− .78	−149.1	−.080	45.0	−.051	− 3.8	−.0224	15.9
†(19) Subbotin (1949)	−5.5															
†(20) Niskanen (1945)	−7.67	− 1 0														
†(21) Jeffreys (1942)	−4.1	J	−2 1	0	−.66	0	− .24	33.3								
†(12) Heiskanen (1938)	−6.34	0														
†(23) Heiskanen (1924)	−9.0	+18.0														
†(24) Helmert (1915)	−6.0	−17 6														

*The J_{nn} and λ_{nn} reported in this table except in one or two instances, have been converted from the original author's set of gravity coefficients of which there are all too many forms in the literature. The blanks indicate the author did not consider that particular tesseral harmonic in fitting an earth potential to the observed data.

†Gravimetric or astro-geodetic tesseral geoids. All other geoid's are either pure satellite geoids or use all three kinds of gravity observations.

COMMENTS

(17), (20), (22), (23), (24) Heiskanen and Meinesz: 1958, The Earth and Its Gravity Field (McGraw Hill, N.Y., p. 79).

(16) Jeffreys, H.: 1959, The Earth (Camb. Univ. Press, N.Y., 4th Ed. Ch. IV and V).

(18) Zongolovitch, D.: 1957, Bull. Inst. Theoret. Astron. 6,505.

(15) Kaula, W. M.: 1959, J. Geophys. Research 64, 2401.

(19) Subbotin, M. F.: 1959, Course in Celestial Mechanics III (Gostekhizdat, Moscow; p. 278).

(12) Izsak, I. G.: 1961, Smithsonian Astronomical Journal 66, #5, 1961 (June) p. 226–229. Uses 2 satellites of medium inclination

(6) Newton, R. R.: 1962, Paper Presented at the Symposium on the Use of Artificial Satellites for Geodesy, April 26–28, 1962, U.S. Naval Observatory, Wash. D.C. (In Press).

(13) Kaula, W. M.: 1961, in Rech. Astr. Van De Hulst, C. and De Jager and A. F. Moore, ed., ed., Space Research II, pp. 360–372, North Holland Publishing Co., Amsterdam.

(1), (5) In: "Non Zonal Harmonic Coefficients of the Geopotential from Satellite Doppler Data," W. H. Guier and R. R. Newton, Johns Hopkins [illegible] (NASA-GSFC), Nov. 1963

(3) Kaula, W. M.: "Tesseral Harmonics of the Gravitational Field and Datum Shifts Derived from Camera Observations of Satellites" *Journal of Geophysical Research*, Vol. 68 #2, Jan. 15, 1963.

(4) Kaula, W. M.: "Improved Geodetic Results from Camera Observations of Satellites" (NASA-GSFC) Paper in Press for the Journal of Geophysical Research—May 1963).

(2) Izsak, I. G.: "Tesseral Harmonics in the Geopotential:" (*Nature*; July 13, 1963; pp. 137–139). Uses 5 satellites, 1 of very high inclination: with Kaula, probably the most accurate satellite tesseral geoid to high order (through 1963).

(9) Newton, R. R.: "Ellipticity of the Earth's Equator Deduced from the Motion of Transit 4A" (*Journal of Geophysical Research*; Vol. 67 #1, Jan. 1962).

(14) In: "Passive Dynamics in Space Flight" A Bureau of Naval Weapons Paper by J. D. Nicolaides and M. M. [illegible] December—1962.

(7) Kozai, Y. (Oct. 1962): Private Communication to the Author.

(10) Kozai, Y. (June 1961: "Tesseral Harmonics of the Geopotential of the Earth as Derived from Satellite Motions" (The Astronomical Journal, 66, #7; Sept. 1961). Uses 2 satellites of medium inclination

(3) Kaula, W. M. (Sept. 1963): "Improved Geodetic Results from Camera Observations of Satellites" (In: *The Journal of Geophysical Research*; Vol. 68, #18, Sept. 1963). Uses 5 satellites of medium inclination

(11) Kaula, W. M. (June 1961): "A Geoid and World Geodetic System Based on a Combination of Gravimetric, Astro-geodetic and Satellite Data" (*Journal of Geophysical Research*; Vol. 66, #6, p. 1807). Probably the most comprehensive "Combined Geoid" to date

(21) Jeffreys, H. (1942): (In-Monthly Notices of the Royal Astronomical Society, Geophys. Supplements, Vol. 5, 55).

$$F_\phi = \frac{\mu}{r^2}(R_0/r)^2\{-3J_{20}\sin\phi\cos\phi + 6J_{22}\cos\phi\sin\phi\cos 2(\lambda-\lambda_{22}) \quad .$$

$$-3/2(R_0/r)J_{30}(5\sin^2\phi-1)\cos\phi + 3/2(R_0/r)J_{31}(15\sin^2\phi-11)\sin\phi\cos(\lambda-\lambda_{31})$$

$$+15(R_0/r)J_{32}(3\sin^2\phi-1)\cos\phi\cos 2(\lambda-\lambda_{32})$$

$$+45(R_0/r)J_{33}\cos^2\phi\sin\phi\cos 3(\lambda-\lambda_{33}) - 5/2(R_0/r)^2J_{40}(7\sin^2\phi-3)\sin\phi\cos\phi$$

$$+5/2(R_0/r)^2J_{41}(28\sin^4\phi-27\sin^2\phi+3)\cos(\lambda-\lambda_{41})$$

$$+30(R_0/r)^2J_{42}(7\sin^2\phi-4)\cos\phi\sin\phi\cos 2(\lambda-\lambda_{42})$$

$$+105(R_0/r)^2J_{43}(4\sin^2\phi-1)\cos^2\phi\cos 3(\lambda-\lambda_{43})$$

$$+420(R_0/r)^2J_{44}\cos^3\phi\sin\phi\cos 4(\lambda-\lambda_{44})\}. \qquad \text{(B-5)}$$

The actual sea-level surface of the earth is to be conceptualized through (B–1) as a sphere of radius 6378 km, around which are superimposed the sum of the separate spherical harmonic deviations illustrated. To these static gravity deviations, of course, must be added a centrifugal earth-rotation potential at the earth's surface, to get the true sea level surface (see Reference 10).

From Table B–1 and equation (B–3), the fourth-order tesseral geoids of Kaula (September 1963), Kozai (1962), Izsak (July 1963), and Zhongolovitch (1957) have been evaluated for the longitudinal perturbation force on a 24-hour satellite with zero inclination at $\lambda = -54.75°$ over the earth's surface (see Table B–2). The harmonics contributing to this perturbation are J_{22}, J_{31}, J_{33}, J_{42} and J_{44}. The results of this comparison are:

TABLE B–2.—*Comparison of Longitudinal Perturbation Forces on a 24-Hour Satellite, From Five Tesseral Geoids.*

	Full Field Longitude Acceleration [Units of 10^{-7} g_s*]	Ratio of Triaxial (J_{22}) Longitude Acceleration to Full Field Longitude Acceleration ($\lambda = -54.75°$)
Zhongolovitch (1957)	7.71	1.06
Kozai (1962)	1.08	1.28
Izsak (July 1963)	1.27	1.19
Kaula (Sept. 1963)	1.77	1.11
Wagner (this reduction: March 1964)	2.21	?

*g_s is the radial acceleration of earth gravity at the "synchronous" altitude ($g_s = 0.735$ ft/sec^2).

Judging from the consistency of the acceleration ratios among these investigators, the "actual" value of J_{22} (separated from higher order gravity effects) is probably somewhat higher than the -1.7×10^{-6} reported herein. All the geodesists reporting in Table B–2 agree that the next most influential earth tesseral at "synchronous" altitudes over most of the equator is J_{33}.

Appendix C

EXPRESSIONS FOR THE INCLINATION FACTOR

Equation (25) gives the inclination factor in the drift causing tangential perturbation (due to equatorial ellipticity) on a 24-hour satellite with a near-circular orbit, as

$$F(i) = \cos(i) + \frac{\Delta\lambda(\max)\,\sin^2(i)}{2} \tag{C-1}$$

$\Delta\lambda$(max.) is the absolute value of the maximum longitude excursion of the figure-8 ground track of the 24-hour satellite (with a near-circular orbit) from the geographic longitude of the nodes.

From (18), this longitude excursion function is

$$\Delta\lambda = \lambda - \lambda_0 = \tan^{-1}[\cos(i)\,\tan\theta] - \theta \tag{C-2}$$

Differentiating (C-2) with respect to the argument angle θ, the minimax excursion arguments are found from

$$\frac{d(\Delta\lambda)}{d\theta} = 0 = \frac{\cos(i)\,\sec^2\theta}{1+\cos^2(i)\,\tan^2\theta} - 1. \tag{C-3}$$

Solving (C-3) for sin θ at $\Delta\lambda$ (minimax),

$$\sin\theta_{\Delta\lambda(\text{minimax})} = [\cos(i)+1]^{-1/2}, \text{ from which}$$

$$\tan\theta_{\Delta\lambda(\text{minimax})} = \sec(i), \tag{C-4}$$

(C-4) in (C-2) gives

$$\Delta\lambda(\text{minimax}) = \tan^{-1}[\cos(i)\,\sec(i)] - \tan^{-1}\sec(i).$$

Thus, since only the absolute value of $\Delta\lambda$ (minimax) is required,

$$\Delta\lambda(\max) = \tan^{-1}[\sec(i)] - 45°, \tag{C-5}$$

where the $\tan^{-1}$ is to be taken in the first quadrant. For example: for $i = 30°$, (C-5) evaluates the maximum excursion as

$$\Delta\lambda(\max) = 49.1° - 45° = 4.1°.$$

The nodal argument angle at this maximum longitude excursion is

$$\theta[\text{at } \Delta\lambda(\max)] = \pm 49.1° \text{ from the nodes.}$$

The assumption in (20) that the excursion in longitude from the ascending node could be approximated by

$$\Delta\lambda - \Delta\lambda(\max)\,\sin 2\theta,$$

predicts the maximum excursion argument as

$$\theta[\text{at } \Delta\lambda(\max)] = \pm 45° \text{ from the nodes.}$$

This discrepancy in the assumed longitude excursion function is not serious until $i > 45°$, as simulated trajectories with variable inclination have borne out.

(C–5) can be written as

$$\Delta\lambda(\max) + 45° = \tan^{-1}[\sec(i)], \text{ from which}$$

$$\tan[\Delta\lambda(\max) + 45°] = \sec(i) \doteq \frac{1+\Delta\lambda(\max)}{1-\Delta\lambda(\max)}, \tag{C-6}$$

for $i < 45°$. Solving (C–6) for $\Delta\lambda$ (max),

$$\Delta\lambda(\max) = \frac{1-\cos(i)}{1+\cos(i)}, \tag{C-7}$$

approximately.

Thus the inclination factor becomes approximately

$$\begin{aligned} F(i) &= \cos(i) + \frac{\sin^2(i)[1-\cos(i)]}{2[1+\cos(i)]} \\ &= \cos(i) + \frac{[1-\cos(i)]^2}{2} = \frac{2\cos(i) + 1 - 2\cos(i) + \cos^2(i)}{2} \\ &= \frac{\cos^2(i)+1}{2} \end{aligned} \tag{C-8}$$

For example: For $i = 30°$,

$$F(i)_{\text{from (C-1)}} = .86603 + 4.1/8 \times 57.3 = .8750$$

$$F(i)_{\text{from (C-8)}} = .8750.$$

The agreement of $F(i)$ from forms (C–1) or (C–8) is good to the third decimal place as long as the inclination is less than 45 degrees. At inclinations higher than 45 degrees, however, the drift theory following (20) begins to break down because $\Delta\lambda(\max)$ is no longer a small angle.

Appendix D

DERIVATION OF THE EXACT ELLIPTIC INTEGRAL OF DRIFT MOTION FOR A 24-HOUR SATELLITE WITH A NEAR-CIRCULAR ORBIT: COMPARISON OF THE EXACT SOLUTION WITH THE APPROXIMATE SOLUTIONS FOR PERIODS VERY CLOSE TO SYNCHRONOUS

The differential equation (33) of 24-hour satellite drift is analogous to the equation describing the large-angle oscillations of a mathematical pendulum (see Reference 11), as in Figure D–1.

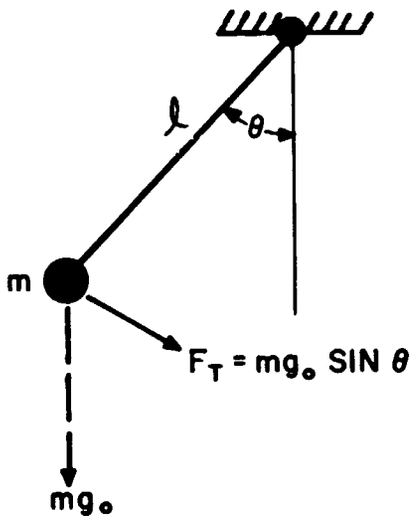


FIGURE D–1.—Configuration of a "mathematical pendulum."

The equation of angular motion of the mass m under the constant gravity force mg_0 is

$$F_T = \mathrm{mg}_0 \sin\theta = m(\dot{\ell}\dot{\theta}) = m\ell\ddot{\theta} \tag{D-1}$$

(D–1) can be rewritten as

$$\ddot{\theta} + (g_0/\ell)\sin\theta = 0. \tag{D-2}$$

From the theory developed in Reference 11 (pp. 327–335), (D–2) has an integral

$$t \text{ (time from } \theta = 0) = (\ell/g_0)^{1/2} F(k, \phi). \tag{D-3}$$

where $F(k, \phi)$ is the elliptic integral of the first kind with argument (or amplitude) $\phi = \sin^{-1}\left[\dfrac{\sin\theta/2}{\sin\theta\ (\max)/2}\right]$, and modulus $k = \sin\theta\ (\max)/2$.

Equation (33):

$$\ddot{\gamma}+A_{22}\sin 2\gamma=0,$$

with maximum libration angle γ_0, can be put in the form of (D–2) by the transformation of the dependent variable

$$\theta=2\gamma, \tag{D-4}$$

with the parameter identification

$$g_0/\ell=2A_{22}. \tag{D-4A}$$

(D–4) implies the identification

$$k=\sin\gamma_0, \qquad \phi=\sin^{-1}[\sin\gamma/\sin\gamma_0]. \tag{D-4B}$$

The pendulum solution (D–3), under the transformation (D–4) and identifications (D–4A) and (D–4B) becomes

$$t \text{ (time of drift libration from } \gamma=0)=(1/2A_{22})^{1/2}\,F(\sin\gamma_0,\,[\sin^{-1}\sin\gamma/\sin\gamma_0]), \tag{D-5}$$

$F(k, \phi)$, in its full integral form, is

$$F=\int_0^{\phi}\frac{d\phi}{(1-k^2\sin^2\phi)^{1/2}}, \tag{D-6}$$

(where $k^2=\sin^2\gamma_0$, $\sin^2\phi=\sin^2\gamma/\sin^2\gamma_0$) for the drift libration. In particular, when $\phi=\pi/2$; then $\gamma=\gamma_0;\dot{\gamma}_0=0$, and the pendulum-drift libration has completed a quarter-period.

Thus, from (D–5) and (D–6), the full period of the long-term drift libration of the 24-hour satellite ground track about the nearest minor equatorial axis longitude is

$$T(\gamma_0)=[8/A_{22}]^{1/2}\int_0^{\pi/2}\frac{d\psi}{(1-\sin^2\gamma_0\sin^2\psi)^{1/2}}. \tag{D-7}$$

The adequacy of the Taylor series expansion approximation of the drift motion in the neighborhood of γ_0, given in equation (51), may be tested against the exact drift solution implicit in (D–5). Table D–1 below gives the evaluation of F for arguments within 5° of $\gamma_0=60°$, using the integral tables in Reference 12.

TABLE D-1.—*Exact and Approximate Drifts of a 24-Hour Satellite from a Stationary Configuration 60° East of the Earth's Minor Equatorial Axis.*

$\gamma_0=60°$	γ (degrees)	γ' (degrees)	γ'' (degrees)	ϕ (degrees)	F (rad.)	ΔF (rad.)	Δt (days from $\gamma=60°$)
$A_{22}=23.2\times10^{-6}$ rad/day^2	60.0	60.0	60	90	2.1565	----------	----------
	59.0	59.003	59.000	81.7967	1.8730	.2835	41.619
	58.0	58.014	58.001	78.3056	1.7564	.4001	58.737
	57.0	57.029	56.999	75.5595	1.6671	.4894	71.846
	56.0	56.051	56.999	73.1938	1.5923	.5642	82.827
	55.0	55.077	54.996	71.0617	1.5265	.6300	92.487

In Table D–1, ΔF is the change in the elliptic integral from the "stationary" configuration at $\gamma=60°$ or $\phi=90°$. $\Delta t=(1/2A_{22})^{1-2}\Delta F$. A_{22} was computed from (30A) with the following gravity-earth

constants and for the inclination of Syncom II:

$$R_0 = 6378.2 \text{ km}$$

$$a_s = 42166 \text{ km}$$

$$J_{22} = -1.7 \times 10^{-6}$$

$$i = 33°.$$

γ' gives the drift position as calculated from the first righthand term of (51) alone (the $(\Delta t)^2$ term). γ'' gives the drift position as calculated from the first two righthand terms of (51). The "actual" Syncom II drift in mid-August 1963 began, apparently, at a γ_0 between 48° and 58° east of the minor axis. Thus, the 16 orbits chosen for the first drift period all should be well represented by the $(\Delta t)^2$-only theory, within the rms error of the longitude observations. Similar exact calculations for $\gamma_0 = 45°$, 50°, and 55° confirm the adequacy of the $(\Delta t)^2$-only theory to apply to the second drift-period orbits. They also prove the contention in section 5 that, for reasonably small excursions from "synchronism," the convergence of the Taylor series (51) is adequate if additional terms are included only when they become of a certain minimum significance to the total drift.

Appendix E

BASIC ORBIT DATA USED IN THIS REPORT

TABLE E-1.—*Syncom II Orbital Elements, August 1963 to Febrary 1964.*

Orbit #	Epoch (Universal Time) Year-Month-Day-Hour-Min	Semimajor Axis (km)	Eccentricity	Inclination (degrees)	Mean Anomaly (degrees)	Argument of Perigee (degrees)	Right Ascension of the Ascending Node (degrees)
1-1	63- 8-22- 6-12-14	42164.58	.00018	33.083	24.126	26.285	317.569
1-2	63- 8-26-17- 0	42164.52	.00016	33.090	190.841	26.099	317.454
1-3	63- 8-31- 0- 0	42166.02	.00018	33.062	296.125	30.073	317.475
1-4	63- 9- 5- 0-0	42166.39	.00012	33.064	333.521	357.756	317.362
1-5	63- 9- 9- 0-0	42166.35	.00015	33.048	326.207	9.077	317.272
1-6	63- 9-12- 2-0	42166.55	.00015	33.079	3.657	4.697	317.224
1-7	63- 9-17- 2-0	42166.70	.00018	33.043	12.694	0.581	317.165
1-8	63- 9-20- 2-0	42167.42	.00018	33.010	359.970	16.282	317.098
1-9	63- 9-27- 2-0	42167.51	.00022	33.046	38.922	344.162	316.996
1-10	63-10- 1- 2-0	42168.88	.00024	33.039	26.615	0.433	316.944
1-11	63-10 -8- 2-0	42169.14	.00020	33.013	42.889	350.866	316.780
1-12	63-10-14- 2-0	42169.78	.00028	32.982	36.727	2.673	316.813
1-13	63-10-22- 2-0	42171.51	.00026	32.993	62.833	344.246	316.603
1-14	63-10-30- 0-0	42171.09	.00028	32.948	29.865	354.548	316.570
1-15	63-11- 6- 0-0	42172.15	.00025	32.952	36.699	354.313	316.328
1-16	63-11-12- 0-0	42172.51	.00031	32.920	108.239	3.425	316.308
2-1	63-11-28- 1-0	42165.89	.00005	32.920	222.170	203.901	315.976
2-2	63-12- 4- 0-0	42167.20	.00009	32.892	39.435	17.564	315.919
2-3	63-12-10- 0-0	42167.18	.00010	32.881	51.942	10.958	315.877
2-4	63-12-16-17-0	42168.17	.00007	32.872	300.000	24.505	315.735
2-5	64- 1- 6-17-0	42168.01	.00013	32.867	332.997	11.625	315.544
2-6	64- 1- 9- 6-0	42169.90	.00015	32.857	165.031	16.992	315.469
2-7	64- 1-20- 0-0	42171.43	.00012	32.826	29.098	28.842	315.300
2-8	64- 1-29-20-0	42171.91	.00019	32.859	37.956	13.171	315.212
2-9	64- 2- 5-16-0	42172.89	.00019	32.800	321.168	36.275	315.075
2-10	64- 2-10-19-0	42173.31	.00014	32.833	32.517	14.553	314.982
2-11	64- 2-17- 0-0	42174.89	.00019	32.762	347.774	35.551	314.883

The Orbit elements for Syncom II in Table E-1 were calculated at the Goddard Space Flight Center from radar and Minitrack observations made during the slow drift periods from mid-August 1963 to February 1964.

As an example of the estimation of the ascending equator crossing nearest to epoch, consider the orbit geometry at epoch (Figure E-1).

6 January 1964 at 17.0 hours Universal Time (orbit 2-5).

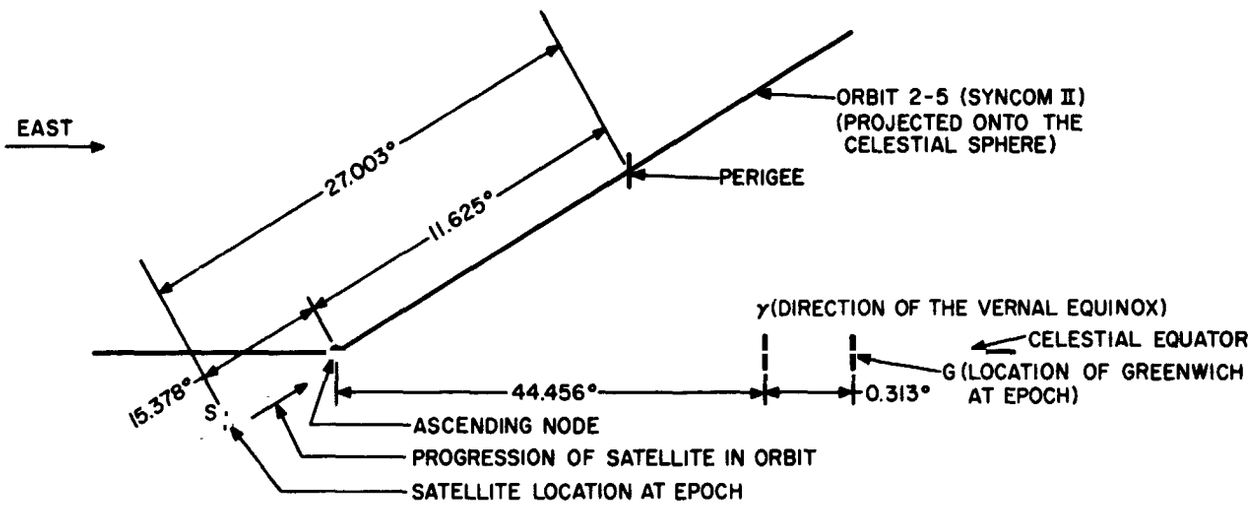


FIGURE E-1.—A portion of the celestial sphere at the epoch of orbit 2-5.

On 6.0 January 1964, the hour angle of the vernal equinox Λ *west* of Greenwich (expressed in hours, with 24 hours = 360°) was

6 hrs. 58 minutes 27.484 seconds (from the Nautical Almanac)

On 7.0 January 1964, the hour angle of Λ was

7 hrs. 2 min. 24.036 sec.

Interpolating, the hour angle of Λ on 6 January at 17 hours Universal Time was

0 hours 1 minute 15.042 seconds, or

0.313 degrees west of Greenwich.

In Figure E–1, the orbit angle 27.003° is taken directly as 360° − the mean anomaly, because the orbit is nearly circular. The reported period for this orbit was

$$T_p = 1436.21696 \text{ minutes.}$$

The earth's sidereal rotation period is taken to be

$$T_{\text{earth}} = 1436.06817 \text{ minutes.}$$

Thus, if the satellite is assumed to traverse orbit 2–5 at a nearly uniform rate, it will reach the celestial equator at a time when the Greenwich meridian has proceeded eastward from the epoch

$$15.378 \times 1436.21696/1436.06817 = 15.380°.$$

Thus, the estimated geographic longitude of the ascending equator crossing nearest to the epoch of orbit 2–5 is

$$\text{Ascending equator crossing longitude} = -(44.456+0.313+15.380)$$
$$= -60.149°.$$

The estimated time of this crossing is

$$15.380°/15°^{/hr.} = 1.025 \text{ hours after the epoch.}$$

The crossing time (Table 1) is thus estimated to be at

6.751 January 1964 (18.025/24+6.0, January 1964).

Appendix F

CALCULATION OF THE RADIATION PRESSURE ON SYMCOM II

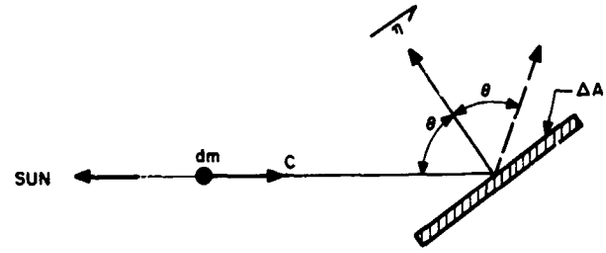


FIGURE F-1.—Flat plate ΔA of satellite surface, with normal η at angle θ with respect to sun's rays.

Consider a flat-plate element ΔA, of the surface of Syncom II, whose normal $\bar{n}$ is at angle θ with respect to the sun's rays (Figure F–1). The sun's radiant energy can be thought of as being made up of a stream of material particles such as dm, moving at the speed of light c. If the energy of each particle is dE, then from Einstein's classical energy-mass equivalence statement,

$$dm = dE/c^2 \tag{F-1}$$

Upon striking the surface, the incident radiation may:

- Reflect completely off the surface at an equal "reflection" angle, undegraded in energy
- Be absorbed into the body of the plate as thermal energy and partially reradiated in all directions from the surface at a reduced flux, depending on the surface and on the thermal properties of the plate and body of the spacecraft
- Be partially "reflected" and absorbed and reradiated, depending on the surface properties of the plate

An estimate of the radiation pressure on Syncom II will be calculated assuming complete light absorption with no reradiation. This is not the most conservative condition but will serve to show the order of magnitude of the effect.

Light-particle dm has $c\,dm$ momentum in the direction of the sun's rays before striking the plate element ΔA. Thus $(c)\,dm$ momentum is transferred to the plate with each light-particle collision. From Newton's second-law, the impulse transferred to the plate in the time of action dt, of dm alone, is;

$$\int_0^{dt} (dF)_{dm}\,dt = (c)dm, \tag{F-2}$$

dF acts on the plate element in the direction of the sun's rays. Assume that the discontinuous collision processes of (F–2) are so frequent as to amount to a smooth transfer of momentum between the stream of light particles and the plate; dt can then be replaced by Δt, a small but finite time interval; $(dF)_{dm}$ by ΔF, a smooth, constant small reactive force on the plate element ΔA; and dm can be replaced by Δm, a small but finite light particle mass impinging on the plate surface ΔA in Δt time. (F–2) then becomes

$$\Delta F(\text{radiation force})\ \Delta t = (c)\Delta m, \text{ or}$$

$$\Delta F(\text{radiation force}) = (c)\frac{\Delta m}{\Delta t}. \tag{F-3}$$

By the mass-energy equivalence relation (F–1) written for the finite small elements involved in the continuous momentum transfer,

$$\Delta m = \Delta E/c^2, \tag{F–4}$$

where ΔE is the energy flux falling on plate element ΔA in Δt time. Clearly,

$$\Delta E = \rho \Delta A \cos\theta(\Delta t), \tag{F–5}$$

where ρ is the sun's energy flux in units of energy/time-area, and $\Delta A \cos\theta$ is the projected area of the element in the direction of the sun's rays. Combining (F–3), (F–4) and (F–5);

$$\Delta F/\Delta A_p = \text{projected area-radiation pressure in the direction of the sun's rays} = p_s = \frac{\rho}{c}, \tag{F-6}$$

($\Delta A_p = \Delta A \cos\theta$ = projected area of the plate element in the direction of the sun's rays.)

The value of ρ outside the earth's atmosphere is estimated to be (see Reference 9)

$$\rho = 95.5 \text{ ft-pounds/ft.}^2\text{-sec.}$$

$$c = 9.835 \times 10^8 \text{ ft./sec.}$$

(F–6) thus becomes;

$$\rho_s \text{ (projected area-solar pressure against a fully absorbing surface in the direction of the sun's rays)} = .9725 \times 10^{-7} \text{ pounds/ft.}^2 \tag{F–7}$$

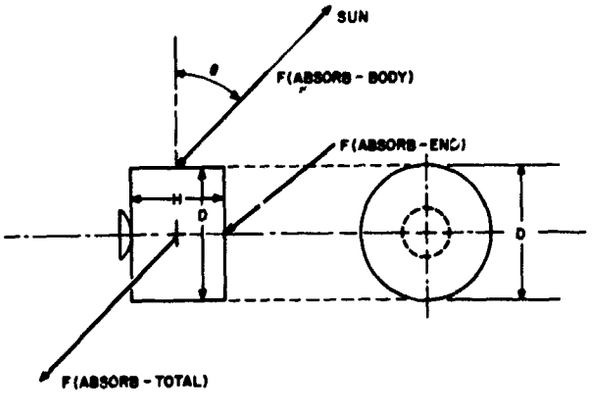


FIGURE F-2.—Configuration of Syncom II with respect to the sun during drift.

Figure F–2 shows the configuration of Syncom II with respect to the sun during the drift. For the cylindrical configuration in Figure F–2, from (F–7);

$$F(\text{absorb-total}) = F(\text{absorb-body}) + F(\text{absorb-end}) = .9725 \times 10^{-7}(HD\cos\theta + \pi D^2 \sin\theta/4) \tag{F–8}$$

The weight of Syncom II in the 24-hour orbit (including the apogee motor) is about 75 pounds. Other parameters are:

$$H = 15''$$

$$D = 28''$$

$$\theta = 21° \text{ (in late August 1963).}$$

N64-33625

INTRODUCTION TO THE APPLICATION OF VON ZEIPEL'S METHOD

WILLIAM J. WICKLESS, JR.

INTRODUCTION

To understand the application of Von Zeipel's method in the case of a satellite orbiting under the influence of a gravitational field plus small perturbating forces, it is first necessary to summarize the development of several concepts of classical mechanics. Therefore, the first portion of this paper will be a brief resume of material which may be found in greater detail and slightly modified form in Chapters 7–8 of *Classical Dynamics of Particles and Systems* by Jerry B. Marion, Dept. of Physics, University of Maryland. A good knowledge of ordinary differential and integral calculus is assumed.

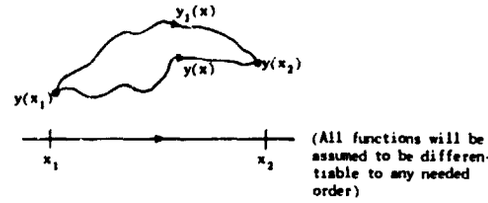
1. AN INTRODUCTION TO THE CALCULUS OF VARIATIONS

The basic problem of the calculus of variations is determining the unknown function $y(x)$ describing a path between two fixed end points we call $y(x_1)$ and $y(x_2)$ such that

$$I=\int_{x_1}^{x_2} f[y(x), y'(x), x]dx$$

takes on a maximum or minimum. f is a given function of the functions $y(x)$, $y'(x)$ and the independent variable x and the limits of integration are fixed. That is, if we wish to minimize I (say), we wish to find a function $y(x)$ such that if $y_1(x)$ is any other continuous function such that $y_1(x_1)=y(x_1)$, $y_1(x_2)=y(x_2)$ – any other path between $y(x_1)$ and $y(x_2)$ – then:

$$\int_{x_1}^{x_2} f[y(x), y'(x), x]dx \leq \int_{x_1}^{x_2} f[y_1(x), y_1'(x), x]dx$$



Two possible paths between $y(x_1)$ and $y(x_2)$ are sketched above.

Two possible paths between $y(x_1)$ and $y(x_2)$ are sketched above.

We will consider families of functions, giving possible paths between two fixed points $y(x_1)$, and $y(x_2)$, indexed by a parameter α, α running over a suitable segment of the real line. That is, we will consider families of the form

$$\{y(\alpha, x)\} \quad \begin{matrix} x\epsilon[x_1, x_2] \\ \alpha\epsilon[a, b] \end{matrix} \quad \text{with}$$

$$\left.\begin{matrix} y(\alpha, x_1)=y(x_1)=\text{constant} \\ y(\alpha, x_2)=y(x_2)=\text{constant} \end{matrix}\right\} \text{for all } \alpha\epsilon\,[a, b].$$

Then the integral I becomes a function of the parameter α:

$$I(\alpha)=\int_{x_1}^{x_2} f[y(\alpha, x), y'(\alpha,x), x]dx.$$

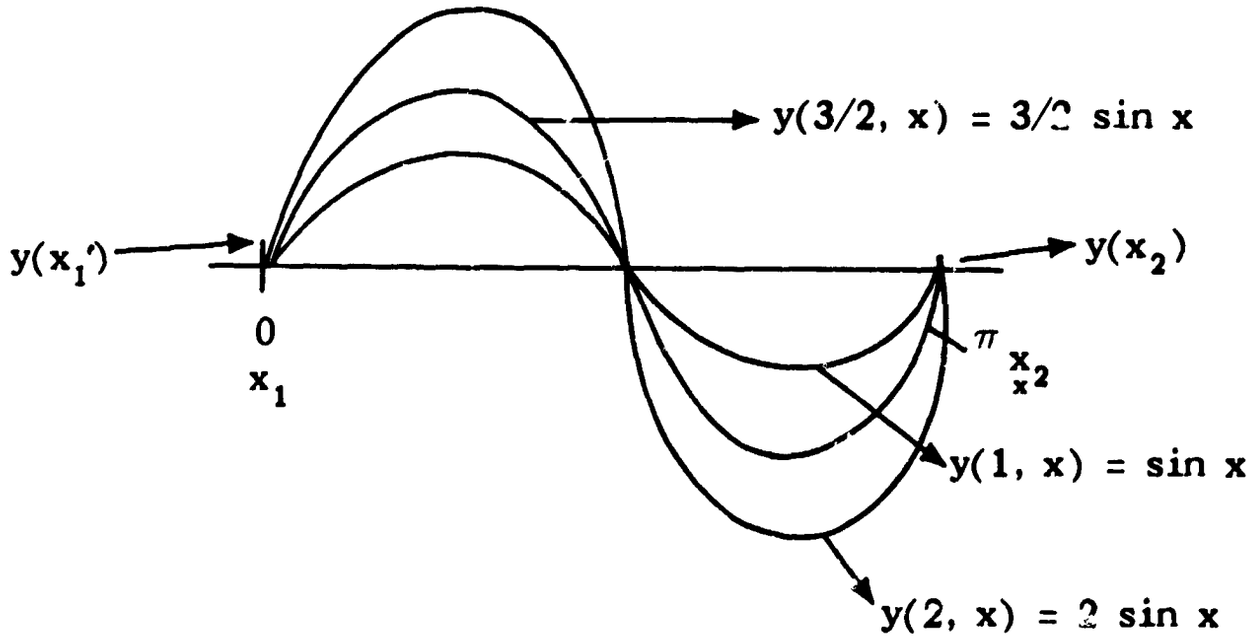
For example

$$I(\alpha)=\int_0^{2\pi} x(\alpha \sin x)dx \qquad \alpha\epsilon[1, 2]$$

where

$$y(\alpha, x)=\alpha \sin x$$
$$f[x, y(\alpha, x), y'(\alpha, x)]=xy(\alpha, x)$$
$$x_1=0,\ x_2=2\pi$$
$$y(x_1)=0\ y(x_2)=0$$

*Published as *Goddard Space Flight Center Document X-547-64-347*, September 1964



NOTE: $\alpha \sin(0) = 0$; $\alpha \sin 2\pi = 0$ for all $\alpha \epsilon [1, 2]$; $\{\alpha \sin x\}$ represents the family of paths between the pts $(0, 0) \equiv y(x_1)$, $(2\pi, 0) \equiv y(x_2)$ shown at top.

Now if the integral so written is to have an extremum along some path $y(\alpha_0, x)$

$$\alpha_0 \epsilon [a, b]$$

then

$$\left.\frac{\partial I(\alpha)}{\partial \alpha}\right|_{\alpha=\alpha_0} = 0.$$

This means

$$\left.\frac{\partial}{\partial \alpha}\int_{x_1}^{x_2} f[y(\alpha, x), y'(\alpha, y), x]dx\right|_{\alpha=\alpha_0} = 0.$$

since x is independent of α, the partial may be taken under the integral sign:

$$\left.\int_{x_1}^{x_2} \left[\frac{\partial f}{\partial y}\frac{\partial y}{\partial \alpha} + \frac{\partial f}{\partial y'}\frac{\partial y'}{\partial \alpha}\right]dx\right|_{\alpha=\alpha_0} = 0.$$

Now the above equation may be integrated by parts (see (1), Sec. 7-3) to obtain

$$\left.\int_{x_1}^{x_2} \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}\right]\frac{\partial y}{\partial \alpha}dx\right|_{\alpha=\alpha_0} = 0.$$

As f is assumed to be an arbitrary function of $y(\alpha, x)y'(\alpha, x)x$, the integrand itself must vanish:

$$\left[\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}\right]\left.\frac{\partial y}{\partial \alpha}\right|_{\alpha=\alpha_0} = 0$$

for all $x \epsilon [x_1, x_2]$.

Also since $\{y(\alpha, x)\}$ is an arbitrary family of paths, in general

$$\left.\frac{\partial y(\alpha, x)}{\partial \alpha}\right|_{\alpha=\alpha_0} \neq 0 \text{ along the curve}$$

$$y = y(\alpha_0, x)$$
$$x \epsilon [x_1, x_2]$$

therefore

$$\frac{\partial f}{\partial y)\alpha_0, x)} - \frac{d}{dx}\frac{\partial f}{\partial y'(\alpha_0, x)} = 0$$

for all $x \epsilon [x_1, x_2]$.

This is known as the Euler-Lagrange equation.

II. APPLICATION OF THE EULER-LAGRANGE EQUATION TO PHYSICAL SYSTEMS

It is known that if a particle is allowed to move between two fixed pts in the plane $x(t_1)$ and $x(t_2)$ in the time interval $[t_1, t_2]$ in any of the possible paths given by the family of curves $\{x(\alpha, t)\}$

$$t \epsilon [t_1, t_2] \quad \alpha \epsilon [a, b]$$

it will actually move in the path $x=x(\alpha_0, t)$ for which

$\int_{t_1}^{t_2}(T-U)dt$ is a relative extreme. (Note that this is a local condition.)

(see (1), Sec. 8.3)

Here T is the kinetic energy of the particle at the point $x(\alpha, t)-T=\frac{1}{2}m\dot{x}^2(\alpha, t)$—and U is the potential energy at the same pt $U=U(x(\alpha, t))$.

$\therefore$ along the path of the particle's motion $\bar{x}(t)=x(\alpha_0, t)$, the Euler-Lagrange equation must be satisfied by the quantity $L=(T-U)$. L is called the *Lagrangian* of the system.

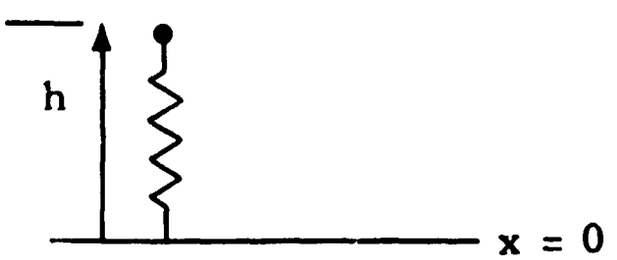
Thus we have

$$\frac{\partial L}{\partial \bar{x}(t)}-\frac{d}{dt}\frac{\partial L}{\partial \bar{x}'(t)}=0.$$

If the Lagrangian of a system is known, the above equation may be integrated giving the path of the particles motion $x=\bar{x}(t)$.

EXAMPLE: Determination of the Equation of Motion for a Harmonic Oscillator

Let a particle of mass m be attached to a spring with spring constant K and relaxed length 0. Choose a co-ordinate system such that the point of equilibrium of the spring is at $x=0$. Let the mass be pulled to a distance h above the $x=0$ level and released at a time $t=0$. Find $x(t)$ (the bar is hereafter dropped for convenience)



Now for any time t, we have $T=\frac{1}{2}m\dot{x}^2(t)$ and $U=\frac{1}{2}Kx^2(t)$.

so $L=T-U=\frac{1}{2}m\dot{x}^2(t)-\frac{1}{2}Kx^2(t)$.

Now if $x(t)$ is the path followed by the particle, the Euler-Lagrange equation must be satisfied, i.e.

$$\frac{\partial L}{\partial x(t)}-\frac{d}{dt}\frac{\partial L}{\partial \dot{x}(t)}=0.$$

Now

$$\frac{\partial L}{\partial x(t)}=-Kx(t);\ \frac{\partial L}{\partial \dot{x}(t)}=m\dot{x}(t)$$

$$\therefore -Kx(t)=\frac{d}{dt}[m\dot{x}(t)]$$

or

$$\ddot{x}(t)+w_0^2x(t)=0 \text{ with } w_0\equiv\sqrt{K/m}.$$

This differential equation may be solved (see any elementary text) to obtain

$$x(t)=A\sin(w_0t+\delta)$$

where $A+\delta$ are constants of integration to be fitted to the initial data—in this case $A=h$, $\delta=\pi/2$.

Now an important thing to notice is that the Lagrangian is a scalar function (being the difference of kinetic and potential energies—both scalar functions). Thus we can, if desired, transform one co-ordinate variable $x(t)$ to a new variable $q(t)$ under some co-ordinate transformation ϕ and we have

$$L(x(t), \dot{x}(t), t)=\bar{L}(q(t), \dot{q}(t), t).$$

(This is actually the defining property of a scalar function).

The Euler-Lagrange equation is still valid:

$$\frac{\partial \bar{L}}{\partial q(t)}-\frac{d}{dt}\frac{\partial \bar{L}}{\partial \dot{q}(t)}=0.$$

Now as one final generalization, it is clear the original analysis could have been carried out to treat the case of a particle with more than one degree of freedom. The Lagrangian then would have been a function of several co-ordinate variables, their derivatives, and possibly the time, i.e.

$$L=L(x_1(t)\cdots x_n(t), \dot{x}_1(t)\cdots\dot{x}_n(t), t).$$

We then could proceed in a similar fashion [see (1), Sec. 7-7] to obtain, instead of the single equation

$$\frac{\partial L}{\partial x}-\frac{d}{dt}\frac{\partial L}{\partial \dot{x}}=0$$

the system of equations:

$$\frac{\partial L}{\partial xj}-\frac{d}{dt}\frac{\partial L}{\partial \dot{x}j}=0$$

$$j=1, 2\cdots N.$$

Now as this Lagrangian is still a scalar function, we may make if desired, a co-ordinate transformation ϕ from the position variables $\{x_1(t)\cdots x_n(t)\}$ to some new set of variables $\{q_1(t)\cdots q_n(t)\}$ and we still have:

$$\bar{L}(q_1\cdots q_n, \dot{q}_1\cdots\dot{q}_n, t)=L(x_1\cdots x_n, \dot{x}_1\cdots\dot{x}_n, t)$$

And also

$$\frac{\partial \bar{L}}{\partial qj}-\frac{d}{dt}\frac{\partial \bar{L}}{\partial \dot{q}j}=0. \quad j=1\cdots n(a)$$

It is important to note that the new variables may be chosen as a set of *any N functions* $q_1(t)\cdots q_n(t)$ at all as long as the Lagrangian may be expressed as $L=L(q_1(t)\cdots q_n(t), q_1(t)\cdots\dot{q}_n(t)t)$. In a real problem, a set of variables $q_1(t)\cdots q_n(t)$ are usually chosen so that L may be easily expressed as $\bar{L}(q_1\cdots q_n\dot{q}_1\cdots\dot{q}_n\, t)$. Then the system (a) is integrated, giving $qj=qj(t,$ constants of integration$)$ $j=1\cdots n(b)$.

Finally the transformation ϕ is inverted and ϕ^{-1} is applied to (b), yielding

$$xj=xj(t) \quad j=1\cdots N.$$

The vector $\vec{r}(t)=(x_1(t)\cdots x_n(t))$ is then the position vector of the particle under consideration.

III. CONSTRUCTION OF THE HAMILTONIAN; THE CANONICAL EQUATIONS OF MOTION

Now let us consider a system in which (1) the Lagrangian is not an explicit function of the time, i.e. $L=L(q_1(t)\cdots q_n(t), \dot{q}_1(t)\cdots\dot{q}_n(t))$ and (2) the q_j's have been obtained from the x_j's by some transformation ϕ that does not explicitly involve the time.

Let us define N new quantities:

$$Pj\equiv\frac{\partial L}{\partial \dot{q}_j} j=1, 2\cdots N. \quad (I)$$

P_j is called the *conjugate momentum* associated with q_j.

Using this definition, the Euler-Lagrange equations for our system may be rewritten in the form

$$\dot{P}j=\frac{\partial L}{\partial q_j} j=1, 2\cdots N. \quad (II)$$

Now let us define a new quantity

$$H\equiv \Sigma P_j\dot{q}_j-L.$$

H is called the *Hamiltonian* of the system, and under assumptions (1) and (2) above, it can be shown that H is equal to the total energy of the particle under consideration. $H=T+U$. (See (1), Sec. 8–8).

Now if H is considered to be a function of q_j, P_j $j=1\cdots N$, then

$$dH=\sum_j\frac{\partial H}{\partial q_j}dq_j+\frac{\partial H}{\partial P_j}dP_j \quad (A)$$

also since

$$H=\sum_j P_j\dot{q}_j-L$$

$$dH=\sum_j P_j\, d\dot{q}_j+\dot{q}_j\, dP_j-\frac{\partial L}{\partial q_j}dq_j-\frac{\partial L}{\partial \dot{q}_j}dq_j$$

using (I) and (II), the above becomes

$$dH=\sum_j[P_j\, d\dot{q}_j-P_j\, d\dot{q}_j]+\dot{q}_j dP_j-\dot{P}_j dq_j$$
$$=\sum_j\dot{q}_j\, dP_j-\dot{P}_j\, dq_j. \quad (B)$$

Now identifying the coefficients of dP_j & dq_j in (A) and (B) we have

$$-\dot{P}_j=\frac{\partial H}{\partial q_j}\dot{q}_j=\frac{\partial H}{\partial P_j} \quad j=1, 2\cdots N.$$

These are called the *cyclic equations of motion.* As in the case of the Euler-Lagrange equations, they may be integrated to obtain the equations of motion of the particle under consideration.

We will now consider our problem—determination of the equations of motion for a satellite orbiting under the influence of a gravitational field plus small perturbations (e.g. the effects of the earth's oblateness.) The following (sec. IV) is a summary of a portion of the material found in the paper "Notes on Von Zeipel's Method" by Giorgio E. O. Giacaglia, published at the Goddard Space Flight Center.

IV. VON ZEIPEL'S METHOD

In the case we are considering, the Hamiltonian may be expressed as a function of six variables (L, G, H, ℓ, g, h) where L, G, H are the conjugate momenta associated with ℓ, g, h respectively. These six variables are called the *Delaunay variables.*

The equations to be integrated for our problem then are:

$$\dot{\ell}=\frac{\partial H}{\partial L}\ \dot{g}=\frac{\partial H}{\partial G}\ \dot{h}=\frac{\partial H}{\partial H}\quad -\dot{L}=\frac{\partial H}{\partial \ell}-\dot{G}=\frac{\partial H}{\partial g}-\dot{H}=\frac{\partial H}{\partial h}$$

here we have just taken the cyclic equations of motion and set $L=P_1$, $G=P_2$, $H=P_3$, $\ell=q_1$, $g=q_2$, $h=q_3$.

In many cases the negative of the Hamiltonian $F=-H$ is introduced; the cyclic equations of motion for our system then become:

$$\dot{\ell}=-\frac{\partial F}{\partial L}\quad \dot{g}=-\frac{\partial F}{\partial G}\quad \dot{h}=-\frac{\partial F}{\partial H}$$

$$\dot{L}=\frac{\partial F}{\partial \ell}\quad \dot{G}=\frac{\partial F}{\partial g}\quad \dot{H}=\frac{\partial F}{\partial h}.$$

Now in solving the above system, it would greatly simplify our problem if we could transform our variables (L, G, H, ℓ, g, h) to a new set of variables $(L', G', H', \ell', g', h')$ in which the new function F could be expressed in a form not explicitly containing one of the new variables (ℓ' say), i.e.

$$F'=F'(L', G', H'-g', h')$$

then

$$\dot{L}'=\frac{\partial F'}{\partial \ell'}=0;\qquad L'=\text{constant}.$$

Here we must remember that, for the cyclic equations of motion to be valid, the new variables must satisfy

$$L'=\frac{\partial L}{\partial \dot{\ell}'}\quad G'=\frac{\partial L}{\partial \dot{g}'}\quad H'=\frac{\partial L}{\partial \dot{h}'}$$

i.e. L', G', H' must be the conjugate momenta with respect to ℓ', g', h'. (This assumption was used to derive the cyclic equations of motion).

If the above equations are satisfied, then $(L', G', H', \ell', g', h')$ are said to be a *canonical set* of variables, and the transformation connecting the primed and unprimed variables is called a *canonical transformation* [assuming $F(L, G, H, \ell, g, h)=F'(L', G', H', -g', h')$ for all times t].

Now finding the desired canonical transformation connecting (L, G, H, ℓ, g, h) to a new set of variables $(L', G', H', \ell', g', h')$ for which $\partial F'/\partial \ell'=0$ is equivalent to finding a function $S=S(\ell, g, h, L', G', H')$, called a *generating function*, satisfying:

$$L=\frac{\partial S}{\partial \ell}\quad G=\frac{\partial S}{\partial g}\quad H=\frac{\partial S}{\partial h}$$

$$\ell'=\frac{\partial S}{\partial L'}\quad g'=\frac{\partial S}{\partial G'}\quad h'=\frac{\partial S}{\partial H'}.$$

This function S will determine an implicit transformation connecting (L, G, H, ℓ, g, h) and $(L', G', H', \ell', g', h')$ and it is known that the new set of variables will be canonical. (See *Planetary Theory* by Brown and Shook, sec. 5–3.)

We now shall give explicit formulas for determining S and F'. Here we first employ the assumption that our satellite is orbiting in a gravitational field, subject to small perturbing forces. We assume the function $F(L, G, H, \ell, g, h)$ can be expanded in a series

$$F=F_0+\sum_{j=1}^{\infty}F_j\lambda$$

where F_0 is the negative of the Hamiltonian calculated in the case of a body orbiting under the influence of gravity alone and λ is a small parameter. (For the calculation of $F_0=\mu^2/2L^2$ where $\mu=K^2M$, with K the Gaussian constant and M the mass of the orbited body; see any intermediate textbook in celestial mechanics.)

We assume similarly:

$$F=F_0'+\sum_{j=1}^{\infty}F_j\lambda^j$$

$$S=S_0+\sum_{j=1}^{\infty}S_j\lambda^j$$

where S_0 is the generating function determining the identity transformation, $S_0=L'\ell+G'g+H'h$. (In the Keplerian case the original set of variables are the ones desired; no transformation is necessary.)

Now to any order in λ, $F(L, G, H, \ell, g, h)=F'(L', G', H'-g', h')$. (Remember F' is not to be an explicit function of ℓ'.)

Look at this equation up to second order in λ, i.e.

$$F_0+\lambda F_1+\lambda^2F_2=F_0'+\lambda F_1'+\lambda^2F_2' \qquad (A)$$

now

$$L=\frac{\partial S}{\partial \ell}=\frac{\partial S_0}{\partial \ell}+\sum_{j=1}^{\infty}\lambda^j\frac{\partial S_j}{\partial \ell}=L'+\sum_{j=1}^{\infty}\lambda^j\frac{\partial S_j}{\partial \ell}.$$

$$g'=\frac{\partial S}{\partial G'}=\frac{\partial S_0}{\partial G'}+\sum_{j=1}^{\infty}\lambda^j\frac{\partial S_j}{\partial G'}=g+\sum_{j=1}^{\infty}\lambda^j\frac{\partial S_j}{\partial G}$$

with similar expressions for G, H, h'.

Substituting the above into (A), we have

$$F_0\left(L'+\sum_{j=1}^{\infty}\lambda^j\frac{\partial S_j}{\partial \ell}\right)+\lambda F_1\left(L'+\sum_{j=1}^{\infty}\lambda^j\frac{\partial S_j}{\partial \ell},\right.$$

$$\left.G'+\sum_{j=1}^{\infty}\lambda^j\frac{\partial S_j}{\partial g},\ H'+\sum_{j=1}^{\infty}\lambda^j\frac{\partial S_j}{\partial h},\ \ell,\ g,\ h\right)$$

$$+\lambda^2F_2\left(L'+\sum,\ G'+\sum,\ H'+\sum,\ \ell,\ g,\ h\right)$$

$$=F_0'L',\ G',\ H'-,\ g+\sum_{j=1}^{\infty}\lambda^j\frac{\partial S_j}{\partial G'}h'+\sum_{j=1}^{\infty}\lambda^j\frac{\partial S_j}{\partial H'}$$

$$+\lambda F_1'(L',\ G',\ H'-g+\Sigma,\ h+\Sigma)+\lambda^2F_2'(L',$$

$$G',\ H'-g+\Sigma,\ h+\Sigma.) \quad \text{(B)}$$

Now expanding the left side of (B) in a Taylor series around L, G, H and F'_1 and F'_2 on the right hand side in a Taylor series around g', h' and keeping only terms up to and including the second order in λ, we have

$$F_0(L')+\frac{\partial F_0}{\partial L}\left[\frac{\partial S_1}{\partial \ell}\lambda+\frac{\partial S_2}{\partial \ell}\lambda^2\right]+\lambda^2F_1(L',\ G',\ H',$$

$$\ell,\ g,\ h)+\lambda^2\left[\frac{\partial F_1}{\partial L}\frac{\partial S_1}{\partial \ell}+\frac{\partial F_1}{\partial G}\frac{\partial S_1}{\partial g}+\frac{\partial F_1}{\partial H}\frac{\partial S_1}{\partial h}\right]$$

$$+\lambda^2F_2[L',\ G',\ H',\ \ell,\ g,\ h]=F_0'(L',\ G',\ H'-g',\ h')$$

$$+\lambda F_1'(L',\ G',\ H'-g,\ h)+\lambda^2\left[\frac{\partial F_1'}{\partial g'}\frac{\partial S_1}{\partial G'}+\frac{\partial F_1}{\partial h'}\frac{\partial S_1}{\partial H'}\right]$$

$$+\lambda^2F_1'[L',\ G',\ H'-g,\ h] \quad \text{(C)}$$

Now equating the coefficients of like powers of λ in (C) we obtain

$$F_0'(L',\ G',\ H'-g',\ l')=F_0(L')=\frac{U^2}{2L'^2} \quad \text{(D)}$$

$$F_1(L',\ G',\ H',\ \ell,\ g,\ h)+\frac{\partial F_0}{\partial L}\frac{\partial S_1}{\partial \ell}$$

$$=F_1'(L',\ G',\ H'-g,\ h) \quad \text{(E)}$$

$$\frac{\partial F_0}{\partial L}\frac{\partial S_2}{F\ell}+\frac{\partial F_1}{\partial L}\frac{\partial S_1}{\partial \ell}+\frac{\partial F_1}{\partial G}\frac{\partial S_1}{\partial G'}+\frac{\partial F_1}{\partial H}\frac{\partial S_1}{\partial h}$$

$$+F_2(L',\ G',\ H',\ \ell,\ g,\ h)=\frac{\partial F_1}{\partial g'}\frac{\partial S_1}{\partial G'}+\frac{\partial F_1}{\partial h}\frac{\partial S_1}{\partial H'}$$

$$+F_2(L',\ G',\ H'-g,\ h). \quad \text{(F)}$$

Now F'_0 is determined by (D) so (E) may be solved by putting F'_1 = part of F_1 independent of ℓ. This means $(\partial F_0/\partial L)(\partial S_1/\partial \ell)=-(F_1-F'_1)$ $=-$(part of F_1 dependent on ℓ). Now F'_1 and S_1 are determined with the desired properties, and therefore (F) may be solved for S_2 by putting F'_2 = part of F_2 independent of ℓ.

This determines F'_2 and S_2 with the required properties. This process may be continued to determine F'_n and S_n to any desired order.

REFERENCES

1. Classical Dynamics of Particles and Systems; JERRY B. MARION; University of Maryland.
2. Planetary Theory; BROWN and SHOOK; Cambridge University Press.
3. Notes on Von Zeipel's Method; GIORGIO E. O. GIACAGLIA; GSFC—X-547-64-161.

III. SOLAR PHYSICS

COMMENTS ON HIGH ENERGY X-RAYS BURSTS OBSERVED BY OSO-1*

KENNETH J. FROST

See N64-30227

Instrumentation aboard the first Orbiting Solar Observatory observed five separate instances of high energy X-ray bursts accompanying solar flares. The X-ray bursts were observed in the 20 to 100 Kev range. Data on the types of flares producing the X-ray bursts and the associated microwave radio bursts are presented.

INTRODUCTION

Among the experiments carried by OSO–1 in the pointed section was a scintillation counter designed to detect hard solar X-ray bursts in the 20 to 100 Kev energy range. The scintillation counter consisted of a cylindrical NaI(Tl) crystal 0.3 cm high and 2.24 cm in diameter and an R.C.A. C-7151 photomultiplier tube. The photomultiplier tube and crystal assembly was placed in a cylindrical copper shield which through an opening in the front end provided a field of view of 0.3 steradian. The copper shield had a wall thickness of 1 cm. The 3.8 cm^2 area of the front surface of the crystal was held normal to the solar direction by the pointing of the satellite. An aluminum disc of 0.08 cm thickness was placed over the front surface of the crystal. The efficiency of the detector in the 20 to 100 Kev interval was approximately 0.90.

The electronics associated with the detector consisted of lower and upper level discriminator circuits and a logarithmic ratemeter. The lower and upper level discriminator circuits were set to select only those pulses from the photomultitube which correspond to an energy loss of between 20 and 100 Kev. in the crystal. The pulses accepted by the discriminator circuits where then passed on to the logarithmic ratemeter. The output of the ratemeter was a voltage level between zero and five volts which corresponded to an input count rate between 10 and 10^4 counts per second. The output of the ratemeter was able to respond to changes in count rate that occurred in a time longer than 10 milliseconds.

The satellite telemetry system sampled the voltage level at the output of the ratemeter continuously for two seconds once every twenty seconds. This sampling format was operative during the entire daylight part of an orbit.

DATA

With the return of the first hundred orbits of reduced data the response of the detector to the general spacecraft X-ray background and the response to the regions ot trapped radiation was determined. In addition it was found that the experiment responded to the electron warm spots detected by an electron-proton spectrometer placed aboard the spacecraft by Schrader et al[1]. When passing through regions not including trapped radiation or electron warm spots the detector was found to produce a count rate for a 2 second sample that consistently fell between limits of 10.5 to 14.5 counts per second.

Only 17 March 1962 a short lived increase in count rate occurred while the satellite was passing through a region where the count rate normally fell in the background range of 10.5 to 14.5 counts per second. The increase in count rate started at 19:39:49 U.T. and ended at 19:40:08 U.T. A record of the satellite telemetry signal for the orbit containing this time period indicated conclusively that the fluctuation in count rate was real and could not be attributed to telemetry noise that

*Published as *Goddard Space Flight Center Document X-610-64-60*, May 1964.

This paper is an abridged version of the paper published in the proceedings of the 5th International Space Science Symposium, Florence, Italy, 1964.

had been erroneously digitized by the automatic data reduction system. Reference to the available solar data revealed that a solar flare had been observed by the McMath-Hulbert and Lockheed Observatories between the times of 1934 to 1959 U.T. and 1936 to 2003 U.T. respectively. The flare was observed on the east limb of the sun and was classified by McMath-Hulbert as an importance 1 flare and by Lockheed as an importance 2 flare. Also at this time an impulsive radio burst of three minutes duration beginning at 19:39 U.T. and reaching maximum at 19:40:12 U.T. was observed at 2800 Mc/sec by Ottawa. Furthermore an SID accompanied the flare beginning at 19:40 U.T. and reaching maximum at 19:44 U.T. The very close correlation in the time of these events and the fluctuation of the 20 to 100 Kev count rate suggested that the most plausible explanation of the fluctuation was a burst of energetic X-rays produced by the flare.

During succeeding orbits of the spacecraft four more instances of rapid excursions in count rate were observed during the presence of a flare on the sun which produced an impulsive radio burst in the microwave range and an SID. These excursions in count rate were also interpreted as solar flare X-ray bursts. The history of these five events is presented graphically in Figures 1 through 5[2].

DISCUSSION

Prior to the launch of OSO-1 eight occurrences of high energy X-ray bursts accompanying solar

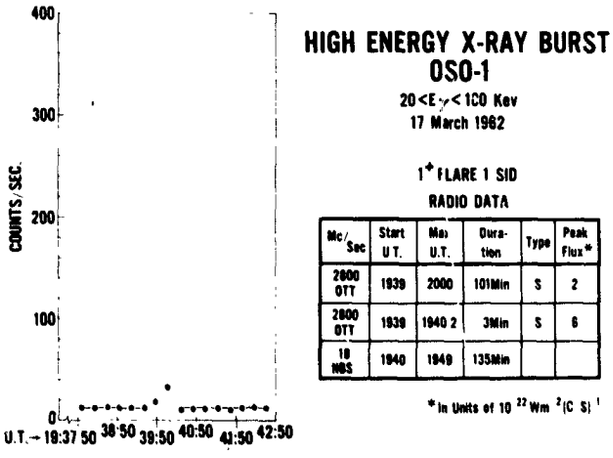


FIGURE 1.—High energy x-ray burst, OSO1; $20 < E_\gamma < 100$ kev, March 17, 1962.

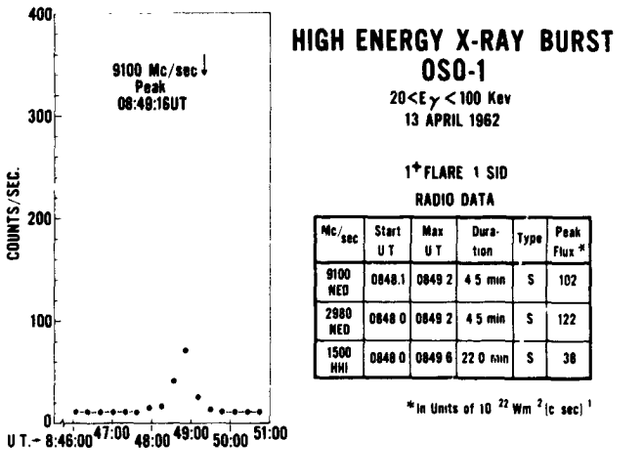


FIGURE 2.—High energy x-ray burst, OSO I; $20 < E_\gamma < 100$ kev, April 13, 1962.

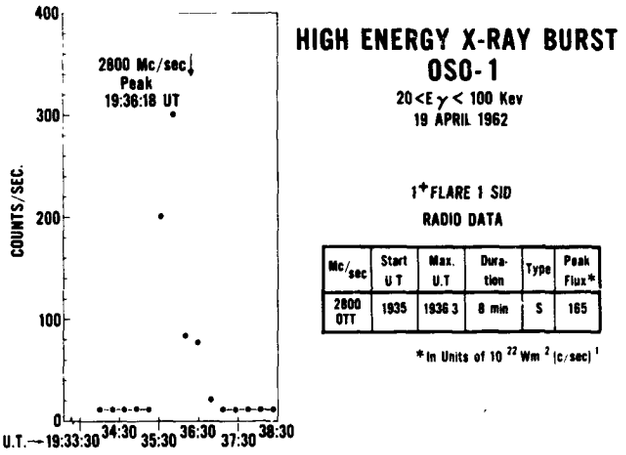


FIGURE 3.—High energy x-ray burst, OSO I; $20 < E_\gamma < 100$ kev, April 19, 1962.

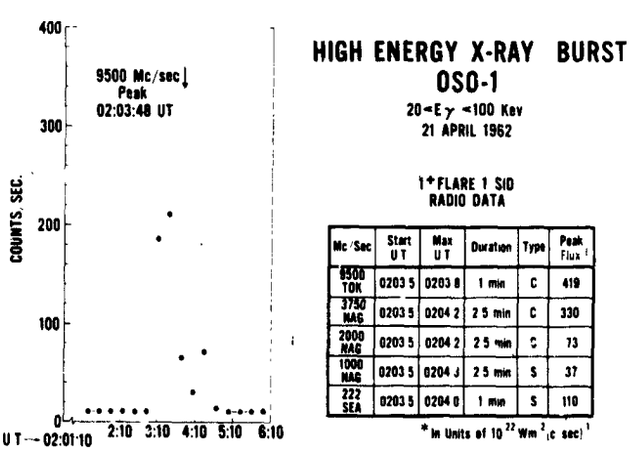
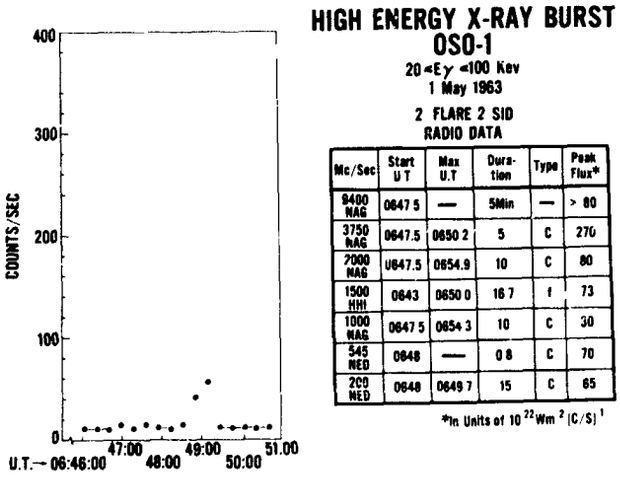


FIGURE 4.—High energy x-ray burst, OSO I; $20 < E_\gamma < 100$ kev, April 21, 1962.



2 FLARE 2 SID
RADIO DATA

Mc/Sec	Start U T	Max U.T	Duration	Type	Peak Flux*
9400 NAG	0647 5	—	5Min	—	> 80
3750 NAG	0647.5	0650 2	5	C	270
2000 NAG	0647.5	0654.9	10	C	80
1500 HM	0643	0650 0	16 7	f	73
1000 NAG	0647 5	0654 3	10	C	30
545 NED	0648	—	0 8	C	70
200 NED	0648	0649 7	15	C	65

*In Units of 10^{22} Wm^{-2} (C/S)$^{-1}$

FIGURE 5.—High energy x-ray burst, OSO I; $20 < E_\gamma < 100$ kev, May 1, 1963.

flares had been observed[3–7]. Kundu has analyzed the radio data associated with each of these events[8]. He has found that an impulsive microwave burst was associated with every X-ray burst whereas type III bursts and other meter wave events did not show as strong a correlation. Furthermore the general character of the microwave spectrum was found to be such that the peak flux observed at wavelengths in the 3 to 10 cm range was stronger than that observed at wavelengths in the 10 to 20 cm range. The radio data associated with the OSO-1 X-ray bursts as indicated in Fig. 1 through 5 appears to lend strong support to Kundu's analysis. Additional information of interest to this analysis is the fact that as far as can be determined from spectral observations in the meter wave range no type III bursts occurred during the X-ray events.

REFERENCES

1. SCHRADER, C., KAIFER, R. C., WAGNER, J. A., ZINGER, J. H., and BLOOM, S., University of California, Lawrence Radiation Laboratory—"Proton, Electron Fluxes (75 Kev) from OSO"—Presented at AGU Spring Meeting, April 1963.
2. Flare, SID, and Radio Data taken from Quarterly Bulletin on Solar Activity No. 137, January/March 1962 and No. 138, April/June 1962.
3. PETERSON, L. E., and WINCKLER, J. R.—J. Geophys. Research, 64, 697, 1959.
4. CHUBB, T. A., FRIEDMAN, and KREPLIN, R. W.—International Space Science Symposium, Nice, p. 695, 1960.
5. WINCKLER, J. R., MAY, T. C. and MASLEY, A. J.—J. Geophys. Research, 66, 316, 1961.
6. VETTE, J. I. and CASAL, F. G.—Phys. Rev. Letters, 6, 334, 1961.
7. Anderson, K. A., and WINCKLER, J. R.—J. Geophys. Res. 67, 4103, 1962.
8. KUNDA, M. R.—Space Science Reviews 2 (1963) 438-469.

N64-27061

TEMPERATURE OF A GRAY BODY MOST CLOSELY FITTING THE SOLAR EXTRATERRESTRIAL SPECTRUM*

WILLIAM B. FUSSELL AND JOHN B. SCHUTT

A definition of "closeness" of fit of a gray body curve to the sun's extraterrestrial spectrum is presented. F. S. Johnson's data are used for the solar spectrum. For the wavelength interval 0.3 to 2.4 microns (pertinent to thermal design), the magnitude of C_1 calculated for use in Planck's black body formula is 8.6504×10^{-17} watt-cm² at a temperature of 5742 ± 0.5 °K.

INTRODUCTION

The problem of simulating the extraterrestrial solar illumination for testing the thermal design of satellites has recently become of wide interest. If the materials for the satellite surface have been selected, the problem reduces to one of matching the absorbed energy from the simulating source—for each material composing the surface—to the absorbed solar energy for that material. However, many materials have spectral absorptivity curves which vary slowly with wavelength between wavelength limits including most of the sun's energy. A solar simulator which matches the sun's spectrum fairly well on a finely resolved basis, and which matches the total energy exactly, will automatically match the solar absorptivities and the simulated absorptivities closely for these materials. To this end, the black body function was used in constructing a gray body curve for which a solar fit was obtained by parametrically varying the temperature.

METHOD

In establishing a closeness of fit criterion for approximating the sun's extraterrestrial spectrum by that of a black body, it seems pertinent at this point to recall the geometric interpretation of the quantities in the black body equation which can be treated as parameters. Consequently, any criterion which is established by means of parametric variations can only be defined as ultimate when it is the most valid one from geometric considerations.

From the black body definition

$$J_\lambda(T;\, C_1,\, C_2) = C_1\, J_\lambda(T;\, C_2) = \frac{C_1}{\lambda^5\, e^{C_2/\lambda T} - 1}, \qquad (1)$$

where λ is the wavelength in cm, T is the temperature in °K, and C_1 and C_2 are constants (or parameters). Any of the quantities T, C_1, and C_2 may be chosen as a parameter. First consider variations in C_1. It is clear that any change in the magnitude of C_1 shifts the spectral distribution up or down along the ordinate depending on the sign of the chosen change for C_1. Since the parameters C_2 and T are geometrically equivalent (in a reciprocal manner), only one of them need be considered; T was chosen because of its experimental accessibility. Variations in T affect the shape of the black body distribution and, of course, the wavelength at which the maximum spectral intensity occurs. Therefore, any criterion which is to be established for fitting the sun's extraterrestrial spectral distribution to Equation 1 can be determined by means of parametric variations in T and C_1, or a gray body fit.

The most straightforward approach is to find the temperature at which the black body maximum matches the sun's maximum, 0.475 micron. This temperature is readily found to be 6060°K. But this method gives a black body curve which contains a greater integrated energy than the sun's integrated spectral distribution, since by

*Published as *NASA Technical Note* D-1845, August 1964

definition the black body distribution gives an upper limit to the emissive ability of any body at a given temperature. Consequently, an upper bound on the required temperature has been established from this matching process. Next, an estimate of the lower bound can be found. To accomplish this, C_1 and T may be varied in an effort to minimize the residual

$$R = \int |J_\lambda(T;\ C_1,\ C_2) - H_\lambda| d\lambda \quad , \qquad (2)$$

where H_λ and the vertical lines represent the sun's spectral distribution and absolute magnitude, respectively. From Equation 2 the lower bound is found to be approximately 5738°K. It is apparent that Equation 2 alone cannot give the required closeness of fit to the desired accuracy because one criterion cannot be used to minimize a bivariant distribution. The final criterion, the one defined as geometrically ultimate and, consequently, most pertinent from the thermal design viewpoint, is the system:

$$R_{Min} = \text{Min} \int |J_\lambda(T;\ C_1,\ C_2) - H_\lambda| d\lambda \qquad (3)$$

and

$$\int J_\lambda(T;\ C_1,\ C_2) d\lambda = \int H_\lambda \, d\lambda \quad . \qquad (4)$$

The addition of Equation 4, to establish the most meaningful criterion, allows a temperature to be found by using matched areas as the basis. The absolute magnitude in Equation 3 is necessary in order to insure that the total deviation between the two curves is minimized. From Equation 1 it is clear that C_1 can be eliminated in Equation 3 by means of Equation 4, which gives a single relation for the determination of T. These equations give a temperature of 5742 ± 0.5°K and a value of 8.65040×10^{-17} watt-cm² for C_1 for the 0.3 to 2.4 micron (μ) region.

It remains to show that Equation 2 should give results identical to Equations 3 and 4, indicating therefore that 5742°K is the desired temperature. If the spectral region of interest is broken into intervals of $J_\lambda(T;\ C_1) > H_\lambda$ and $J_\lambda(T;\ C_1) < H_\lambda$, where the intervals of the first kind are designated by $\Delta_{1,1}$ and those of the second kind by $\Delta_{2,1}$, then Equations 2 and 3 may be rewritten:

$$\int_{\Delta_{1,1}} |J_\lambda(T;\ C_1) - H_\lambda| d\lambda + \int_{\Delta_{2,1}} |J_\lambda(T;\ C_1) - H_\lambda| d\lambda$$

$$= \int_{\Delta_{1,1}} C_1 \bar{J}_\lambda(T) d\lambda - \int_{\Delta_{1,1}} H_\lambda \, d\lambda - \int_{\Delta_{2,1}} C_1 \bar{J}_\lambda(T) d\lambda + \int_{\Delta_{2,1}} H_\lambda \, d\lambda \quad . \qquad (5)$$

where the first subscript of Δ refers to the interval type and the second is an index which enumerates subintervals. When the derivatives of the right-hand side of Equation 5 (ignore Equation 4) are taken with respect to C_1 and T in an effort to obtain simultaneous equations for C_1 and T, the following results are obtained:

$$\int_{\Delta_{1,1} \cdot \Delta_{2,1}} \bar{J}_\lambda(T) d\lambda = 0$$

and

$$C_1 \int_{\Delta_{1,1} \cdot \Delta_{2,1}} \frac{\partial}{\partial T} \bar{J}_\lambda(T) = 0 \quad .$$

Consequently, Equation 4 must be included in order to define C_1 and obtain the desired results. This implies the equivalence of Equation 2 and the system of Equations 3 and 4. By taking the derivative of the right-hand side of Equation 5 with respect to T and equating the result to zero in order to obtain a minimum (make use of Equation 4), Equation 5 may be rewritten:

$$\int_{\Delta_{1,1}} |J_\lambda(T;\ C_1) - H_\lambda| d\lambda + \int_{\Delta_{2,1}} |J_\lambda(T;\ C_1) - H_\lambda| d\lambda \leq \epsilon \quad . \qquad (6)$$

where ϵ is a number characterizing the fit of the black body curve to the sun's distribution and $\Delta_{1,i}$ and $\Delta_{2,i}$ become the sets $\{\Delta_{1,i}\}$ and $\{\Delta_{2,j}\}$, respectively, with $1 \leq i \leq N$ and $1 \leq j \leq M$. If the fit is exact, $\epsilon = 0$ and the equality sign holds; however, if the fit is not exact, the left-hand side of Equation 5 gives the best fit, which means that Equations 3 and 4 give the desired temperature.

RESULTS

The results given after each method of calculation described in the previous section are for the wavelength range of $0.3-2.4\ \mu$. Calculations were also made for the intervals $0.4-1.2$ and $0.220-7.0\mu$; the former region is of interest for purposes of solar cell illumination, and the latter is a complete cover of Johnson's data.* All results are given in Table 1. The auxiliary range of $0.3-2.0\mu$ is included for comparison purposes.

It is apparent that the best fit (2.97 percent) was obtained for the $0.4-1.2\mu$ region, which indicates that the adjusted black body curve conforms reasonably well to Johnson's data throughout the visible and into the near infrared. Less satisfactory agreement (6.08 percent) was obtained for the region of $0.3-2.0\mu$ because a single black body formula cannot be made to fit the steep rise in emissiveness characteristic of the sun's radiation in the region of $0.34-0.4\mu$. Agreement is enhanced, however, when the region is extended to 2.4μ because the additional $2.0-2.4\mu$ region, where the fit is good, makes the subvisible region ($0.34-0.4\mu$) of somewhat less weight. The fit for the complete range covered by Johnson's data, $0.220-7.0\mu$, is least satisfactory. The black body curve (with only two adjustable parameters) cannot be made to fit the steep rise in the $0.220-0.4\mu$ region as well as the tail of the sun's spectral distribution in the infrared.

Table 1

Results of Gray Body Fit to Solar Extraterrestrial Spectrum.

Range (microns)	T (°K)	C_1 (watt-cm^2)	$\int H_\lambda d\lambda$ (watt/cm^2)	$\int \lvert J_\lambda - H_\lambda \rvert d\lambda$ (watt/cm^2)	$\frac{\int \lvert J_\lambda - H_\lambda \rvert d\lambda}{\int H_\lambda d\lambda} \times 100$
0.4 - 1.2	6142 ± 0.5	6.76111×10^{-17}	9.86137×10^{-2}	2.92699×10^{-3}	2.97
0.3 - 2.0	5730 ± 0.5	8.72019×10^{-17}	1.29419×10^{-1}	7.86669×10^{-3}	6.08
0.3 - 2.4	5742 ± 0.5	8.65040×10^{-17}	1.32629×10^{-1}	7.88560×10^{-3}	5.94
0.220 - 7.0	5693 ± 0.5	8.80334×10^{-17}	1.39424×10^{-1}	1.03376×10^{-2}	7.41

In order to clarify the fit to the solar extraterrestrial spectrum by the above gray body matching technique, a gray body curve was calculated for the $0.3-2.4\mu$ region. The gray body curve of this region for a temperature of 5742°K is compared with the solar data in Figure 1. According to column 6 of Table 1 the areal mismatch of these curves is 5.94 percent when computed relative to the solar constant. It can be concluded that a gray body curve agrees well with the solar data

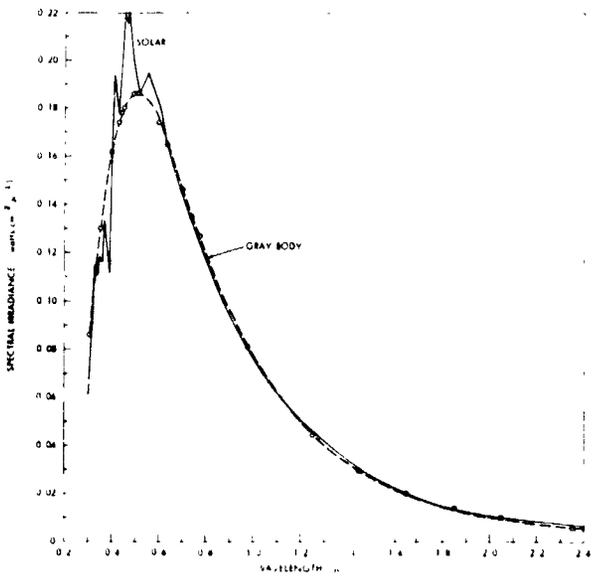


FIGURE 1.—Comparison of a 5742° K gray body curve with the solar extra terrestrial spectrum.

*Johnson, F. S., "The Solar Constant," *J. Meteorology* 11(6): 431–439, December 1954.

curve, except forthe irregular region of 0.41–0.55μ. What has happened is that virtually exact fits occurred at 0.43 and 0.52μ, and the intermediate irregularities could not be fit by a single gray body curve.

DISCUSSION

If the black body approximation to the solar spectrum is to be useful from the thermal design viewpoint, the average absorptance of a spacecraft coating material, as calculated from the black body spectrum, should be fairly close in value to the average solar absorptance of the material. Mathematically expressed, it is desirable that $\bar{a}_{BB} \approx \bar{a}_s$ where

$$\bar{a}_{BB} = \frac{\int_0^\infty a_\lambda J_\lambda(T)d\lambda}{\int_0^\infty J_\lambda(T)d\lambda} , \qquad (7)$$

and

$$\bar{a}_s = \frac{\int_0^\infty a_\lambda H_\lambda d\lambda}{\int_0^\infty H_\lambda d\lambda} , \qquad (8)$$

with

$\bar{a}_{BB}$ = the average absorptance for a black body,
$\bar{a}_s$ = the average solar absorptance,
H_λ = the solar spectral intensity per unit wavelength bandwidth, as given by Johnson,
$J_\lambda(T)$ = the Planck black body spectral intensity for absolute temperature T,
a_λ = the absorptance of the material at wavelength λ.

For simplicity, it is desirable to normalize H_λ and J_λ to H'_λ and J'_λ:

$$H'_\lambda = \frac{H_\lambda}{\int_0^\infty H_\lambda \, d\lambda} ,$$

$$J'_\lambda = \frac{J_\lambda}{\int_0^\infty J_\lambda \, d\lambda} .$$

Equations 7 and 8 now become

$$\bar{a}_{BB} = \int_0^\infty a_\lambda J'_\lambda(T)d\lambda , \qquad (9)$$

and

$$\bar{a}_s = \int_0^\infty a_\lambda H'_\lambda d\lambda . \qquad (10)$$

Arbitrarily, a 10 percent proportional difference in $\bar{a}_{BB}$ and $\bar{a}_s$ will be chosen as the criterion of usefulness; that is, if

$$\frac{|\bar{a}_{BB} - \bar{a}_s|}{\bar{a}_s} = E_p \leq 0.1 , \qquad (11)$$

the black body approximation will be considered useful. Equation 11 is actually

$$\frac{\left|\int_0^\infty \left[J'_\lambda(T)d\lambda - H'_\lambda\right] a_\lambda d\lambda\right|}{\int_0^\infty a_\lambda H'_\lambda \, d\lambda} \leq 0.1 . \qquad (12)$$

It can be seen that the results given in column 5, Table 1 for the quantity

$$\int |J'_\lambda(T) - H'_\lambda| d\lambda = \Delta_p \qquad (13)$$

do not indicate the magnitude of E_p (Equation 11). Consider now the following two cases:

1. If a_λ is approximately constant, then the proportional error E_p becomes

$$E_p = \frac{a_\lambda \left|\int_0^\infty \left[J'_\lambda(T) - H'_\lambda\right] d\lambda\right|}{a_\lambda \int_0^\infty H'_\lambda d\lambda} , \qquad (14)$$

and, since

$$\int_0^\infty J'_\lambda(T)d\lambda = \int_0^\infty H'_\lambda d\lambda = 1 , \qquad (15)$$

it is clear that

$$\int_0^\infty \left[J'_\lambda(T) - H'_\lambda\right] d\lambda = 0 \qquad (16)$$

and $E_p = 0$.

2. If a_λ can be represented by a power series in λ, that is,

$$a_\lambda = (a_\lambda)_0 + (\lambda - \lambda_0)\left(\frac{da_\lambda}{d\lambda}\right)_0 + \frac{(\lambda - \lambda_0)^2}{2!}\left(\frac{d^2 a_\lambda}{d\lambda^2}\right)_0 + \cdots , \tag{17}$$

then

$$E_p = \frac{\left| \int_0^\infty \left[J'_\lambda(T) - H_\lambda \right] \lambda \left(\frac{da_\lambda}{d\lambda}\right)_{0,1} d\lambda \right|}{\int_0^\infty H'_\lambda a_\lambda d\lambda} , \tag{18}$$

in which the error is now weighted by the slope.

See N64-33628

DISTRIBUTION IN HELIOGRAPHIC LONGITUDE OF FLARES WHICH PRODUCE ENERGETIC SOLAR PARTICLES*

DONALD E. GUSS

A number of evidences have been found in the past which point to the longitudinal persistence of solar phenomena. The life of active regions on the sun extends over several solar rotations and centers of sunspot formation exist which seem to outlive individual spot groups.[1,2] Also, there are 27-day recurrences of geomagnetic disturbances[3] and neutron-monitor counting-rate data,[4] which sometimes last for as long as 10 solar rotations. More recently Bryant *et al.*[5] and Simpson[6] have found 27-day recurrences in the enhancement of low-energy proton intensity in free space as measured on satellites. It is the purpose of this Letter to point out a similar longitudinal persistence in the occurrence of solar flares which result in solar particle events. In particular, flares from a single 10° interval in heliographic longitude caused most of the large solar particle events over the last solar cycle. This indicates the existence of a center for the formation of active regions which has persisted for more than 73 solar rotations.

Figure 1 shows the heliographic longitude for flares which produced solar-particle events between 16 January 1955 and 23 October 1962 during the last solar cycle. The list of 56 events used is from that compiled by Malitson[7] plus the 23 October 1962 event. The series of flares between 80° and 90° in Fig. 1 produced the largest particle events of the last solar cycle, the event of 23 February 1956, the multiple events of July 1959, the multiple events of November 1960, and the multiple events of July 1961. Using the integrated particle intensities calculated by Malitson and Webber,[8] these events contributed about 75 percent of the integrated particle intensity with energy greater than 30 MeV and about 90 percent of that with kinetic energy >100 MeV detected at the earth during the last solar cycle. In addition, of the ten solar particle events which produced neutron-monitor increases, indicating the presence of significant numbers of particles with kinetic energy greater than about 500 MeV, seven are in this narrow longitude region.

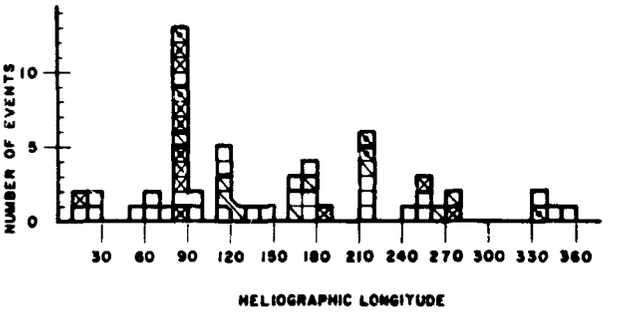


FIGURE 1.—Number of solar particles events vs heliographic longitude with the central meridian during the 23 February 1956 event set to 0°, and assuming a rotation period of 27.04 days. The particle intensity with kinetic energy >30 MeV detected at the earth[8] integrated over the particle event is ⊠, $I \geq 10^8$p/cm²; ⊡, $I \geq 5\times10^6$p/cm²; open squares, $I < 5\times10^6$p/cm². A dot in a square signifies that there was a neutron-monitor rate increase, indicating the presence of a significant number of particles with kinetic energy greater than about 500 MeV. Light lines are used to separate individual events and heavy lines to separate individual active regions.

Trotter and Roberts[9] have noted that the vicinity of the sun where the 23 February 1956 flare occurred had had a prior history of active regions lasting for several rotations, dying out, and returning in new bursts of activity. Hence, this center of activity existed prior to the 23 February 1956 event and persisted beyond the event of 20 July 1961, a period of more than 5.5 years. In his theory of the sun's field Babcock[10] proposes a mechanism to explain the longitudinal

*Published in *Physical Review Letters*, 13(12):363-364, September 21, 1964.

persistence whereby new active regions are likely to arise near old regions. However, it is not clear whether this mechanism is sufficient to explain an active site of such long duration.

The remainder of the events during the last solar cycle also fall into longitude bands, but not so sharply defined. The events between 240° and 280° with one exception occurred between 20 January 1957 and 23 March 1958, indicating the existence of an active site which lasted for more than a year. The events between 210° and 220° are those of March through September 1960, two of which caused ground-level neutron-monitor increases. The interval between 160° and 190° contains events which occurred between 9 August 1957 and 10 May 1959. The large event is that of 10 May 1959. The interval between 110° and 140° includes events from 6 June 1958 to 22 August 1958 and the two small events of 10 September and 28 September 1961.

Finally, there is a dearth of events in the longitude interval between 280° and 80°. Of the 26 larger particle events during the last solar cycle only two, the events of 31 August 1956 and 7 July 1958, resulted from flares in this 160° interval.

In conclusion, flares which produce energetic particles arise predominantly in narrow longitudinal regions which outlive visible active regions. This points to the existence of long-lived active centers beneath the photosphere which periodically manifest themselves on the surface as active regions which produce flares that accelerate solar particles. A single well-defined longitude region was responsible for the most intense particle events of the last solar cycle. The fact that this region can be compressed into a 10° band of longitude with a suitable choice of the period of solar rotation would indicate that this site rotated at constant rate as opposed to the variable rotation of the photosphere. The shorter lived regions, responsible for the smaller events, in general show a wider spread in longitude which could indicate a drift of the producing centers. A study of the meridional distribution of a larger body of data such as solar flares would perhaps result in a further definition of the active centers.

---

[1] K. O. KIEPENHEUER, *The Sun*, edited by G. P. Kuiper (University of Chicago Press, Chicago, Illinois, 1953), p. 338.

[2] H. M. LOSH, Publ. Obs. Univ. Mich. **7**, 127 (1938).

[3] J. BARTELS, Terrest. Magnetism Atmospheric Elec. **37**, 48 (1932).

[4] J. A. SIMPSON, W. FONGER, and L. WILCOX, Phys. Rev. **85**, 366 (1952).

[5] D. A. BRYANT, T. L. CLINE, U. D. DESAI, and F. B. MCDONALD, Phys. Rev. Letters **11**, 144 (1963).

[6] J. A. SIMPSON, IMP Symposium, Goddard Space Flight Center, Greenbelt, Maryland, March 1964 (unpublished).

[7] H. H. MALITSON, National Aeronautics and Space Administration Technical Report No. R-169, edited by F. B. McDonald, 1963, p. 109 (unpublished).

[8] H. H. MALITSON and W. R. WEBBER, National Aeronautics and Space Administration Technical Report No. R-169, edited by F. B. McDonald, 1963, pp. 12, 13 (unpublished).

[9] D. E. TROTTER and W. O. ROBERTS, Solar Activity Summary I, High Altitude Observatory (1956) p. 8.

[10] H. W. BABCOCK, Astrophys. J. **133**, 581 (1961).

See-X63-15346

N66 37953

THE EFFECT OF A CORONAL EXPANSION ON THE COMPUTATION OF AN ELECTRON TEMPERATURE FROM EMISSION-LINE INTENSITIES*

WERNER M. NEUPERT

Certain assumptions made in the calculation of a coronal electron temperature from the relative populations in two stages of ionization of an element are examined, with the conclusion that this calculation may not be valid for the particular case of the solar corona, yielding a fictitiously low value if there is an outward expansion of ions from the base of the corona into a region of higher electron temperature. A non-equilibrium formulation for the distribution of ions is presented that may better describe the actual ion distribution in the corona. An approximate calculation shows that an outward radial velocity which is small compared with the average thermal velocity of the ions is sufficient to explain the apparent observed increase in height of the maxima of emission of the Fe ion sequence with increasing stage of ionization.

I. INTRODUCTION

Past observations of the solar corona have resulted, with considerable consistency, in a discrepancy between the electron temperature of the "quiet" or "normal" corona, undisturbed by centers of activity, as obtained (*a*) from the electron density gradient inferred from scattering of photospheric light by electrons and (*b*) from ionization theory. The first method yields values of about 1.6×10^6 °K, whereas the second, from the ratio Fe XIV/Fe X, gives values half as large. Recent measurements by Orbiting Solar Observatory-I indicate that the latter method may lead to inconsistent results, for the ratio Fe XVI/Fe XV, as measured by the satellite, implies an average electron temperature of $1.7-1.8\times10^6$ °K (Neupert 1963). In addition to these values for the electron temperature, a high value of the kinetic temperature, of the order of 2.0×10^6 °K, is obtained from observed line-broadening if it is assumed that the broadening is due entirely to thermal motions. It has often been suggested that the electron temperature computed from the ratio Fe XIV/Fe X is the more accurate value, while that (kinetic) temperature determined from line-broadening is inaccurate because of the presence of nonthermal motions. Although this possibility exists, it has yet to receive observational verification.

Line-broadening measurements also suggest that any temperature gradient in the quiet corona is quite small at heights between $r=1.03\ R\odot$ ($R\odot$ being the solar radius) and a few solar radii, being negative and of the order of 3°K per km for the Fe XIV (5303 Å) line (Billings and Lilliequist 1963). Under such circumstances it becomes difficult to account for the apparent simultaneous existence of low stages of ionization, such as Fe X and Fe XI, and also a stage as high as Fe XVI, observed by OSO-I to be the emitter of two of the more intense emission lines of the extreme ultraviolet spectrum between 170 Å and 400 Å, even in the absence of visible centers of activity. The suggestion has been made (Shklovsky 1951) that limb observations may include regions of differing electron temperature, thereby giving the spectrum a misleading appearance. Such an argument is without doubt correct near centers of activity which manifest an increase in electron temperature by the appearance of the Ca XV (5694 Å) line. However, for the undisturbed corona, the existence of regions having the

*Published in *The Astrophysical Journal*, 139(3):935–940, April 1, 1964.

required range of electron temperatures would require tangential temperature gradients considerably greater than the radial gradient reported by Billings and Lilliequist.

Since it is the ionization theory which leads to an electron temperature considerably less than that obtained by other observational methods, it is of interest to re-examine certain assumptions made in the theory in light of current knowledge of the physical state of the quiet corona. It is the purpose of this paper to suggest that a combination of physical circumstances, at least in the quiet solar corona, may alter the distribution of ions in the various stages of ionization, thereby making the ionization theory invalid as a means of computing an electron temperature for the coronal plasma, and also permitting a broader distribution of ions than is allowed by ionization theory in a nearly isothermal corona. A steady state non-equilibrium formulation, valid for an optically thin plasma, is proposed which appears to be more compatible not only with present optical information about the corona at visible wavelengths but also with certain observations in the extreme ultraviolet spectrum made from rockets and satellites.

II. EFFECT OF A CORONAL EXPANSION

A combination of two physical conditions may lead to a discrepancy in the coronal temperatures derived from line intensities and from line-broadening. These two factors are (1) an average time between ionizations (of heavy ions) which is large compared with the time for equipartition of energy between the ions and the coronal plasma; (2) a mass motion of ions through a region having a high electron-temperature gradient in a period of time equal to or less than the above-mentioned time between ionizations. We will first consider the existence of each of these conditions in the lower corona and then discuss their influence on a theory which assumes ionization equilibrium.

An estimate of the time between ionizations may be obtained from ionization rates iron for ions computed from cross-sections given by Elwert (Massey 1956). These rates increase with increasing electron temperature, from 7.4×10^{-13} sec^{-1} cm^3 at 0.8×10^6 °K to 1.95×10^{-11} sec^{-1} cm^3 at 2.0×10^6 °K for Fe XIII, and decrease with increasing stage of ionization from 2.5×10^{-11} sec^{-1} cm^3 for Fe XI to 3.5×10^{-12} sec^{-1} cm^3 for Fe XV at 1.6×10^6 °K. These rates are higher (by a factor of 4 or more) than those used by Firor and Zirin (1962) but slightly lower than those used by Seaton (1962). We will assume, as do Firor and Zirin, that the dependence of ionization rates on the number of valence electrons is small and may be neglected. A measure of the time between successive ionizations (disregarding recombinations) is then given by

$$\tau_{ion}=\frac{1}{N_e S(T)},$$

where N_e is the electron density and $S(T)$ is the ionization rate in sec^{-1} cm^3. Taking a value for N_e of 2.0×10^8 cm^{-3} and a value for $S(T)$ of 1.0×10^{-11} sec^{-1} cm^3, one finds that a time interval of the order of 500–1000 sec (one to two time constants) is required to ionize, without recombinations, an assemblage of ions from stage i (taken as Fe XIII) to stage $i+1$ in the lower corona at an electron temperature of 1.6×10^6 °K, a value consistent with that obtained from the electron-density gradient. On the other hand, the time required to establish equipartition of energy between protons and heavy ions in the coronal plasma is considerably shorter. This characteristic time can be computed from equations given by Spitzer (1956) and is of the order of 2.0 sec for Fe XIII–proton equilibrium at an electron (and, approximately, proton) density of 2.0×10^8 cm^{-3} and a temperature of 1.6×10^6 °K. It is of interest to note that the formula predicts the heavy ion-proton equipartition time to be less than that for proton-proton equipartition by a factor of 2 (Fe XI) to 4 (Fe XVI) because of the high charge of the heavy ions. The time for equipartition is again inversely proportional to the electron density, leading to the relation

$$\tau_{ion}/\tau_{equi}\sim\text{constant}\gg1$$

for a given ion anywhere in the solar corona.

In considering the second of the two conditions stated above we must restrict ourselves to regions of low magnetic field, thus permitting an expansion of the corona which results in the corpuscular stream deduced by Biermann (1951) from his work on comet tails and was more recently discussed by Parker (1960). Using the continuity equation one can compute the average solar-wind

velocity at any height in the atmosphere once its velocity and particle density at another height have been defined. Using as initial parameters (Parker 1961) a proton density of 1.8×10^7 cm^{-3} and a solar-wind velocity of 66 km/sec at $r=1.43R\odot$ one then finds, near the base of the corona, assuming an electron density of 2.0×10^8 cm^{-3}, a velocity of about 12 km/sec. A recent calculation by Billings and Lilliequist produces a value about half as great. One direct measurement of the solar wind in the lower corona, a radar observation, has been made which yields a velocity similar to the computed values. The Massachusetts Institute of Technology El Campo station has observed, at 38 Mc, a Doppler shift of 4 kc, corresponding to a radially outward velocity of the reflecting electrons of 16 km/sec (James 1963). Knowing the electron density at which a critical frequency is reached, one may compute the velocity at other densities using the continuity equation. At an electron density of 2.0×10^8 cm^{-3} one then obtains 3.5 km/sec. It must be pointed out that these velocities are at variance with those obtained by extrapolation of recent observations made by Mariner II between the orbits of Venus and Earth. From this satellite data one would predict a wind, for the quiet sun, of only 0.06 km/sec in the lower corona. It remains to be seen, however, whether an extrapolation over nearly 1 a.u. using a limited sampling of the interplanetary plasma is valid. For this reason the radar results will be used in this paper. Using a solar-wind velocity of several kilometers per second, one concludes that an ion may travel radially outward a distance of 10^3–10^4 km in the lower corona before undergoing further ionization.

The result of fulfilling both of the above conditions is that an ion, in passing through a region having a steep positive electron-temperature gradient, very quickly attains a thermal velocity characteristic of the higher temperature (condition 1) but carries with it, for a much longer time and radial distance, a history of any previous ionization equilibrium (condition 2). Under such circumstances no theory which postulates equilibrium between ionization and recombinations can furnish a valid value for the electron temperature of the region even if the predominant ionization mechanism is electron impact, and recombination is by radiative capture. A kinetic temperature derived from line-broadening may still be valid, however. Because of the large increase in temperature over a distance of 10000 km or less in the lower solar corona, it appears that the situation here described may well occur throughout most of the region of the corona from which radiation is observed.

At any height in the corona one must then replace the equilibrium equation (Woolley and Allen 1948),

$$N_iN_eS_i^{i+1}(T)=N_{i+1}N_eQ_{i+1}^i(T),$$

where $Q(T)$ is the recombination rate in sec^{-1} cm^3, with the more general equation

$$\nabla\cdot F_i\gtrless 0,$$

where F_i is the flux of ions, in stage i, of each element. In general,

$$\begin{aligned}\nabla\cdot F_i=&(\text{ionizations from } i-1)\\&-(\text{recombinations to } i-1)\\&+(\text{recombinations from } i+1)\\&-(\text{ionizations from } i).\end{aligned}$$

For an optically thin corona of low density in which photoionizations and three-body recombinations are assumed to be negligible one obtains

$$\begin{aligned}\nabla\cdot F_i=&N_{i-1}N_eS_{i-1}^i(T)-N_iN_e[Q_i^{i-1}(T)+S_i^{i+1}(T)]\\&+N_{i+1}N_eQ_{i+1}^i(T),\end{aligned}$$

for each stage of ionization, i, of each element present, together with

$$N(r)=\sum_i N_i(r),$$

where N_i is the number density in each stage of ionization. The summation is made over all stages of ionization of the particular element being considered.

For a spherically symmetric corona expanding with velocity of magnitude $v(r)$ we have, for the total flux of a particular ion crossing a shell of radius r,

$$F_i=4\pi r^2N_i(r)v(r).$$

By requiring a steady-state outward flow, all ions moving outward with the same average velocity, we obtain

$$F_0=\sum_i F_i\approx 4\pi r^2N_e(r)v(r)=\text{constant}.$$

In this case the summation includes all elements present. Using the notation P_i, where P_i is the relative population in each stage of ionization i,

$$P_i \equiv \frac{N_i(r)}{Nr)},$$

we obtain

$$\frac{d}{dr}P_i = \frac{N_e^2}{F_0/4\pi r^2}\{P_{i-1}S_{i-1}^i(T) - P_i[Q_i^{i-1}(T) + S_i^{i+1}(T)] + P_{i+1}Q_{i+1}^i(T)\}$$

for each element and each stage of ionization.

We note that this equation is valid for any local areas, e.g., plages, in which the outward expansion may be different from that of the undisturbed corona, since the coefficient of the right-hand side can be evaluated locally in the corona. This equation is similar in form to that obtained by Burton and Wilson (1961) for a fully contained plasma except for a change in the independent variable from time to position and a change in the coefficient of the right-hand side. An analytic solution is possible under the assumption that only three adjacent stages of ionization are significantly changing for any value of the independent variable, all lower stages having reached a steady-state relative population and all higher states having yet to be populated. Although such an assumption is apparently valid for the plasma treated by Burton and Wilson and permits an analytic solution to be obtained for any stage of ionization, this is not the case for the solar atmosphere. In the latter, ions in the lowest stages of ionization may exist in regions of high optical thickness where the equation given is not even approximately valid. In order to apply the equation to the corona, where it is valid, one therefore requires boundary conditions at the base of the corona where more than three stages of ionization may already be significantly populated. At this boundary the solar atmosphere exhibits considerable fine structure, on which insufficient data are available for a precise evaluation of the equations.

However, it is possible to discuss the general behavior of the system by considering the expansion of an assemblage of ions from a region of low electron temperature of 5×10^5 °K to a hotter region of 1.6×10^6 °K, the change in electron temperature taking place over a distance of only a few thousand kilometers. Examination of the equation for the predominant stage of ionization present at 5×10^5 °K indicates that its relative population will decrease, in the absence of recombinations, by a factor of $1/e$ in a distance of approximately

$$\frac{F_o}{4\pi r^2}\frac{1}{N_e^2 S(T)},$$

which is of the order of 10^3 km. Higher stages of ionization will become more populated with increasing height above the photosphere, reaching an ionization equilibrium characteristic of the higher temperature only in a distance of a few tens of thousands of kilometers. The reliability of this conclusion depends, of course, on the validity of the data we have used. If the solar proton measurements of Mariner II are used, rather than the radar measurement, together with the higher ionization rates of Seaton, the effect being discussed here will be entirely negligible.

III. COMPARISON WITH OBSERVATIONS

A small amount of observational evidence is available which appears to be consistent with the foregoing arguments. Beginning with the lower stages of ionization observed in the corona, it has been found, for the undisturbed corona, that (*a*) the maximum surface brightness of Fe XI emission occurs below 10000 km above the photosphere (February 25, 1952 eclipse) (Athay and Roberts 1955), (*b*) the maximum surface brightness of Fe XIV occurs (statistically) at about 18000 km (Petri 1952), (*c*) the intensity of the Fe XV line at 7059Å decreases much more slowly with distance from the limb than do the lines of Fe X (6374 Å) and Fe XI (7891 Å) (February 15, 1961, eclipse) (Wlerick and Fehrenbach 1961). This observation supports those of Allen (1946) who noted that the equivalent widths of coronal lines with relatively low ionization potentials showed a faster decrease with height above the photosphere than did lines of high ionization potential. Based on the observation by OSO-I that the intensity of emission from Fe XVI requires emission from 0.1 to 0.5 of the solar corona, one might hypothesize, in accordance with the model being presented here, that the maximum in surface brightness of Fe XVI occurs at about 40000 km.

Because of the rapidly decreasing electron density with height in the lower corona, the maximum in intensity of an ion may be very broad and need not be coincident with the maximum in relative population of the ion.

The above observations are, of course, partially consistent with a coronal model, based on ionization equilibrium, which incorporates a positive temperature gradient outward from the chromosphere. Such a model could not be employed beyond 1.03 $R\odot$, however, since at distances greater than this the corona exhibits a negative temperature gradient, in contradiction to what would be expected from observation (*c*).

It is of interest to apply the arguments which have been developed to these observational data, thereby estimating the radial ion velocity in the lower corona. Assuming an average time between ionizations of 500 sec for stages Fe XI–XIV, we obtain, for the average radial velocity of these ions

$$\bar{v}=\frac{10000}{1500}=6.6 \text{ km/sec.}$$

Taking recombinations into account, we can reduce this velocity but by an amount estimated to be no more than 10 to 30 percent. This velocity is comparable to the solar-wind velocity in the lower corona inferred from radar observations. By comparison, the average thermal velocity for Fe ions at 1.6×10^6 °K is 15 km/sec. Since only limb observations of the corona are possible at visible wavelengths, little shift in the forbidden emission lines due to the outward ion velocity should be observable.

IV. CONCLUSION

It appears that even at the low electron densities found in the corona a computation of the electron temperature based on ionization theory may be invalid in regions of outward expansion through an electron-temperature gradient. Such a condition may be responsible, in part, for the difference in coronal temperatures obtained from ionization theory and from line-broadening measurements. Its eventual importance will depend on better estimation of ionization and recombination rates and the velocity of the solar wind. The presence of a coronal expansion does not rule out the existence of turbulence in the quiet corona, which may still account for the difference in temperatures obtained from electron-density gradients and from Doppler broadening of coronal lines.

REFERENCES

ALLEN, C. W. 1946, *M.N.*, **106**, 137.

———. 1961, *Proceedings of IAU Symposium No.* **16**, ed. J. W. Evans (New York:Academic Press), p. 1.

ATHAY, R. G., and ROBERTS, W. O. 1955, *Ap. J.*, **121**, 231.

BIERMANN, L. 1951, *Zs. f. Ap.*, **29**, 274.

BILLINGS, D. E., and LILLIEQUIST, C. G. 1963, *Ap. J.*, **137**, 16.

BURTON, W. M., and WILSON, R. 1961, *Proc. Phys. Soc.*, **78**, 1416.

FIROR, J., and ZIRIN, H. 1962, *Ap. J.*, **135**, 122.

JAMES, J. 1963, private communication.

MASSEY, H. S. W. 1956, *Hdb. d. Phys.*, Vol. **36**, ed. S. Flügge (Berlin:Springer-Verlag), p. 307.

NEUPERT, W. M. 1963, *Space Research IV* (to be published).

PARKER, E. N. 1960, *Ap. J.*, **132**, 821.

———1961, *Space Astrophysics*, ed. W. Liller (New York: McGraw-Hill Book Co.), p. 157.

PETRI, W. 1952, *Forschungsber. Ob . Wendelstein*, Vol. 2.

SEATON, M. J. 1962, *Observatory*, **82**, 111.

SHKLOVSKY, I. S. 1951, *Izv. Krymsk. Ap. Obs.*, **6**, 105.

SPITZER, L. 1956, *Physics of Fully Ionized Gases* (New York:Interscience Publishers).

WLERICK, G., and FEHRENBACH, C. 1961, *Proceedings of IAU Symposium No.* **16**, ed. J. W. Evans (New York: Academic Press), 199.

WOOLLEY, R. v. d. R., and ALLEN, C. W. 1948, *M.N.*, **108**, 292.

Sec - 64-30257

TIME VARIATIONS IN THE SOLAR EXTREME ULTRAVIOLET SPECTRUM OBSERVED BY OSO-1*

W. M. NEUPERT

1. INTRODUCTION

A grazing incidence spectrometer[1] has been flown as a pointed experiment on OSO-1 to monitor the extreme ultraviolet (EUV) spectrum of the sun. Over six thousand spectra were obtained at the rate of about 100 per day over a period of time from March through May 1962, corresponding to nearly three solar revolutions. Intermittent operation has subsequently provided spectra for a period of greater than one year. This paper presents, in a condensed form, the variations observed in three of the more reliably identified lines of the spectrum: 284-A (Fe XV), 304-A (He II) and 335-A (Fe XVI).

One may consider the solar EUV observations as falling into three categories: a) quiet corona, b) slowly varying components, and c) transient phenomena, such as flares. These categories are also listed in the order of increasing difficulty of making reliable observations. This paper will review some of the observations of slowly varying components that have been made and then present several instances in which transient phenomena have been observed.

2. DESCRIPTION OF SPECTROMETER

During operation the spectrometer was pointed at the center of the solar disk within approximately two min of arc. In this orientation, radiation from the entire solar disk and inner corona passed directly through the entrance slit and struck a concave grating mounted in grazing incidence, the angle of incidence being 88°. The exit slit and detector were mounted on a carriage which was driven on a circular rail so that the exit slit scanned along the Rowland Circle, where the spectrum was focused, from 10–400-A. The 50 micron entrance and exit slits provided a spectral passband of 1.7-A and permitted resolution of lines 0.85-A apart. The detector was a windowless photomultiplier developed by the Bendix Corporation specifically for use in this spectrometer. A tungsten photocathode was chosen to minimize response to wavelengths above 1500-A, and to reduce changes in sensitivity due to variations of the emission properties of the cathode. The instrument calibration was performed (1) by exposing the entire spectrometer to a beam of monochromatic radiation of known intensity and (2) by evaluating the essential components of the spectrometer (grating, detector, etc.), and then computing the sensitivity of the instrument. The first method was used at 44-A using a proportional counter for determining the source intensity. The second method was applied at longer wavelengths (80–400-A). In addition, a comparison was made of the solar fluxes obtained against the fluxes measured by Hall, Damon and Hinteregger[2] with a calibrated rocket instrument. The comparison could only be made in the region of overlap, 250-A to 400-A, and would be meaningful only if the solar radiation was the same. The 2800-Mc mean daily flux recorded by the National Research Council, Ottawa, Canada, was used as an independent estimate of solar flux to choose the satellite data for the comparison. A best fit was made between these methods of obtaining a calibration which yielded values of 8.0×10^{5} photons cm^{-2} $count^{-1}$ at 335-A and 4.2×10^{5} photon cm^{-2} $count^{-1}$ at 284-A. These values are useful for comparing the intensities of the lines at these two wavelengths, but their use to obtain absolute solar fluxes is not recommended.

*Published as *Goddard Space Flight Center Document X-614-64-24*, January, 1964.

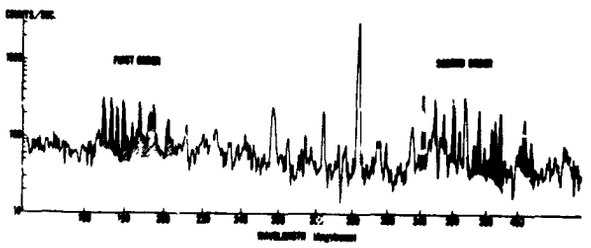


FIGURE 1.—Solar spectrum from 170 A to 400A with second order line images indicated

3. PRESENTATION OF DATA

Figure 1 presents a typical scan obtained over the region of 170–A to 400–A. Several factors reduce the usable spectral range of the spectrometer from its nominal range of 10–A to 400–A. At wavelengths below 100–A the decreasing sensitivity of the spectrometer combined with an increase in scattered light upon approaching the central image made it impossible to distinguish a reliable spectrum on a single scan. However, a combination of several scans, using cross-correlation techniques, has shown the apparent existence of a line emission spectrum in the region from 50–A to 100–A (Neupert, Behring, White and Lindsay[3]). At wavelengths above 342–A the second order images of intense spectral lines observed in first order above 171–A obscure a considerable amount of interesting data. This effect is illustrated in Figure 1 which indicates, by crosshatching, those first order lines which have strong second order counterparts in the recorded spectrum. These second order images were less prominent in earlier rocket results, leading to considerable dissimilarity between the rocket and satellite spectra above 342–A (Neupert and Behring[4]). For these reasons the obviously usable range of the instrument can be considered to be from 170–A to 340–A, a range which can be extended, with special precautions, to cover the interval from 50–A to 400–A.

The brightest emission line in the region from 170–A to 340–A is the Lyman-alpha line of ionized helium at 304–A. In addition, numerous other emission lines appear with combined flux comparable to, or somewhat greater than, that of the helium line. Resonance lines of heavy ions (Mg through Fe) are expected in this region, leading to attempts (Zirin, Hall and Hinteregger[5]), (Neupert and Behring[4]) to identify the more prominent features of the spectrum in terms of such lines. The difficulty of positive identification can be appreciated more if, for example, we realize that the resonance lines of all stages of ionization of iron from Fe X through Fe XVI, with the exception of Fe XV, are expected in the small wavelength interval from 335–A to 390–A. The overlapping of these resonance multiplets combined with second order radiation of shorter wavelengths makes positive identification a formidable task. The resonance lines of Fe X through Fe XIV have not been identified in the extreme ultraviolet (EUV) spectrum although their intensity as predicted by Ivanov-Kholodny and Nikolsky[6] should permit their observation. Tousey[7], working with a spectrum having higher resolution, reports that no Fe XIV multiplet can be found in his record. Because of these difficulties, discussion will center around three of the more intense and more reliably identified lines of the spectrum, the Fe XV ($3s^2\ {}^1S_0 - 3s\,3p\ {}^1P1_0$) line at 304–A. The location of these lines is indicated in Figure 2.

4. SLOWLY VARYING COMPONENTS OF THE EUV SPECTRUM

The months of March and April of 1962, were auspicious for a study of the solar EUV spectra in that observations could be made on both a quiescent and a disturbed solar atmosphere. During the second week in March the sun was especially quiet, the Zurich Provisional Relative Sunspot number being zero on 11 March. As the month progressed solar rotation carried several centers of activity across the visible hemisphere of the sun. Definite enhancements in the solar spectrum were associated with these centers of activity.

Figure 2 presents two scans of the EUV spectrum taken approximately ten days apart. During the first of these observations (13 March) only one small region of activity was present on the solar disk. In spite of this low level of activity it is observed that the Fe XV and Fe XVI lines persist as two of the more prominent features of the spectrum. The second spectrum (23 March) was obtained while several large and well-developed centers of activity were present on the disk.

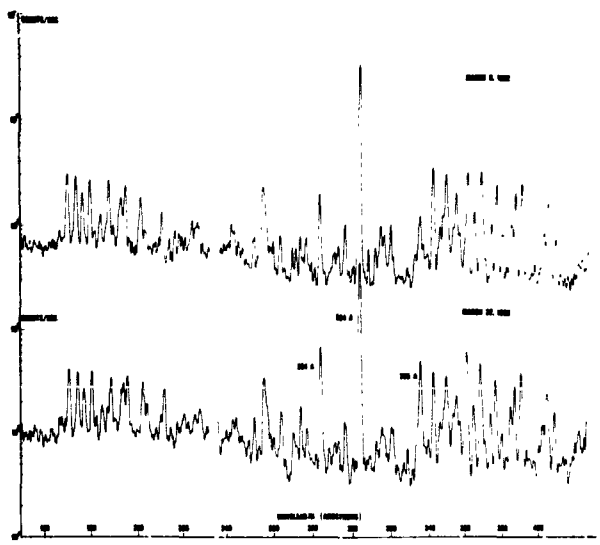


FIGURE 2.—Comparison of two spectral scans taken under different solar conditions

Comparing these two spectra we observe that the emission lines have increased in intensity but not all by the same amount. The Fe XV and XVI lines, already prominent even in the near absence of solar activity, have increased in intensity appreciably more than any other line observed with certainty in this spectral range. The He II line has also increased but by a lesser amount.

Time variations of these three lines as observed for the first 1066 orbits of operation are given in Figure 3, in which each datum point represents the average of one orbit's observations (six to eight scans). Error bars indicating one standard deviation in the data are given for each wavelength. Also indicated are the times of appearance and disappearance of the major plage areas as cataloged by the McMath-Hulbert Observatory. The first increase in the counting rates above the "quiet sun" values is associated with the appearance of McMath Plage No. 6366 on the east limb on 11 March. The counting rate for the He II (304–A) line increases gradually to a maximum on 23 March, after which it drops rather abruptly to an apparent plateau. This sudden drop cannot be unambiguously attributed to the disappearance of plage region 6366 on the west limb, for it follows quite directly after an unusual flare of importance 3 observed in plage region 6370 on 22 March, for which data are not given in the figure. A gradual decline in readings is observed from 29 March to 7 April. A similar enhancement, followed by a return to near preplage rates, occurred in April. Since the spectrometer did not carry an internal standard it is impossible to state positively that its counting characteristics did not change over the period of observations. The consistent behavior of the data, however, as demonstrated in Figure 3, leads to the conclusion that no significant change in sensitivity occurred throughout the period of observation. (Subsequent analysis of real time data after one year in orbit supports this conclusion, even for the longer time interval).

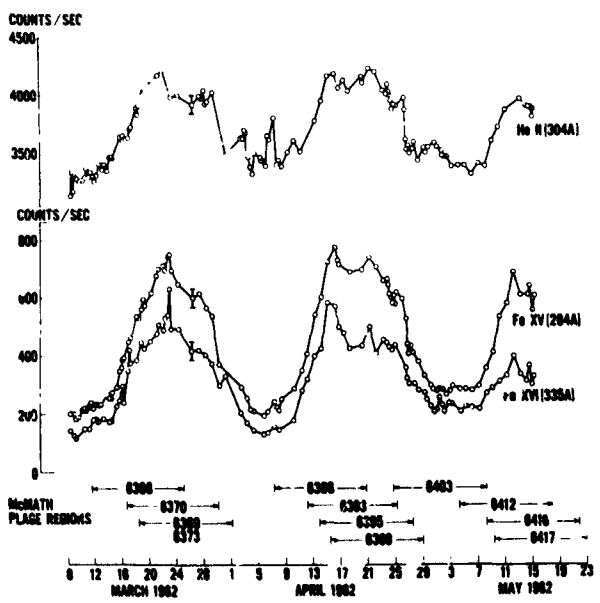


FIGURE 3.—OSO–1 data for three spectral lines compared with appearance and disappearance of major plage areas.

Also shown in the figure are the variations with solar activity of the coronal Fe XV (284–A) and Fe XVI (335–A) lines, the relative increases being considerably larger than for the He II (304–A) line. A definite, although small, enhancement associated with the appearance of plage No. 6366 is observed in both the Fe XVI and Fe XV lines. The possibility exists that the coronal emissions appear slightly earlier as might be expected since the coronal emission must take place at greater heights than the emission of ionized calcium radiation. The most obvious increase in intensity of these lines is associated with the appearance, in succession, of plage areas numbered 6370, 6369 and 6373 on 17 March through 19 March. The counting rates then increased gradually throughout this period indicating that the enhanced emission was not from a point source but rather from

an extended volume having as its base an area at least as large as the underlying calcium plage. The disappearance of area 6366 on 25 March results in a slight decrease in counting rates, followed by a general decline from 29 March through 3 April as the other plage areas are carried off the visible hemisphere of the sun by the solar rotation.

5. DISCUSSION OF SLOWLY VARYING COMPONENTS

The relationship of these observed counting rates to several ground-based measurements of solar activity is presented in Figures 4 and 6. In Figure 4 the He II radiation is compared with

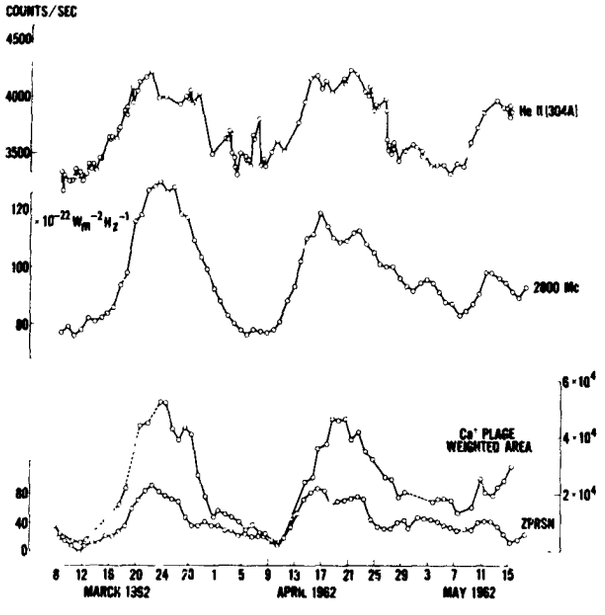


FIGURE 4.—Relationship of observed counting rates for the He II Lyman-alpha line to ground-based measurements of solar activity.

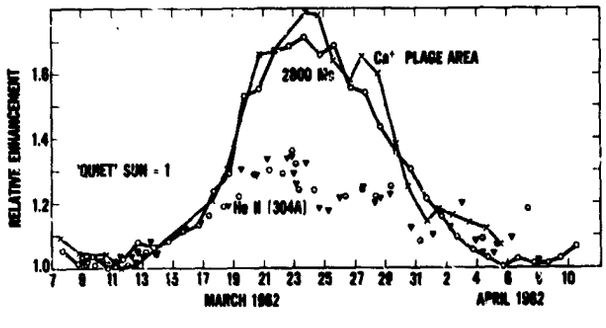


FIGURE 5.—Comparison of 304 A data, normalized to a "quiet sun" value of one, with weighted calcium plage data and 2800 Mc radio observations.

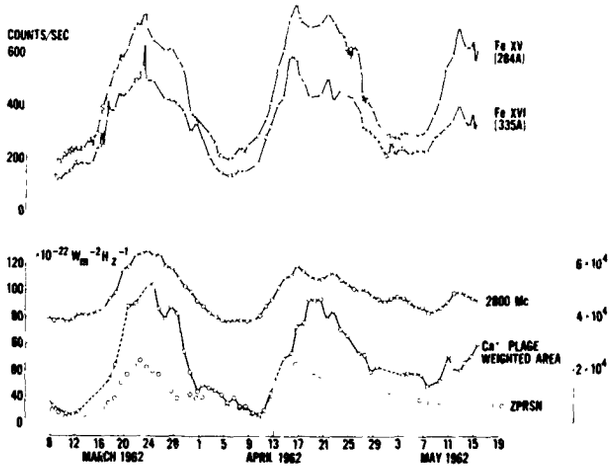


FIGURE 6.—Relationship of observed counting rates for the 284 A and 335 A line to ground-based observations.

daily values of the solar flux at 2800–Mc, and with the Zurich Provisional Relative Sunspot Number (ZPRSN). Also shown is an estimate of the calcium plage area, each area being weighted by the estimated intensity of the area on a scale from 1 to 5 and by the cosine of its longitude from the central meridian. Values for this computation were supplied by the McMath-Hulbert Observatory.

A more detailed comparison is made in Figure 5 in which the radio and He II fluxes have been adjusted by dividing each observation by the rate observed for the "quiet sun." In addition, the weighted Ca^+ plage intensity has been adjusted to coincide with the 2800 Mc enhancement on 21 March. This plot shows more clearly the close relationship which has been known for some time between the 2800 Mc radiation and the plage areas and also shows the divergences of the He II (304–A) line from both of these. A better agreement between plages and the 304–A line is obtained if only the plage areas, with no weighting factors for intensity or longitude from central meridian passage is used. In Figure 6 the daily values of solar flux at 2800 Mc and the Zurich Provisional Relative Sunspot Number are compared with the coronal lines of Fe XV (284–A) and Fe XVI (335–A). The estimated weighted calcium plage area is also shown.

The spectral lines chosen for presentation here were selected because they are reliably identified with particular ions, not because they convey, more than any other line in the spectrum, changes

in solar flux which occur with the appearance of plage areas. These three lines represent the extremes which have thus far been observed in the spectral region from 171–A to 400–A; only a few faint lines have smaller non-flare variations than the He II Lyman-alpha line, while no other outstanding lines have increases as great as those observed for 284–A and 335–A.

An attempt has been made to compute an electron temperature for the corona, as nearly as possible undisturbed by centers of activity, from the observed ratio of intensities of the Fe XV and Fe XVI lines. Such a calculation has validity only if both radiations originate in the same region of the corona at some well-defined electron temperature and only if an ionization-recombination equilibrium, as assumed by theory, is actually present. It is found that an electron temperature of 1.75 ± 0.10 million degrees in that part of the corona assumed to be the source of both Fe XV and Fe XVI best fits the observed data. This result contrasts sharply with the value of 800,000 °K usually obtained from observation of the forbidden lines of Fe X and Fe XIV.

It has been estimated that the emission of Fe XV and Fe XVI radiation associated with a plage area is approximately 150 to 200 times that of an equivalent area of the quiet sun. The corresponding value for the 304–A line is 15. Assuming that the increase in emission is due to a density increase only, one finds it necessary to hypothesize an increase in electron density by a factor of 12 to 14 over each plage for the two iron lines. The fact that a large range of intensity variations is observed demonstrates, however, that an increase in electron density alone is not sufficient to explain the behavior of the spectrum. On the other hand, the constancy of the ratio Fe XVI/Fe XV as one extrapolates to zero plage area argues against a significant change in electron temperature over plages. One can only conclude, at this time, that those regions (or region) of the corona emitting Fe XV and Fe XVI radiation occupy a relatively greater volume of the corona over plages and centers of activity.

The initial analysis of only three lines (He II 304–A, Fe XV 284–A, Fe XVI 335–A) indicates that the relative prominence of spectral lines may depend upon the age of the center of activity which is responsible for the increased radiation. As an example of this, one may observe (Figure 6) that the maximum emission in the Fe XV line apparently occurs later in time than the maximum for the 2800 Mc flux, or for the plage areas, observed during March, April and May. In addition to such a slowly changing effect, one may note that localized perturbations appear (7–9 March and 16–17 April) for which the relative

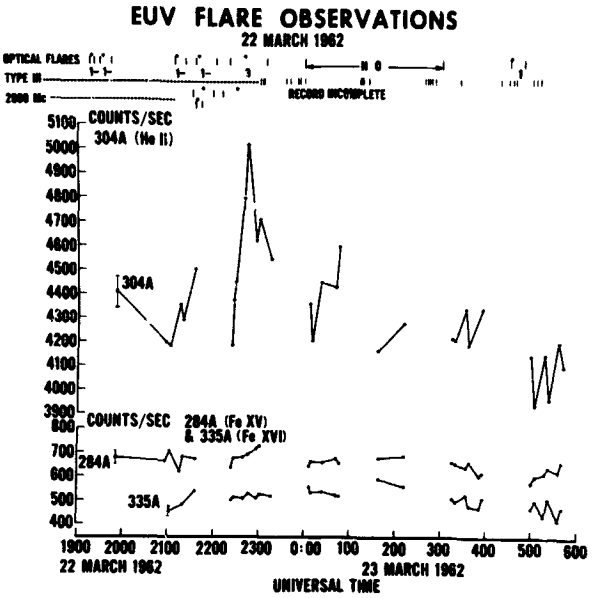


FIGURE 7a.—Increases in EUV line emission associated with optical flares.

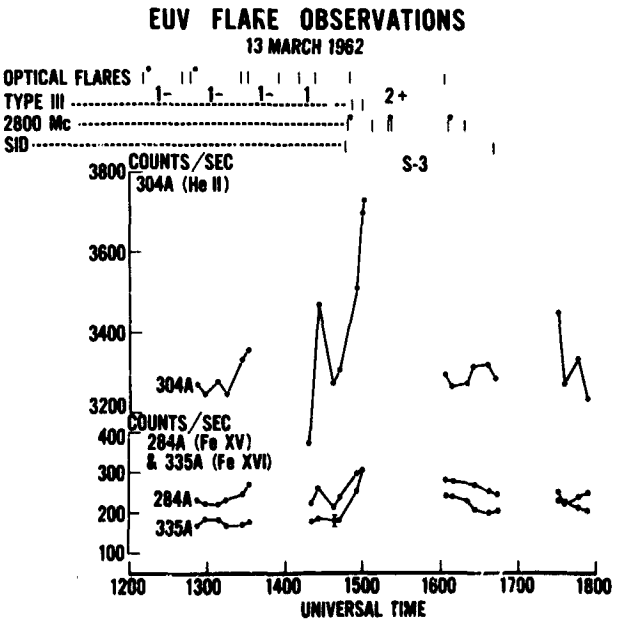


FIGURE 7b.—Increases in EUV line emission associated with optical flares.

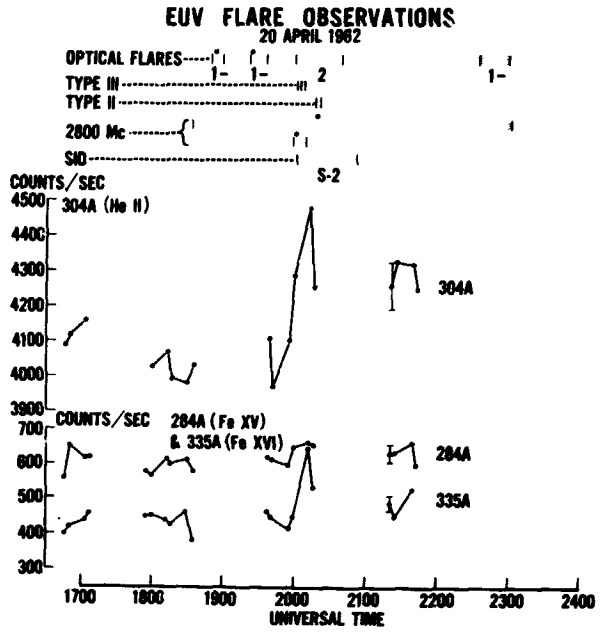


FIGURE 7c.—Increases in EUV line emission associated with optical flares.

increases in the helium and the iron lines are considerably different. In these instances we are apparently observing phenomena localized at particular levels in the solar atmosphere.

6. EUV FLARE OBSERVATIONS

It is difficult to obtain reliable EUV flare observations with the type of spectrometer used in OSO–1 because the flare radiation is seen as only a small increase in the radiation from the entire solar disk, and because the long time required to scan the spectrum (8 min) severely limits the time resolution with which the flare can be observed. In spite of these difficulties, several increases in the EUV radiation have been associated with flares. Three such instances are shown in Figures 7 a,b,c. Supplementary information for these events is given in Table I.

6.1. 22 March, 1962 (Flare of Importance 3)

From the observations of the He II line at 304–A made one minute before and three minutes after maximum in H-alpha one may conclude that the maximum at 304–A certainly occurs no earlier than H-alpha maximum and may occur later. No changes at 284–A and 335–A can be associated with this flare, optically of importance 3 but geophysically of much lesser importance.

TABLE 1.—*Supplementary Flare Observations*

A. 22 March 1962

	Start	Max	End	Importance
Optical Flare	22:20?	22:41	23:10?	3
2800 Mc Bursts	Gradual rise 22:00 to sunset (22:40). Short Bursts at 22:31, 22:33.5, 22:38.			35 Flux Units
Type III Bursts	None Observed			
Type II Bursts	None Observed			
SID (HAO Preliminary Report)				
SCNA	22:33		23:02	1–
SEA	22:35		22:35	1–

APPROXIMATE EUV OBSERVATION TIMES (UT) AT PEAK OF FLARE
March 22–23

		(Peak)		
304A (He II)	22:40:23	22:44:10	22:58:00	
284A (Fe XV)	22:40:03	22:43:50	22:57:40	
335A (Fe XVI)	22:40:53	22:43:20	22:58:30	

TABLE 1.—*Supplementary Flare Observations*—Continued

B. 13 March 1962

	Start	Max	End	Importance
Optical Flare	14:48 (earlier than)	15:05	16:01 (later than)	2+
SID (CRPL–F Bulletin, NBS)				
SCNA	14:50	15:00	15:45	1
SEA	14:50	----------	15:55	2
SPA	14:45	15:05	16:40	
2800 Mc Bursts	14:47.5	14:50.5	15:07	470 Flux units
Type III Bursts	14:50	----------	14:58	Group of Bursts

APPROXIMATE EUV OBSERVATION TIMES (UT)

				(Peak)
304A (He II)	14:40:05	14:53:55	14:57:10	
284A (Fe XV)	14:40:25	14:53:35	14:57:30	
335A (Fe XVI)	14:39:35	14:54:25	14:56:40	

C. 20 April 1962

Optical Flare	19:58	20:02	20:40	2+
SID (CRPL–F Bulletin, NBS)				
SCNA	20:00	20:04	20:35	2
SEA	20:01	----------	20:50	2
SPA	20:02	20:07	20:50	2
2800 Mc Burst	19:57.3	19:59	20:09	80 Flux Units
Type III Bursts	20:00.6	----------	20:02	(Harvard)
	20:00	----------	20:04.15	(HAO)
	20:05.15	----------	20:07.30	(HAO)
Type II Bursts	20:04.6	----------	20:19	(Harvard)
	20:16	----------	20:20.3	(HAO)

APPROXIMATE EUV OBSERVATION TIMES (UT)

		(Peak)		
304A (He II)	19:58:00	20:11:50	20:15:00	
284A (Fe XV)	19:58:20	20:11:30	20:15:20	
335A (Fe XVI)	19:57:30	20:12:20	20:14:30	

6.2. 13 March, 1962 (Flare of Importance 2+)

Peak phase of this flare apparently occurred near satellite sunset. The last valid observations before sunset indicate that the EUV flare radiation reaches a maximum after the 2800 Mc radio burst. Increases are noted for all three lines, with the ratio 335–A/284–A apparently increasing at or near maximum phase of the optical flare. The enhancement in the coronal lines may continue after cessation of the visible flare.

6.3. 20 April, 1962 (Flare of Importance 2+)

EUV observations are poorly situated in time relative to the optical maximum of this flare. It appears, therefore, that the EUV emission maxima may have been significantly greater than the maximum reading actually recorded. In spite of this shortcoming, an increase in the Fe XVI radiation, relative to Fe XV can again be observed.

7. DISCUSSION OF FLARE OBSERVATIONS

A considerable number of additional events have apparently been observed, some associated with optical flares, others not. However, the data for these events have not been sufficiently analyzed to establish their validity. The increase of the ratio Fe XVI/Fe XV is most frequently associated with solar events producing ionospheric disturbances. The increase in this ratio suggests that some volume of the solar atmosphere, associated with the flare, has been raised to a higher electron temperature.

It is interesting to note that the EUV maxima coincide more closely with the optical maxima and with ionospheric affects than they do with centimetric radio bursts. The association with ionispheric effects is understandable if we assume that the hard X-rays, which themselves may be line emission at short wavelengths, and which are causing these ionispheric disturbances, have the same common thermal origin in flares as does EUV flare radiation observed by the OSO–1 spectrometer. On the other hand, the coincidence with optical H-alpha maxima cannot be explained in the same manner since the hydrogen radiation must originate in a cooler region of the solar atmosphere.

DISCUSSION

Maron: In comparing the active sun and the quiet sun, is your finding that the variations in fluxes from the lines in the higher ionization states of iron have greater amplitude than the lower states? Is this not most simply interpreted as saying that the relative density change is greatest in the regions of highest temperature?

Dr. Neupert: Yes, that is a possible conclusion. However, we cannot discriminate between this hypothesis and one in which the electron temperature profile of the corona changes above the active region in such a manner that a greater relative volume is at the higher temperature.

Zirim: Just one remark. I have the impression that you have essentially a factor of two increase in the total radiation from the whole sun coming from a flare which essentially is .1 percent of the area, or something like that, which impresses me with a very large change. Is that so, would you say?

Dr. Neupert: We can associate changes of 50 percent or more during flares only with the Fe XVI and Fe XV lines. Many lines show no statistically reliable change. For the event on March 13 we can demonstrate, by averaging our data over one complete orbit, that several other lines do show changes. Rapid variations of 10 percent to 20 percent in the faint lines would be unobservable because of statistical fluctuations. We should probably use the H-alpha area of a flare only as a lower limit in estimating the size of the EUV emitting region.

Guerjoy: Would you clarify for me how long, how large an area of the sun you are looking at?

Dr. Neupert: We are looking at the entire solar disc and out to about a quarter of a solar radius from the limb. So we are looking at the entire disc plus the inner corona with this instrument.

Chubb: Can you say something about how the variability of the lines around 170 angstroms is as compared to, say, the 304 and the iron lines?

Dr. Neupert: The variations with solar plage area of the lines between 170–A and 200–A are less than those of the 284–A and 335–A lines. They are comparable to or somewhat more than that of the helium line. We have not analyzed flare data for these lines but it is our experience that lines which increase most due to the presence

of a center of activity also increase the most during flares.

Dr. Lindsay: I note that you have averaged about 16 lines in this region over a period of about a month, as I recall. Why don't you report at least the integrated or the average of these lines varies?

Dr. Neupert: We have made integrations over the spectrum for two days. Not for the 13th of March and 23rd of March, but for two days on which activity was similar to those which we had in the second slide. The increase in counting rate was of the order of 50 to 80 percent. This is an increase due to the presence of centers of activities on the sun's disc, not increase due to flares.

REFERENCES

1. W. E. Behring, W. M. Neupert and W. A. Nichols, J. Opt. Soc. Amer. **52** (5):597, May, 1962.
2. L. A. Hall, K. R. Damon and H. E. Hinteregger, Space Research III, ed. W. Priester (Amsterdam, North-Holland Publ. Co., 1963) 745.
3. W. M. Neupert, W. E. Behring, W. A. White and J. C. Lindsay, GSFC Document X-614-62-156 (1962)
4. W. M. Neupert and W. E. Behring, J. Quant. Rad. Transf. **2** (1962) 527.
5. H. Zirin, L. A. Hall and H. E. Hinteregger, Space Research III, ed. W. Priester (Amsterdam, North-Holland Publ. Co., 1963) 760.
6. G. S. Ivanov-Kholodny and G. M. Nikolsky, Astronomicheskii Zhurnal **38** (1961) 828.
7. R. Tousey, Space Science Reviews (to be published) (1963).

THE UV LINES OF FE XIV IN THE SOLAR CORONA

R. STOCKHAUSEN

Knowledge of the structure of the solar corona can be increased by observations of the many emission lines in the far ultraviolet. However, identification of these lines is made difficult by the lack of knowledge of many of the atomic parameters relevant to these lines. Garstang's (1962) calculations of energy levels and oscillator strengths for Fe XIV allow predictions of lines strengths and wave lengths for the lines of this ion.

Collision cross-sections of the optically permitted lines are related to the oscillator strengths by the expression given by Van Regemorter (1962)

$$Q(n, m) = \frac{1}{k_n^2} \frac{8\pi}{\sqrt{3}} \frac{I_H}{\Delta E} f(nm)\ g\pi a_o^2 \qquad (1)$$

For levels not connected to the ground level by an optically allowed transition, electron exchange collisions are the only excitation mechanism. In a manner similar to Osterbrock (1963) these exchange cross-sections can be estimated by

$$Q_e(in, jm) < Q_o(in, jm)$$

$$= \frac{\pi}{k_n^2} \begin{cases} (2\ell+1)\ [g_{jm}/\Sigma_m g_{jm}] \\ (2\ell'+1)\ [g_{jm}/\Sigma_m g_{jm}](g_j/g_i) \end{cases} \qquad (2)$$

$$\text{or } \frac{\pi}{k_n^2} \begin{cases} (2\ell+1) \\ (2\ell'+1)\,(g_{jm}/g_{in}) \end{cases}$$

whichever of the four is smaller. Here $i\ n$ is the n^{th} level of term i and $j\ m$ is the m^{th} level of term j. The term in square brackets comes from the assumption that the levels of term j are populated in proportion to this weight factor.

In his work with nebular ions, Osterbrock has taken $Q_e = 0.1 Q_0$. On the basis of the rapid decrease, with increasing energy of the colliding electron, of the cross-sections for exchange collisions in He I (St. John, Miller, and Lin 1964) and the decrease of exchange cross-sections along the iso-electronic sequences shown in Table VIII of Seaton (1958), Dr. Osterbrock has suggested that a value of Q_e nearer to $Q_e = 0.01 Q_0$ may be more appropriate for Fe XIV. With the uncertainties involved, parallel calculations were made under three assumptions:

$$Q_e = \begin{cases} .0 \\ .1Q_0 \\ .01Q_0 \end{cases} \qquad (3)$$

The total cross-section is then the sum of (1) and (3).

Having the collision cross-sections, production rates of the UV lines can be calculated. Following the notation of Seaton (1963), the rate of excitation from level i to level j followed by spontaneous transition to level k is given by

$$\alpha_{ik} = q_{ij} C_{jk} \qquad (4)$$

where q_{ij} is the collisional excitation rate of j from i and C_{jk} is the probability of going to k from j. For the UV lines of Fe XIV, C_{jk} does not involve collisional de-excitation due to the high spontaneous transition probabilities, and the low coronal electron densities.

Table 1 gives the production rates for UV lines due to transitions ending in the two levels (1 denotes $^2P^\circ_{1/2}$ and 2 denotes $^2P^\circ_{3/2}$) of the ground term. Results are given for two temperatures and three values of the exchange cross-section. Excitations are assumed to take place only from the ground level.

The sum of the production rates to the upper level ($^2P^\circ_{3/2}$) of the ground term is important in that it gives the excitation rate (due to cascade) of the coronal green line, λ5303. It can be seen that there is a slight discrepancy between these

*Published as *Goddard Space Flight Center Document X-510-64-202*, July 1964.

TABLE 1.—*Production Rates*

Levels	$Q_e=.0$		$Q_e=.01\ Q_0$		$Q_e=.1\ Q_0$	
	a_{11}	a_{12}	a_{11}	a_{12}	a_{11}	a_{12}
$T=1\times10^6$						
$sp^2\ {}^4P_{1/2}$	$.01\cdot10^{-9}$	----------	$.03\cdot10^{-9}$	$.01\cdot10^{-9}$	$.17\cdot10^{-9}$	$.05\cdot10^{-9}$
${}^4P_{3/2}$	----------	----------	----------	.04	.02	.40
${}^4P_{5/2}$	----------	----------	----------	.06	----------	.62
${}^2D_{3/2}$	.66	$.02\cdot10^{-9}$	.70	.02	1.03	.03
${}^2D_{5/2}$	----------	----------	----------	.06	----------	.57
${}^2S_{1/2}$	1.19	.05	1.20	.05	1.35	.05
${}^2P_{1/2}$	.34	.49	.35	.50	.41	.59
${}^2P_{3/2}$	.17	.17	.18	.72	.24	.96
$s^2d\ {}^2D_{3/2}$	1.59	.39	1.65	.40	2.30	.56
${}^2D_{5/2}$	----------	----------	----------	.13	----------	1.32
		$1.65\cdot10^{-9}$		$1.99\cdot10^{-9}$		$5.15\cdot10^{-9}$
$T=2\times10^6$						
$sp^2\ {}^4P_{1/2}$	$.01\cdot10^{-9}$	----------	$.02\cdot10^{-9}$	$.01\cdot10^{-9}$	$.14\cdot10^{-9}$	$.04\cdot10^{-9}$
${}^4P_{3/2}$	----------	----------	----------	.03	.02	.33
${}^4P_{5/2}$	----------	----------	----------	.05	----------	.52
${}^2D_{3/2}$	.77	$.02\cdot10^{-9}$	.80	.02	1.09	.03
${}^2D_{5/2}$	----------	----------	----------	.02	----------	.50
${}^2S_{1/2}$	1.51	.06	1.53	.06	1.66	.07
${}^2P_{1/2}$	.43	.62	.44	.64	.55	.80
${}^2P_{3/2}$	.22	.89	.23	.91	.25	1.02
$s^2d\ {}^2D_{3/2}$	2.04	.49	2.11	.51	2.74	.66
${}^2D_{5/2}$	----------	----------	----------	.13	----------	1.31
		$2.08\cdot10^{-9}$		$2.38\cdot10^{-9}$		$5.28\cdot10^{-9}$

cascade excitation rates and those given in Table 6 of Seaton (1963). In a private communication, Dr. Seaton has noted a numerical eror in his work.

Inclusion of the exchange collisions has a significant effect on the production rates only for those levels which have low f values with respect to the ground level, and then only for the case $Q_e=0.1Q_0$.

The product of these production rates and the energy of the respective emitted photons, normalized so that the ${}^2P^o_{1/2}-{}^2P_{3/2}$ transition is unity, gives the predicted relative line strengths of the expected UV lines. These strengths are given in Table 2. Table 3 gives the predicted wavelengths and the wavelengths assigned by Neupert and Smith (1964) from observations of the NASA satellite OSO-I. These identifications were made on the basis of time variations of the line strengths and wavelength separations of the multiplet members. Also included are the observed relative line strengths for two periods of observation in 1962. Quiet Sun (average of March 8, 9, 10) is a period of low solar activity and April 20 is a period of high solar activity. These observed line strengths are also normalized to the ${}^2P^o_{1/2}-{}^2P_{3/2}$ line. This line was about 1.5 times more intense on April 20 than during the Quiet Sun period. At this stage of the reductions, the observed relative line strengths are accurate to only a few tenths (in these units).

TABLE 2.—*Predicted Line Strengths*

Levels	J_L	J_U	I					
			$T=1\times10^6$			$T=2\times10^6$		
$3s^2p-3sp^2$			$Q_e=.0$	$Q_e=.01Q_0$	$Q_e=.1Q_0$	$Q_e=.0$	$Q_e=.01Q_0$	$Qe=.1Q_0$
$^2P^\circ-{}^4P$	1/2	1/2	----	.1	.5	----	----	.3
	1/2	3/2	----	----	----	----	----	----
	3/2	1/2	----	----	.1	----	----	.1
	3/2	3/2	----	.1	.8	----	.1	.7
	3/2	5/2	----	.2	1.4	----	.1	1.1
$^2P^\circ-{}^2D$	1/2	3/2	2.8	2.8	3.1	2.5	2.5	3.2
	3/2	3/2	.1	.1	.1	.1	.1	.1
	3/2	5/2	----	.2	1.6	----	.1	1.4
$^2P^\circ-{}^2S$	1/2	1/2	6.4	6.1	5.2	6.3	6.1	6.1
	3/2	1/2	.3	.2	.2	.2	.2	.2
$^2P^\circ-{}^2P$	1/2	1/2	2.3	1.9	1.7	1.9	1.9	2.2
	1/2	3/2	1.0	1.0	1.0	1.0	1.0	1.0
	3/2	1/2	2.7	2.6	2.3	2.6	2.6	3.0
	3/2	3/2	3.9	3.8	3.8	3.9	3.8	3.9
$3s^2p-3s^2d$	----	----	----	----	----	----	----	----
$^2P^\circ-{}^2D$	1/2	3/2	10.9	10.7	11.2	10.8	10.7	12.8
	3/2	3/2	2.6	2.5	2.6	2.5	2.5	3.0
	3/2	5/2	----	.8	6.2	----	.6	5.9

TABLE 3.—*Observed Relative Intensities and Wavelengths*

Levels $3s^2p-3sp^2$	J_L	J_U	λ_P	λ_0	I Quiet Sun	April 20
$^2P^\circ$ ^2S	1/2	1/2	272.9	274.2	1.1	1.9
	3/2	1/2	287.7	288.6	.6	.6
$^2P^\circ$ ^2P	1/2	1/2	256.0	Blend	----	----
	1/2	3/2	250.5	251.8	1.0	1.0
	3/2	1/2	269.0	270.5	.7	1.0
	3/2	3/2	262.9	264.5	1.2	1.5
$3s^2p-3s^2d$	----	----	----	----	----	----
$^2P^\circ$ ^2D	1/2	3/2	214.1	203.9	.9	1.3
	3/2	3/2	223.1	211.5	1.3	1.8
	3/2	5/2	222.0			

In comparing the observations and predictions, it must be noted that the satellite observations do not cover the spectral region occupied by the $^2P^\circ-{}^4P$ multiplet. For the $s^2P(^2P^\circ)$ multiplet, $-sp^2(^2D)$ no lines of the predicted strengths or with the correct time behavior appear to be present at the predicted wavelengths, however the instrumental sensitivity is decreasing in this region.

Since these observed strengths are a result of contributions from the entire solar disc, and due to the uncertainties in the data, only semi-quantitative comparisons with the calculated strengths will be made. Most apparent is the relatively large spread in predicted strengths as compared

to the observed strengths. Using a temperature of 5×10^6 does not significantly reduce this spread. Exchange collisions do not seem important for most cases. However, for the $s^2p(^2P^o)-s^2d(^2D)$ multiplet the ratio of the ratio of the sum of the predicted intensities of the 3/2−3/2 and 3/2−5/2 lines to the 1/2−3/2 line (ratio=8.8:11.2 for $T=10^6$ and $Q_e=0.1Q_0$) is closer (but still in the wrong sense) than for the case where $Q_e=0$.

The possibility of blends cannot be ruled out absolutely, but since the time variations of all these line strengths was quite similar, it is assumed that blends have a small effect. One possibility for the discrepancies in relative strengths is some kind of self-absorption. Due to the very low rate of collisional de-excitation (relative to spontaneous transitions), "true" absorption is not to be expected. Repeated reabsorptions of the UV photons in the corona can change the relative strengths since after each absorption the atom has a choice of making a transition to either of the ground levels. This has the effect of making the relative strengths more nearly equal, which is the desired effect in this case.

This effect has been discussed for the corona by Pecker and Thomas (1963) and Pottasch (1964). Since the number of times a photon is scattered in leaving the corona is τ^2, the optical depth must be greater than unity to have any effect at all. Pottasch (1964) has calculated optical depths for some of the transitions involved here (using an electron density of $2\times10^8\text{cm}^{-3}$). These values ($\tau\approx10^{-1}$ to 10^{-2}) are too small to have any effect on the relative line strengths.

Neupert, Behring and Lindsay (1963) have shown that the UV line strengths seem to be correlated with plage regions and other solar activity. It may be that a higher electron density than that used in Pottasch's calculations may be more appropriate. However, even the electron densities measured by Aly, Evans, and Orrall (1962) in a bright coronal condensation ($\lesssim3\text{X}10^9\text{cm}^{-3}$) are too small to give self-absorption for most of the lines. One possible solution to the problem would be an inhomogeneous corona in which the dense elements had electron densities in the neighborhood of 10^{10}cm^{-3}.

There is one other interesting observation from OSO-1 that needs to be explained. The UV line intensities vary by a factor of two or three for different conditions of solar activity, while the intensity ratios vary by about 50 percent or less. In terms of an inhomogeneous model, this might be explained as each dense element being optically thick (thereby giving the correct intensity ratios). An increase in solar activity would merely mean an increase in the number of dense elements (thereby increasing the intensities of the lines, but having less effect on the intensity ratios. Further reduction of data and calculations should allow a more quantitative discussion of the observations

I am grateful to W. Neupert and E. Smith for discussions of the observations and for the relative line strengths prior to publication, to S. Kastner for helpful discussions and to D. Osterbrock for several discussions, especially those concerning the exchange cross-sections and the self-absorption mechanism.

REFERENCES

ALY, M. K., EVANS, J. W., and ORRALL, F. W. 1962, Ap. J., **136**, 956.

GARSTANG, R. H. 1962, *Ann. d'ap.*, 25, 109.

NEUPERT, W. M., BEHRING, W. E., and LINDSAY, J. C. 1963, Paper given at the Fourth International Space Science Symposium, Warsaw, Poland.

NEUPERT, W. M., and SMITH, E. v. P. 1964, Paper given at June 24-27 meeting of the American Astronomical Society.

OSTERBROCK, D. E. 1963, *Planet, Space Sci.*, 11, 621.

PECKER, C., and THOMAS, R. N. 1963, Ap. J., **137**, 967.

POTTASCH, S. R. 1964, M. N. R. A. S., **128**, 73.

REGEMORTER, H. VAN. 1962, *Ap. J.*, **136**, 906.

SEATON, M. J. 1958, *Rev. Mod. Phys.*, 30, 979.

———1963, *Planet, Space Sci.*, **12**, 55.

ST. JOHN, R. M., MILLER, F. L., and LIN, C. C. 1964, *Phys. Rev.*, 134, A888.

N66 37954

SOLAR X-RAYS: A COMPARISON WITH MICROWAVE RADIATION

W. A. WHITE

Analysis of data from the OSO–1 satellite shows that solar X-rays shortward of 10 Angstroms exhibit a violent dynamic behavior which however contains within counterparts of the somewhat less dramtic behavior of the solar microwave flux. Specifically, the solar X-ray flux contains: (1) a slowly-varying component, (2) "gradual rise and fall" bursts, and (3) impulsive bursts. The OSO–1 data shows the first two to be related to their microwave counterparts, and the emperical relations are given in this paper. Observations at 10 Angstroms of the third component (impulsive bursts) could not be associated directly with microwave observations owing to the extreme disparity in sensitivities of the two methods of observation of this phenomenon. However, observations by Frost (reported at this Symposium) of solar X-rays in the 20 Kev—100 Kev region do show a direct counterpart of the microwave impulsive burst.

1. INTRODUCTION

With the launching of the OSO–1 on March 7, 1962, it became possible for the first time to point instruments at the sun accurately and continuously for entire daylight portions of a satellite orbit; for the 550-Km orbit of OSO–1 these observing time intervals were the order of one hour each, separated by darkness intervals of about two-thirds of an hour. The observing periods were long enough to disclose some interesting dynamic effects which would be difficult to study otherwise.

2. SENSOR CHARACTERISTICS

OSO–1 provided coverage of the solar X-ray radiation near a wavelength of 10 Angstroms by means of an ion chamber whose characteristics are given in Table 1.

TABLE 1

Window Material	Beryllium
Window Thickness	0.005 inch
Total Window Area (Two chambers in parallel)	3.38 cm^2
Absorbing Gas	Xenon
Gas Pressure	780 mm
Ion Chamber Dept at Normal Incidence	2.19 cm
Ion Pairs per erg	2.8 $(10)^{10}$

*Published as *Goddard Space Flight Center Document X-610-64-42*, February 1964.

The conversion efficiency as a function of wavelength is shown in Fig. 1.

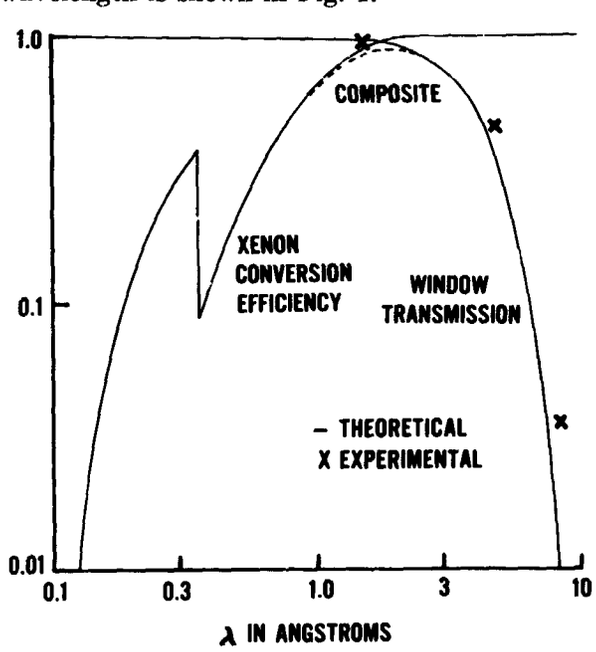


FIGURE 1.—OSO–1 ion chamber conversion efficiency as function of wavelength.

Amplified ion-chamber current was sampled for 2 seconds every 20 seconds by the telemetry.

The full-scale sensitivity of such a broad-band sensor is, of course, dependent upon the shape of the input spectrum. It has been shown[1] that

for non-flare periods it is reasonable to assume a spectral shape consistent with a $2.8(10)^{6}K$ plasma composed of ionized hydrogen and helium in the ratio of their solar abundances. Over the frequently used band limits of 2–8Å, the full-scale sensitivity is then $3.6(10)^{-4}$ ergs cm^{-2} sec^{-1}.

3. COMPONENTS OF THE SOLAR X-RAY FLUX

Changes in the solar X-ray flux as observed by OSO–1 occurred over a broad range of time scales; from fractions of a second to many weeks. In particular, counterparts to the microwave slowly-varying component and to the microwave "gradual rise and fall" burst are clearly seen.

The markedly dynamic character of the solar X-ray flux as contrasted with the 2800–Mc radiation should be emphasized; for instance, on only 3 occasions out of 70 days (7 March through 15 May, 1962) was the X-ray flux "quiet") i.e., free from bursts) for an interval of as much as 8 hours.

3.1 Slowly-Varying Component

Solar rotation markedly affects the X-ray flux. In fact, if one defines the non-burst X-ray flux background as the minimum flux reached in a 6-hour interval and evaluates this for a succession of such intervals, one finds that for the spring of 1962 the OSO–1 data empirically fits the relation

$$\Phi_{2-8A} = k_1 \quad (\Phi - \Phi_0)_{2800Mc} \text{ erg cm}^{-2} \text{ sec}^{-1} \qquad (1)$$

That is, the daily mean non-burst X-ray flux is proportional to the difference between the daily mean of the 2800–Mc flux and a value of 2800–Mc flux which is presumably a "quiet sun" asymptote for the appropriate phase of the solar cycle.[2] Note that any such quiet sun asymptote in the X-ray flux is *zero*, at least to the sensitivity and accuracy of this empirical fit. Apparently a truly quiet sun, stripped of plage, would be dark in X-rays while still emitting an appreciable 2800–Mc flux. Fig. 2 shows the measured mean non-burst X-ray flux and the index values calculated from Eqn. (1) with $K_1 = 1.4(10)^{-5}$, $(\Phi)_{2800\ Mc}$ in units of $10^{-22} Wm^{-2}(c/s)^{-1}$, and $(\Phi_0)_{2800Mc} = 73$ such units.

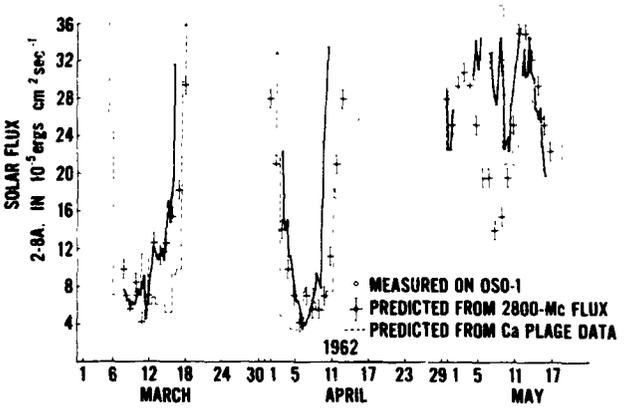


FIGURE 2.—Slowly-varying component of solar x-ray flux

It is fruitful to look in more detail at the relationship between the calcium plages and the X-ray flux. For example, is the brightness of a plage in Ca emission a measure of its brightness in X-rays? If so, we might expect a relationship somewhat as follows:

The brightness of a plage in Ca *K* emission is expressed by an Intensity Number assigned by the McMath-Hulburt Observatory apparently on the basis of a subjective estimate of relative densities on a patrol photograph; as such, it is likely to be approximately logarithmic.

The other routinely reported parameter characterizing a Ca plage is its area. Since we are dealing at X-ray wavelengths with a volume source, an assumption has to be made regarding the dependence of the X-ray source volume upon the area of the plage. It is found from the OSO–1 data that it is much better to assume that the thickness of the X-ray source is proportional to the mean plage diameter, than to assume a constant thickness. The volume of the i^{th} plage is then given by

$$(dV)_i \propto (A_i)^{3/2}$$

where A_i is the reported area.

The hypothesis that the X-ray flux arises from such a plage-associated volume sources with X-ray brightness proportional to Ca *K* brightness can be written as

$$\Phi_{2-8A} = K_2 \sum_i (A_i)^{3/2}(b) I_i \text{ erg cm}^{-2} \text{ sec}^{-1} \qquad (2)$$

where A_i is the volume of the i^{th} plage in thousandths, I is the McMath Ca *K* intensity number, and k_2 and b are constants to be empirically determined from the OSO–1 data.

[1] White, W. A., "Solar X-rays: Slow Variations and Transient Events" —COSPAR IV, Warsaw, June 1963. NASA GSFC Document X-614-63-195.

[2] A. E. Covington and G. A. Harvey, Astrophys. Jour. *132* (1960) 435.

The best fit to date is plotted in Fig. 2, for plages of area $\geq$ 1 thousandth, for

$$k_2 = 0.134\ (10)^{-5},$$

and for

$$b = 2.$$

Plage data reported by the McMath-Hulburt Observatory and published in the CRPL Bulletin was used, with times of rising and setting of the X-ray sources calculated from plage Central Meridian Passage minus and plus 7.8 days respectively. This corresponds to an assumed source height extending to some 21,000 Km above the photosphere.

Two types of gross discrepancy can be seen between the X-ray flux and the hypothetical indices derived from plage and 2800–Mc data. The first type, exemplified by the periods March 12–14, April 10–14, and May 1–2, show too low a value for the plage-derived index. Each of these periods saw the birth of one or more small plages upon the disc which were not included in the index computation of Fig. 2. The second type of discrepancy is seen during the period May 3–7 for which the index derived from the 2800–Mc flux is low. This period included a large plage (McMath 6412) which was reported to be declining in Ca *K* intensity; this plage contained a single sunspot which was seen to die on the disc.

3.2 Gradual Rise and Fall Bursts

Superimposed upon the X-ray slowly-varying component and occurring so frequently as to make difficult the finding of intervals completely free from their effects, are X-ray bursts analogous to the "gradual rise and fall" bursts seen in solar microwave radiation. A complete statistical analysis of the X-ray bursts observed by OSO–1 has not yet been made, but typical on-scale bursts (see Fig. 3) have rise times of 2 to 4 minutes and fall times of 10–20 minutes. However, many bursts carry the experiment output offscale; for these, the duration of the event can be as long as 6 hours. A comparison of the X-ray data with 2800–Mc records (kindly made available by A. E. Covington) taken at Ottawa shows that all 2800–Mc gradual bursts had a time-coincident counterpart in the X-ray data. A peak amplitude on only 1 or 2 units of 10^{-22} $W\ m^{-2}$ sec^{-1} at 2800-Mc

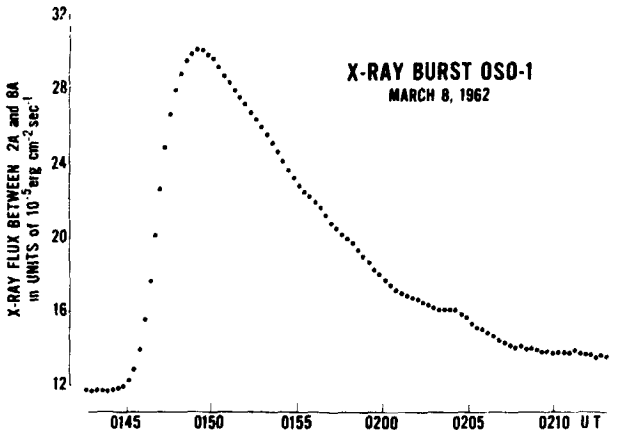


FIGURE 3.—X-ray burst OSO–1 March 8, 1962

was in all cases accompanied by an off-scale X-ray burst; larger bursts at 2800–Mc were of course also off-scale in X-rays, but again times of onset and recovery were found to agree with the corresponding times for the accompanying X-ray burst.

It is important to note that the relationship between burst flux at 2800–Mc and at 2–8 A is different from the relationship between the slowly-varying component at 2800–Mc and at 2–8 A. As has been shown (Eqn. 1), a change in 2800–Mc slowly-varying component of 1 flux unit is accompanied by a change in 2–8 A flux of about 1.4 $(10)^{-5}$ erg cm^{-2} sec^{-1}; whereas a burst type change at 2800–Mc of 1 flux unit is accompanied by a change in 2–8 A flux of at least $30.(10)^{-5}$ erg cm^{-2} sec^{-1}. This increase in the X-ray/microwave ratio of more than 20 times is suggestive of a different physical process being responsible for the two phenomena.

Several apparent associations of certain X-ray bursts into groups displaying a definite pattern were observed; Fig. 4 shows such grouping. Similar groupings are present in the data for the first week in April 1962; in fact, the one particular March group shown in Fig. 4 appears to have an exact April counterpart 27.1 days later with identical time-separations between events and with identical peak excursions above mean background level. In addition to these two groups, two others each having 4 members and one other having 3 members were found. For all five groups, the envelope joining the peaks of the bursts within a group is a straight line of slope $\pm 1.51(10)^{-9}$ erg cm^{-2} sec^{-2} for wavelengths be-

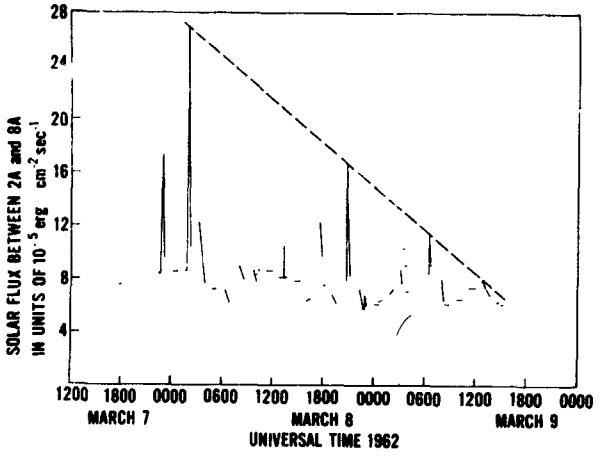


FIGURE 4.—Solar x-rays (2–8A) showing grouping of x-ray bursts.

tween 2 and 8A; three groups displayed positive slopes, two groups negative.

3.3 Impulsive Bursts

Several instances of very fast bursts (entire duration less than 2 minutes) were seen by Frost (reported at this Symposium) from an OSO–1 sensor responsive to photons of energy greater than 20 Kev. For every one of these instances the lower-energy ion-chamber channel was off-scale prior to and during the time interval of the impulsive burst. It is indeed unfortunate that we are thus unable to obtain an X-ray/microwave ratio for an impulsive burst to compare with the values obtained for gradual bursts and for the slowly-varying component.

IV. IONOSPHERE AND RADIO PHYSICS

N66 37955

D REGION FORMATION*

ARTHUR C. AIKIN

I. INTRODUCTION

That part of the terrestrial ionosphere which lies below 85 km is termed the "*D* region." Observations of ion and electron densities have been made at altitudes as low as 40 km so that this altitude can be taken as the lower boundary of the region. The mean molecular mass of the neutral atmosphere is essentially constant throughout the *D* layer and the distribution of the principal constituents, molecular oxygen and nitrogen, can be described by the hydrostatic equations relating density, pressure, and temperature. There are also several minor constituents of importance such as atomic oxygen, which reaches its peak concentration above 85 km, ozone and various oxides of nitrogen.

One of these minor constituents nitric oxide has an ionization potential such that it will be ionized by the strong hydrogen Lyman alpha line of the solar spectrum at 1216A. Lyman alpha together with short wavelength X-rays and galactic cosmic radiation are considered to be the ionizing sources of the undisturbed *D* region. Enhancement of 2 to 8A X-rays during solar flares causes large increases of electron density. These increases are termed sudden ionospheric disturbances or "SID's." Other equivalent terms in common use are cosmic noise anomaly (CNA) and sudden phase anomaly (SPA), although these terms refer to the effect on a specific type of measurement.

Electrons in the energy range 10 to 200 kev are precipitated in the auroral zone, which extends between 60° and 70° geomagnetic latitude. These electrons are the primary source of the aurora and cause large increases in *D* and lower *E* region ionization. Protons of 10 to 100 Mev produced in the course of large solar flares arrive at the earth a few hours after onset of the flare. The resultant increases in ionization occurring over the entire polar cap and extending south to a geomagnetic latitude of 65° are termed "Polar Cap Absorption" events or PCA. After the sudden commencement of a magnetic storm, which occurs about 20 hours after a strong flare, the polar cap radio wave absorbing region is extended to a geomagnetic latitude of 55°. Some sudden commencement events are accompanied by dumping of electrons in the auroral zone.

Electron density distributions, observed during some of the ionizing events discussed here, are illustrated elsewhere in this book (Electron Density Distribution by R. E. Bourdeau). A description of the ionization and recombination processes taking place in the homosphere under the influence of different radiations is given in the following sections.

II. ATMOSPHERIC ABSORPTION OF PHOTONS AND CHARGED PARTICLES

In order for radiation to penetrate to the *D* region it must have an absorption cross section of less than 10^{-19}cm^2 in air. Molecular oxygen is the limiting agent in this regard. In fact ultraviolet radiation having a wavelength shorter than the first ionization potential of O_2 at 1026.5A is precluded. At longer wavelengths the penetration of photons to sufficiently low altitudes can occur only in the atmospheric windows near 1216, 1187, 1167, 1157, 1143 and 1108A. For these wavelengths unit optical depth occurs near 10^{20} molecules cm^{-2} of vertical column. When the number of molecules in a vertical column is sufficient to reduce a flux of photons to 1/e of its incident value, unit optical depth is attained.

At wavelengths longer than 1800A the penetration of radiation is also possible between the bands of the Schumann-Runge system of O_2.

*Published as *Goddard Space Flight Center Document X-615-64-64*, March 1964. This work is a contribution to the "International Dictionary of Geophysics" Pergamon Press, Oxford, England.

However, ionization by these wavelengths requires constituents for which the ionization potential (IP) is low, such as sodium (IP–2413A) and calcium (IP–2028A).

For wavelengths shorter than 12.4A (1 kev) the absorption cross section rapidly decreases below 10^{-19}cm^2. It is found that the wavelength region most useful in the production of ionization between 70 and 85 km is 2 to 8A. The wavelength interval between 8 and 13A as well as near 35A does not penetrate to the *D* region since it is absorbed between 85 and 100 km resulting in the base of the *E* region. Radiations more energetic than 10 kev penetrate below 70 km as illustrated in figure 1. Here unit optical depth is plotted as a function of photon energy.

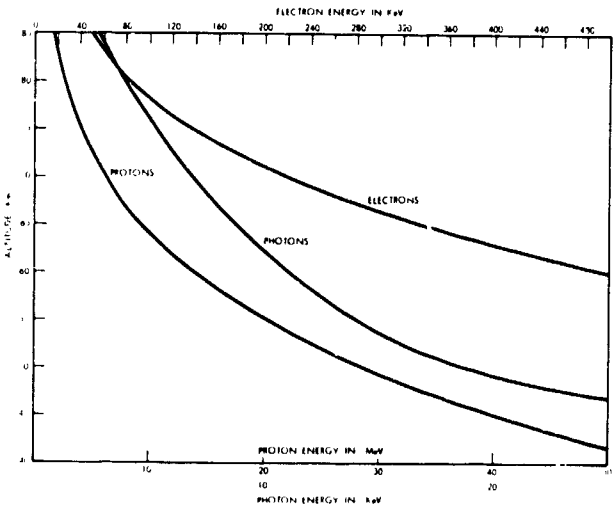


FIGURE 1.—Depth of penetration in the terrestrial atmosphere as a function of energy for photons, electrons and protons.

Nicolet and Aikin (1960) have compiled a table of the total absorption cross sections in O_2, N_2, A, and air for wavelengths between 10 and 0.1A. Both the photoelectric and Compton scattering cross sections for photons with energies greater than 10 kev, have been tabulated by Grodstein (1957). Above 1A, Compton scattering can be neglected and the absorption of a photon releases an electron whose energy depends on the ionization potential of the atom. If the electron is ejected from the *K* shell of the atom, the Auger effect must also be considered as a mechanism for producing additional electrons.

When an incident photon has an energy, $h\nu$, which exceeds the threshold of ionization for the *K* shell of O_2 and N_2 ($\sim$500 ev), the ejected photoelectrons will have a kinetic energy of $\frac{1}{2}$ mv^2 $=h\nu-500$. Since the average kinetic energy that is required of an electron to produce an ion pair is 35 ev, the efficiency of ionization, ϵ, of a photon which interacts by the photoelectric effect can be computed.

The behavior of charged particles in the earth's atmosphere is complicated by the presence of the earth's magnetic field. This field determines the latitude dependence for the energy deposition point of charged particles and modifies the incident particle energy spectrum. However, in order to know what energies are required of various charged particles to contribute to the ionization of the *D* region, it is sufficient to consider the range in gm/cm^2 of air of the particles.

Experimental data on the range of protons in air has been given by Bethe and Ashkin (1953). For energies less than 500 Mev Maeda and Singer (1961) have employed the following empirical formula

$$R(\text{gm/cm}^2) = E^{1.78}/420 \qquad (1)$$

where E is the proton energy in Mev. This expression can be used to estimate the depth of penetration as a function of proton energy as illustrated in figure 1. It is seen that protons in the energy range 2 to 40 Mev dissipate most of their energy in the altitude region 85 to 45 km. Maeda and Singer (1961) have also given an approximate range for electrons of energy less than 0.2 Mev

$$R(\text{gm/cm}^2) = E^{1.96}/0.75 \qquad (2)$$

based on the data of Katz and Penfold (1953).

III. THE PRODUCTION OF IONIZATION BY SOLAR PHOTONS

In the previous section a discussion was given of the absorption of photons in air and the subsequent production of ion pairs by the photoelectric and Compton processes. For the case of monochromatic solar radiation, the number of ion pairs produced cm^{-3}sec^{-1} at an altitude z is

$$q_p = n_j \epsilon'_j \sigma Q_\infty \exp(-\tau) \qquad (3)$$

where σ'_j is the effective ionization cross section in

cm^3 of the j constituent, n_j is the number density cm^{-3} of that constituent, Q_∞ is the number of photons/cm^2/sec vertically incident on the earth's atmosphere and ϵ is the efficiency of ionization. The optical depth, τ, is the sum of the products of the absorption cross section, σ_{Ak}, and the total number of absorbing molecules of species k between the height z and the sun. It is given by the expression

$$\tau = \sum_k n_k \sigma_{Ak} H_k \sec \chi \qquad (4)$$

where χ is the solar zenith angle, n_k is the number of molecules of species k at height z and H_k is the local scale height of molecule k. The quantity $n_k H_k \sec \chi$ represents the total number of absorbing molecules between the height, z and the sun for a plane earth. However, for solar zenith angles greater than 85° it is necessary to take into account the curvature of the earth. This has been considered in various calculations for the sunrise effect in the ionosphere. Chapman (1931) has derived a function which can be used in place of sec χ in equation 4 provided that the scale height is constant. The peak altitude for the production of ion pairs can be obtained by differentiating (3). For an overhead sun this occurs for the condition

$$\tau = 1 + \beta \qquad (5)$$

where $\beta = dH/dz$ is the gradient of the scale height H_k. Since β is small at the height levels of interest, the altitude of the electron production peak is approximately the altitude of unit optical depth which is shown in figure 1.

Both satellite and rocket observations have been made of solar X-rays in the 2 to 8A region. Measurements employing rocket-borne Geiger counters sensitive to the total intensity in this wavelength range, have been conducted for several years by the U.S. Naval Research Laboratory, (Friedman (1962)). The results have demonstrated that the flux varies considerably depending on the degree of solar disturbance. Utilizing this data Nicolet and Aikin (1960) divided the 2 to 8A range into 3 wavelength bands centered at 2, 4, and 6A. Their results are given in table 1 for various solar conditions ranging from solar sunspot minimum to class 3 solar flares. Measurements of the spectral distribution in the 4 to 8A range have been carried out on rockets by Pounds and Sanford (1962) and on the Ariel I satellite by Pounds, et al (1963). The results of these measurements are in agreement with table 1.

TABLE 1.—*Variations in the X-Ray Intensities According to Various Solar Conditions*
Energies in erg cm^{-2} sec^{-1}

Condition of the Sun	2A	4A	6A
Completely Quiet	10^{-8}	10^{-7}	10^{-6}
Solar Minimum			
Quiet	10^{-7}	10^{-6}	10^{-5}
Lightly Disturbed	10^{-6}	10^{-5}	10^{-4}
Solar Maximum			
Disturbed	10^{-5}	10^{-4}	10^{-3}
Special Events	10^{-4}	10^{-3}	10^{-2}
Flares			
Class 3 Flares	10^{-3}	10^{-2}	10^{-1}

The cause of sudden ionospheric disturbances was at one time attributed to enhancement of hydrogen Lyman alpha radiation (1216A). However, a rocket experiment by Chubb et al (1957), during the course of a flare, established that although the intensity of 2 to 8A X-rays was enhanced, no appreciable change occurred in the Lyman alpha flux. Satellite observations have shown that the 2 to 8A flux enhancement is quite different from one event to another. Furthermore the enhancement of the X-ray flux may last much longer than the optical portion of the flare and take hours to decay to its preflare value. Also increases in the short wavelength range below 2A have been observed coincident with radio noise bursts at the commencement of a strong flare by various workers including Peterson and Winckler (1959). Nicolet and Aikin (1960) have shown that the radiation below 2A will be important only if its ionizing effect is greater than the effect of cosmic rays, which are the normal source of ionization below 70 km.

Since the X-ray flux emitted by the sun at sunspot minimum is insufficient to cause any appreciable ionization, another mechanism must be found to explain the formation of the D region during quiet sun conditions. The hypothesis that the D region might be formed by the ionization of nitric oxide by Lyman alpha radiation was postulated by Nicolet (1945). It was shown in

section II that Lyman alpha is absorbed in the *D* region. Watanabe (1954) has measured the ionization cross section of nitric oxide at 1216A and found it to be 2×10^{-18}cm^2. It then remains to establish the presence of a sufficient quantity of nitric oxide in the mesosphere.

Nitric oxide is a minor constituent of the mesosphere. The most important processes determining the nitric oxide distribution involve atomic nitrogen, which is produced by dissociative recombination of N_2^+ and NO^+ (see section VI). When the concentration of atomic nitrogen is sufficiently great, the density of NO will be determined by the reactions of N with O_2 and NO namely

$$N+O_2 \rightarrow NO+O \qquad (6)$$

$$N+NO \rightarrow N_2+O \qquad (7)$$

The ratio of the rate coefficients of these reactions has been measured by Clyne and Thrush (1961) and found to be 0.28 exp(−3450/T), where T is the gas temperature in °K. The density of nitric oxide is then

$$n(NO)=0.28n(O_2)\exp(-3450/T) \qquad (8)$$

Barth (private communication) has detected nitric oxide bands in the dayglow spectrum. Nitric oxide concentrations at least as large as those given by (8) can be deduced from the intensity of the dayglow. It has been demonstrated by Nicolet and Aikin (1960) that even concentrations of the order of 10^{-10} of the total number density are sufficient to account for *D* region ionization in the absence of X-rays. Inn (1961 a and b) has suggested that the photoionization of vibrationally excited O_2 by Lyman alpha may contribute significantly to the ion density of the *D* region.

IV. THE COSMIC RAY PRODUCTION FUNCTION

Galactic cosmic radiation is also an important factor in the formation of the *D* region. The ionizing effect of cosmic rays was first considered by Nicolet (1958). If $q_0(\phi)$ is the ionization rate for $n_0=2.6\times10^{19}$ molecules cm^{-3} at geomagnetic latitude ϕ, then the ionization rate for a number density n is

$$q(\phi)=q_0(\phi)n/n_0 \text{ cm}^{-3}\text{ sec}^{-1} \qquad (9)$$

Van Allen (1952) has shown that $q_0(\phi)$ is reduced by a factor of 10 at the geomagnetic equator compared to a latitude of 70° namely between 30 and 300 ion pairs cm^{-3} atm^{-1} sec^{-1}. The time variation of cosmic rays has been summarized by Webber (1962). The important factor to bear in mind is that $q_0(\phi,z)$ is 50 percent larger at sunspot minimum compared to sunspot maximum. Figure 2 shows the results of a calculation of the ion pair production function $q(\phi,z)$ for a $q_0(\phi)$ equal to 269 ion pairs cm^{-3} atm^{-1} sec^{-1}.

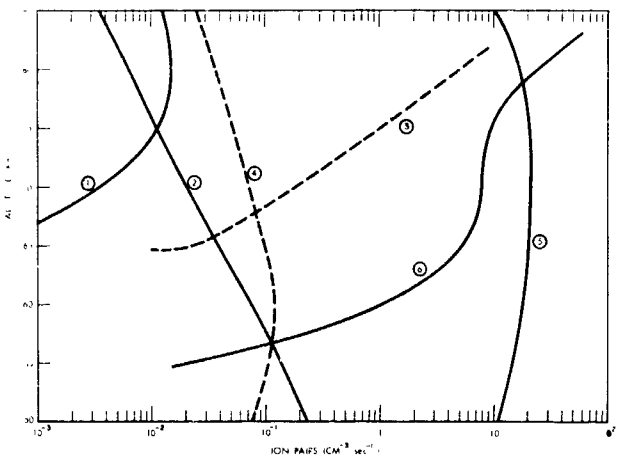


FIGURE 2.—Ion-pair production functions for different sources of ionization.
(1) Solar Lyman alpha ionizing nitric oxide.
(2) Galactic cosmic rays.
(3) Electrons precipitated in the auroral zone.
(4) Bremsstrahlung due to electrons precipitated in the auroral zone.
(5) Polar cap protons.
(6) Enhanced x-ray emission during a solar flare.

V. IONIZATION BY LOW ENERGY ELECTRONS AND PROTONS

The number of ion-pairs, q_β, cm^{-3} sec^{-1} produced at a height, z, by charged particles with an energy distribution between E_{min} and E_{max} may be expressed as

$$q_\beta=\frac{\rho(z)}{0.032}\int_{E_{min}}^{E_{max}} K_e(E)i(E,z)dE \qquad (10)$$

where $\rho(z)$ is the atmospheric density in gm cm^{-3} at altitude z. The formula for energy loss by

ionization, $K_e(E)$, has been approximated by Maeda and Singer (1961) to be

$$K_e(\mathrm{E}) = \frac{-dE}{dR} = 0.38E^{-0.96} \qquad (11)$$

for electrons and

$$K_p(\mathrm{E}) = \frac{-dE}{dR} = 236E^{-0.78} \qquad (12)$$

for protons, where R is the range in gm/cm^2 and E is expressed in Mev. The number of charged particles of energy, E, arriving at a height z is $i(E,z)$ cm^{-2} sec^{-1} kev^{-1}. It can be found if the incident spectrum of the particles is known. Since the particles enter the atmosphere at various angles with respect to the earth's magnetic field and have a spatial distribution, it is necessary to calculate $i(E,z)$ by integrating over both the pitch and solid angles.

Maeda (1963) has shown that

$$i(E,\ z) = \int_0^{\theta} \int_0^{\pi/2} j(E)\ \exp\left[-\frac{x}{\lambda(E)\cos\alpha}\right] \sin\alpha d\alpha d\theta \qquad (13)$$

where θ is the solid angle, α is the pitch angle, x is the atmospheric depth at height z, and $\lambda(E) = (3.15)10^{-7}E^{2.2}$ is the attenuation mean free path. For the case of an auroral absorption event a possible representation of the differential energy spectrum is $j(E) = (1.6)10^{12}E^{-5.2}$ particles cm^{-2} sec^{-1} ster^{-1} kev^{-1}. This type of spectrum and flux of particles has been observed by McIlwain (1960). However, there were no accompanying measurements of electron density in the D region. The production function q_β for this type of input spectrum is illustrated in figure 2.

Aikin and Maier (1963) have calculated the ionizing effect of bremsstrahlung produced by the electron differential energy spectrum described above. The cross section per atom for production of photons of energy hν by an electron of energy E is

$$\phi(E,\ h\nu)d\nu = (8/3)\alpha\ \rho_0^2 Z^2 mc^2 \frac{1}{\nu E} \ln\frac{\{\sqrt{E}+\sqrt{E-h\nu}\}^2}{h\nu} d\nu \qquad (14)$$

for energies less than 510 kev (Heitler, 1954). Here α represents the fine structure constant, ρ_0 the classical radius of the electron and Z the charge of the nucleus. The differential energy spectrum of photons $dQ(h\nu)$ emitted at an atmospheric depth x in a thickness dx is

$$dQ(h\nu)d\nu = K\int_{E=h\nu}^{\infty} \phi(E_{h\nu},h\nu)i(E,x)dE\ dxd\nu \qquad (15)$$

where K is the number of atoms per gram. Note that only electrons of energy $E \geq h\nu$ contribute to the photon flux at energy $h\nu$. The resulting ion-pair production function is also illustrated in figure 2.

Although most auroral absorption events are caused by electrons, accompanying proton fluxes have also been observed. Large solar flares are sources of large fluxes of protons and heavier particles, which produce polar cap absorption events. These events have been discussed by Bailey (1957, 1959) and Webber (1962). Calculations of the type discussed above for electrons can be carried out to estimate the ionizing effect of solar flare protons incident on the atmosphere. Reid (1961) has performed such a calculation for a power law differential energy spectrum of the form $j(E) = K\,E^{-4}$. However, Webber (1962) finds a better fit to the experimental data if use is made of an exponential spectrum of the form

$$i(P) = Ke^{-P/P_0} \qquad (16)$$

where P is the rigidity of the particles. P_0 may range from 50 to 200 Mv, depending on the geomagnetic latitude at which the ionization is observed. For instance observations of the absorption of cosmic noise for latitudes greater than 70° require a rather steep particle spectrum or a $P_0 < 100$ Mv. For geomagnetic latitudes of 65° or less a $P_0 > 150$ Mv is required. This requirement can also be met by introducing a geomagnetic cutoff into a power law spectrum. Curve 5 of figure 2 is taken from Webber (1962) and illustrates the electron production function for an integral intensity of solar protons of 0.05 particles/cm^2 ster sec for energies greater than 100 Mev. A rigidity spectrum is assumed and P_0 is taken to be 50 Mv. More recently Webber and Freier (1963) have shown that α particles should be included in the calculation of the production function for polar cap events.

VI. LOSS PROCESSES

The distribution of ions and electrons in the *D* region depends on the interactions of these particles with one another as in the case of recombination and on interactions with constituents of the neutral atmosphere. In the latter instance, negative ions may be formed through the process of electron attachment, also charge transfer and ion-atom interchange reactions modify the distribution of various species of ions. At the present time the problem is complicated by a lack of experimental data. Although some measurements of positive ion density have been reported, the technique is presently in the development stage. No data are available on the species of positive and negative ions found in the mesosphere. The problem is further complicated by the difficulty of duplicating lower ionospheric conditions in laboratory experiments for the determination of reaction cross sections. There are a large number of possible reactions, but only those which are currently in use will be discussed here.

Dissociative recombination involves an electron and ion through the reaction

$$XY^+ + e \rightarrow XY^* \rightarrow X' + Y' \qquad (17)$$

where * and ′ denote excited states of the molecule and atoms respectively. The three ions of importance in the *D* region are O_2^+, N_2^+, and NO^+. The rates of dissociative recombination of O_2^+ and N_2^+ have been measured by Kasner, et al (1961), who employed a mass spectrometer for the identification of ions. Values of $\alpha_D(O_2^+) = (1.7 \pm 1)10^{-7}$ cm^3 sec^{-1} and $\alpha_D(N_2^+) = (3 \pm 1)10^{-7}$ cm^3 sec^{-1} were reported for a nitrogen partial pressure $\lesssim 1$ mmHg at 300°K. Although the rate coefficient for dissociative recombination is temperature dependent, see Bates and Dalgarno (1962), the temperature employed in the laboratory measurements is close enough to the mesospheric temperature of about 200°K that this dependence can be ignored.

The measurements of the dissociative recombination coefficient $\alpha_D(NO^+)$ are not so well developed as those for O_2^+ and N_2^+. Guntan and Inn (1961) have reported a value of $(1.3)10^{-6}$ cm^{3}sec^{-1} and Doering and Mahan (1962) a value of $(3^{+17}_{-2})10^{-7}$cm^{3}sec^{-1}. Since both these measurements were conducted at partial pressures of NO > 0.1 *mmHg*, where the formation of complex ions is possible and no mass spectrometer measurements were made of the ions, there is some doubt as to whether these measurements really refer to $\alpha_D(NO^+)$. Ionospheric experiments tend to support a slower rate of recombination and indicate that $\alpha_D(NO^+)$ should be less than either $\alpha_D(O_2^+)$ or $\alpha_D(N_2^+)$. Nicolet and Aikin (1960) employed the value of (3) 10^{-9} cm^3 sec^{-1} for $\alpha_D(NO^+)$ but revised estimates of the nitric oxide densities and measurements of electron densities together with solar radiations would tend to support a rate in the neighborhood of $1-5\ 10^{-8}$ cm^3 sec^{-1}, (Aikin, et al, (1963)).

Gali, et al (1963) have investigated the ion-molecule reactions

$$N_2^+ + O_2 \rightarrow NO^+ + NO(\gamma) \qquad (18)$$

and

$$O_2^+ + N_2 \rightarrow NO^+ + NO(\gamma_2) \qquad (19)$$

where the rate coefficients γ_1 and γ_2 were assigned an upper limit of $(2.1)10^{-13}$cm^3/sec for both these reactions. The ratio of the neutral particle to electron density in the *D* region is such that these would be important reactions for rate coefficients as low as 10^{-18} cm^2/sec. The net result of these processes is to make NO^+ the dominant species of positive ion in the *D* region. Therefore, an in situ ion mass spectrometer measurement cannot be used to establish the validity of the Lyman alpha hypothesis for the formation of the quiet sun *D* layer. The reaction

$$N_2^+ + O_2 \rightarrow O_2^+ + N_2(\gamma^3) \qquad (20)$$

is also energetically possible and may have a rate coefficient as high as 10^{-16} cm^3 sec^{-1}.

Electrons can attach themselves to oxygen molecules through the three-body process

$$e + O_2 + O_2 \rightarrow O_2^- + O_2 \qquad (21)$$

to form negative ions of O_2. Chanin, et al (1962) have measured the rate of this reaction. At *D* region temperatures a value of $a(O_2) = 1.5 \times 10^{-30}$cm^6 sec^{-1} is found. Burch, et al (1958) have studied the rate of photodetachment of electrons from O_2 as a function of wavelength of the radiation. They estimate that for the solar spectrum a photodetachment coefficient of

$$d(O_2) = 0.44 \text{ sec}^{-1}$$

is applicable for O_2^-. Branscomb, et al (1958) have made a similar study for atomic oxygen from which they deduce a radiative attachment coefficient, $a(O)$, of electrons to atomic oxygen of $(1.3)10^{-15}$ cm^3 sec^{-1} and a photodetachment rate, $d(O)$, of 1.4 sec^{-1}.

The problem of the removal of negative ions in the absence of photodetachment is unclear at the present time. Phelps and Pack (1961) have measured the rate of collisional detachment of O_2^- in the presence of O_2 and find a rate coefficient, η, of (4) 10^{-20} cm^3 sec^{-1}. It was also found that N_2 was ineffective as a detaching agent.

Associative detachment of the form

$$O_2^- + O \rightarrow O^3 + e \qquad (22)$$

has also been proposed as a possible mechanism for the removal of negative ions. Possible values of the rate coefficient, ρ, of this reaction range from 10^{-14} cm^3 sec^{-1} to 10^{-10} cm^3 sec^{-1}. A similar process involving O^- may take place.

The remaining process to be considered is mutual neutralization of a positive and negative ion. This process is of the form

$$XY^+ + WZ^- \rightarrow XY + WZ \qquad (23)$$

The rate coefficient, α_i, has been measured in the laboratory for the mutual neutralization of NO^+ and NO_2^- by Mahan and Person (1964). A value of (2.1) 10^{-7} cm^3 sec^{-1} was obtained. At altitudes below 45 km the Thompson three-body process must be considered. This reaction is of the form

$$XY^+ + WZ^- + M \rightarrow XY + WZ + M \qquad (24)$$

and has a rate coefficient of

$$\alpha_T = 8 \times 10^{-3} p / T^{5/2} \text{cm}^3\text{sec}^{-1}$$

here p is the pressure in *mmHg* and T the temperature in °*K*.

The continuity equations governing the altitude distribution of each species of ion and electrons in the *D* region can be written so that the processes discussed above are included together with the appropriate ion-pair production function. Such a set of simultaneous differential equations can be solved with the aid of a computer, however, usually the approximation is made that $dN/d^3 = 9$, i.e., that quasi-equilibrium exists throughout the *D* region.

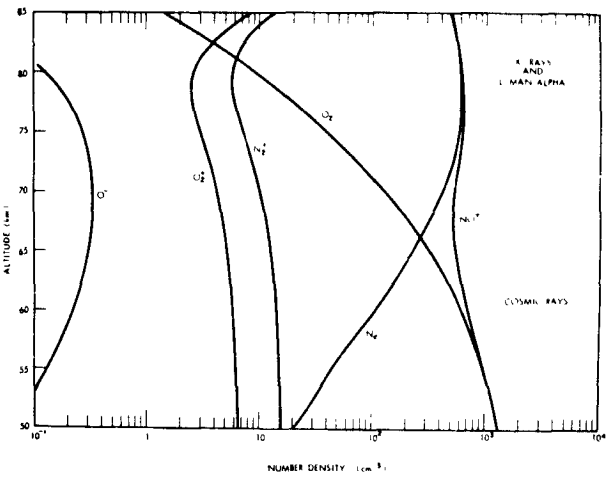


FIGURE 3.—Ion and electron density distributions for the rate coefficients listed in table II.

The results of an exact solution of such a set of equations are illustrated in figure 3 for a geographic latitude of 38°. Here the relative importance of Lyman alpha, cosmic rays, and x-radiation in the formation of the *D* region under conditions of a completely quiet sun is shown for a solar zenith angle of 55°. The effect of reactions (18) through (20) in changing N_2^+ and O_2^+ into NO^+ when the rate coefficients are 10^{-18} cm^3 sec^{-1} is clearly demonstrated. This will enhance the electron density, since the dissociative recombination coefficient of NO^+ has been taken to be almost an order of magnitude below that of O_2^+ and N_2^+. Table 2 summarizes the numerical factors employed in the calculation.

TABLE 2.—*Rate Coefficients Employed in the Calculations for Figure 3*

Symbol	*Rate*
αDO_2	1.7×10^{-7}cm^3/sec
αDN_2	2.8×10^{-7}cm^3/sec
αDNO	3×10^{-5}cm^3/sec
γ_1	10^{-18}cm^3/sec
γ_2	10^{-18}cm^3/sec
γ_3	10^{-18}cm^3/sec
$a(O_2)$	1.5×10^{-30}cm^6/sec
$d(O_2)$	0.44 sec^{-1}
$a(O)$	1.3×10^{-15}cm^3/sec
$d(O)$	1.4 sec^{-1}
η	4×10^{-20}cm^3/sec
ρ	10^{-13}cm^3/sec
α_i	10^{-7}cm^3/sec

The nitric oxide density is 10^{-10}n(M).

VII. THE NOCTURNAL *D* REGION AND THE SUNRISE EFFECT

In the absence of ionization sources other than cosmic rays, the *D* region electron density decays to low values at night. Thus, it is extremely difficult to obtain electron density measurements during this period. However, observations are possible in the auroral zone with the aid of cosmic noise absorption techniques. This has led to a comparison of the sunlit and dark *D* regions under the action of the same flux of ionizing particles. In this way the different processes affecting the negative ion distribution can be examined. Let us first discuss the behavior of the undisturbed *D* region at sunrise.

A good deal of information concerning the electron density distribution has been determined from the phase and amplitude of 16 kc/sec waves reflected from the ionosphere at near vertical incidence. Bracewell (1952) has noted that the amplitude of such a wave begins to decrease at a time when the sun's rays grazing the earth illuminate an area 60 km above the ground. There is no corresponding decrease in phase at this time. Rather a decrease in phase path begins just before ground sunrise with the major portion of the drop occurring after ground sunrise, the total change in height amounting to 15 km. These results were interpreted by Bracewell and Bain (1952) as being due to a double-layered *D* region.

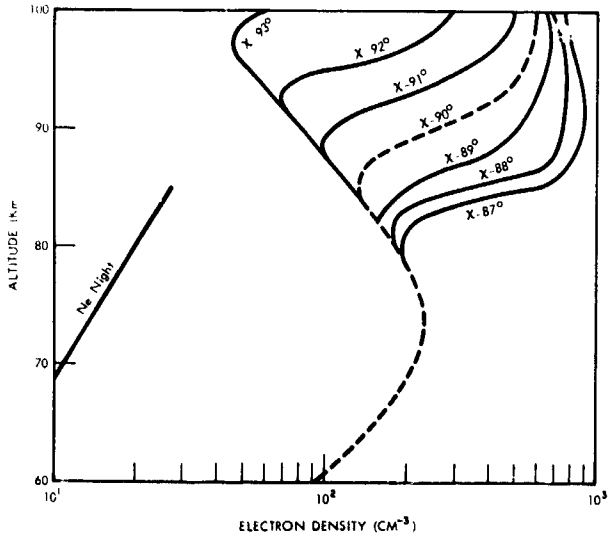


FIGURE 4.—The electron density distribution at sunrise for different solar zenith angles and the nocturnal electron density distribution.

The origin of the double layer lies in the ability of solar radiation in the visible wavelength region to penetrate to the lower *D* region and photodetach electrons from O_2^- ions, which are produced as the result of cosmic ray ionization. This layer centered around 73 km will thus appear before Lyman alpha and X-rays have had a chance to penetrate into the *D* region. As the solar zenith angle decreases these radiations produce a second *D* region above the cosmic ray layer thus forming a two-layered *D* region. Eventually the two layers merge to form the normal daytime *D* region. The theoretical variation of the electron density distribution at sunrise is illustrated in figure 4 which is taken from Aikin (1962).

As noted previously most *D* region calculations are performed using the equilibrium solutions of the continuity equations for electrons and the various species of ions. This has led to the definition of the negative ion to electron ratio λ. During the day, λ can be approximated as the ratio of the rate of attachment to photodetachment which for O_2^- is

$$\lambda_{O_2^-} = (3.4)10^{-30}[n(O_2)]^2 \qquad (25)$$

and for O^-

$$\lambda_{O^-} = 10^{-15}n(O) \qquad (26)$$

Even in the upper *D* region where atomic oxygen reaches the peak of its concentration, λ_{O^-} is less than 10^{-2}. On the other hand $\lambda_{O_2^-}$ is equal to 1 near 70 km so that O_2^- is the primary negative ion in the *D* region as shown in figure 3.

The diurnal variation of cosmic noise absorption during auroral and polar cap events has recently been used to derive information on the rate coefficients affecting the ratio λ. For a constant flux of ionizing particles it has been demonstrated (see e.g., Bailey (1957)), Hultquist (1963), Webber (1962), that the ratio of day to night absorption $A_D/A_N = 1+\lambda_N/1+\lambda_D$ where λ_N is the night average value of λ at the height where the absorption occurs and λ_D is the day average for the same altitude range. This expression is valid only for the case where the absorption occurs in a relative narrow altitude range of about 5 km. The more general case has been considered by Reid (1961).

Because the fluxes of auroral particles are highly variable, a statistical analysis must be employed to find the ratio A_D/A_N for auroral absorption events. Holt and Landmark (1963) deduce a value of 2, while Hultquist (1962) deduced yearly averages between 0.89 and 1.38 with a 4 year average of 1.16.

Polar cap absorption events yield higher values of A_D/A_N than auroral absorption observations. Webber (1962) has summarized riometer data from a large number of polar stations and finds that for geomagnetic latitudes greater than 70° $A_D/A_N \cong 6.5 \pm 1$ and for 65° or less $A_D/A_N \cong 5 \pm 1$. The difference in the two cases being accounted for in the rigidity and therefore atmospheric penetration of the protons. Since the protons penetrate to greater atmospheric depths than auroral electrons, the cosmic noise absorption during polar cap events occurs at altitudes where λ is more of a determining factor for the absorption. Various authors have shown that collisional detachment values must exceed the laboratory estimates by 3 orders of magnitude to explain the observations. Whitten and Poppoff (1962) have suggested that associative detachment may replace collisional detachment as the negative ion removal process during darkness.

The variation of polar cap absorption events during twilight periods has given rise to speculation that a negative ion other than O_2^- is found in the *D* region. It has been observed that nighttime values of absorption are reached despite the fact that the sun did not set at heights greater than 50 km (see for example Reid and Collins (1959), and Eriksen, et al (1960)). These observations led Reid and Leinbach (1962) and Eriksen et al (1960) to suggest that the dominant negative ion during twilight was not O_2^- but rather an ion with a much larger electron affinity requiring ultraviolet light for photodetachment.

Clearly the explanation of the diurnal variation of auroral and polar cap absorption depends on a more exact knowledge of the particle spectrum and species of negative ion as well as the inclusion of solar radiations in the calculations of cosmic noise absorption. There is, however, laboratory evidence for the existence of a negative ion with a high value of electron affinity. Curran (1962) has demonstrated that the electron affinity of NO_2^- is greater than 3.82 ev. For photodetachment to occur by solar radiation whose wavelength range is in the near ultraviolet an electron affinity of about 4 ev is required. Thus NO_2^- is a possible candidate. However, since the NO_2 concentration is predicted to be only a fraction of the nitric oxide concentration and Curran (1962) has found that direct attachment of electrons to No_2 is unimportant, charge exchange processes involving O^- and O_2^- must be operative in order to form an appreciable amount of NO_2^-. Curran (1962) has given a rate coefficient of 10^{-9} to 10^{-10}cm^3/sec for the formation of NO_2^- by

$$O^- + NO_2 - NO_2^- + O \qquad (27)$$

Inasmuch as O^- is thought to be a negative ion of minor importance, the reactions

$$O_2^- + O \rightarrow O^- + O_2 \qquad (28)$$

and

$$O_2^- + NO_2 \rightarrow NO_2^- + O_2 \qquad (29)$$

should also be considered.

The theory of formation of the D region is by no means complete. While the sources of ionization and their efficiency of ion pair production have been reasonably well established, the chemistry of a nitrogen-oxygen system under the influence of different radiations is not well understood. A great deal more investigation is required in order to establish the role of the various reactions discussed here. In situ measurements are becoming increasingly more important not only because of the promise of measuring difficult parameters such as negative ions, but also because the necessary s multaneous determination of several parameters can be effected. Such phenomena as auroral absorption events and ionization by X-ray enhancements during the course of solar flares cannot be adequately explained without recourse to such experiments.

REFERENCES

AIKIN, A. C. in *Electron Density Profiles in the Ionosphere and Exosphere*, Ed. by B. Machlum Pergamon Press (1962).

AIKIN, A. C., Kane, J. A. and TROIM, J., Space Research IV (to be published) (1964).

AIKIN, A. C. and MAIER, E. J., to be published in Proceedings of NATO Advanced Study Institute, Skeikampen, Norway (1963).

BAILEY, D. K., J. Geophys. Res. **62**, 431, (1957).
BAILEY, D. K., Proc. of the I.R.E. **47**, No. 2, 255, (1959)
BARTH, C. A., Private communication.
BATES, D. R. and DALGARNO, A. in *Atomic and Molecular Processes*, Ed. by D. R. Bates, Academic Press (1962).
BETHE, H. and ASHKIN, in *Experimental Nuclear Physics*, Volume 1, Ed. by Segré John Weiley and Sons (1951).
BIONDI, M. A. in *Advances in Electronics and Electron Physics*, Vol. 18, Ed. by L. Marton Academic Press (1963).
BRACEWELL, R. N., Proc. I.E.E., **99** III, 217 (1952)
BRACEWELL, R. N. and BAINE, W. C., J. Atmos. and Terr. Phys. **2**, 216, (1952).
BRANSCOMB, L. M., BURCH, D. S., SMITH, S. J. and GELTMAN, S., Phys. Rev., **111**, 504, (1958).
BURCH, D. S., SMITH, S. J. and BRANSCOMB, L. M., Phys. Rev. **112**, 171, (1958).
CHANIN. L. M., PHELPS, A. V. and BIONDI, M. A., Phys. Rev. **128**, No. 1, 219, (1962).
CHAPMAN, S., Proc. Roy. Soc. **43**, 484, (1931).
CHUBB, T. A., FRIEDMAN, H., KREPLIN, R. W. and KUPPERIAN, JR., J. E., J. Geophys. Res. **62**, 389 (1957).
CLYNE, M. A. A. and THRUSH, B. A., Proc. of the Royal Soc. *A*, 261, #1305, 259, (1961).
CURREN, R. K., Phys. Rec. **125**, No. 3, 910, (1962).
DOERING, J. P. and MAHAN, B. H., J. Chem. Phys. **36**, 669, (1962).
ERIKSEN, K., HOLT, O. and LANDMARK, B., J. Atm. and Terr. Phys. **18**, 78, (1960).
*FRIEDMAN, H. in Reports on Progress in Physics, Vol. XXV, Ed. by A. C. Stickland, The Institute of Physics and the Physical Society (1962).
GALLI, A., GEARDINE-GUIDONI, A. and VOLPI, G. G., J. Chem. Phys. **30**, 518, (1963).
GRODSTEIN, G. W., National Bureau of Standards Circular 533, (1957).
GUNTON, R. C. and INN, E. C. Y., J. Chem. Phys. **35**, 1896, (1961).
HEITLER, W., *The Quantum Theory of Radiation*, Oxford University Press, 3rd Edition, 1954.
HOLT, O. and LANDMARK, B. in *Radio Astronomical and and Satellite Studies of the Atmosphere*, Ed. by J. Aarons North Holland (1963).
HULTQUIST, B., Plan. and Space Sci. **11**, 371 (1963).
HULTQUIST, B. J., Atm. & Terr. Phys. **25**, 225, (1963).
INN, E. C. Y., Plant. Space Sci., **5**, 70, (1961).
INN, E. C. Y., Plant. Space Sci., **8**, 200, (1961).
KASNER, W. H. ROGERS, and BIONDI, M.A., Phys. Rev. Let. **1**, 321, (1961).
KATZ, L. and PENFOLD, A. S., Rev. Mod. Phys., **24**, 28, (1952).
MAEDA, K. J. Geophys. Res. **68**, No. 1, 185 (1963).
MAEDA, K. and SINGER, S. F., Arkif För Geofysik **21**, No. 21, 531, (1960).
MAHAN, B. H. and PERSON, J. C., J. Chem. Phys., **40** No. 2, 392, (1964).
MCILWAIN, C. E., J. Geophys. Res. **65**, 2727, (1960).
NICOLET, M., Mem. Inst. Meteor. Belg. No. 19, 83, (1945).
NICOLET, M., Penn State University Ionospheric Res. Lab. Sci. Report No. 102 (1958).
NICOLET, M. and AIKIN, A. C., J. Geophys. Res. **65**, 1469 (1960).
NICOLET M., to be published in Handbuch der Physik Springer-Verlag.
PETERSON, L. E. and WINCKLER, J. R., J. Geophys. Res. **64**, 667, (1959).
PHELPS, A. V. and PACK, J. L., Phys. Rev. Lett., **6**, 111, (1961).
†POPPOFF, I. G. and WHITTEN, R. C., to be published J. of Atmospheric Sciences (1964).
POUNDS, K. A. and SANFORD, P. W., "Proceedings of the International Conference on the Ionosphere," held at Imperial College, London, pp. 50–56 (1962).
POUNDS, K. A. WILMORE, A. P., BOWEN, P. J., NORMAN, K. and SANFORD, P. W., to be published Proceedings of the Royal Society.
REID, G. C., J. Geophys. Res. **66**, No. 12 (1961).
REID, G. C. and COLLINS, C., J. Atm. & Terr. Phys. **14**, 63, (1959).
REID, G. C. and LEINBACH, H., J. Atm. and Terr. Phys. Vol. 23, (1962).
VAN ALLEN, J. A., *Physics and Medicine of the Upper Atmosphere*, Edited by C. S. White and O. O. Benson, Jr., University of New Mexico Press, (1952).
WATANABE, K., Cham. Phys., **22**, 1564, (1954).
WEBBER, W. R., *Progress in Elementary Particle and Cosmic Ray Physics*, Volume VI, Edited by J. G. Wilson and S. A. Wouthuysen North Holland Publishing Company (1962).
WEBBER, W. R., J. Geophys. Res. **67**, 5091, (1962).
WEBBER, W. R. and FREIER, P. S., J. Geophys. Res. **68**, No. 23, 6223, (1963).
WHITTEN, R. C. and POPPOFF, I. G., J. Geophys. Res. **67**, 1183, (1962).

*This is an excellent review of rocket and satellite measurements of the solar spectrum.

†A more detailed discussion of loss processes is given in this article.

SOME RESULTS OF ROCKET EXPERIMENTS IN THE QUIET *D* REGION*

A. C. AIKIN, J. A. KANE AND J. TROIM†

Two electron density profiles have been obtained for the quiet day mid-latitude *D* region by means of ground to rocket radio propagation experiments. These profiles are interpreted between 85 and 70 km in terms of solar Lyman alpha ionization of nitric oxide. The role of 2 to 10 Ångstrom X-rays is shown to be unimportant to the formation of the *D* region for the quiet solar conditions present during the two rocket flights. The electron collision frequency in the *D* region is shown to be subject to significant variations which are correlated to pressure varia- of the stratosphere.

INTRODUCTION

A theory of the formation of the *D* region was proposed by Nicolet and Aikin in 1960. The sources of *D* region ionizations under conditions of a quiet sun were explained by solar Lyman alpha ionizing nitric oxide in the altitude region between 85 and 70 km, while ionization below 70 km was attributed to cosmic rays. In the present paper, the results of two rocket experiments are presented together with an interpretation of these results which is consistent with that theory.

Measurements of the atmospheric attenuation of Lyman alpha flux as a function of altitude are used to determine the electron collision frequency profile for each flight. It is shown that the collision frequency is subject to significant variations which are correlated with meteorological processes occurring at the 30 km level of the stratosphere. With a knowledge of the collision frequency, electron density profiles are deduced from the results of radio propagation experiments. Based on satellite data the contribution of 2 to 8 Ångstrom solar X-rays to the ionization of the *D* region is estimated. Finally a discussion of the different nitric oxide data available since 1910 is given in terms of the effect on the *D* region recombination coefficient.

The experimental method is based upon the Nike-Apache sounding rocket. The results from two such rocket flights at Wallops Island, Virginia, latitude 38°, are reported here. The first, labeled 14.107, was fired at 1430 hours EST on March 8, 1963, and the second, 14.108, was fired at 1530 hours EST on April 9, 1963. The solar zenith angles were 53° and 55° respectively.

LYMAN ALPHA FLUX

On both flights Lyman alpha flux was measured by means of a lithium fluoride window ionization chamber filled with carbon disulfide gas. For Q_∞, the Lyman alpha flux incident upon the earth's atmosphere, a value of $(3\pm1)\times10^{11}$ photons cm^{-2} sec^{-1} was obtained. This value is consistent with the results of Friedman *et al.*, (1963) who have shown that the Lyman alpha flux is not strongly dependent upon solar activity.

PRESSURE AND ELECTRON COLLISION FREQUENCY DETERMINATION

The Lyman alpha flux $Q(z)$ that penetrates the atmosphere to an altitude z is given by

$$Q(z)=Q(\infty)e^{-\sigma\int_z^\infty N(z)dz} \qquad (1)$$

where, assuming O_2 to be the dominant source of absorption, σ is the absorption cross section of an O_2 molecule and $N(z)$ is the O_2 number density.

*Published as *Goddard Space Flight Center Document X-615-64-173*, June 1964.

†Norwegian Defense Research Establishment, Kjeller, Norway.

Since the partial pressure of O_2 at an altitude z is defined as

$$PO_2(z) = \int_z^\infty N(z) M g dz \qquad (2)$$

equation (1) can be rewritten as

$$Q(z) = Q(\infty) e^{-\frac{\sigma PO_2(z)}{Mg}} \qquad (3)$$

where M is the mass of an O_2 molecule and g is the gravitational acceleration. A measurement of the ratio $Q(\infty)/Q(z)$ yields therefore the partial pressure of O_2 from which, assuming a constant ratio between oxygen and nitrogen partial pressures, the total pressure is derived.

The derived pressures can be used to obtain the collision frequency of electrons with neutral molecules. Phelps (1960) from laboratory measurements has shown that ν, the electron collision frequency in air for monoenergetic electrons, is proportional to the electron energy. From this and the gas law it follows that for monoenergetic electrons of energy kT, the collision frequency is proportional to the atmospheric pressure. The relationship is

$$[\nu]_{AIR} = 9 \times 10^7 p(\text{mm } Hg) \text{sec}^{-1} \qquad (4)$$

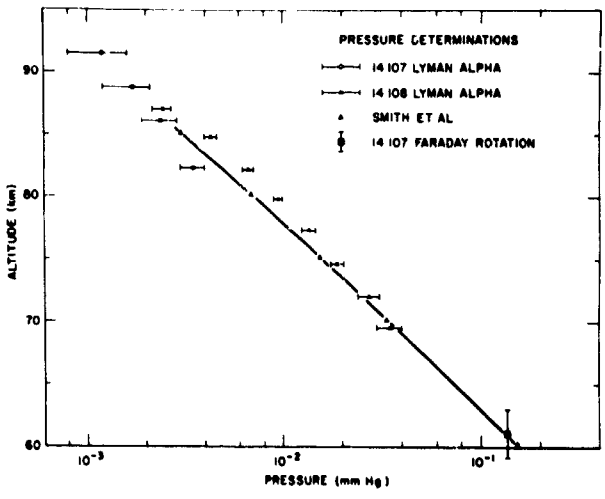


FIGURE 1.—Atmospheric pressure versus altitude at Wallops Island measured at 14:30 hours (14.107) and 19:00 hours (Smith et al.,), March 8, 1963, and at 15:30 hours, April 9, 1963 (14.108).

The results of the pressure determinations on flights 14.107 and 14.108 are shown on Figure 1. The resolution in the pressure measurements, indicated by the horizontal bars, is determined by the precision to which the Lyman alpha ion chamber aspect angle is known. Included in this figure is the pressure profile determined by Smith *et al.*, (1964) who performed a rocket-borne grenade experiment at Wallops Island five hours after flight 14.107. It is believed that the indicated pressure differences in these three measurements are real and indicate the presence of meteorological effects in the mesosphere.

METEOROLOGY

From Equation (4) of the previous section it can be expected that the D region collision frequency profile will be subject to the diurnal, seasonal and latitudinal variation of mesospheric pressure that the work of Stroud and Nordberg (1961) and others has revealed.

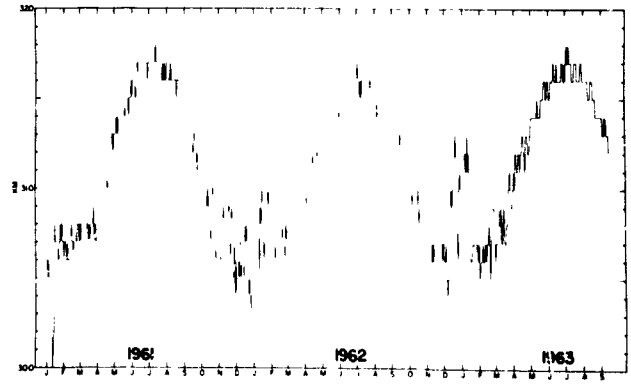


FIGURE 2.—Seasonal variation of the altitude of the 10 millibar pressure level at Wallops Island.

In Figure 2 is shown for Wallops Island a plot of the altitude of the 10 millibar pressure level for a three year period. In addition to the very apparent seasonal variation comparable short term variations are present, particularly during winter. The Meteorologische Abhandlunge Vol. XL, 1963, from which these data are taken shows that a sudden warming event occurred in the stratosphere over Canada on April 4. By April 9 the effects of this event could be noticed at Wallops Island where the altitude of the 10 millibar level increased from 30.80 km on March 8 to

31.16 km on April 9 while the temperature of this level increased from −47°C to −36°C. These stratospheric changes lend support to our April 9 collision frequency profile since the mesosphere is linked to the stratosphere through the pressure relation

$$p(z) = p_0 e^{-\int_0^z dz/H(z)} \quad (5)$$

where the scale height H(z) is proportional to the temperature T(z). Although no measurements of T(z) exist for 14.108 a reasonable model can be constructed which leads to mesospheric pressures of the required magnitude.

Convincing evidence for the existence of an ionosphere-stratosphere interaction has been given by Bossolosco and Elena (1963) who have shown that a strong correlation exists between the temperature at the 10 millibar level and mid-latitude winter radio wave absorption at frequencies around 2 Mc/s. Their results show a temperature increase of 20° at the 10 millibar level correlated with a factor 2 increase in absorption. Our results would seem to indicate that this effect, the so called winter absorption anomaly, is due simply to an increase in mesospheric pressure. However, since radio wave absorption involves the electron density as well as collision frequency, and since both the electron production and loss mechanisms are pressure dependent, such a conclusion must be withheld until quantitative calculations have been made.

ELECTRON DENSITY PROFILES

The electron density profiles reported here were obtained by means of radio propagation techniques employing ground to rocket transmissions. In flight 14.107 linearly polarized signals were transmitted from the ground at frequencies of 3.0 and 4.9 Mc/s. The mechanical spin of the rocket of about 3 cycles per second was used to rotate the rocket-borne receiving antennas through the polarization patterns of the arriving waves. The telemetered signal strengths exhibited a fading pattern, the frequency of which was the sum of the rocket spin frequency and the ionospheric Faraday rotation frequency. A comparison of the period of this fading pattern with the mechanical spin period, independently measured by means of a solar aspect sensor, yielded the Faraday rotation of the plane of polarization.

Under the condition of quasi-longitudinal propagation, the plane of linear polarization, defined by the angle ψ, rotates with rocket altitude z

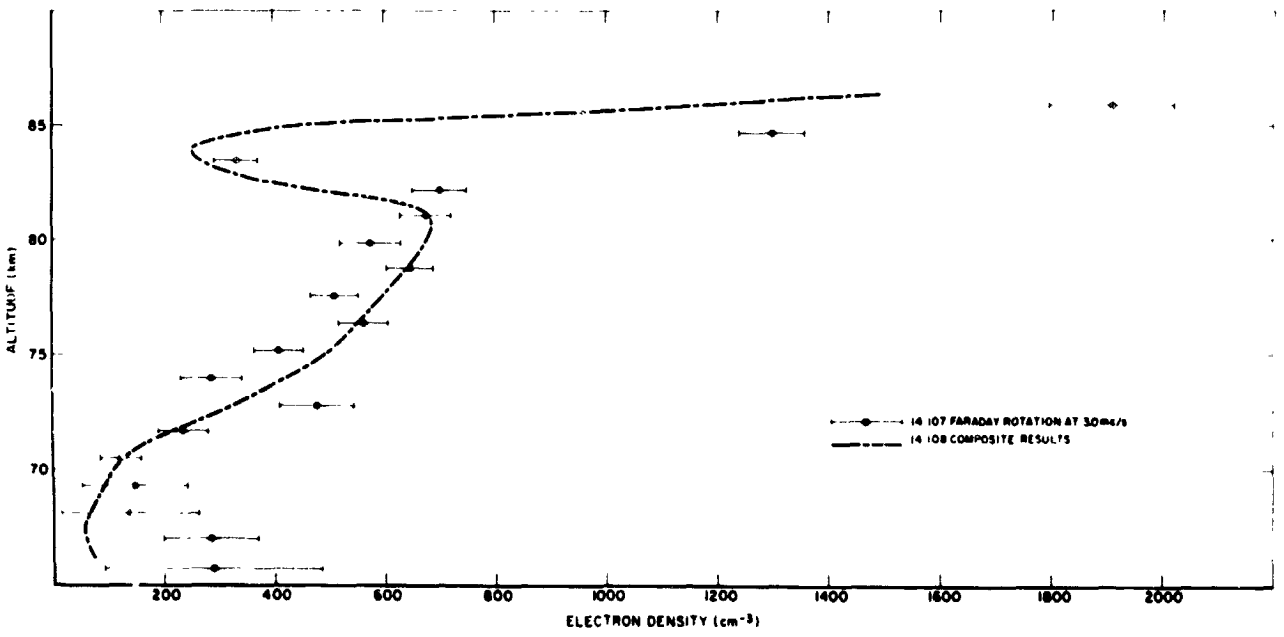


FIGURE 3.—Electron density versus altitude for March 8, 1963 (14.107). Dashed curve is a composite of the results shown in Figure 4.

according to an expression of the form

$$\frac{d\psi}{dz} = N_e(z)F(\omega, \omega_H, \nu(z)) \qquad (6)$$

where $N_e(z)$ is the electron density and F is a function of the exploring frequency ω, the gyrofrequency ω_H and the collision frequency ν. The explicit form of F involves the Dingle integrals of the generalized Appleton-Hartree formula (Sen and Wyler, 1960). Before $F(z)$ can be evaluated it is necessary to have a collision frequency model, the determination of which was described in section 2. The Faraday rotation experiment by itself allows the collision frequency to be determined at a single altitude. This follows from the fact that the function $F(\omega, \omega_H, \nu(z))$ changes sign at a unique value of ν/ω. By noting the altitude at which the reversal in the sense of the Faraday rotation occurred on flight 14.107 a value of $\nu = 13.8 \times 10^6 \text{sec}^{-1}$ was deduced for an altitude of 61 ± 2 km. As seen in Figure 1 this value is consistent with the March 8 pressure profile obtained from the grenade experiment of Smith *et* al., (1964).

The electron density results of the 14.107 Faraday rotation experiment at 3.0 Mc/s are shown in Figure 3. Each point is the average electron density in an altitude interval of approximately one kilometer. The horizontal bar indicates the probable uncertainty in the determination of this average value. This uncertainty is due to random echoes from above the D region distorting the Faraday pattern. This effect was more severe on the 4.9 Mc/s Faraday experiment which yielded electron densities in agreement with those shown in Figure 3, but with uncertainties two to three times as large.

In flight 14.108 linearly polarized signals were transmitted at frequencies of 1.8 and 4.9 Mc/s. A 3.0 Mc/s signal was also transmitted from ground to rocket on flight 14.108, but in this case the transmitted pattern was alternately switched between opposite circularly polarized modes, which in the ionosphere were differentially absorbed. Denoting the received signal strength of the two polarization modes as E_0 and E_X, the altitude variation of the logarithmic ratio $ln(E_0/E_X)$ can be expressed as

$$\frac{d}{dz}\ln(E_0/E_X) = N_e(z)G(\omega, \omega_H, \nu(z)) \qquad (7)$$

where again $N_e(z)$ is the electron density and $G(z)$ is an altitude dependent function involving the Dingle integrals and requiring a collision frequency model for explicit evaluation.

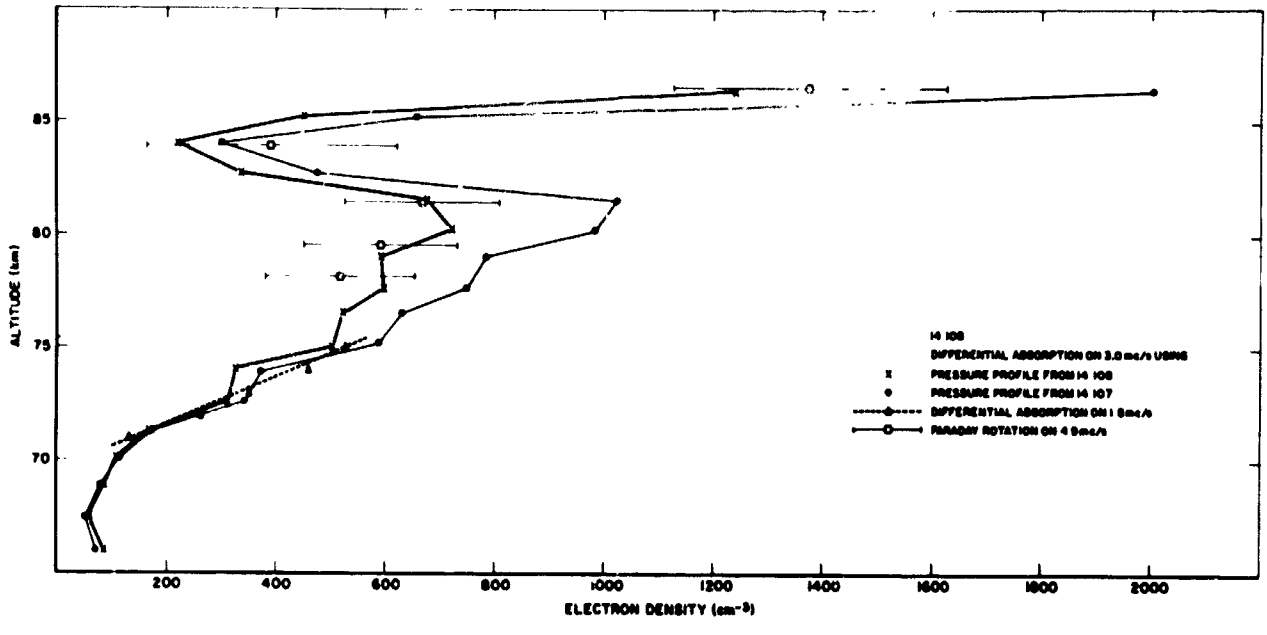


FIGURE 4.—Electron density versus altitude for April 9, 1963 (14.108) showing the effect of the choice of collision frequency model used to interpret radio absorption data.

In addition to a direct comparison of circularly polarized E_0 and E_X signal strengths, it is possible to obtain differential absorption data also from the Faraday pattern of a linearly polarized signal. This follows from the fact that the maximums in the Faraday pattern represent the sum, while the nulls represent the difference of the E_0 and E_X signal strengths. This technique was the basis of the 1.8 Mc/s experiment.

The results of the three propagation experiments on flight 14.108 are shown in Figure 4. The electron densities deduced from the 4.9 and 1.8 Mc/s propagation experiments are relatively insensitive to the choice of collision frequency profile. This is not true however for the 3.0 Mc/s differential absorption experiment. In Figure 4 are shown two electron density profiles deduced from the 3.0 Mc/s differential absorption measurement, the open circle points were computed using the Figure 1 pressure values of Smith, while the values shown as crosses were computed using a 50 percent higher pressure profile. It is seen that consistency between the results of the 14.108 Faraday experiment on 4.9 Mc/s and the 14.108 differential absorption experiment on 3.0 Mc/s requires collision frequency values which between 75 and 80 km are approximately 50 percent higher than those observed one month previously. This requirement is satisfied by the Lyman alpha data shown in Figure 1. For comparison with the 14.107 electron densities the results from 14.108 are plotted as the dashed line in the previous Figure 3. The similarity of the two profiles is apparent, particularly with respect to the minimum at the mesopause (83 km). Possible causes of the minimum at the mesopause include a temperature dependent nitric oxide distribution, and the attachment of electrons to dust. This latter hypothesis has some support from the work of Witt *et al.*, (1962) who reported the detection of dust in the vicinity of the mesopause, and the work of Frocco and Smullin (1963) who reported scattering of a laser beam in this altitude region.

In order to interpret the electron density profiles presented here it is necessary to have a knowledge of the ion pair production function for each of the ionizing radiations affecting the *D* region. This is the subject of the following sections.

COSMIC RAY IONIZATION

The contribution of cosmic rays to the ionization content of the normal *D* region has been considered by Nicolet and Aikin (1960). They showed an ion production rate which was important below 70 km and which involved a variation of a factor 10 between the geomagnetic latitudes of 0° and 70°. For a geomagnetic latitude of 50°, based on the work of Webber (1962), a slightly revised ionization rate of 180 ± 30 ion pairs/sec/atmosphere can be assumed for the altitude region between 60 and 85 km for the 1963 portion of the solar cycle. Combining this value with a model atmosphere yields the altitude dependence of the cosmic ray ionization rates shown as the lower portion of the curves labelled q_{O_2} and q_{N_2} in Figure 5.

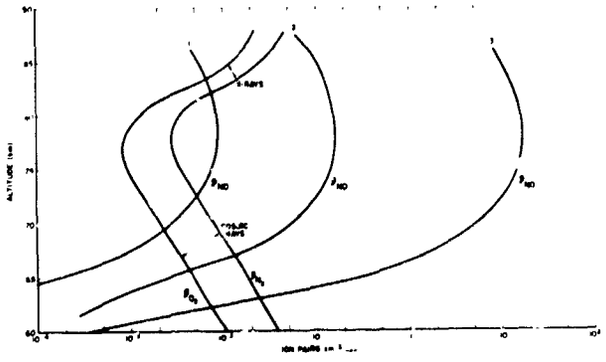


FIGURE 5.—Calculated ionization production function versus altitude. Curves labeled 1, 2, and 3 show the effect of Lyman alpha flux on three different nitric oxide distributions.

X-RAY IONIZATION

The major atmospheric constituents of the *D* region can be ionized by solar X-rays in the 2 to 8Å wavelength region. The importance of this process will depend, however, upon the degree of solar activity. Instrumentation failure prevented a direct measurement of 2–8Å X-rays during our flights. Thus it is necessary to rely on estimates of the flux. The basis of these estimates are the 2800 Mc flux as measured at Ottawa, the McMath-Hulbert calcium plage data, and the Zurich provisional sunspot number. White (1964) has compared direct measurements of X-ray flux from the OSO–1 satellite with these indices and derived an empirical relation for

estimating the X-ray intensity. From these data White estimates that the 2 to 8Å flux was the same to within 50 percent during both our rocket flights and had an integral value of 1.9×10^{-4} ergs/cm²sec.

From Ariel I satellite, Pounds *et al.*, (1961)

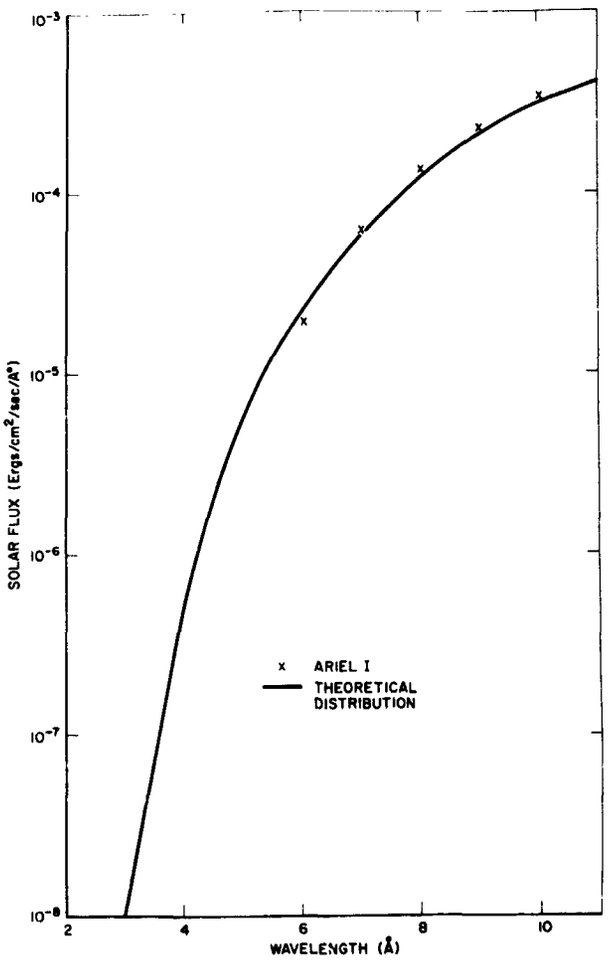


FIGURE 6.—Quiet sun spectrum in the 2 to 10 Ångstrom wavelength interval. Solid curve is the spectral distribution $J(\nu)d\nu = Ae^{-h\nu/kT}d\nu$, with T equal to 2.8×10^6°K.

have obtained solar spectra between 6 and 11Å. The crosses in Figure 6 show a typical non-flare spectrum in which the integrated intensity below 8Å was 1.3×10^{-4}ergs/cm²/sec. According to White (1963) both the Ariel I and OSO-1 non-flare data can be fitted to a frequency spectrum of the form $J(\nu)d\nu \propto e^{-h\nu/kT}d\nu$ with T equal to 2.8×10^6 °K. This is shown as the solid curve of Figure 6.

Using the ionization cross sections given by Nicolet and Aikin (1960) together with a model atmosphere derived from the Smith pressure profile and the spectral distribution of Figure 6 (normalized to an integral value of 1.9×10^{-4}ergs/cm²/sec), the ion pair production function due to X-rays can be calculated as a function of altitude. The results are shown as the upper portion of the curves labelled q_{O2} and q_{N_2} in Figure 5. Before the relative importance of ion pair production by X-rays can be evalutated, it is necessary to have some knowledge of the concentration of nitric oxide, which is the trace constituent ionized by the Lyman alpha flux.

NITRIC OXIDE

For sufficiently large concentrations of atomic nitrogen estimates of the nitric oxide distribution can be based upon the processes

$$N+O_2 \rightarrow NO+O \qquad (8)$$

and

$$N+NO \rightarrow N_2+O \qquad (9)$$

which have rate coefficients b_1 and b_2 respectively. The density of NO in the D region is then determined by

$$n(NO) = (b_1/b_2)n(O_2) \qquad (10)$$

Nicolet and Aikin (1960) adopted 5×10^{-10} as the ratio b_1/b_2. The ion-pair production function resulting from such a distribution is shown as curve 1 of Figure 5. A laboratory measurement of b_1/b_2 by Clyne and Thrush (1961) yielded a value of 9×10^{-9}. The ion-pair production function for this nitric oxide distribution is labeled curve 2 in Figure 5. Barth (1964) has made a direct measurement of the nitric oxide concentration in the upper atmosphere and has obtained a value of 6.2×10^7cm^{-3} between 75 and

TABLE 1

Curve #	b_1/b_2	n(NO)(cm^{-3})	α(cm³/sec)
1	5×10^{-10}	4×10^4	2×10^{-8}
2	9×10^{-9}	6×10^5	3×10^{-7}
3	9×10^{-7}	6×10^7	3×10^{-5}

85 km. If this density is used in Equation (10) for an altitude of 80 km, then a $b./b_2$ ratio of 9×10^{-7} is obtained. The resulting production function is shown as curve 3 in Figure 5. Table 1 summarizes the nitric oxide information for an altitude of 80 km. Also tabulated for this altitude are three derived values of the recombination coefficient α. This parameter is the topic of the following section.

RECOMBINATION COEFFICIENT

The electron densities of Figure 3 are elated to the ion production functions of Figure 5 by the equation

$$N_e^2 = \Sigma_i q_i/\alpha_i \qquad (11)$$

where α_i is the effective recombination coefficient of the i^{th} species of positive ion. Biondi (1964) has indicated that O_2^+ and N_2^+ recombine dissociatively with a rate coefficient of 2×10^{-7}cm^3 sec^{-1}, while NO^+ has a recombination coefficient of less than 5×10^{-7}cm^{3}sec^{-1}. The values of recombination coefficient derived from curves 1 and 2 of Figure 5 and given in Table 1 are consistent with dissociative recombination of NO^+ as the electron loss process of the D region. Curve 3 requires the introduction of loss processes not previously considered in *D* region theory since 10^{-6}cm^{3}sec^{-1} is the upper limit assigned to dissociative recombination by both laboratory measurements and theoretical determinations.

By using in Equation (11) the recombination coefficients given by Biondi together with the ion production functions shown in Figure 5, it is seen that the ionization due to X-rays can dominate the quiet *D* region only above 83 km and even than only for the case of a nitric oxide distribution corresponding to curve 1 of Figure 5.

CONCLUSION

Electron density and collision frequency profiles have been obtained for the quiet day mid-latitude *D* region. The electron collision frequency is subject to significant variations of a meteorological nature. The electron density profiles agree with the theory that for a reasonably quiet sun solar X-rays are unimportant, and Lyman alpha ionization of nitric oxide is the major contributor to the ionization content of the *D* region. Although existing theory can explain these measured profiles, the recent nitric oxide measurements of Barth (1964) would, if accepted, require a reevaluation of the electron-ion less processes.

ACKNOWLEDGMENTS

The authors acknowledge helpful conversations with Goddard colleages—W. A. White of the Solar Physics Branch and G. Warnecke of the Meteorology Division. The cooperation of C. A. Barth of the Jet Propulsion Laboratory in furnishing data prior to publication is appreciated.

REFERENCES

Barth, C. A., Rocket Measurement of Nitric Oxide in the Upper Atmosphere. Presented at COSPAR meeting, Florence, Italy, (1964).

Biondi, M. A., Electron-ion and ion-ion recombination, Annales de Géophysique, T 20 No. 1, pp. 24-46 (1964).

Bossoclosco, M. and A. Elena, Absorption de la couche D et température de la mésophère. Comptes Rendus de l'Academie des Sciences, T256, No. 21, 4491-4493, (1963).

Clyne, M. A. A. and B. A. Thrush, Kinetics of the Reactions of Active Nitrogen with Oxygen and with Nitric Oxide. Proc. Roy. Soc. A, **261**, No. 1305, 259-273, (1961).

Figocco, G. and L. D. Smullin, Dectection of Scattering Layers in the Upper Atmosphere (60-140 km) by Optical Radar. Nature, Vol. 199, No. 4900, Sept. 28, 1963.

Friedman, H., Ultraviolet and X-rays from the Sun. Annual Review of Astronomy and Astrophysics, Vol. 1, (1963).

Nicolet, M. and A. C. Aikin, The formation of the D Region of the Ionosphere. J. Geophys. Res., **65**, 1469-1483, (1960).

Phelps, A. V., Propagation Constants for Electromagnetic Waves in Weakly Ionized Dry Air. J. Appl. Phys., **31**, 1723-1729, (1960).

Pounds, K. A., A. P. Willmore, P. J. Bower, K. Norman and P. W. Sanford, Measurements of the Solar Spectrum in the Wavelength Band 4-14Å. Proc. Roy. Soc. (In press).

Sen, H. K. and A. A. Wyller, On the Generalization of the Appleton-Hartree Magnetoionic Formulas. J. Geophy. Res. **65**, 3931-3950, (1960).

Smith, W., L. Katchen, P. Sacher, P. Swartz and J. Theon, Temperature, Pressure, Density and Wind Measurements with the Rocket Grenade Experiment 1960-1963. NASA-GSFC Report X-651-64-106, (1964).

Stroud, W. G. and W. Nordberg, Seasonal, Latitudinal and Diurnal Variations in the Upper Atmosphere. NASA Tech. Note D-703, (1961).

WEBBER, W., The Production of Free Electrons in the Ionospheric D Layer by Solar and Galactic Cosmic Rays and the Resultant Absorption of Radio Waves. J. Geophys. Res., 67, 5091–5106, (1962).

WHITE, W. A., Solar X-rays; Slow Variation and Transient Events. NASA Tech. Note G–545, (1963).

WHITE, W. A. Solar X-rays; A Comparison with Microwave Radiation. NASA–GSFC Report X–610–64–42, (1964).

WITT, G., C. L. HEMENWAY and R. K. SOBERMAN, Collection and Analyses of Particles from the Mesopause Presented at COSPAR meeting, Warsaw, Poland, (1963)

N66 37956

HYDROGEN IONS*

SIEGFRIED J. BAUER

Whistler observations which first indicated surprisingly large electron densities at several earth radii (cf. "Whistlers and Exospheric Structure" by L. R. O. Storey) led to the interpretation that the outer ionosphere consists predominantly of hydrogen ions (protons) and electrons. Consequently, the name "protonosphere" has been suggested for this region of the earth's outer atmosphere (Johnson, 1960). The hydrogen ions of the protonosphere are now considered to be the result of charge exchange between ionospheric oxygen ions and neutral hydrogen of the telluric geocorona (Johnson, 1961), rather than due to photoionization. While the protonosphere was originally envisioned (Johnson, 1960) to be a direct extension of the ionospheric F region, where oxygen ions predominate, it has been established recently that a region of helium ions whose extent varies with temperature and solar cycle, intervenes between the regular ionosphere and the protonosphere (cf. "Helium Ions" by W. B. Hanson).

It was first pointed out by Dungey (1955), that the charge exchange process

$$H+O^+ \rightleftarrows H^+ + O \qquad (1)$$

may be of importance in the upper atmosphere. Johnson, (1960) was the first to argue that this process represents the major source of hydrogen ions constituting the protonosphere. Since oxygen and hydrogen have nearly the same ionization potential, reaction (1), which is of the accidental asymmetric resonance type, occurs about as rapidly either way and therefore presents the source and sink for H^+. It has been suggested (Rapp, 1963) that the cross-section for process (1) is large enough (10^{-15} to 10^{-16}cm^2) so that (1) occurs rapidly enough to be the most important chemical reaction involving hydrogen ions in the upper atmosphere. (In interplanetary space, photoionization and radiative recombination obviously will become the controlling processes.) Hanson, Patterson and Degaonkar (1963) have recently inferred from experimental data a rate coefficient $K=3.9\times10^{-10}$cm^{3}sec^{-1} for reaction (1), corresponding to a reaction cross-section of 7.6×10^{-16}cm^2; these values are considered to be correct within a factor of three.

The altitude distribution of an ion, such as H^+, is governed by the equation of continuity

$$\frac{\partial(H^+)}{\partial t}=q(H^+)-L(H^+)-\text{div } F(H^+) \qquad (2)$$

where $q(H^+)$ is the production rate and $L(H^+)$ is the loss rate of hydrogen ions and $F(H^+)$ is the flux of H^+ resulting from diffusion under gravity.

For the steady state, $\frac{\partial(H^+)}{\partial t}=0$, and considering mainly an upward or downward flux of hydrogen ions (i.e. in the z direction), equation (2) can be expressed by

$$q(H^+)-L(H^+)-\frac{\partial F(H^+)}{\partial z}=0 \qquad (3)$$

While equation (3) is not readily amenable to an analytic solution, but has to be solved numerically (cf. Bates and Patterson, 1961; Hanson, 1963, Hanson et al, 1963) there are two simple limiting cases to (3): 1) where diffusion of hydrogen ions can be neglected, *chemical equilibrium* ($q=L$) will prevail, and 2) at high enough altitudes where diffusion is the predominant process, *diffusive equilibrium* will prevail $\left(\frac{\partial F(H^+)}{\partial z}=0\right)$

*Published as *Goddard Space Flight Center Document X-615-64-48*, March 1964. This work is a contribution to the "International Dictionary of of Geophysics" Pergamon Press, Oxford, England.

1. Chemical Equilibrium Distribution of H^+

Noting that both production and loss of H^+ are controlled by the charge exchange reaction (1), we can write (cf. Bates and Patterson, 1961)

$$Kn(H)n(O^+)=\sigma Kn(H^+)n(O) \tag{4}$$

where K is the rate coefficient and $\sigma=8/9$ is the value for the ratio of the products of the statistical weights. Thus, the chemical equilibrium distribution for hydrogen ions is given by:

$$n(H^+)=\frac{9}{8}\frac{n(H)}{n(O)}n(O^+) \tag{5}$$

where $n(X)$ is the number density of the neutral constituent X which is distributed exponentially according to its scale height $\underline{H}(X)=kT/m(X)g$, with k Boltzmann's constant, T the absolute temperature, $m(X)$ the mass of the constituent and g the acceleration of gravity; the number density of oxygen ions, $n(O^+)$ is distributed according to its scale height (assuming predominance of O^+) $\underline{H}(O^+)=k(T_e+T_i)/m(O)g$, where T_e and T_i are the electron and ion temperature, respectively, which may be different from each other. The distribution above a reference level h_0 where $n(H^+)=n_0(H^+)=9n_0(H)n_0(O^+)/8n_0(O)$ and above which all constituents are distributed exponentially, is given by

$$n(H^+)=n_0(H^+)\exp\left[z/\underline{H}(7)\right] \tag{6}$$

where Z is the geopotential altitude above the reference level $\left(z=\int_{h_0}^{h}(g/g_0)dh\right)$ which takes into account the variation of the acceleration of gravity g with altitude, and $\underline{H}(7)$ is the scale height corresponding to an effective mass of 7 AMU, (resulting from the combined scaled heights), and to the acceleration of gravity at the earth's surface g_0. (It is assumed that all constituents entering into the *R.H.S.* of (5) are in diffusive and thermal equilibrium; e.g., $\underline{H}(O^+)=2\underline{H}(O)$). Thus, it is obvious that H^+ in chemical equilibrium is *increasing* with altitude.

2. Diffusive Equilibrium Distribution of H^+

It is now generally known that the diffusive equilibrium distribution of an ion (such as H^+) in a mixture is not independent of the other ionic constituents because of the presence of an electrostatic field which is set up to prevent further charge separation between electrons and ions diffusing under gravity. This electrostatic field, which depends on the mean ionic mass of the mixture and on the electron and ion temperatures, is sensed by all ions in the mixture and tends to counteract the gravitational force. This fact has been known to astrophysicists for many years (cf. Eddington, 1926), but has been revived more recently (Dungey, 1955; Mange, 1960) in connection with the terrestrial ionosphere. According to this concept, the distribution of H^+ in diffusive equilibrium is given by

$$n(H^+)=n_0(H^+)\exp\left[-\int_0^z\left(\frac{1}{\underline{H}(H^+)}-\frac{\tau}{1+\tau}\frac{m_+g_0}{kT_i}\right)dz\right] \tag{7}$$

where $\underline{H}(H^+)=\dfrac{kT_i}{m(H^+)g_0}$, $\tau=T_e/T_i$ is the ratio of electron to ion temperature and $m_+=\Sigma m_i n_i/\Sigma n_i$ is the mean ionic mass of the mixture. Since the upper ionosphere consists of a ternary ion mixture (H^+, He^+, O^+) we also can write (7) as

$$n(H^+)=n_0(H^+)\exp\left[-\int_0^z\frac{dz}{H(H^+)}\right] \tag{8}$$

where $H(H^+)=\underline{H}(H)(1+\tau)\dfrac{n(O^+)+n(He^+)+n(H^+)}{n(H^+)-\tau[14n(O^+)+2n(He^+)]}$

Thus, it is easily seen that, e.g. for the case of thermal equilibrium ($\tau=1$), $H(H^+)<0$, if $n(H^+)<2[n(He^+)+7\ n(O^+)]$ and approaches $2\underline{H}(H)$ when H^+ becomes the predominant ion. As long as H^+ is a minor constituent, it is *increasing* with altitude. In fact, in a predominantly O^+ ionosphere the initial increase of H^+ in diffusive equilibrium follows almost the same scale height as the one corresponding to chemical equilibrium (6) and in this case, the predominance of chemical or diffusive equilibrium is undistinguishable from the behavior of the altitude distribution alone. Figure 1 shows the behavior of the distribution of H^+ in diffusive (and thermal) equilibrium in the presence of He^+ and O^+ at a temperature of $T=1200°K$. For the simple computation of the diffusive and thermal equilibrium distribution of an ion in a mixture, the following formula is quite

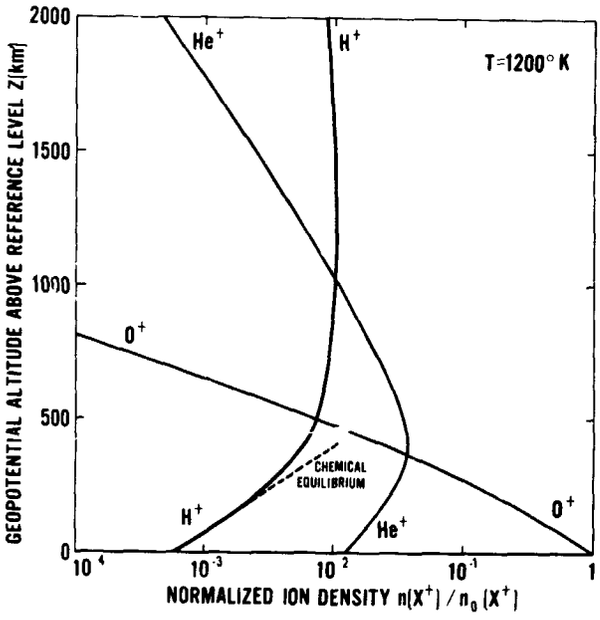


FIGURE 1

useful (cf. Hanson, 1962; Bauer, 1962):

$$n(X_k^+) = n_0(X_k^+) \exp[-z/H_k] \cdot \left[\frac{\Sigma_t n_0(X^+)}{\Sigma_t n_0(X^+) \exp(-z/H_t)}\right]^{1/2} \quad (9)$$

(Although the electron temperature often differs from the ion temperature, calculations based on thermal equilibrium generally can provide an adequate representation of the altitude distribution of an ion in diffusive equilibrium.)

The actual distribution of hydrogen ions in the upper atmosphere is one which more closely corresponds to a steady-state distribution (3) with a non-vanishing flux of ions. Figure 2 shows an observed H^+ distribution (Taylor, Brace, Brinton and Smith, 1963) which has been interpreted by Hanson et al, (1963) in terms of a steady-state distribution, corresponding to a rate coefficient for the charge exchange reaction of 3.9×10^{-10}cm^{3}sec^{-1} and an upward flux of hydrogen ions at 900 km of $F(H^+) = 1.3\times10^7$cm^{-2}sec^{-1}.

The altitude where diffusion becomes predominant compared to chemical processes in controlling the distribution of H^+ can be estimated by comparing the rate of the downward flux of hydrogen ions $\left(-\frac{\partial F(H^+)}{\partial z}\right)$, with the loss rate due to charge exchange with atomic oxygen (cf. Bates and Patterson, 1961) leading to the criterion

$$Kn(O) = \frac{D(H^+, O^+)}{2[\underline{H}(O)]^2} \quad (10)$$

where $D(H^+, O^+) = 8.4\times10^{15}(T/1500°)^{5/2}/[n(H^+) + n(O^+)]$, (Hanson, 1963) is the diffusion coefficient for Coulomb diffusion of protons through a predominantly O^+ ionosphere, and the other quantities have their previously defined meaning. For the condition representing the data of Taylor et al, (1963) shown in Figure 2, the transition from predominantly chemical to diffusive control would occur at an altitude of about 600 km and since the diffusion process is controlled by oxygen ions one would expect that one scale height, $\underline{H}(O^+) = 2\,\underline{H}(O)$, above and below this level diffusive and chemical equilibrium respectively, would prevail. This estimate is in good agreement with the conclusions of an analysis of the behavior of the ratio of $n(H^+)/n(He^+)$ of the same data by Bauer (1964). It was first shown by Hanson and Ortenburger (1961) that oxygen ions represent an effective barrier to the diffusion of H^+ because of Coulomb interaction between ions. Thus, the "critical level" defined by the criterion

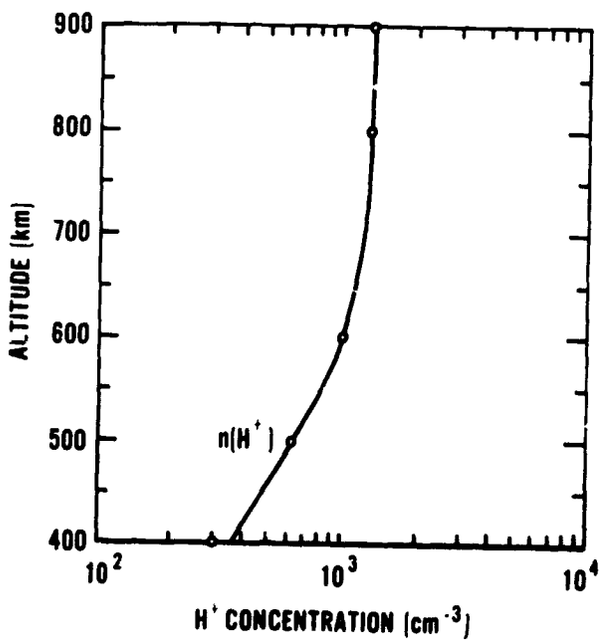


FIGURE 2

in (10) may be considered the base of the protonosphere.

Since the charge exchange process (1) controls the hydrogen ion concentration at the lower altitudes, the time-dependent boundary condition for the proton distribution at higher altitudes is given to a first approximation by the ratio of $n(H)/n(O)$ (cf. (5)). It is well known that the neutral hydrogen concentration is quite sensitive to the atmospheric temperature because of the thermal escape of neutral hydrogen. Thus, the concentration of hydrogen at a fixed level of say 500 km will be higher at low temperatures than at high temperatures. Consequently, the ratio $n(H)/n(O)$ at 500 km may vary by a factor of 100 for a temperature change of about 1000°K (Bates and Patterson, 1961; Kockarts and Nicolet, 1963). Considering a ternary ion mixture (0^+, He^+ and H^+) and boundary conditions varying with temperature according to the variation of the neutral species corresponding to these ions at 500 km (Bauer, 1963; Kockarts and Nicolet, 1963) the altitude domain where H^+ is the predominant ion can be described qualitatively as shown in Figure 3.

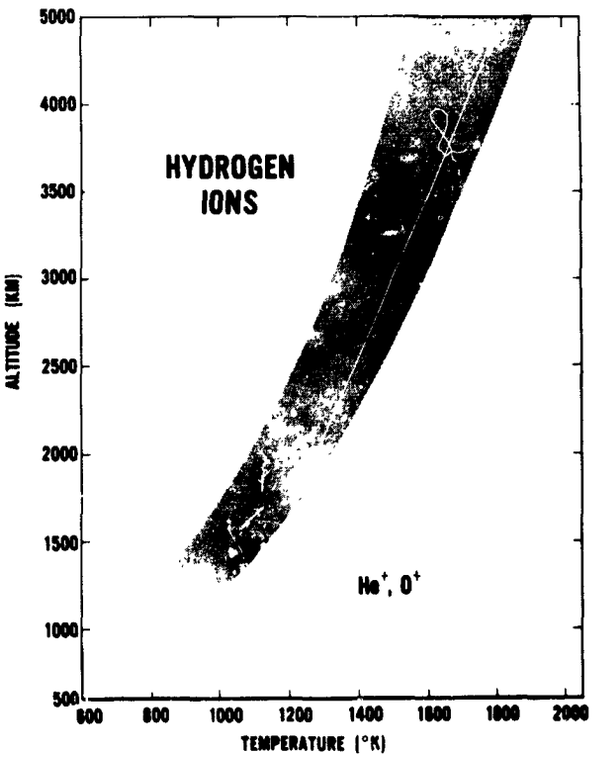


FIGURE 3

While such a concept is roughly indicative of the longer term temperature variations it may not be representative for diurnal variations, since the response time of the change in the hydrogen concentration to a change in temperature, at least at the lower temperatures, is of the order of tens of hours (Bates and Patterson, 1961). Although there is some indirect evidence of large diurnal variation in the concentration of hydrogen based on nighttime profiles of ion and electron density (Donley, 1963; King et al, 1963) which indicate a preponderance of H^+ already at altitudes below 1000 km, a large diurnal variation is difficult to explain in view of the long response time of hydrogen. Hanson and Ortenburger, (1961) have pointed out that because of the inhibiting effect of Coulomb diffusion, the protons cannot follow diurnal variations of the oxygen ions in the ionospheric F region and thus, the protonosphere is essentially decoupled from the upper ionosphere. In fact, the diurnal variation in the protonosphere based on whistler observations is only of the order of at most 30 percent. Since the total content of H^+ throughout the protonosphere is estimated to be about 7.5×10^{12} protons cm^{-2} (Hanson and Ortenburger, 1961), a 30 percent variation would imply an average flux of 5×10^7 ions $cm^{-2}sec^{-1}$ into the ionospheric F region during a 12-hour period. This flux would be, in general, upward during the day and downward during the night. The upward flux during the daytime, however, seems to be limited by the flux of neutral hydrogen from the lower atmosphere which is of the order of 2.5×10^7 atoms $cm^{-2}sec^{-1}$ (cf. Hanson and Patterson, 1963). As mentioned before, an upward flux of 1.3×10^7 protons $cm^{-2}sec^{-1}$ has been inferred by Hanson et al (1963) from the daytime proton distribution measured by Taylor et al (1963), (cf. Figure 2). Consequently, it appears (Hanson and Patterson, 1963) that, although nighttime condition may allow a sufficient flux of protons to account for a 30 percent change in the hydrogen ion content of the protonosphere, the return flow required during the day does not appear to be feasible. Thus, it also seems unlikely that fluxes of protons between hemispheres affect the ion density at the $F2$ peak as has been suggested (Rothwell, 1962) since the observed proton fluxes are small compared with the normal flux through the $F2$ peak

which is estimated to be of the order of 2×10^8 to 10^9 ions $cm^{-2}sec^{-1}$ (Hanson and Patterson, 1963).

While the general behavior of hydrogen ions is already reasonably well understood, quantitative data especially those concerning its time variation are still inadequate. However, this situation will hopefully be remedied as the result of additional measurements with rockets and satellites.

REFERENCES

Bauer, S. J. (1962), J. Atmos. Sci. **19**, 276–278.

Bauer, S. J. (1963), Nature **197**, 37–38.

Bauer, S. J. (1964), J. Geophys. Res. 69, 553–555.

*Bates, D. R. and Patterson, T. N. L., (1961) Planet. Space Sci., 5 257–273, 1963.

Donley, J. L. (1963), J. Geophys. Res. **68**, 2058–2060.

Dungey, J. W. (1955), *The Physics of the Ionosphere*, 229–236, Physical Society, London.

Eddington, Sir Arthur, (1926), *The Internal Constitution of Stars*, 272–274, London.

Hanson, W. B. (1962), J. Geophys. Res. **67**, 183–188.

*Hanson, W. B. (1963), *Electron Density Profiles in the Ionosphere and Exosphere*, Pergamon Press.

*Hanson, W. B. and Ortenburger, I. B. (1961), J. Geophys. Res. 66, 1425–1436.

*Hanson, W. B. and Patterson, T. N. L. (1963), Planet. Space Sci. 11, 1035–1052.

Hanson, W. B., Patterson, T. N. L. and Degaonkar, S. S. (1963), J. Geophys. Res. **68**, 6203–6205.

Johnson, F. S. (1960), J. Geophys. Res. **65**, 577.

Johnson, F. S. (1961) Astrophys. Journ. **133**, 701.

King, J. W., Eccles D., Smith, P. A., *et al*, 1963, DSIR Radio Res. Station Report RRS/IM 112, Slough, England.

*Kockarts, G. and Nicolet, M., (1963), Ann. Geophys. **19**, 370–385.

Mange, P. (1960), J. Geophys. Res. **65**, 3833–3834.

Rapp, D. (1963), J. Geophys. Res. **68**, 1773–1775.

Rothwell, P., (1962), *Proc. Kyoto Intern. Conf. Cosmic Rays and Earth Storms*, Phys. Soc. of Japan I. 263.

Taylor, H. A., Brace, L. L., Brinton, H. C., and Smith, C. R. (1963), J. Geophys. Res. **68**, 5339–5348.

*Suggested as source for a more comprehensive treatment of the subject of this chapter.

MEAN DIURNAL VARIATION OF THE TOPSIDE IONOSPHERE AT MID-LATITUDES

S. J. BAUER AND L. J. BLUMLE

About 500 ionograms obtained with the Alouette satellite have been used to construct the mean diurnal variation of the mid-latitude topside ionosphere. Since this diurnal variation is based on observations from October to December 1962, quasiseasonal effects had to be removed. The data are presented in form of electron density contours at fixed altitudes for two latitude ranges 35°N–40°N and 40°N–45°N. These two latitude ranges, which are approximately separated at the 75°W meridian by the 70° magnetic dip line, show remarkable differences in the detailed behavior of electron density as a function of time and altitude. From the scale heights of these mean distributions, it is apparent that there is absence of thermal equilibrium, at least during the day, and that light ionic constituents become important in the altitude region from 500 km to 1000 km, especially during the night. In addition, there is the suggestion of latitude gradients in the electron temperature and mean ionic mass.

INTRODUCTION

Prior to the launching of the Canadian topside sounder satellite *Alouette*, our knowledge of the vertical structure of the ionosphere above the F_2 peak was based on occasional rocket flights and incoherent radar backscatter observations from the ground, both of which are limited to specific locations. The topside sounder satellite provides us for the first time with a tool for the study of the topside ionosphere and its variation with latitude *and* time. Since the topside sounder operates only on command, observations are restricted to times when the satellite is within appropriate range of a telemetry and command station. We have made use of soundings of the topside ionosphere taken while the satellite was within range of the NASA telemetry station at Blossom Point, Maryland. The latitude coverage obtainable from this station is between 20°N and 55°N. We have concentrated on the study of ionograms obtained at mid-latitudes (35°N to 45°N), for which our data were most complete. As the result of the orbital characteristics of the Alouette satellite and combining north- and southward passes, about three months are required to obtain a complete diurnal variation of the ionospheric parameters. This paper presents the mean diurnal behavior of the mid-latitude topside ionosphere for the three-month period from October to December 1962.

DIURNAL VARIATION OF ELECTRON DENSITY AT CONSTANT ALTITUDES

About 500 ionograms taken on magnetically quiet days during the first three months of Alouette's operation and covering the latitude range from 30°N to 50°N (and longitudes 50°W to 100°W) were selected for the present study. The ionograms were converted to electron-density profiles by means of the exponential lamination method described by *Fitzenreiter and Blumle*, (1964). For the study of the diurnal behavior of the topside ionosphere, a presentation in terms of electron density at constant altitudes was chosen. In this way a mean behavior could best be defined from the large number of data points. It became apparent that because of the latitude variation of the topside ionosphere, latitude ranges of not more than 5° latitude should be used, otherwise the spread in the electron density data points at constant altitudes would often be greater than the separation of electron density contours at successive 100 km altitude increments. Since below 35°N and above 45°N the data available were not sufficient to define a complete diurnal variation, only the two latitude ranges from 35°N to 40°N

*Published as *Goddard Space Flight Center Document X-615-64-99*, April 1964

and from 40° to 45°N will be presented in this paper. The "mean" contours of electron density at 100 km height intervals are based on individual data points; the maximum spread of individual data points around the density contours at fixed heights is of the order of ±25 percent. Figure 1 shows an example of such mean contours every 200 km up to the satellite altitude for the latitude range 40°N to 45°N. It can be seen that these mean contours have some oscillations superimposed which need not represent actual diurnal behavior, since the entire "diurnal" variation is based on three months of data. Thus, the mean

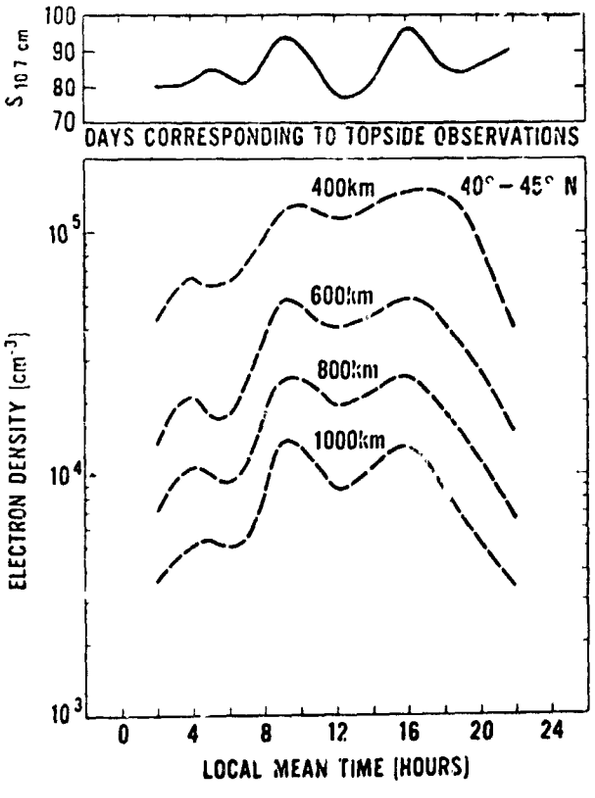


FIGURE 1.—Mean diurnal variation of electron density at constant altitudes based on actual data points and flux of 10.7 cm solar radiation (in units of $10^{-22} W/m^2/cps$) at times corresponding to the ionospheric measurements, indicating "quasi-seasonal" effects (27 day cycle).

contours presented in Figure 1 will also include quasi-seasonal effects, e.g., the 27 day cycle. This becomes quiet apparent, if the flux of the solar 10.7 cm radiation, $S_{10.7}$, which is an indicator of solar activity, is plotted for the days corresponding to the individual data points for electron density at the indicated local times. The smoothed variation of $S_{10.7}$ is plotted in the top portion of Figure 1 and it is evident that the electron density at fixed altitudes behaves in a similar fashion. At first, it may seem surprising that the topside electron density should be closely related to the 10.7 cm flux from the sun, since there is no obvious correlation of this parameter with the density at the F_2 peak, or below. However, it should be realized that at lower altitudes where production and loss of ionization play a predominant role, the detailed correlation of electron density with short term solar variations as indicated by the 10.7 cm flux may be obscured by the fact that both, production and loss are affected by the variation of the appropriate neutral constituents and is thus effectively cancelled out. At altitudes above the F_2 peak, where the distribution is controlled by diffusion, and thus by the scale height, this effect may appear more directly since the scale height and temperature are known to be directly correlated with the 10.7 cm flux. Correspondingly, the "seasonal" effect superimposed on the diurnal variation was removed using the behavior of the

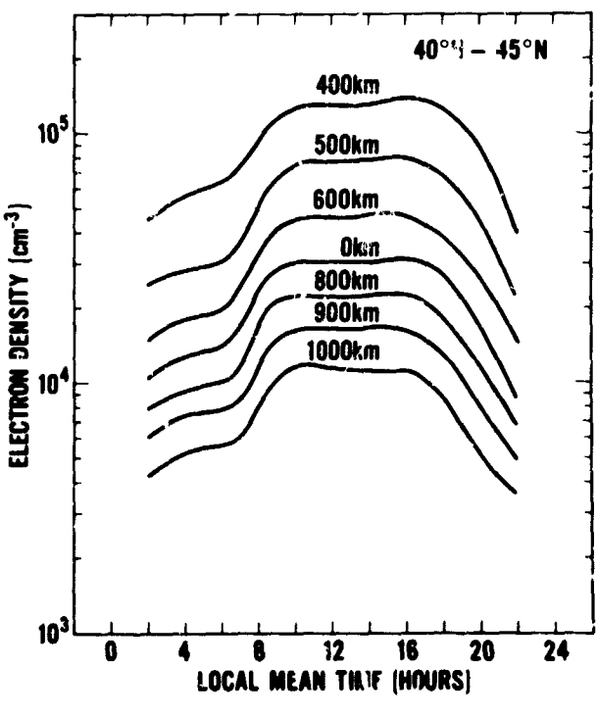


FIGURE 2.—Diurnal variation of electron density at constant altitudes for the latitude range 40° N to 45° N. (Corrected for quasi-seasonal effects).

10.7 cm solar flux as an indicator and normalizing the mean diurnal variation to a condition of $S_{10.7}=85\times10^{-22}W/m^2/cps$. This was accomplished by making a percentage correction of the electron density contours appropriate to the variation of $S_{10.7}$. The mean diurnal variation of electron density at fixed altitudes, corrected for quasi-seasonal effects, is shown in Figure 2 for the latitude range 35°N to 40°N and in Figure 3 for

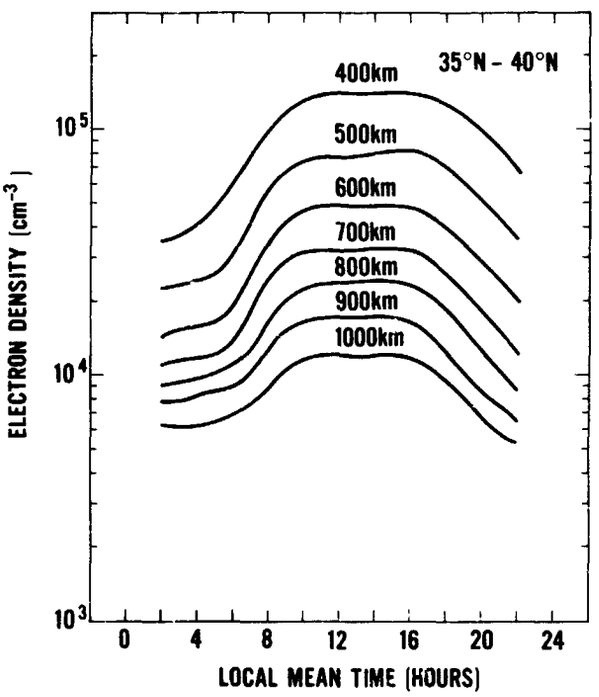


FIGURE 3.—Diurnal variation of electron density at constant altitudes for the latitude range 35° N to 40° N. (Corrected for quasi-seasonal effects).

40°N to 45°N. While the daytime values for the two latitude ranges are virtually identical, the electron density contours show a quite different behavior during the rest of the day. (The fact that there are no data between 2200 and 2400 LMT is the result of poor performance of the topside sounder during these night hours as the result of leakage of terrestrial noise which causes blocking of the receiver (*Atkins and Chapman*, 1963)). In comparing the behavior of the mean diurnal variation at the two latitude ranges, which can be termed in the geographical sense as *mid*-latitudes, one has to keep in mind, that magnetically this latitude range at the 75°W meridian is already representative of rather high dip angles; e.g., at 39°N, 75°W, the magnetic dip is I=70°. It appears that this dip represents somewhat of a boundary between "normal" and "auroral" type of ionosphere. This is definitely evidenced in the behavior of spread F (*Calvert and Schmid*, 1964) and thus is possibly also indicative of the role of corpuscular effects in the ionosphere (cf. *Mariani*, 1963). Similar conclusions have been drawn on the basis of a comprehensive study of the behavior of the "bottomside" ionosphere (*Wright*, 1962). The different behavior of the topside in the two latitude ranges, which are only 5° apart, yet include the "dip-boundary" can best be illustrated by comparing the amplitude of the diurnal variation at selected fixed heights at the 35°N to 40°N latitude range with that at 40°N to 45°N. This

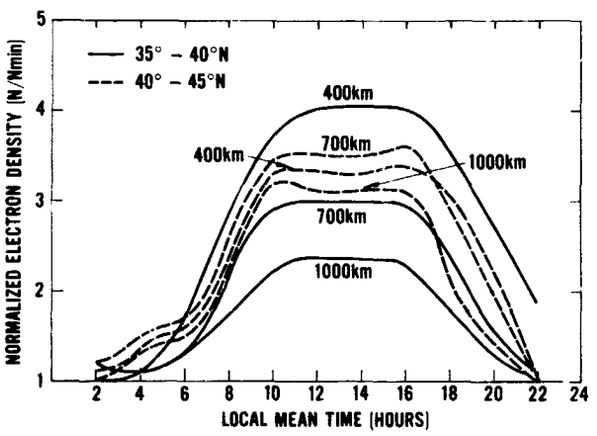


FIGURE 4.—Time variation of normalized electron density at fixed altitudes illustrating the variation with altitude of the diurnal amplitude for the two latitude ranges.

is shown in Figure 4, where the normalized electron density, i.e. the ratio of electron density at a particular local time to the minimum value of electron density at selected altitudes is plotted as a function of local mean time for both latitude ranges. It is quite obvious that the diurnal amplitude is decreasing with altitude for the latitudes 35°N to 40°N, while for 40°N to 45°N the diurnal amplitude is virtually independent of altitude. Furthermore, the absolute value of the amplitude at 400 km is higher at 35°N to 40°N than at 40°N to 45°N. The variation of the diurnal amplitude at latitudes 35°N to 40°N as a function of altitude is qualitatively what is expected from theoretical considerations (*Gliddon and Kendall*, 1962), how-

ever the actual decrease with altitude of the ratio of maximum to minimum density from a value of 4 at 400 km to a value of 2.3 at 1000 km is still much slower than theoretical estimates. It should be noted, however, that although the theoretical model of the F_2 region by Gliddon and Kendall is based on the solution of the time-dependent continuity equation, it is not strictly applicable to the topside ionosphere because it assumes an isothermal atmosphere whose temperature also does not show a diurnal variation. It is quite evident from Figure 4 that the diurnal behavior of the topside ionosphere, especially at latitudes greater than 40°N, is quite different from simple model concepts. The reason for the discrepancies is obviously the importance of a variable scale height as the result of diurnal variation of temperature, absence of thermal equilibrium between electrons and ions and a varying ion composition.

The above conclusion that a variable scale height is indeed responsible for the altitude behavior of the diurnal amplitude of electron density at fixed levels, is more readily apparent in "mean" vertical profiles of electron density (derived from cross sections of the electron density contours at fixed heights, as shown in Figures 2 and 3).

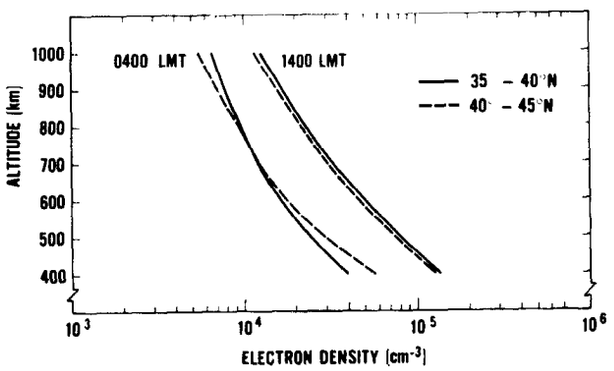


FIGURE 5.—"Mean" electron density profiles of the mid-latitude topside ionosphere corresponding to times of the diurnal temperature maximum (1400 LMT) and minimum (0400 LMT).

Figure 5 shows such "mean" profiles for the two latitude ranges at 0400 LMT and 1400 LMT, the times corresponding to the diurnal minimum and maximum of the atmospheric temperature. The problems concerning a variable scale height of the topside ionosphere and the implications from the observations presented here will be discussed in the following section.

DERIVED QUANTITIES: SCALE HEIGHT, ELECTRON TEMPERATURE AND ION COMPOSITION

The spacing of electron density contours at fixed altitudes shown in Figures 2 and 3 is a measure of the scale height of the electron-ion gas. It is well known that in the topside ionosphere this scale height is given by

$$H' = \frac{k(T_i+T_e)}{m_+ g} = \left[-\frac{1}{N}\frac{\partial N}{\partial z} \right]^{-1}$$

where k is Boltzmann's constant, T_e and T_i are the electron and ion temperature respectively, m_+ is the mean ionic mass and g is the acceleration of gravity. From the diurnal variation of the mean electron density contours it is thus also possible to infer the diurnal variation of the scale

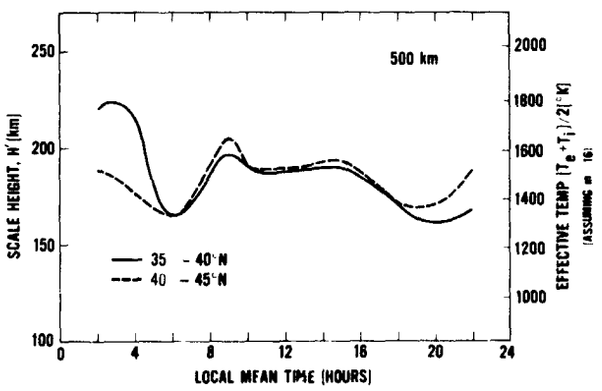


FIGURE 6.—Electron-ion scale heights at 500 km for the two mid-latitude ranges as a function of local mean time, together with corresponding scale for the effective charged particle temperature $(T_e+T_i)/2$ assuming $m_+ = 16$. The interpretation of scale heights in terms of electron temperature T_e and mean ionic mass m_+ is discussed in the text.

height. Figure 6 shows the scale height H' at an altitude of 500 km for the two latitude ranges under consideration as a function of local mean time. In addition, a scale is given for the effective charged particle temperature $(T_e+T_i)/2$, assuming a mean ionic mass $m_+ = 16$. The latter assumption may be justifiable at the given alti-

tude, at least during the daytime. In examining the variation of H' as a function of local time one has to keep in mind that this variation is the result of variation in *electron- and ion-temperature, and ion composition.* Referring to the temperature scale, it is obvious that the effective charged particle temperature is in excess of the neutral gas temperature T_n for the given three-month period, when T_n was about 1000°K during the day, and about 700°K at night (cf. *Nicolet*, 1963). Thus, it is quite evident that thermal equilibrium is absent, at least during the day (when $m_+ = 16$ may be a reasonably good assumption). If we assume that $T_i = T_n \approx 1000°$, the electron temperature T_e is of the order of 2000°K during most of the day with an even higher value around 0900 LMT. (If m_+ is less than 16, then T_e is reduced proportionately; thus, the above value represents an *upper* limit). It is interesting to note, that actual measurements of T_e near the F2 peak for the same latitude range obtained from Blossom Point telemetry read-outs of the S-6 aeronomy satellite (*Brace, Spencer and Dalgarno*, 1964) during spring and summer of 1963, are in excellent agreement with our inferred electron temperatures. Even the maximum at 0900 LMT and the subsequent plateau in the electron temperature appear in both data. Thus, we are confident, that our assumptions are reasonable. Furthermore, there is the suggestion of a latitude gradient in the scale height during the day when we are able to infer electron temperature. The electron temperature appears to be higher at latitudes greater than 40°N than below that, a conclusion, again in agreement with the actual measurements of electron temperature by *Brace et al* (1964). During the early morning hours (0200 0500 LMT) the electron-ion scale height H' is higher than, or at least comparable to, the daytime values. This cannot be explained in terms of temperature since there is good reason to believe that atmospheric heating *is* lower at night (assuming that the main heat source is solar EUV with a possibly constant, corpuscular component super imposed on it). The more plausible explanation is that the large scale height during this time period is the result of a change in the ion composition, indicating the presence of the lighter ionic constituents He^+ and H^+. The latitudinal gradient in scale height is reversed during this time period, implying that in terms of ion composition, the preponderance of the light ions is weighted towards lower latitudes. This inference finds support in observations of ion composition on the Ariel satellite (*Bowen et al*, 1964). If such a latitude gradient of ion composition does indeed exist even during the day, then the latitude-gradient in electron temperature would be even larger than that apparent from the daytime scale heights for the two latitude ranges. Ion composition, or rather the mean ionic mass m_+ at 500 km, can be inferred from the scale height if assumptions are made concerning the electron and ion temperatures. Assuming first, that thermal equilibrium prevails during the night, then at 0400 LMT (corresponding to the diurnal temperature minimum) $T_e = T_i = T_n \approx 700°K$ for the time period under consideration. Accordingly, the mean ionic mass at 500 km is $m_+ \approx 8$ *AMU* at latitudes 40°N to 45°N and $m_+ \approx 7$ *AMU* at latitudes 35°N to 40°N. The trend seems to be reversed in the late evening hours when the scale heights are greater at higher latitudes than at the lower latitudes. However, the assumption of thermal equilibrium may not hold even during the night, since electron temperatures $T_e \approx 1000°K$ have been observed at nighttime (*Brace et al*, 1964) which are in excess of the estimated neutral gas temperature for the corresponding time period. In this case the mean ionic masses quoted above would be increased to $m_+ \approx 10$ *AMU* (40°N to 45°N) and $m_+ \approx 8$ *AMU* (35°N to 40°N). In any event, it is quite obvious that the light ionic constituents must become important during the night even at altitudes as low as 500 km.

It is interesting to compare the scale heights at 500 km with the scale heights at 800 km derived in the same fashion. During the night (0400 LMT) the ratio of the scale heights $H'(800)/H'(500) \approx 2.5$ at 35°N to 40°N and about 2.1 at 40°N to 45°N, while during the day (1400 LMT), this ratio is about 1.9 at the lower latitudes and 1.6 at the higher latitudes. Taking into account the altitude variation of the acceleration of gravity g, and making the same assumptions concerning the electron and ion temperature as before, it is again possible to infer the mean ionic mass at 800 km. Table 1 summarizes the inferred mean ionic mass at the two latitude ranges, considered here.

TABLE 1.—*Inferred Mean Ionic Mass* m_+

		0400 LMT		1400 LMT	
Altitude	$T_e/T_{i,n}$	35°–40°N	40°–45°N	35°–40°N	40°–45°N
500 km	1	~7 AMU	~8 AMU		
	>1	~8 AMU	~10 AMU	(16 AMU)	(16 AMU)
800 km	1	~3 AMU	~4 AMU	~6 AMU	~7 AMU
	>1	~3.5 AMU	~5 AMU	~9 AMU	~11 AMU

It is obvious from Table 1 that light ionic constituents (H_e^+, H^+) must become important at 800 km even during the day; during the night H^+ must be an important constituent, at least at the lower latitudes. To infer the relative abundance of ionic constituents becomes somewhat hazardous, since the additional assumption would have to be made that the light ions are in diffusive equilibrium, an assumption which may not be justified even at altitudes as high as 700 km (*Bauer*, 1964). Thus, the mean ionic mass m_+ can only be used as an *indicator* of the presence of ions lighter than O^+, while no reliable and unique data on the abundance of the individual light ions can be inferred from the electron-ion scale height under the present circumstances.

CONCLUSIONS

Although it is not possible by means of satellite observations to obtain directly the true diurnal variation of the topside ionosphere, a *mean* diurnal variation can be constructed with proper precautions from daily observations over an extended period of time. The mean diurnal variation of the topside ionosphere at mid-latitudes based on Alouette observations from October to December 1962 exhibits the following features:

1. There is a significant difference in the diurnal behavior between the latitude ranges 35°N to 40°N and 40°N to 45°N, the boundary of which appears to be associated with the 70° magnetic dip line.

2. The vertical cross section of the topside ionosphere is indicative of an electron-ion scale height *variable* with altitude, latitude and time.

3. There is definite evidence of absence of thermal equilibrium, at least during the daytime, with $T_e/T_{i,n} \approx 2$ at 500 km.

4. The mean ionic mass inferred from the scale height indicates that the light ionic constituents H_e^+ and H^+ are already of importance during the day at 800 km and during the night at an altitude as low as 500 km.

5. There is some indication of opposite latitude gradients in electron temperature and mean ionic mass, the electron temperature increasing with latitude.

REFERENCES

ATKINS, E. A. and CHAPMAN, J. H., Satellite measurements of radio breakthrough, *Can. J. Phys.* **41**, 1388–1393, (1963).

BAUER, S. J., Some implications of a direct measurement of the hydrogen and helium ion distribution in the upper atmosphere, *J. Geophys. Res.* **69**, 3, 553–555, (1964).

BRACE, L., SPENCER, N. W. and DALGARNO, A., Electrostatic probe data and interpretation, **45***th Annual Meeting, AGU*, Washington, D. C. to be published, (1964).

CALVERT, W. and SCHMID, C. W., Topside spread F, to be published, (1964).

BOWEN, P. J., BOYD, R. L. F., RAITT, W. J. and WILLMORE, A. P., Ion composition in the upper F-region, *Proc. Roy. Soc. A.*, (1963).

FITZENREITER, R. J. and BLUMLE, L. J., Analysis of topside sounder records, *J. Geophys. Res.*, **69**, 3, 407–415, (1964).

GLIDDON, J. E. C. and KENDALL, P. C., A mathematical model of the F2 region, *J. Atmo. Terr. Phys.*, **24**, 1073–1099, (1962).

MARIANI, F., Evidence for the effect of corpuscular radiation on the ionosphere, *J. Atmos. Sci.*, **20**, 6, 479–491, (1963).

NICOLET, M., Solar radio flux and temperature in the upper atmosphere, *J. Geophys. Res.*, **68**, 22, 6121–6144. (1963).

WRIGHT, J. W., Diurnal and seasonal changes in the structure of the mid-latitude quiet ionosphere, *J. Res., NBS*, **66D**, 3, 297–312, (1962).

ELECTRON DISTRIBUTION IN THE IONOSPHERE*

N66 37957

ROBERT E. BOURDEAU

1. INTRODUCTION

Under average conditions, the electrons with thermal energies that form the ionosphere result from photoionization of the upper atmosphere by ultraviolet and x-radiation from the sun and possibly by corpuscular radiation. These electrons are lost by recombination with ions simultaneously produced. The loss rate is slow enough that the ionosphere persists throughout the night especially at the higher altitudes.

In addition to the above-described production and loss mechanisms, gravitational and electromagnetic forces contibute to the determination of the altitude distribution of electrons the combination of all these factors is such that several regions with distinctive features are formed. This natural subdivision permits a discussion of the ionospheric electron distribution to be subdivided accordingly. Specifically, as will be done here, it allows for separate discussions of the D (50–85 kilometers), the E (85–140 km) and the F (140–1000 km) regions and of the outer ionosphere which extends from 1000 km to several earth radii.

The electron density (N_e) typically varies from $10^3 cm^{-3}$ in the D region, to $10^5 cm^{-3}$ in the E region, reaches a maximum of $10^6 cm^{-3}$ at 300 km in the F region and then decreases monotonically to $10 cm^{-3}$ at a distance of several earth radii. The principal changes from a given electron density profile will occur with time of day, with season, with temporal position within the solar cycle and with latitude. The relative importance of each of these effects depends upon the ionospheric subdivision under discussion. The dynamic nature of the ionosphere is evidenced by the fact that one of these effects can produce an order of magnitude variation in N_e at specified altitudes. The problem of orderly discussion is complicated further by the very large variations that result from solar storms and from energetic particle precipitation into the auroral region.

In organizing the discussion, it would have been natural to choose a condition of maximum or minimum solar activity and then treat the variations about it. However, observational evidence at least for the regions below 85 and above 300 km is weighted heavily toward the middle of the current solar cycle. For this reason, the term "average profile" is used here for a noontime, mid-latitude ionosphere at an epoch midway between solar maximum and solar minimum conditions. "Average profiles" will be developed for each of the four subdivisions prior to discussions of temporal and latitudinal changes about them. Solar storm and auroral phenomena shall be discussed but treated as special events.

2. AN AVERAGE D REGION ELECTRON DENSITY PROFILE

The D region, even though the most accessible, is the most difficult to study experimentally because it contains the lowest ratio of electron to neutral gas density. The difficulty principally is embodied in the separation of the electron density and collision frequency parameters, which simultaneously contribute to measured phenomena. Separation of these two parameters has been accomplished by experiments involving radio propagation between the earth and a rocket. Ground-based experiments have been carried out by cross-modulation and partial-reflection techniques. One common denominator of ground-based methods is that the altitude dependence of electron density is extracted from the experimental data only by assuming an electron collision

*Published as *Goddard Space Flight Center Document X-615-64-29*, January 1964. This work was also published in "International Dictionary of Geophysics" Pergamon Press, Oxford, England

frequency profile, usually that measured with rocket experiments by Kane (1961). Rocket-borne dc probes have been used but the theoretical analysis is based on the assumption of a collision-free plasma sheath, an unlikely condition at D region altitudes.

Few D region profiles have been reported for the "average" conditions. An average profile applicable to a quiet midday ionosphere is pre-

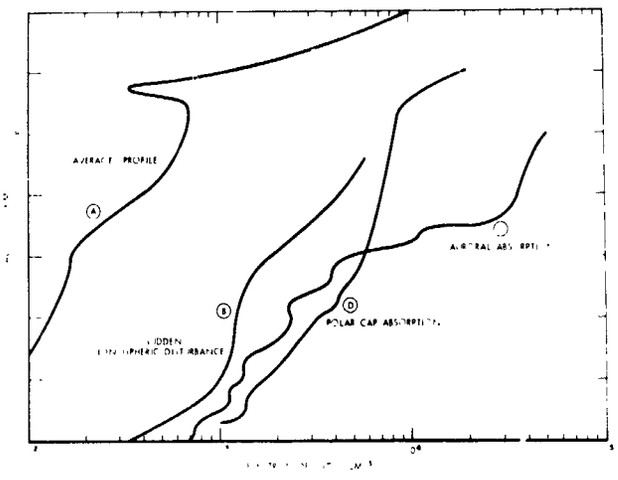


FIGURE 1.—D region electron density profiles

sented as curve A in Figure 1. It was constructed from the space flight results of Aikin et al (1963) and of Hall (1963) and from the ground-based observations of Belrose (1961) and of Barrington et al (1962). Except for the latter results above 75 km, there is agreement to better than a factor of two between all data sets. The agreement is encouraging in view of the somewhat different times and locations at which the data were obtained.

The "average" profile is characterized by two separate gradients which could permit different nomenclature for the region below 70 km and the region between 70 and 85 km. It additionally is characterized by an often seen minimum located close to the mesopause (about 84 km) which tends to separate the D and E regions. The maximum value of N_e in the D region (N_mD) for the stated conditions averages 700 electrons cm^{-3}.

A note of caution needs to be injected in the discussion of this average D region profile. The data on which it is based were obtained at what would be considered mid-latitudes from the standpoint of a neutral atmosphere. However, it must be emphasized that most of the results were obtained at magnetic dips greater than 70°. Consequently, it justifiably could be argued from the viewpoint of ionospheric physics that the profile is more representative of polar conditions, in the absence of solar storms and of auroral and magnetic disturbances.

3. VARIATIONS FROM THE AVERAGE D REGION ELECTRON DENSITY PROFILE

Observational D region data are so sparse that temporal and latitudinal variations are principally hypothetical. Excluding special events, the *diurnal* variation is the major effect, the D region being essentially non-existent at night, at least from the standpoint of the N_e parameter. The results of Barrington et al (1962) indicate rapid formation of the D region just after layer sunrise.

We only can conjecture about the *solar cycle* variation at this time. If the prevailing theory* for D region formation is correct, then a solar cycle change of less than a factor of two in the entire reference N_e profile would be expected because of the relative constancy of the responsible ionizing radiation (cosmic rays below 70 km and Lyman alpha radiation between 70 and 85 km). Moderate (less than a factor of 2) *latitude* variations should exist in the lower D region, in accordance with the latitudinal dependence of the flux of cosmic rays.

We turn now to the large electron density enhancements that occur in the D region during special events and which generally are associated with radio-blackouts. Belrose and Cetiner (1962) have measured the enhancement during a Sudden Ionosphere Disturbance* (Curve B of Figure 1), an event occurring on the sunlit side of the earth simultaneously with the appearance of a flare and having a time duration of about one hour. It is seen that there is a general increase in N_e of about a factor of 10 during such periods of enhanced X-ray activity. Even larger electron density increases have been observed with rocket experiments flown in connection with high latitude phenomena. Jespersen et al (1963) have obobtained N_e data during an auroral absorption

*See D Region Chapter.

event* (curve C of Figure 1) which is associated with a magnetic disturbance and the precipitation of energetic electrons into the D region. Jackson and Kane (1959) obtained an electron density profile (curve D of Figure 1) during a Polar Cap Absorption* (PCA) event which is produced by energetic proton precipitation following certain types of solar flares. It should be noted that the PCA data were obtained during the early phases of the event. Ground-based measurements of radio absorption indicate much larger electron density enhancements during the main phase of a PCA.

Knecht (1963) has called attention to a fourth type of anomalously high D region absorption event observed frequently during the winter at temperate geographic latitudes. Electron density increases of up to a factor of ten have been reported in instances not associated with magnetic disturbances or a PCA event. In general, since these and seasonal variations in N_e have been inferred from ground-based techniques, the N_e profiles likely should be treated as qualitative because the measured phenomena might also represent temporal changes in the electron collision frequency parameter, which was assumed to be constant in deriving N_e.

4. AN AVERAGE E REGION ELECTRON DENSITY PROFILE

The altitude interval ascribed here to the E region lies between 85 and 140 km. The largest body of data concerning its behavior has been obtained by use of the conventional ground-based ionosonde. These results are particularly important to our understanding of the geographic and temporal variations of the E region under sunlit conditions. Unfortunately, the ionosonde experiment is not well suited for the determination of the detailed altitude dependence of E region electron densities, especially under nighttime conditions.

Our knowledge of the fine structure of the E region electron density profile has come from rocket experiments. These have been of two general types: radio propagation between the rocket and the earth and direct-sampling probes. It generally is recognized that the propagation experiments produce the most accurate profiles because they are relatively free of possible errors originating from the disturbance introduced into the medium under study by the rocket carrier. However, quite recently the probe experiments have been "calibrated" by including them on the same rocket flights that contain propagation experiments. This lends confidence to the validity of nighttime electron density profiles discussed in the succeeding section and which have been obtained from probe experiments because of their inherent high sensitivity.

An E region N_e profile representing the average obtained by many rocket experiments is identified as the upper portion of curve A in Figure 2. It is evident that the E region profile contains two distinct altitude intervals. The first (85–100 km) is characterized by one of the largest positive electron density gradients a(bout 10^4 electrons cm^{-3} per kilometer) found in the ionosphere. Most theoreticians ascribe the electrons found in this altitude interval to the combined effects of Lyman β (1026 Å) and x-radiation (100–30 Å).† The second region (100–140 km), believed to result from the combined effects of x-rays and of radiation principally in the 1027–911 Å portion of the ultra-violet spectrum,† is characterized by a relatively constant electron density. In fact, it is the absence of a significant "valley" in the daytime profiles that represents one of the most important results of the early rocket experiments. This feature has made it quiet difficult to discuss separately the E and F regions of the ionosphere.

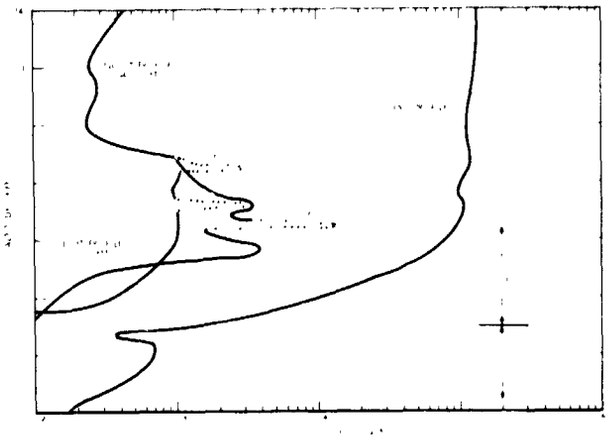


FIGURE 2.—D and E region electron density profiles

†See E Region Chapter.

5. VARIATIONS FROM THE AVERAGE E REGION ELECTRON DENSITY PROFILE

The major change from the average E region profile is *diurnal* in nature. It is possible from ionosonde data to construct a detailed temporal variation at specific latitudes of the maximum electron concentration (N_mE) in the E region for the sunlit hours. There is a symmetrical variation about a noontime value of 1.2×10^5 electrons cm^{-3} down to 5×10^4 electrons cm^{-3} near the sunrise and sunset hours. Two nighttime profiles obtained from rocket experiments by L. G. Smith (1962) are presented in Figure 2, curve B representing an evening condition and curve C a predawn situation. With the exception of the much lower value of N_mE, the features that best distinguish the nignttime profiles from the average profile are (a) the existence of a "pronounced" value above 110 km which extends into the F1 region and (b) the irregular fine structure. Profiles diurnally spaced as in Figure 2 have been used to infer the rate of destruction of electrons in the E region. They are believed to be lost mainly by recombination with diatomic ions.* An effective E region recombination coefficient of $2\times10^{-8}cm^3sec^{-1}$ has been computed from data such as that contained in Figure 2.

Solar cycle, seasonal and *latitude* variations of N_mE are small compared to a complete diurnal variation. This conclusion from ionosonde data is quite important since it tends to narrow the hypotheses concerning formation of that portion of the E region lying above 100 km to those which include ionization sources having quite stable long term characteristics. Additionally, there is an important implication in favor of a time independent structure of the neutral atmosphere. Ionosonde data indicate that the noontime value of N_mE is only about 50 percent larger during sunspot maximum than during sunspot minimum. There are seasonal variations in N_mE of up to a factor of two at midlatitudes with the maximum occurring in June and the minimum in December in the northern hemisphere.

The perturbations in E region electron density during solar flares, magnetic storms and auroral events are known from ionosonde data to be small compared to the above-described diurnal, solar cycle and seasonal effects. This most likely is because such events are associated with energetic radiation which penetrates into the D region before ionization results. The most frequent E region anomaly is not necessarily associated with such events. Rather, it exhibits a rather random temporal variation which has led to the nomenclature of sporadic-E or E_s ionization.

One common form of E_s ionization is a thin layer in which the electron density is considerably higher than the region immediately below and above. Rocket experiments have best described the altitude characteristics of this type of E_s ionization, examples of which are illustrated by the dashed portions of the nighttime profiles shown in Figure 2. It is seen that the N_e enhancement can be as large as 6 and that the thickness of the layer measured at half the peak electron density can be smaller than 0.5 km. Larger N_e enhancements have been observed and there is evidence that the horizontal extent sometimes is measured in hundreds of kilometers. Another common form of E_s ionization is characterized by large electron density gradients lying below a region of relatively constant electron density.

Smith (1957) and Seddon (1962) have reviewed the occurrence probability of E_s ionization. Seddon finds that the preferred altitudes are 100, 105, 111, 117 and 129 km, three of which are approximately confirmed by the nighttime profiles shown in Figure 2. Smith finds that E_s is principally a nighttime phenomenon in and near the auroral zone but conversely that it is more a permanent daytime feature in a narrow band centered at the geomagnetic equator. Here we caution again that auroral characteristics might extend to magnetic dips as low as 70° and consequently in North America down to what normally is considered mid-latitudes. At the middle magnetic latitudes, the E_s occurrence probability is mostly seasonal in nature with the predominance is summer months.

There undoubtedly are several causative mechanisms for E_s phenomena, the most likely being the combined existence of wind shear and magnetic field for the mid-latitudes and two stream plasma instability for the equatorial and perhaps the auroral zones (cf Knecht, 1963).

*See E Region Chapter.

6. AN AVERAGE F REGION ELECTRON DENSITY PROFILE

Of all the ionospheric subdivisions, the F region is the most variable. Diurnal, solar cycle and latitudinal effects can each produce changes of almost an order of magnitude. Additionally, seasonal variations are quite large. Thus, it should not be surprising if the average F region profile presented here is rarely observed.

The average F region profile, constructed from rocket results obtained under the appropriate conditions, is superimposed on the average D and E region profiles as curve A in Figure 3. As

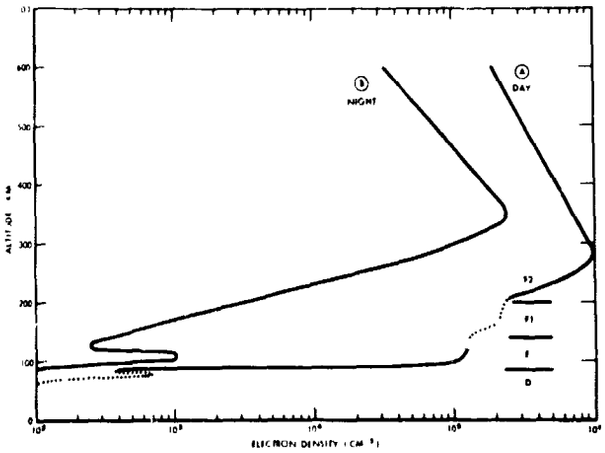


FIGURE 3.—Typical D, E, and F region electron density profiles

shown, the F region contains the altitude ($h_{max}F2$) of the maximum electron concentration ($N_{max}F2$) found in the ionosphere. Average values for $h_{max}F2$ and $N_{max}F2$ are 280 km and 10^6 electrons cm^{-3}, respectively. The daytime sub-peak region can be divided into two altitude intervals, the bases of which are identified by positive electron density gradients. The classical nomenclature for these subdivisions are F1 (approximately 140–200 km) and F2 (above 200 km).

It is likely that the electrons in the F region are produced by the same portion of the ultraviolet spectrum as for the upper E region. The N_e gradients that identify the F region probably are the result of changes from electron loss through direct recombination to loss mechanisms involving an intermediate process of ion-atom interchange prior to recombination.

In addition to electron production and loss, a third process (diffusion) becomes effective at higher F2 altitudes and contributes to the formation of the F2 peak. Diffusion or motion of the electron-ion gas can take place in accordance with gravitational or electro-magnetic forces. Diffusion is more effective than electron production or loss at altitudes above the F2 peak and is the reason for the exponential decrease of N_e with altitude.*

Our average F2 region can be defined in the Northern Hemisphere as lying between 40° and 70° magnetic dip. This permits us to minimize the corpuscular radiation effects that influence auroral ionosphere and also to neglect diffusion of the thermal electrons horizontally along magnetic field lines, a low latitude phenomenon. In this restricted reference latitude region, the tendency is for the upper ionospheric electrons to be distributed in accordance with diffusive equilibrium. Specifically, the profile above $h_{max}F2$ is heavily dependent on the mean ionic mass and the average of the electron and ion temperatures. The reference upper ionosphere profile in Figure 3 is characteristic of that obtained in the presence of a single ionic constituent (O^+) and an average isothermal electron-ion temperature of 1600°K.

7. VARIATIONS FROM THE AVERAGE F REGION ELECTRON DENSITY PROFILE

The existence of a distinct F1 region is only a daytime feature of the ionosphere. Furthermore, there is ionosonde evidence that even its very existence under daytime conditions is seasonally dependent. Its occurrence probability is highest in summer at solar maximum and is extremely low in winter at solar minimum. When it does exist, the maximum electron density values ($N_{max}F1$) show only a small dependence on temporal position within a solar cycle.

The behavior of $N_{max}F2$ and of $h_{max}F2$ is extremely complicated. At mid-latitudes, during winter months, $N_{max}F2$ decreases by a factor of about 4 from day to night. This diurnal effect is accompanied by an increase in $h_{max}F2$ of about 70 km. These features are brought out in Figure 3 by comparison of curves A and B which respectively represent day and night conditions. It

*See F Region Chapter.

should be noted that at night the breadth of the F2 peak is smaller and more importantly, that above the F2 peak there is a more rapid decrease of N_e with altitude. The more rapid decrease is brought about by a corresponding reduction of the electron and ion temperatures.

It is known from ionosonde data that, at mid-latitudes in winter, there is up to an order of magnitude decrease in N_{max} from solar maximum to solar minimum. The solar cycle effect is smaller in summer than in winter. To illustrate further the extreme variability of the F region we are faced with a seasonal anomaly, wherein for a given epoch there is a decrease in N_{max} of up to a factor of four from winter to summer.

Because it only has been recently explored to any extent, detailed temporal variations of the upper F region consist only of diurnal studies based principally on Alouette Topside Sounder Satellite results. Observational Alouette data (Bauer and Blumle, 1964) illustrating the diurnal variation of N_e for altitudes between 400 and 1000 km are shown in Figure 4. They are applicable to October-December, 1962, and thus are more representative of solar minimum conditions. The data apply only to geographic latitudes 35–40°N over North America which lie within our definition for average F2 conditions. The most striking feature is that the diurnal variation decreases with altitude. This may be caused by the relative importance of the O^+, He^+, H^+ ions which vary between night and day in such a way that the lighter constituents become more important at night. The lower mean ionic mass at night tends to overcome the effect of lower charged particle temperatures and lower values of $N_{max}F2$ in determining the electron density at higher altitudes.

As we have implied previously, there is a pronounced latitude effect on the F2 region. The existence of an ionospheric equatorial anomaly was first detected by ionosonde experiments which since have provided details of its bottomside characteristics. A more detailed description of the anomaly has come from the dramatic results of the Alouette Topside Sounder Satellite. An idealized representation of the diurnal behavior of this anomaly over South America is illustrated in Figure 5 (Jackson, 1964). It was prepared by combining the satellite results of Lockwood and Nelms (1964) with ionosonde data by Wright (1962). Plotted are idealized isoelectronic contours as a function of altitude and latitude for

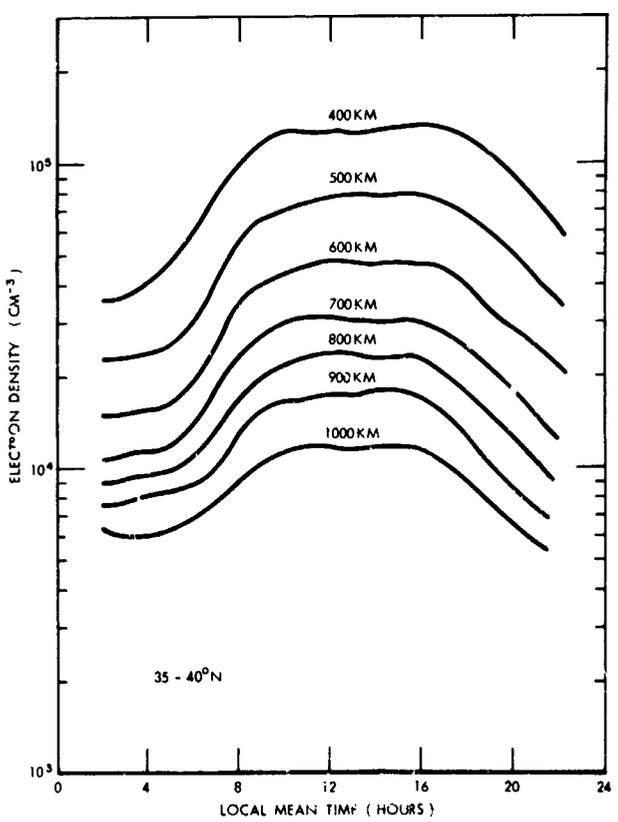


FIGURE 4.—Diurnal variation of electron density in the upper ionosphere (Bauer and Blumle, 1964)

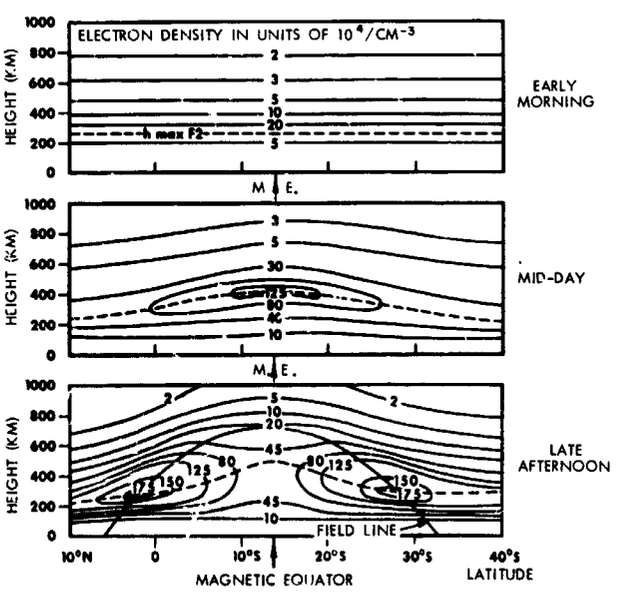


FIGURE 5.—Idealized representation of the equatorial anomaly (Jackson, 1964)

three diurnal conditions representative of the late 1962 period.

It is seen from Figure 5 that the anomaly is mainly a daytime feature. For midday conditions, a pronounced dome or peak in electron density occurs over the geomagnetic equator with a corresponding increase in $h_{max}F2$. In late afternoon, the equatorial dome is the dominant feature only at altitudes above 600 km. Below this altitude, two N_e peaks are located symmetrically about the geomagnetic equator along a specific magnetic field line. King et al (1963) have made satellite studies of the anomaly near Singapore with somewhat different results. They find that the formation of the double peak occurs much earlier in the day for the easterly longitudes. The diurnal behavior of the equatorial anomaly suggests that its characteristics are closely related to the competition between electron production which tends to maximize N_e at the sub-solar latitudes and diffusion along magnetic field lines which tends to produce a symmetrical N_e distribution about the geomagnetic equator.

The equatorial anomaly causes the appearance of small ionization ledges to appear on N_e altitude profiles wherever the appropriate magnetic shell appears. The appropriate shell for this case using the McIlwain (1961) notation occurs at magnetic L values which vary from 1.06 to 1.18. This particular field-aligned ionization enhancement has been varified experimentally by direct measurements of N_e on the Ariel satellite (Sayers et al, 1963) and by Alouette satellite results (King et al, 1963). The latter have identified additional field-aligned ledges detected from the Ariel and/or Alouette satellite results with the region of maximum flux of electrons from the artificial radiation belt ($L=1.22\pm .02$), with the heart of the inner radiation belt ($L=1.6\pm0.1$) and with the most intense portion of the outer radiation belt ($L\cong4.0$). Other ledges appear at $L=2.2\pm0.1$ and at $4.5<L<6.5$.

At higher latitudes, the electron scale height at 500 km increases markedly with latitude from what would be inferred from Figure 4 (King et al, 1963). There also is evidence (Bauer and Blumle, private communication) that at the higher latitudes the diurnal variation of N_e above $h_{max}F2$ is relatively constant with altitude, in contrast with the middle latitude results represented by Figure 4.

8. ELECTRON DENSITIES AT VERY HIGH ALTITUDES

Very few electron density distributions have been obtained for the region above 1000 km at magnetic dips 40°–70°. The few that do exist (cf Hanson, 1962; Bauer and Jackson, 1963; Donley, 1963) have been interpreted in terms of an isothermal ionosphere in diffusive equilibrium.

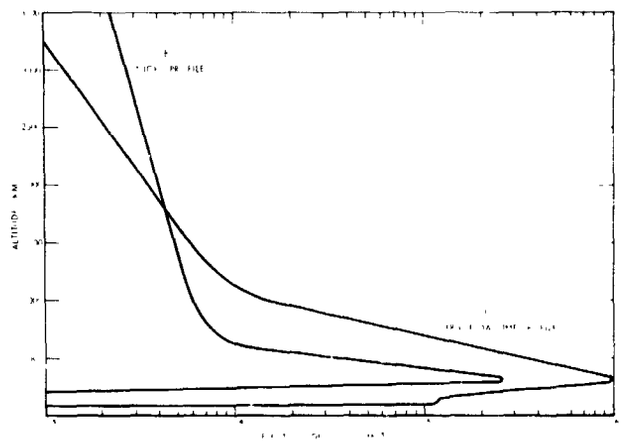


FIGURE 6.—Electron density profiles in the outer ionosphere at mid-latitudes

In Figure 6, two such profiles, one for a daytime (curve A) and the other for a nighttime conditions (curve B) have been superimposed on a typical lower ionosphere profile. The distribution above 1500 km for the daytime case corresponds to what would be expected for a plasma predominated by He^+ ions and for an average electron-ion temperature of 1600°K. This infers that protons for the daytime ionosphere in the middle of the solar cycle do not become effective below 3500 km (cf Hanson, 1962). The distribution above 1000 km for the nighttime case, on the other hand, corresponds to H^+ ions and an average electron-ion temperature of 800°K. The two profiles suggest that electrons are more abundant at the very high altitudes for a nighttime than for a daytime condition. It should be remembered, however, that the available rocket data on which these profiles depend were obtained at different portions of the solar cycle. Still, it is also possible to infer greater electron densities at night at altitudes above 1500 km by extrapolating profiles

constructed from the diurnal F region variations of Figure 4. This conclusion does not necessarily apply at other than mid-latitudes.

More extensive observations of magnetospheric electron densities have resulted from whistler observations. Carpenter (1963) has summarized the whistler measurements of magnetospheric electron densities for equatorial latitudes. His "smoothed" curve is presented in Figure 7.

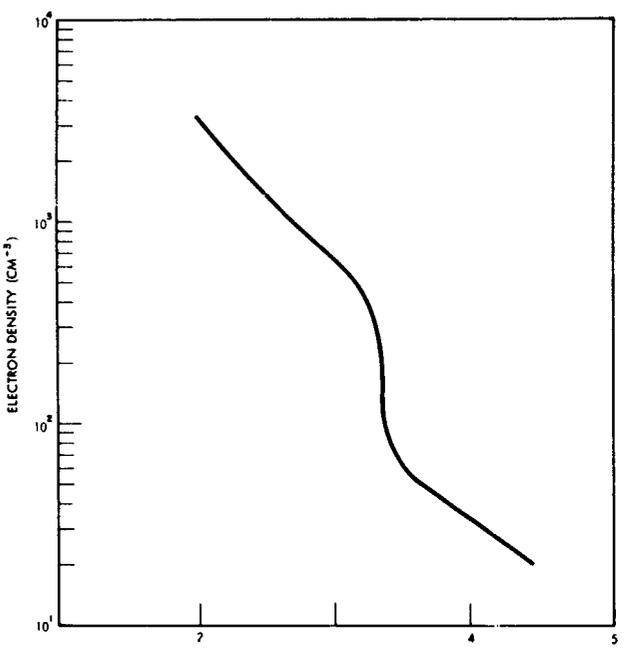


FIGURE 7.—Equatorial profile of magnetospheric electron density (Carpenter, 1963)

The most pronounced characteristic is the knee located at about $3R_E$ where the density drops by a factor of 6 or more. Carpenter finds these results to be consistent with the above-described rocket results, with the USSR results from the Lunik II spacecraft (Gringauz et al, 1960), and with the ground-based incoherent backscatter results of Bowles (1962). He suggests that the position of the "knee" is quite variable, moving from $2R_E$ during high magnetic activity to 7–8 R_E during low magnetic activity. One important implication from Figure 7 is that if one defines the ionosphere as a region containing a significant number of very low energy electrons, then the ionosphere extends out to several earth radii.

9. ACKNOWLEDGEMENTS

The stimulating discussions with J. E. Jackson of the Goddard Space Flight Center during the preparation of this review is gratefully acknowledged. The author also is grateful to S. J. Bauer and L. Blumle for access to their data prior to publication.

10. REFERENCES

1. AIKIN, A. C., J. A. KANE and J. TROIM, Space Research IV, in press, 1963.
2. BARRINGTON, R. E. and E. V. THRANE, J. Atmos. Terr. Phys., 24, 31, 1962.
3. BAUER, S. J. and L. J. BLUMLE, private communication, to be published, 1964.
4. BAUER, S. J. and J. E. JACKSON, J. Geophys. Res., 67, 1675, 1962.
5. BELROSE, J. S., Proc. International Symposia on Ionospheric Sounding, 1961.
6. BELROSE, J. S. and E. CETINER, Nature, 195, 688, 1962.
7. BOWLES, K. L., NBS Report 7633, 1962.
8. CARPENTER, D. L., Proc. of the XIV URSI General Assembly, 1963, in press.
9. DONLEY, J. L., J. Geophys. Res., 68, 2058, 1963.
10. GRINGAUZ, K. I., V. G. KURT, V. I. MOROZ and I. S. SHLOVSKII, Astronomicheskii Zhurnal, 37, 716, 1960.
11. HALL, J. E., Proc. NATO Advanced Study Institute, Skeinkampen, Norway, in press, 1963.
12. HANSON, W. B., J. Geophys. Res., 67, 183, 1962.
13. JACKSON, J. E., NASA Technical Note, 1964, in press.
14. JACKSON, J. E. and J. A. KANE, J. Geophys. Res., 64, 1074, 1959.
15. JESPERSEN, M., O. PETERSEN, J. RYBNER, B. BJELLAND, O. HOLT and B. LANDMARK, Norwegian Space Res. Comm. Rept. No. 3, 1963.
16. KANE, J. A., J. Atmos. Terr. Phys., 23, 338, 1961.
17. KING, J. W., P. A. SMITH, D. ECCLES and H. HELM, DSIR Radio Research Station Report RRS/I.M. 94, 1963.
18. KNECHT, R. W., Proceedings of the XIV URSI General Assembly, 1963, in press.
19. LOCKWOOD, G. E. K. and G. L. NELMS, J. Atmosph. Terr. Phys., 1964, in press.
20. MCILWAIN, J. Geophys. Res., 66, 3621, 1961.
21. SAYERS, J., P. ROTHWELL and J. H. WAGER, Nature, 198, 230, 1963.
22. SEDDON, J. C., Ionospheric Sporadic E, Pergamon Press, Oxford, 909, 1962.
23. SMITH, E. K., NBS Circular 582, 1957.
24. SMITH, L. G., Geophys. Corp. of America., Tech. Rept. 62-1-N, 1962.
25. WRIGHT, J. W., NBS Technical Note 138, 1962.

N65-32074

RESEARCH WITHIN THE IONOSPHERE*

ROBERT E. BOURDEAU

As in the other space sciences, our understanding of the mechanisms which govern the characteristics of the earth's ionosphere has been enriched by the recent, golden opportunities to perform experiments on rockets and satellites which pass through the medium under study. This understanding also has been improved by the use of new and exciting ground-based experiments.

Traditionally, the ionosphere is defined as that portion of the upper atmosphere which contains a significant number of charged particles with thermal energies (tenths of an electron volt or less). Ionospheric charged particles, electrons and ions, result from ionization of the neutral constituents by ultraviolet and X-radiation from the sun and possibly by corpuscular radiation. The electrons are lost by recombination with the positive ions that are simultaneously produced. The loss rate is slow enough that the ionosphere persists throughout the night especially at the higher altitudes. Because of the high electron number density, the ionosphere classically is associated with its effects on radio communication processes.

In addition to the above-described production and loss-mechanisms, gravitational and electromagnetic forces contribute to ionospheric characteristics. The combination of all these factors is such that several regions with unique features are formed. This natural subdivision permits a discussion of the ionosphere to be subdivided. Specifically, as will be done here, it allows for separate discussions of the D (50–85 kilometers), the E (85–140 km) and the F (140–600 km) regions and of the upper ionosphere which extends from 600 km to several earth radii.

The D, E and that portion of the F region which lies below about 300 km were identified and named as a result of research with the classical tool of ionospheric research, the ground-based ionosonde. With this "low-frequency radar," one measures the time between transmission of a radio signal from the earth to reception of the echo reflected from the ionosphere. The electron density at the point of reflection is proportional to the square of the frequency. Therefore, when this is done as a function of frequency it is possible to obtain electron density as a function of altitude. We have learned much about the temporal and latitudinal behavior of the electron distribution in the lower ionosphere through long term world-wide use of ionosonde apparatus.

It is the purpose of this paper to summarize recent advances in our understanding of the ionosphere with emphasis on those that have come about as a result of the ability to place observatories in the medium under study. This necessarily involves relating the detailed altitude, latitude and temporal variations of all characteristics of the thermal-charged particles to gravitational and electromagnetic forces, to possible ionizing sources, and to the nature of the neutral atmosphere from which the ions and electrons are created.

THE NEUTRAL ATMOSPHERE

The neutral atmosphere classically is divided into regions in accordance with the variation of temperature with altitude. The altitude dependence of temperature is represented for average daytime conditions[7] in Fig. 1a. Free electrons are found in significant abundance only above 50 km. Consequently, the regions of the neutral atmosphere that are of major interest to the ionospherici st are the mesosphere, which lies between the temperature maximum at 50 km and the minimum near 85 km, and the thermosphere (above 85 km).

*Published in *Science*, 148 (3670): 585–594, April 30, 1965.

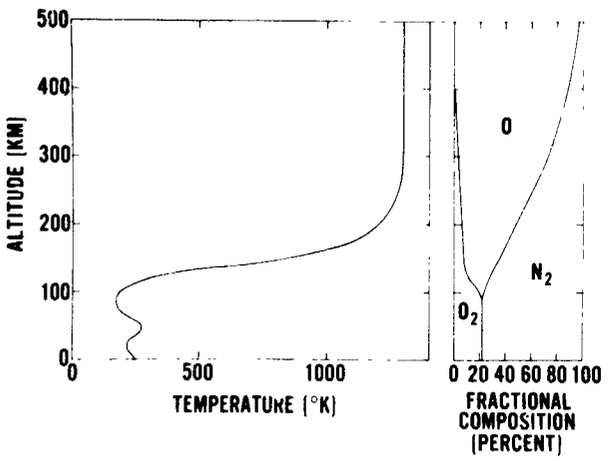


FIGURE 1.—Altitude dependence of neutral gas temperature and fractional composition (Johnson, 1962).

Mesospheric temperatures have been deduced principally from rocket-borne pressure gages and by sound-velocity experiments using rocket-borne grenades. These data show that the mesosphere exhibits large latitudinal and seasonal variations.[1] The large increase of temperature in the lower thermosphere is due principally to the absorption of solar ultraviolet radiation. Heat conduction keeps the temperature nearly constant above 200 km.[2] Thermospheric temperatures have been deduced mainly from atmospheric density measured by use of gauges flown on rockets[3] and satellites[4,5] and indirectly by studying satellite orbital decay.[6]

From the standpoint of theories of formation of the ionosphere, the most important parameter of the neutral atmosphere is its composition. An average percentage distribution of the major constituents is presented[7] as a function of altitude in Fig. 1b. Below 100 km, mixing controls the relative abundance of the neutral constituents and consequently molecular oxygen and nitrogen predominate. Above this altitude, dissociation of atomic oxygen takes place as a result of the absorption of ultraviolet radiation by O_2. At the higher altitude mixing becomes unimportant and the constituents are in diffusive equilibrium, each component being distributed independently of the others. The distribution of the constituents can be calculated theoretically from [illegible] hydrostatic equation using an assumed altitu[illegible] for diffusive separation and assumed atmospheric temperatures.

In Fig. 1b, it is shown that the molecular constituents diminish in importance with increasing altitude so that atomic oxygen dominates the atmosphere at 500 km. Above 500 km, the lighter gases become important. Up until an analysis of atmospheric drag on the ECHO I satellite, it was believed that there is a transition directly from an oxygen to a hydrogen atmosphere. However, this analysis[8] first suggested the existence of an intervening helium layer. The importance of neutral helium at the higher altitudes was first confirmed[5] by mass spectroscopy on the Explorer 17 satellite, some time after ionized helium had been detected from rocket[9] and satellite[10] experiments. The thickness and altitude of the helium region should be a strong function of atmospheric temperature.[11] This is reflected in the graphs (Fig. 2a) of mean molecular weight as a function of altitude for the diurnal

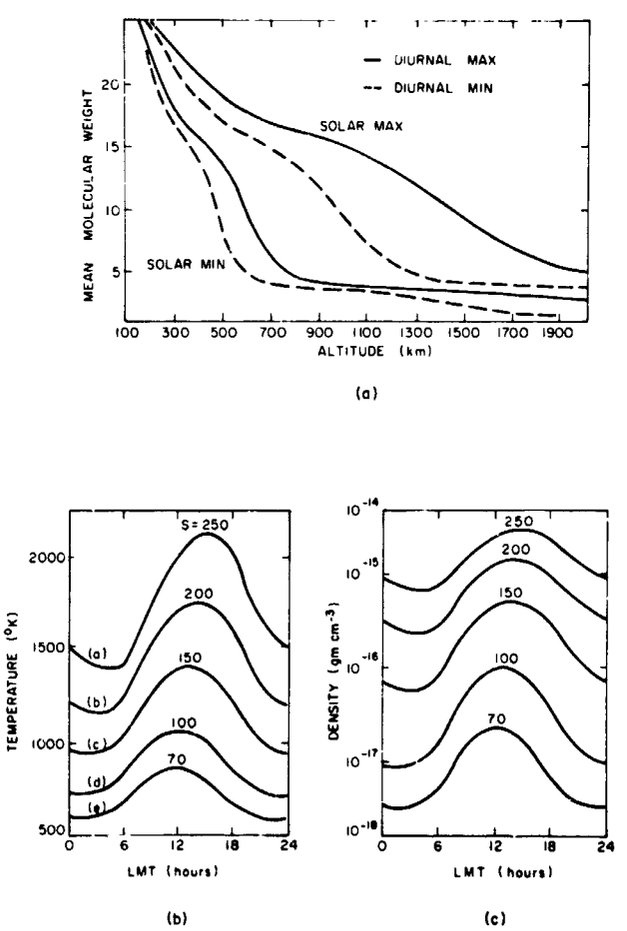


FIGURE 2.—Time-dependent model of neutral gas (a) mean molecular weight, (b) thermospheric temperature, and (c) 600 km density, Harris and Priester (1962).

and solar cycle extremes.[12] The region is believed to be diminishingly thin, for example, at night during the year of minimum solar activity.

The structural behavior of the atmosphere shown in Figs. 1b and 2a has been inferred principally from total density measurements using rocket-borne gages and analyses of satellite drag. Early rocket-borne mass spectrometers[13] qualitatively established that diffusive equilibrium controls the composition at the higher altitudes. However, with the early experiments was associated a high probability of errors due to surface recombination within the instruments. Consequently, it has been only in the last year that quantitatively significant measurements of the O/O_2 and O/N_2 ratios were obtained by direct sampling.[5,14,15]

Satellite drag observations show that the temperature and density of the isothermal region of the thermosphere varies considerably with time of day and with the 11-year solar cycle (Figs. 2b and 2c). The five curves[12] in Figs. 2b and 2c are for different levels of solar activity, an index of which is the 10.7 cm flux (S) measured at the earth's surface. Values for S range from 70 at sunspot minimum to 250×10^{-22} $w\ m^{-2}$ $(\text{cps})^{-1}$ during the year of maximum solar activity.

THE NORMAL D REGION

The D region, which occupies approximately the same altitude interval as the mesosphere (50–85 km), is the lowest region where a significant number of free electrons are found. Here, the relatively dense atmosphere results in a high frequency of collisions between the electrons and the neutral constituents. Consequently, there is a high probability that electromagnetic energy which has been transferred to the electrons will be lost irretrievably in these collisions. Thus, the D region acts as an absorber of radio waves and from this standpoint is the most important ionospheric subdivision. Yet it is the least studied experimentally, primarily for two reasons: (a) the difficulty of devising experiments which are valid in such a weakly ionized medium and (b) the trend on the part of most ionosphericists to perform the more esoteric satellite experiments in the upper ionosphere.

In this section, we shall discuss only the "normal" D region constraining the conditions geographically to mid-latitudes and temporally to times free of solar flare effects. It wasn't until quite recently that even a preliminary model for the altitude distribution of electrons in the normal D region has evolved, despite the fact that the region is accessible with relatively inexpensive rockets.

The normal D region abundance is too low to permit study by use of the conventional ground-based ionosondes. However, breakthroughs have been accomplished as a result of the development of more complex ground-based radio propagation experiments.[16,17] However, one common denominator of ground-based methods is that the altitude dependence of electron density (N_e) is extracted from the data only by assuming an electron collision frequency (ν) profile since both N_e and ν simultaneously affect the measured radio propagation phenomena. On the other hand a collaborative effort on the part of a team of Goddard and Scandinavian investigators[18,19] has resulted in novel experiments involving transmission of radio signals from the ground to rocket-borne receivers. The in-situ reception featured by these complex experiments permits unique separation of N_e and ν with adequate sensitivities for the low densities found in the normal D region.

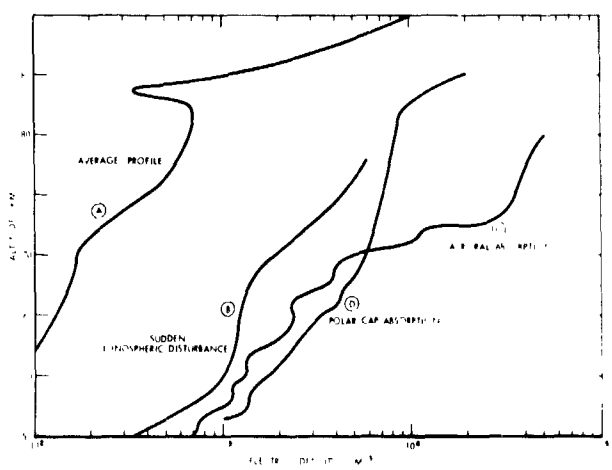


FIGURE 3.—D region electron density profiles.

In Fig. 3, we have combined the few available rocket and ground-based observations to generate an average N_e profile for the normal daytime D region (Curve A). This profile is only one of the

important pieces of information required to explain how the region is formed. To complete the task requires that one relates N_e to the competition between electron production and loss for each discrete altitude.

Electron production (q) can be computed from a knowledge of (a) the intensity of the ionizing radiation, (b) the density of the ionizable constituents responsive to this radiation and (c) the absorption cross-sections of the ionizable constituents. The problem of estimating q is complex because the cross-section of each individual constituent is a different function ofwavelength if solar radiation is the ionizing source or of electron or proton energy for the case of corpuscular radiation.

In the D region, electron loss is believed to occur mainly through dissociative recombination of electrons with positive ions leading to an excited but neutral constituent. To estimate electron loss requires a knowledge of the recombination rates which are different for each ion species.

As a result mainly of rocket measurements of solar radiation, theoretical models[20] narrow the sources of the normal D region to three individual or combined possibilities. It generally is accepted that the lower part (50–70 km) is produced by the action of cosmic rays on the principal neutral constituents (O_2 and N_2). If so, this region should show a strong latitude dependence. There is some disagreement as to the relative roles of the remaining two sources in ionizing the upper part of the D region. Here, one possibility involves the ionization of O_2 and N_2 by 2−8 Angstrom X-rays, an extremely variable source with a very low intensity (not exceeding 10^{-3} erg cm^{-3} sec^{-1} for a flareless sun). The other possibility is Lyman alpha radiation (1216A), the only ultraviolet source for which there is a favorable cross-section and which by rocket tests has been observed to penetrate into the D region. This stable but intense source measured in a few ergs cm^{-3} sec^{-1} acts only upon a trace constituent, nitric oxide. Rocket and satellite measurements indicate that the X-ray fluxes at the extremes of the solar cycle vary by more than two orders of magnitude.[21] Therefore, the relative importance of X-rays and Lyman alpha radiation to the formation of the normal D region may depend on position in the solar cycle.

There have been few spaceflight input-output experiments where the ionization source and the ionization characteristics have been simultaneously measured. According to some observers[18] who measured N_e on a rocket simultaneously with Lyman alpha flux, X-radiation can be ruled out as a significant source of the normal D region at least for the minimum of the solar cycle. To be certain of the relative importance of these two sources at all times, we desperately need laboratory investigations which can resolve existing uncertainties in our knowledge of the absorption cross-sections and recombination rates.

An important tool in ionospheric research is the ion spectrometer, a very difficult experiment to carry out in the D region—so difficult that it has been carried out only once[22] and this during the last year. Other than possible contaminants borne aloft by the rocket, the major ionic constituent observed below 83 km was NO^+. This observation supports the Lyman alpha hypothesis but does not rule out, a priori, ionization by X-radiation because of the possibility of ion-molecule interaction.[23] For example, O_2^+ produced directly by X-rays can react with an N_2 molecule to form an NO^+ ion and an NO molecule. Thus, until the various reaction rates are better known, the ion spectrometer observation does not permit a choice between the Lyman alpha or the X-ray hypothesis.

SPECIAL D REGION EVENTS

There are many phenomena which can enhance the D region electron abundance by up to more than two orders of magnitude with associated electromagnetic wave attenuation strong enough to produce radio blackouts. Simultaneously with the appearance of a solar flare, increased absorption is observed in the D region on the sunlit side of the earth for periods lasting up to approximately one hour. The causative mechanisms for these sudden ionospheric disturbances (S.I.D.) were not established until investigators accurately timed rocket launchings during the course of a flare and observed enhanced X-ray activity penetrating as low as 30 km.[24] The dominant role of solar X-rays in the production of S.I.D.'s has been confirmed by correlating visible flares with satellite observations of enhanced X-ray fluxes and with increased radio absorption.[21] An order of

magnitude increase in D region ionization during an S.I.D. has been measured with ground-based techniques[16] (Curve B of Fig. 3) and estimated theoretically.[20]

At high latitudes, enhanced radio absorption occurs during auroras. This is attributable to enhanced ionization resulting from direct and indirect (bremsstrahlung) effects of precipitating energetic electrons.[23] Evidence for up to a two order of magnitude increase in electron density has been obtained[19] by timing rocket flights to occur during such auroral absorption events (Curve C of Fig. 3). Another type of absorption takes place above the auroral zone and has been correlated with satellite measurements[25] of enhanced energetic proton fluxes during certain types of solar flares. There has been one rocket measurement[26] of enhanced electron densities during such polar cap absorption events (Curve D of Fig. 3).

Increased radio absorption associated with solar flares is often observed in the winter at middle latitudes. The causes of these events are uncertain. It has been suggested from ground-based observations[16] that these are associated with increases in electron density. However, absorption may also be associated with changes in electron collision frequency. Significant changes in ν have been correlated with measured pressure variations in the stratosphere as a result of two rocket flights during which the measured electron densities were approximately the same, thus suggesting a meteorological influence on the D region.[18]

THE E REGION

Although ground-based ionosondes have been valuable in detailing the temporal and latitudinal variations of the maximum amount of ionization found in the E and F regions, the important altitude variation of N_e has come about as a result of rocket probing using both radio propagation[27,28] and plasma probe experiments.[29,30] Typical day and night N_e profiles composed from such measurements are presented in Fig. 4. These show that at night the D region essentially disappears, the E region electron abundance has decreased by a hundredfold but that there is a strong persistence of the upper F region.

In addition to N_e measurements, significant contributions to our understanding of the physics

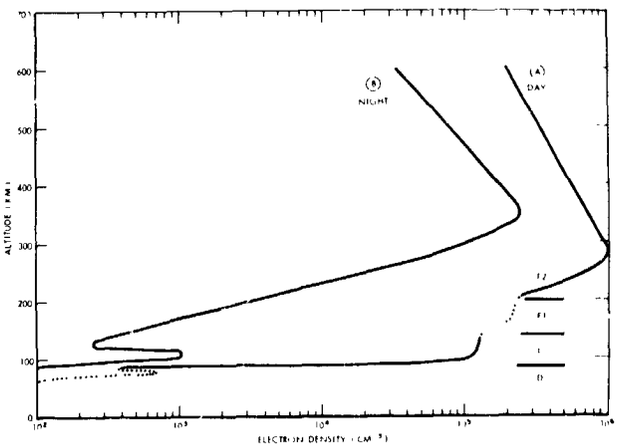


FIGURE 4.—Typical D, E, and F region electron density profiles.

of the ionosphere have come about as a result of vertical cross-sections taken of the intensity of solar radiation[31] and of ionic composition.[32,33] In an important work,[31] solar radiation measurements have been used together with a model neutral atmosphere to estimate the altitude dependence of (a) electron production rate for discrete portions of the X-ray and ultraviolet spectra (Fig. 5a) and (b) the rate of production of each of the

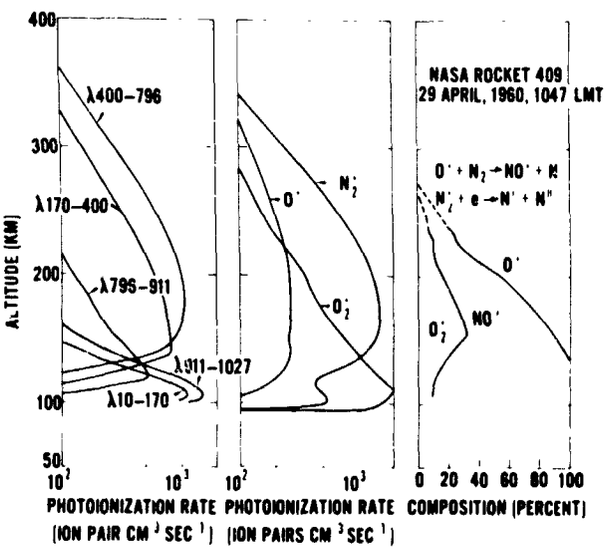


FIGURE 5.—Comparison of ion production rates (Watanabe and Hinteregger, 1962) with rocket measurement of ionic composition (Taylor and Brinton, 1961)

ion species (Fig. 5b). We see that the ions formed at the highest rate are O_2^+, N_2^+ and O^+. Spectrometer observations[32] typically represented in Fig. 5c, on the other hand, show that N_2^+ is a

minor ionic constituent despite the predicted high production rate. They also show that NO^+ is a dominant E region constituent even though it is not significantly produced by direct ionization. Most models[34] attribute the loss of N_2^+ to the combined effects of dissociative recombination with electrons and of ion-atom interchange ($N_2^+ + O \rightarrow NO^+ + N$). The existence of NO^+ generally is explained by ion-atom interchange involving N_2^+ and O^+.

It should be apparent that our understanding of the formation of the daytime E region has been vastly improved by the above described rocket measurements. Yet, models do disagree considerably as to even the relative importance of X-ray and ultraviolet radiation in the E region. Laboratory measurements of the various rate coefficients and additional rocket flights are needed to resolve these differences.

It wasn't until quite recently that rocket experiments were made sensitive enough for extensive studies of the nighttime E region. In Fig. 4, it is shown that N_e drops below 10^3 cm^{-3} and that the region is characterized by a ledge at about 100–110 km with a valley of ionization just above. It is suggested[33] by a comparison of the results from ion spectrometers flown both during the day and for the first time at night that the maintenance of the nighttime E region is explained by slow decay through dissociative recombination without resort to a nighttime source of ionization. There also is experimental evidence from rockets for metallic ions of meteoric origin in the 100–110 km region.

A frequent anomaly of the E region is sporadic E or E_s ionization. One common form of E_s is a layer as thin as 0.5 km in which N_e is considerably higher than the region immediately below and above, leading to reflection of radio signals at (Fig. 6) abnormally high frequencies. These layers exist at preferred altitudes[35] in the region 100–120 km. Some theories[36] explain E_s ionization as the result of the combined effects of wind shear and electromagnetic forces. A correlation has been found between wind shear measured on one rocket carrying a sodium vapor release experiment and E_s ionization detected on a second rocket launched almost simultaneously.[29] However, the exact relationship between E_s ionization and wind shear is not clear at this time.

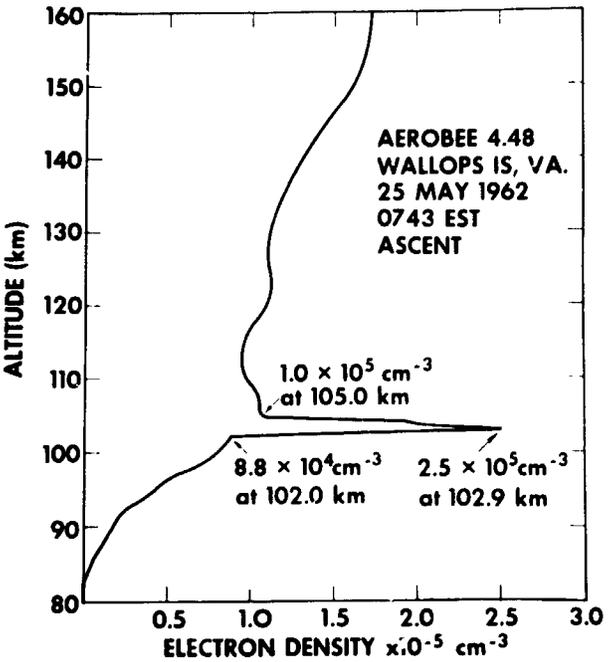


FIGURE 6.—Rocket detection of sporadic-E layer (Smith, 1962).

THE F REGION

As shown in Fig. 4, the F region contains the altitude of maximum electron density (Nm). The region is subdivided into the F1 and F2 regions because a ledge often appears in the daytime 140–200 km region.

It is important to note by comparing Figs. 4 and 5a that the altitude of maximum electron density lies well above the height of maximum electron production. These rocket results confirm previously established theoretical models which ascribe the formation of the F2 peak to charge transport mechanisms of importance comparable to photochemical processes. Specifically, these models ascribe the F2 peak formation to the competition between electron production, a height dependent decrease in electron loss rate and charge transport.

The rocket solar radiation and ion composition measurements (see Fig. 5) show that the origin of F region electrons lies principally in the production of O^+ ions by solar ultraviolet radiation. The dominant F region loss process is believed to be radiative recombination, a two-step process involving firstly ion-atom interchange between O^+ ions and N_2 molecules and secondly recombination of the resulting NO^+ ions with electrons. The slowness of this two-step process is a partial

explanation for the strong persistence of the nighttime F region. The possibility that the electron loss rate decreases more rapidly with altitude than the production rate is an explanation for the experimental observations that N_{max} lies above the altitude of maximum production.

Because photochemical, gravitational and electromagnetic forces all are effective, the behavior of the F region is extremely complicated. Ionosonde data taken over the last three decades show that the diurnal seasonal, solar cycle and latitude effects each produce large variations in the magnitude and altitude of N_m.[34] These observations have been related to time-dependent neutral atmospheres deduced from satellite drag measurements (Fig. 2) to estimate ionization, loss and diffusion rates.[37,38]

As the density of the neutral constituents decrease, the importance of photochemical processes diminish. Thus, charge transport processes become dominant and bring about the observed decrease of N_e with altitude. Gravitational forces, for example, act upon the ions which, by coulomb attraction, cause the electrons to diffuse downward. In this case, the hydrostatic law can be invoked and the electron distribution is controlled by the average electron-ion temperature and the type of ion.

Before rockets and satellites penetrated the upper ionosphere, it was believed that a transition takes place from O^+ ions directly into the protonosphere. However, rocket[9] and satellite[10] experiments indicated that a region of helium ions separated these regions in the daytime during the middle of the solar cycle. Subsequent rocket results[39] suggest that the helium ion region disappears at night. It had been predicted[11] on the basis of the different behavior of the escape rates of hydrogen and helium that both the altitude and thickness of the helium ion region would diminish with decreasing temperature, implying a strong latitudinal and solar cycle dependency.

The role of charged particle temperatures and of ionic composition in controlling the ionosphere is illustrated by rocket measurements of charged particle density taken out to very high altitudes and presented in Fig. 7. Both a daytime and a nighttime result are illustrated. The exponential portions of both profiles provide evidence for a constant electron-ion temperature throughout

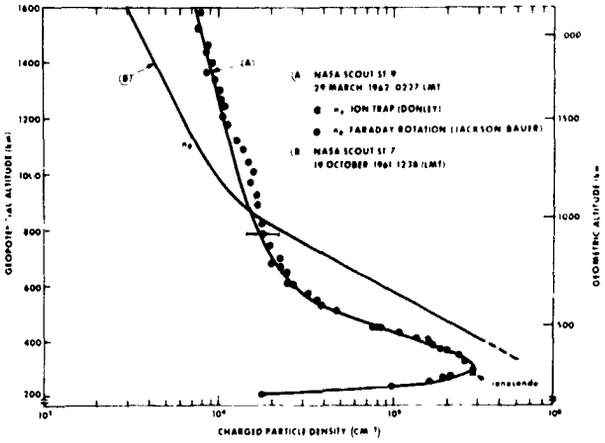


FIGURE 7.—Daytime (Bauer and Jackson, 1962) and nighttime (Donley, 1963) measurements of electron density.

the region with values of 1300° and 800°K, respectively. The change of slope in both cases is interpreted as the result of a transition from O^+ to lighter ions, He^+ being dominant above this transition for the daytime case but H^+ dominant at night. It is seen that the lower altitude of the transition and the relatively high importance of H^+ causes the surprising result that at very high altitudes the nighttime densities exceed those measured in the daytime. This may not be representative of a true diurnal variation because the two sets of data were obtained in different days.

SATELLITE STUDIES

Three different types of satellites have been used for ionospheric research each performing a different task. With Direct Measurements Satellites, one uses environmental sampling techniques to measure many ionospheric parameters, but only in the immediate vicinity of the spacecraft. The US Satellite EXPLORER VIII and the British Satellite ARIEL I each contained three major experiments for such studies. The first of these measures electron temperature by techniques similar to those developed by Langmuir in his laboratory studies of gaseous discharges. The second measures the local electron density by use of the radio-frequency impedance characteristic of a probe immersed in the medium. The third type of experiment involves the use of gridded ion traps whose principle of operation is similar to that flown on SPUTNIK III. Here, the high satellite-to-ion velocity permits the trap to act as

a poor man's ion spectrometer such that both ion composition and temperature are obtained.[10]

The second category of satellite (topside sounding) involves the ingenious use of an orbiting ionosonde. As with the classical tool of ground-based ionospheric research, one measures the time between transmission of a radio signal to reception of the reflected echo. For the satellite case the reflection is from the topside of the ionosphere. By sweeping the transmitted frequency, an electron density profile for the region between the F2 maximum and the satellite altitude is measured continuously along the satellite's path. Soundings of the upper ionospheric regions have been made in a near polar orbit for two years with the Canadian Satellite ALOUETTE I, which features a swept-frequency sounder. With a swept-frequency device, one obtains good altitude resolution. The recently-launched US EXPLORER XX features fixed-frequency sounding where vertical resolution is sacrificed so as to better study horizontal irregularities.

The third type of satellite observation involves the study at the earth's surface of the arrival characteristics of radio signals transmitted from the spacecraft at frequencies which penetrate the ionosphere. Faraday rotation and doppler phenomena permit the measurement of the total electron content in a cross-section between the satellite and the receiving site. The first satellite exclusively devoted to this method is Explorer XXII, launched too recently for results to be reported here.

SATELLITE RESULTS

The major results that have come from satellite studies lie in the unique observations of the latitudinal and diurnal behavior of the regions above the F2 peak and of magnetic-field aligned irregularities.

Evidence for magnetic field control of ionospheric electron densities first came about from the use of ground-based ionosondes, which since have provided details of the bottomside characteristics of the now familiar equatorial anomaly. The morphology of this anomaly has now been obtained to altitudes of about 1000 km. In Fig. 8 are presented results[41] from the Alouette satellite of electron density as a function of magnetic

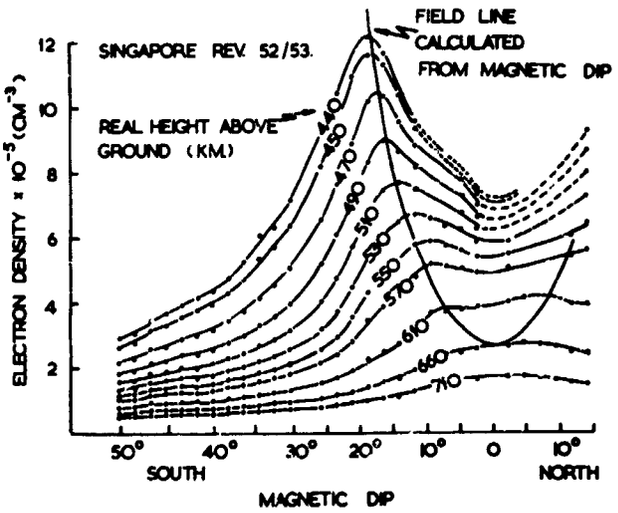


FIGURE 8.—Contours of electron density at constant altitudes measured above Singapore at 1000 LMT (King et al, 1963).

dip for discrete altitudes. These esults specifically are for the eastern hemisphere at 1000 local time. A general increase of electron density in the equatorial regions results from diffusion along the nearly horizontal magnetic field lines. For the particular time of day illustrated, the electron density reaches a maximum at the geomagnetic equator for altitudes above about 600 km. Below this altitude, two peaks are symmetrically located along a specific field line. The equatorial anomaly is predominantly a daytime feature of the ionosphere. The altitude above which only a single peak is formed has been termed the "dome" of the anomaly. The altitude of the dime over Singapore varies from 600 km in the early morning and evening hours to a mid-afternoon maximum of about 1000 km.[41] Similar measurements[42] suggest that the anomaly builds up later in the day along the 75th West Meridian. The diurnal behavior of the equatorial anomaly suggests that its characteristics are closely related to the competition between electron production which tends to maximize N_e at the subsolar point and diffusion along magnetic field lines which tends to produce a symmetrical N_e distribution about the geomagnetic equator.

Another latitude feature of the topside ionosphere are electron density troughs which appear to be a common phenomenon at middle latitudes during ionospheric storms associated with magnetic disturbances. This is illustrated by Fig. 9,

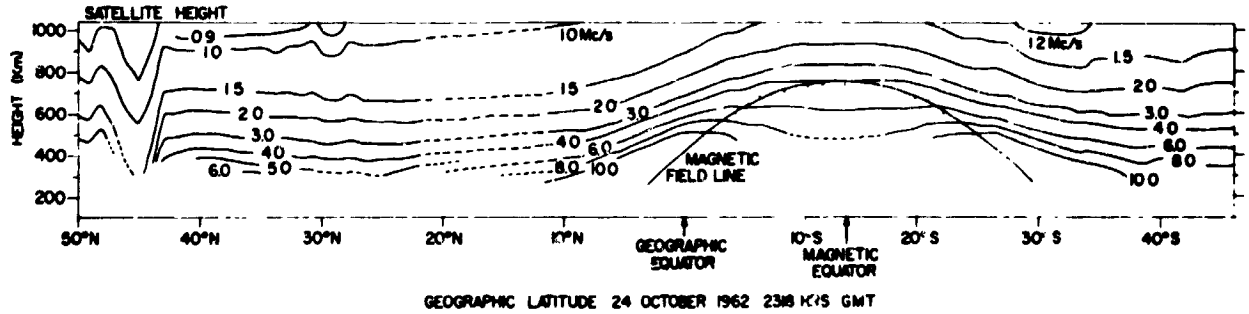


FIGURE 9.—Contours of constant plasma frequency measured with the ALOUETTE satellite (Lockwood and Nelms, 1964)

a plot of Alouette satellite measurements[42] of the frequency of reflections contours. In this illustration, applicable to 65° West Longitude, the trough is less than 5° wide and centered at 45°N.

Evidence for strong magnetic control of the upper ionosphere was first inferred from electron densities measured with an rf probe on the ARIEL satellite.[43] From these observations was inferred the existence of enhanced ionization lying along three specific magnetic field shells, one of which is accounted for by the above-described equatorial anomaly. The existence of these shells of enhanced ionizations were confirmed and others discovered by the ALOUETTE satellite.[41] Some of these have been associated with the American artificial radiation belt, the heart of the inner radiation belt and near the region of maximum flux of energetic particles in the outer radiation belt. All of these observations suggest that ionization by fast particles should now be considered along with ultraviolet radiation as an ionization source for the F2 region.

A very common anomaly of the F region is the "spread F" condition, where the region shows a diffuse character generally attributed to patches of ionization having concentrations different than

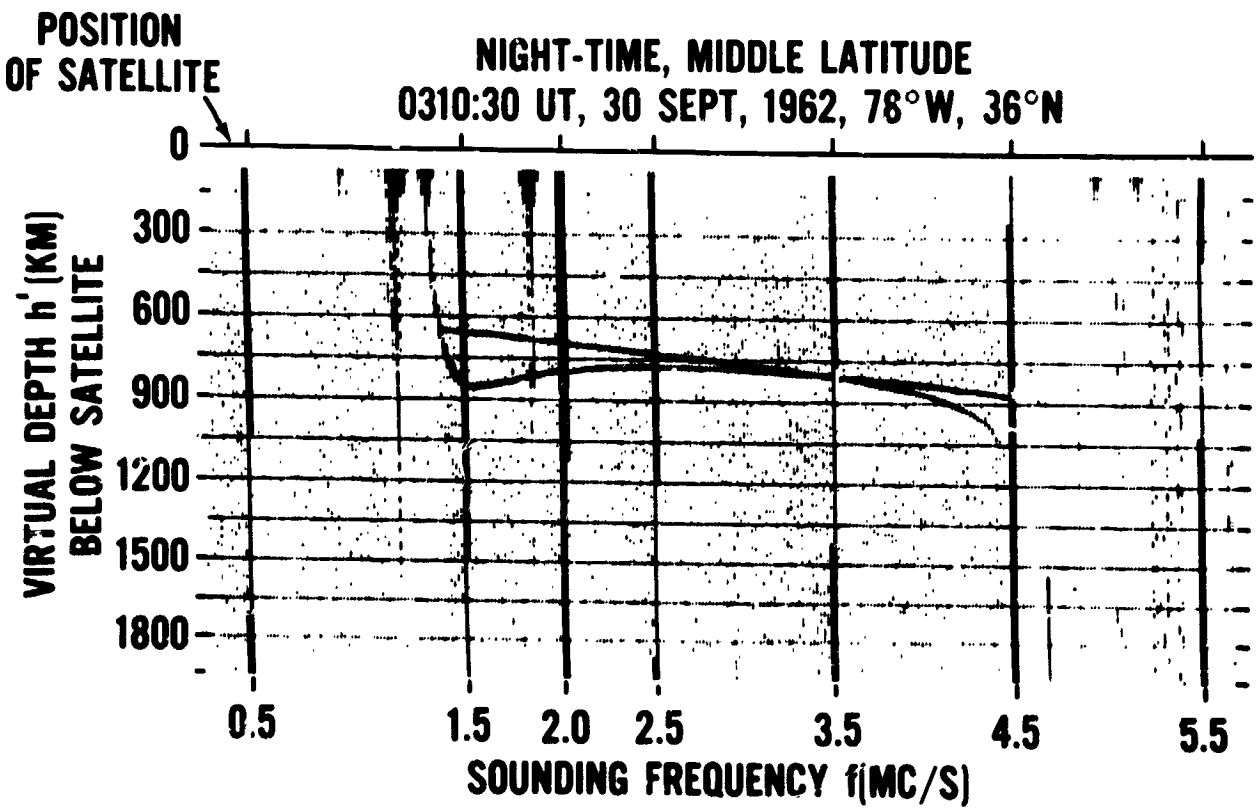


FIGURE 10.—Typical [illegible] from ALOUETTE satellite taken in the absence of spread-F.

the immediate surroundings. An ALOUETTE satellite ionogram typical of a homogeneous ionosphere is presented in Fig. 10, the two traces corresponding to the ordinary and extraordinary

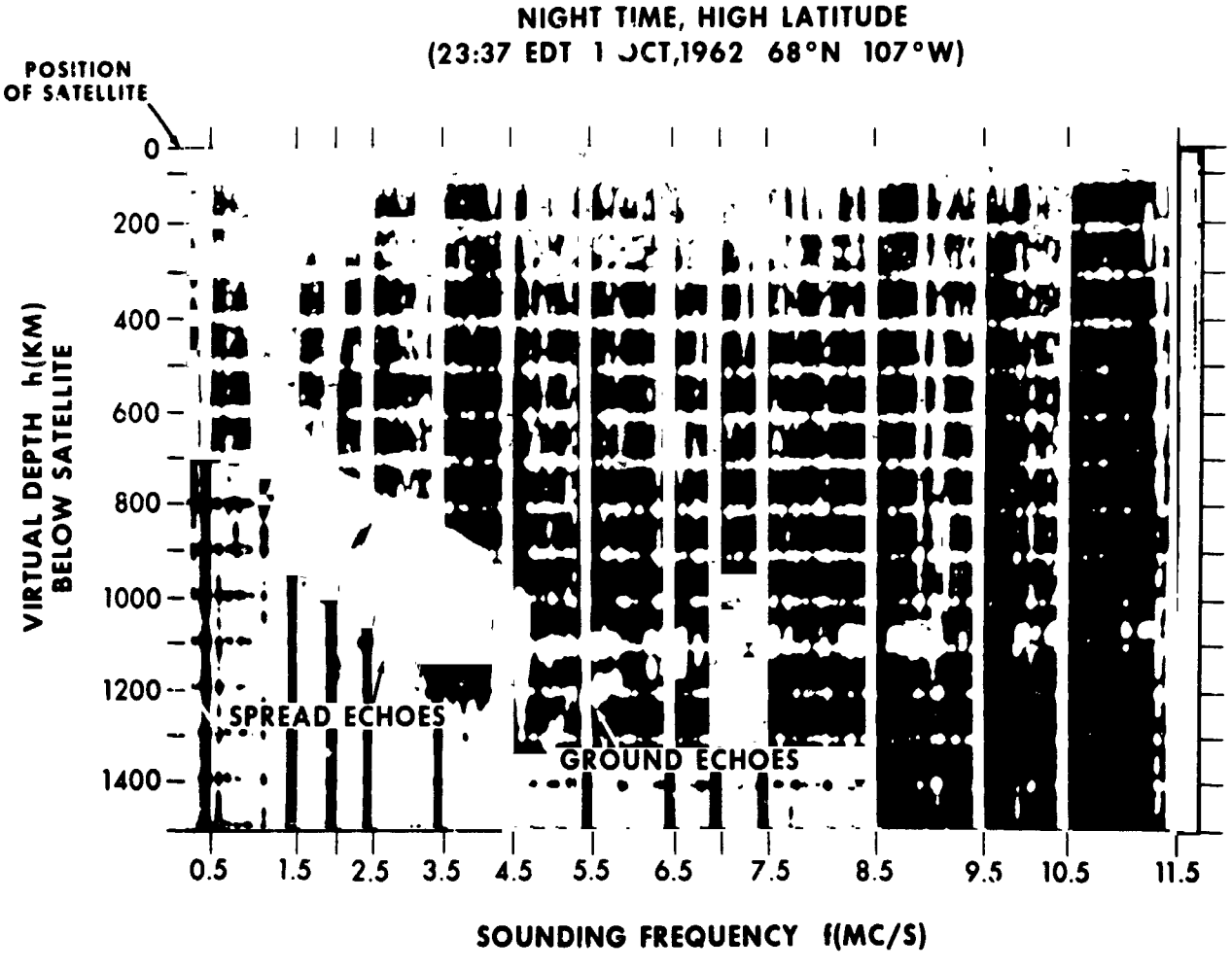


FIGURE 11.—ALOUETTE satellite ionogram taken during the presence of spread-F.

modes of radio-propagation. The effectiveness of spread-F in producing multiple echoes is illustrated by comparing this ionogram with that shown in Fig. 11 which was taken during spread-F conditions. The occurrence probability of spread-F has been the subject of a thorough analysis with the ALOUETTE satellite. This analysis[44] confirms previous ground-based ionosonde studies which showed that the phenomenon is almost a permanent feature of the high latitude ionosphere, occurs only during the night near the equator and that it occurs relatively seldom at mid-latitudes.

Referring back to Fig. 9, we observe that the electron density is not a strong function of magnetic latitude between about 15 and 45°N geographic latitude. In this region, we can expect that an important charge transport mechanism is due to gravity in which case the electron distribution is a strong function of the mean ionic mass and the average electron-ion temperature. The diurnal variation of the electron distribution in this region has been studied[45] by the use of the ALOUETTE satellite. (Fig. 12). These studies have shown the amplitude of the diurnal variation becomes smaller with increasing altitude. This would confirm a conclusion based on the rocket results shown in Fig. 7 that because of the increased relative importance of the light ionic constituents, the electron densities above 1000 km at night are higher than for a daytime condition. The diurnal variation shown in Fig. 12 have been

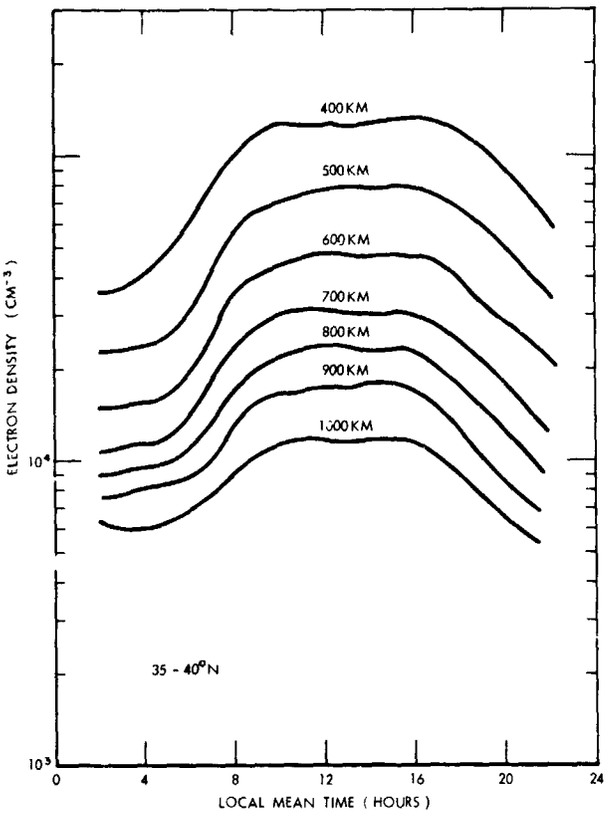


FIGURE 12.—Diurnal variation of electron density in the topside ionosphere (Bauer and Blumle, 1964).

treated in terms of ionic composition and charged particle temperature, one possible result being that for daytime at 500 km the principal ion is O^+ with an average electron-ion temperature of about 1500°K, while at night the lighter ionic constituents become important even at altitudes as low as 500 km.[45]

CHARGED PARTICLE TEMPERATURES IN THE UPPER IONOSPHERE

Because electron (T_e) and ion (T_i) temperatures are important to the behavior particularly of the region above 300 km, it is important to perform experimental observations of these parameters. Upon their creation, photoelectrons will possess energies in excess of the ambient electrons. They attempt to share this excess energy with the ambient electrons either by direct elastic collisions or after some of the excess energy has been lost through previous inelastic collisions with the ambient neutral particles or positive ions. As a result, there is a tendency for the ambient electron gas to be hotter than the neutral gas, at least in the daytime. It generally is accepted that at the lower altitudes where the neutral gas density is relatively high, the ion and neutral gas temperatures will be identical. At the higher altitude where the percentage ionization is becoming appreciable, there is a possibility that the ions are in better thermal contact with electrons than with the neutral particles and consequently the ion temperature can be elevated above the neutral gas temperature.[46]

Measurements made with the EXPLORER VIII satellite first indicated a strong diurnal control of ionospheric electron temperatures.[47] Later results with the ARIEL satellite[48] and with ground-based radar incoherent backscatter apparatus[49] confirmed this diurnal dependency. The ARIEL results and a comparison of radar backscatter observations at two locations[49,50] showed that there also is a strong dependency of T_e upon magnetic latitude. Both the diurnal and latitudinal control of T_e were subsequently confirmed[5] with the EXPLORER 17 satellite. The observations together show that the daytime ratio of T_e to the neutral gas temperature can reach values up to 2 depending on latitudinal and temporal conditions.

Since newly-created electrons share their excess energy with the ambient electrons, it follows that N_e and T_e should be strongly coupled. The latitudinal behavior of T_e for the daytime ionosphere thus can be explained in terms of magnetic field control of the electron density at least at moderate altitudes. But, it has been observed that T_e increases with altitude above 600 km.[48] This would not be expected if direct effects of solar ultraviolet radiation constitute the only source of daytime charged particle heatings. The observations[48] are consistent with the hypothesis[46] of possible indirect effects of ultraviolet radiation, specifically a mechanism whereby all of the newly-created electrons do not deposit their energy in the lower F region where they are created but where some are permitted to diffuse along magnetic field lines and deposit the energy at high altitudes. Another important observation[5,48] is that the electron temperature is somewhat higher than the neutral gas temperature at night. This requires a nighttime source of particle heating having an

intensity which is a small fraction[48] of the daytime ultraviolet source.

THE UPPER IONOSPHERE

In this paper, the higher altitudes have been subdivided into the F region where O^+ ions dominate and the upper ionosphere where light ionic constituents are more abundant. Most observations show that on the average, the transition altitude between these two regions is about 600 km. EXPLORER VIII[51] and ARIEL[52] satellite results show that the transition is from O^+ to He^+ and then to H^+ as the altitude increases, at least during the middle of the solar cycle. These results also show that the thickness and altitude of the helium ion region decrease drastically from day to night. Other results show a more complicated behavior in which for nighttime conditions, and for all diurnal times during the year of minimum solar activity, helium ions are never dominant. The exact morphology of upper ionospheric composition is not yet clear.

It has been suggested[53] that the mean ionic mass is a function of the ion temperature. There already exists evidence[49] that the ion temperature in the upper ionosphere is controlled by the electron temperature which we have shown exhibits a complicated diurnal and latitudinal behavior. It is becoming clear that theoretical models of the F region, the upper ionosphere and the interdependence of these two regions need to be updated to include (a) gravitational forces into which are inserted charged particle temperatures and ionic composition that are temporally and latitudinally variable; (b) the possibility of an ionization source related to fast particles which is superimposed in the ultraviolet source and which contributes to the maintenance of the nighttime ionosphere and (c) the possibility of indirect ionization from ultraviolet radiation, specifically the effect of photoelectrons diffusing from the lower F region along magnetic field lines to deposit their energy at the higher altitudes. Thus although satellite and ground-based studies in recent years have provided a preliminary description of the charged particle parameters at high altitudes, the observations need to be extended and correlated with measurements of photoelectron and fast particle fluxes before adequate theories of formation of these regions can be formulated.

REFERENCES

1. W. Nordberg and W. G. Stroud, NASA Technical Note D-703, (1961).
2. L.Spitzer, Atmospheres of the Earth and Planets, University of Chicago Press, (1952).
3. R. Horowitz, H. E. LaGow and J. F. Guiliani, J. Geophys. Res., 64, 2287-2294, (1959).
4. G. S. Sharp, W. B. Hanson and D. D. McKibbin, J. Geophys. Res., 67, 1375, (1962).
5. N. W. Spencer, G. P. Newton, C. A. Reber, L. H. Brace and R. Horowitz, NASA Goddard Space Flight Center Report X-651-64-114, (1964).
6. L. G. Jacchia, Planetary Space Sciences, 12, 355-378, (1964).
7. F. S. Johnson, Astronautics, 8, 54, (1962).
8. M. Nicolet, J. Geophys. Res., 66. 2263, (1961).
9. W. B. Hanson, J. Geophys. Res., 67, 183-188, (1962).
10. R. E. Bourdeau, W. C. Whipple, Jr., J. L. Donley and S. J. Bauer, J. Geophys. Res., 67, 467-475, (1962).
11. S. J. Bauer, J. Atmos. Sciences, 19, 276-278, (1962).
12. I. Harris and W. Priester, J. Geophys. Res., 67, 4585-4591, (1962).
13. E. B. Meadows and J. W. Townsend, Space Research I, North-Holland Publishing Co., Amsterdam, (1960).
14. E. J. Schaefer, J. Geophys. Res., 68, 1175-1176, (1963).
15. A. O. Nier, J. H. Hoffman, C. Y. Johnson and J. C. Holmes, J. Geophys. Res., 69, 979-989, (1964).
16. J. S. Belrose and E. Cetiner, Nature, 195, 688, (1962).
17. R. E. Barrington and E. V. Thrane, J. Atmos. Terr. Phys., 24, 32, (1962).
18. A. C. Aikin, J. A. Kane and J. Troim, J. Geophys. Res., (November 1964).
19. M. Jespersen, O. Petersen, J. Rybner, B. Bjelland, O. Holt and B. Landmark, Norwegian Space Res. Comm. Rept. No. 3, (1963).
20. M. Nicolet and A. C. Aikin, J. Geophys. Res., 65, 1469, (1960).
21. H. Friedman, Astronautics, 8, 14, (1962).
22. R. S. Narcisi and A. D. Bailey, Space Research V, to be published).
23. A. C. Aikin, International Dictionary of Geophysics, (1964).
24. T. A. Chubb, H. Friedman, R.W. Kreplin and J. E. Kupperian, J. Geophys. Res., 62, 389, (1957).
25. B. Maehlum and B. J. O'Brien, J. Geophys. Res. 67, (1961).
26. J. E. Jackson and J. A. Kane, J. Geophys. Res., 64, 1974, (1959).
27. J. C. Seddon and J. E. Jackson, Ann. de Geophys., 14, 456, (1958).
28. J. E. Jackson and S. J. Bauer, J. Geophys. Res., 66, 3055, (1961).
29. L. G. Smith, Geophys. Corp. America, Tech. Rept. 62-1-N, (1962).

30. W. B. Hanson and D. D. McKibbin, J Geophys. Res., 66, 1667, (1961).
31. K. Watanabe and H. E. Hinteregger, J. Geophys. Res., 67, 999, (1962).
32. H. A. Taylor and H. C. Brinton, J. Geophys. Res., 66, 2587, (1961).
33. J. C. Holmes, C. Y. Johnson and J. M. Young, Space Research V, (in press).
34. T. E. Van Zandt and R. W. Knecht, Space Physics, D. P. Galley and A. Rosen, ed., John Wiley and Sons, New York, (1964).
35. J. C. Seddon, Ionospheric Sporadic E, Pergamon Press, Oxford, 909, (1962).
36. J. D. Whitehead, J. Atmos. Terr. Phys., 20, 49, (1961).
37. H. Rishbeth, J. Atmos. Terr. Phys., 26, 657, (1964).
38. J. S. Nisbet and T. P. Quinn, J. Geophys. Res., 68, 1031, (1963).
39. J. L. Donley, J. Geophys. Res., 68, 2058, (1963).
40. S. J. Bauer and J. E. Jackson, J. Geophys. Res., 67, 1675, (1962).
41. King, J. W., P. A. Smith, D. Eccles and H. Helm, Radio Research Station Report 94, Bucks, England, (1963).
42. Lockwood, G. E. K. and G. L. Nelms, J. Atmos. Terr. Phys., (1964), in press.
43. Sayers, J., P. Rothwell and J. H. Wager, Nature, 198, 230, (1963).
44. Calvert, W. and C. W. Schmid, J. Geophys. Res., 69, 1839–1852 (1964).
45. Bauer, S. J. and L. Blumle, J. Geophys. Res., 69, 3613–3618 (1964).
46. Hanson, W. B., Space Research III, 282, (1962).
47. Serbu, G. P., R. E. Bourdeau and J. L. Donley, J. Geophys. Res., 66, 4313–4319 (1961).
48. Willmore, A. P., Proceedings of the Royal Society Meeting, May 1963, in press.
49. Evans, J. V., J. Geophys. Res., 67, 4914–4920 (1962).
50. Bowles, K. L., J. Research NBS, 65D, 1–14 (1961).
51. Bourdeau, R. E. and J. L. Donley, Proceedings of Royal Society Conference May 1963, in press.
52. Boyd, R. L., Proceedings of the Royal Society Conference, May 1963, in press.
53. Bauer, S. J., J. Geophys. Res., 69, 553–555 (1964).

THE TEMPERATURE OF CHARGED PARTICLES IN THE UPPER ATMOSPHERE*

R. E. BOURDEAU

Three general methods of investigating charged particle temperatures in the upper atmosphere have been used: (a) direct measurements from rockets and satellites; (b) indirect determination using electron scale heights measured from rockets and satellites; (c) ground-based radar incoherent backscatter experiments. The latitude, altitude and temporal trends of these results are reviewed and the implications discussed.

Observations by all three methods are consistent in showing that the electron temperature increases with latitude for both daytime and nighttime conditions. Moderate differences between the daytime electron and neutral gas temperatures are indicated to altitudes well above the F2 peak for a winter mid-latitude ionosphere at an epoch between solar maximum and solar minimum conditions. Much larger daytime differences are observed for summer months and for solar minimum conditions. All of these trends reflect corresponding changes in the electron density.

The daytime observations are consistent with ultraviolet radiation as the predominant heat source if the possibility of photoelectrons diffusing along magnetic field lines and depositing their excess energy elsewhere is included. A nighttime heat source small compared to the daytime ultraviolet effect is required to explain the observations.

INTRODUCTION

At the first Florence meeting of COSPAR, early direct measurements of electron temperature T_e were reported from rockets (Aono et al, 1961) and from the Explorer VIII Satellite (Bourdeau, 1961). The Japanese rocket results suggested temperature equilibrium between electrons and neutral constituents in the E region of the daytime ionosphere. Low values of T_e in the daytime E region subsequently were confirmed by US rocket experiments (Spencer et al, 1962). *However*, other rocket T_e measurements (Smith, 1961; Aono et al, 1962; Brace et al, 1963) showed that significant departures from temperature equilibrium are more often observed than not in the E region even at night. The results of Spencer et al (1962) showed that departures from temperature equilibrium extended well into the daytime F region.

Early measurements of charged particle temperatures applicable to the region considerably above the F2 peak at midlatitudes for solar conditions when the 10.7 cm flux index ($S_{10.7}$) was 150 $WM^{-2}CPS^{-1}$ were reviewed by Bauer and Bourdeau (1962) and Bourdeau (1963). Implications of temperature equilibrium at high altitudes which they optimistically derived from the data critically depended on the assumed neutral gas temperature (T_g). When these early direct measurements of T_e and of the average electron-ion temperature $(T_e+T_i)/2$ obtained from electron density profiles now are compared with more recent reference atmospheres moderate values for T_e/T_g of about 1.3 are indicated at midday even to altitudes above 1000 km for the indicated level of solar activity. These observations are in approximate agreement with the theoretical model of Hanson (1962) who assumed electron density values close to these particular observational conditions.

More extensive charged particle temperature observations for different epochs of the solar cycle now have been made by use of ground-based radar incoherent backscatter experiments, by additional direct measurements from rockets and satellites and indirectly from electron density profiles obtained from rockets and the Alouette Topside Sounder Satellite. In general, the trend is toward

*Published as *Goddard Space Flight Center Document X-615-64-103*, May 1964.

much larger departures from temperature equilibrium than indicated in the theoretical models and in the earlier observations especially at the higher altitudes.

It is timely then, as is done in this report, to compare these trends with the introduction of latitude and temporal electron density variations in the early theoretical models.

FACTORS CONTROLLING THE ELECTRON TEMPERATURE

The principal factors controlling T_e are:

Heat Input

1. Solar Ultraviolet Radiation
   - A. Locally-deposited energy
   - B. Diffusing photoelectrons ($Z>300$km)
2. Corpuscular Radiation

Heat Loss

1. Inelastic collisions with neutral constituents ($Z<250$km)
2. Elastic collisions with ions ($Z>250$ km)
   - A. Ion temperature controlled only by neutral constituents
   - B. Ion temperature also controlled by electrons ($Z>600$ km)
3. Coulomb collisions of photoelectrons with ambient electrons

Thermal conductivity in the electron gas ($Z>600$ km)

Theoretical charged particle temperature models based on solar ultraviolet heating alone have been developed by Hanson and Johnson (1961), Hanson (1962) and Dalgarno et al (1962). Of these, Hanson's model is the most complete in that he introduced the possibility of (a) photoelectrons diffusing along magnetic field lines and depositing their energy elsewhere, (b) the loss of ion temperature control by the neutral constituents and (c) the importance of thermal conductivity in the electron gas at high altitudes. The indicated altitudes where these factors are important represent Hanson's estimates based on his assumed model atmosphere and electron density profile.

The models of Hanson and of Dalgarno et al each used a single electron density profile and both depended on rocket measurements of ultraviolet radiation intensity (Hall et al, 1962) to estimate the heat input, Q. The EUV intensity used applies to a level of solar activity corresponding to $S_{10.7}\cong 100$. The heat input is the product of the EUV intensity, the ionization cross-sections and the density of the ionizable constituents and of the heating efficiency. Both theoretical charged particle temperature models exhibit approximately the same altitude behavior wherein low values of T_e/T_g are indicated in the E region, with the ratio reaching a maximum of about 2.0−2.5 at 200 km and decreasing in the upper F region. A principal and important difference is that the model of Dalgarno et al has T_e-T_g essentially vanishing above 300 km where Hanson's model permits values for T_e/T_g of about 1.2 constant at extremely high altitudes.

Because the observational evidence is most heavily weighted for altitudes above the F2 peak, the most important effect to examine is the efficiency of cooling to positive ions. On the assumption that cooling occurs only by elastic collisions to atomic oxygen ions, the electron temperature is given (Hanson, 1962) by:

$$\frac{T_e-T_i}{T_e^{3/2}}\cong\frac{2.1\times 10^6 Q}{N_e^2} \qquad (1)$$

where Q is the heat input to the electrons expressed in ev cm^{-3} sec^{-1} and N_e is the electron density. For values of Q greater than a critical value (Q_c) given by

$$Q_c\cong 2\times 10^{-7} N_e^2 T_i^{-1/2}, \qquad (2)$$

there is no solution to (1) and T_e is not limited by energy transfer to positive ions (Hanson and Johnson, 1961). If $Q=Q_c$, $T_e>2T_i$ and if $Q\gg Q_c$, very large "runaway" values of T_e will result and heat conduction in the electron gas is an important effect.

SOLAR CYCLE VARIATIONS OF T_e

In Figure 1 are illustrated daytime mid-latitude electron density profiles typical of December 1960 and 1962, respectively. The principal differences to note are the much higher values of N_{max} and h_{max} for the 1960 period, and the constant electron scale height for the 1960 case in contrast to the

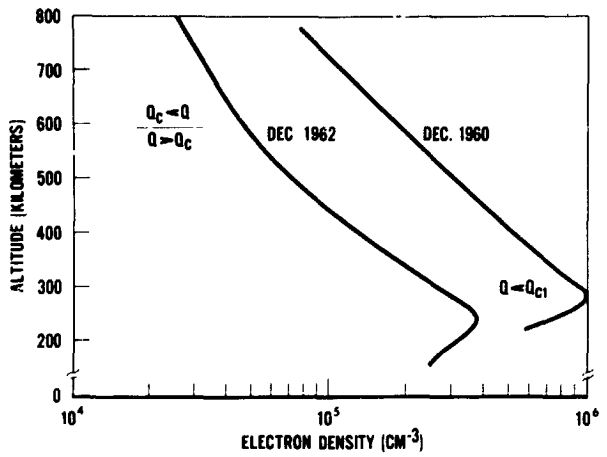


FIGURE 1.—Typical winter mid-latitude electron density profiles.

continually varying scale height for the 1962 profile. The monthly mean of the solar 10.7 cm flux during December 1960 was 150 $WM^{-2}CPS^{-1}$.

In Figure 2 is presented the mid-latitude diurnal electron temperature variation directly measured

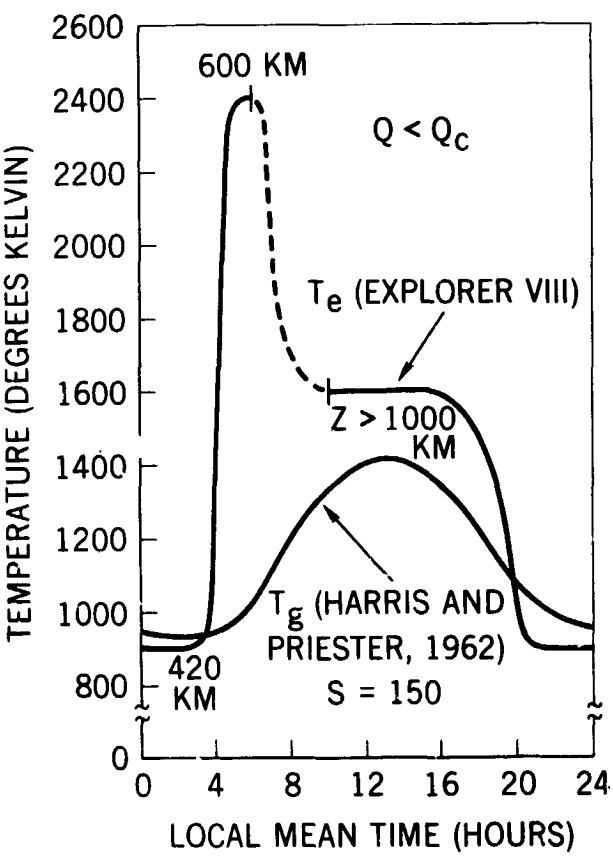


FIGURE 2.—Diurnal electron temperature variation measured during November 1960 from the Explorer VIII satellite for magnetically-quiet days at mid-latitudes assuming T_e is constant with altitude (425–2400 KM).

during November-December 1960 (Bourdeau and Donley, 1963) from the Explorer VIII Satellite for magnetically-quiet days assuming T_e is constant with altitude. Depending on which of the current reference atmospheres for the pertinent level of solar activity is used, the average midday T_e value of 1600°K taken at midday and at altitudes above 1000 km corresponds to an estimated value for T_e/T_g of about 1.15–1.33.

It is seen from Equation (1) that the ratio of the heat input to the square of the electron density (Q/N_e^2) controls the electron temperature at high altitudes. The value for Q/N_e^2 computed at 400 km from the December 1960 N_e profile in Figure 1 closely corresponds to the value used by Hanson as an upper limit in his model. We have taken into account that the heat input Q would be larger than in Hanson's case by assuming a linear change of EUV intensity with the decimetric flux index, $S_{10.7}$, and a corresponding change in the density of the ionizable constituent, $n(O)$. The higher heat input is compensated for largely by the higher electron density than that used by Hanson. Thus we should and do estimate similar values for T_e as did Hanson. Consequently, we find excellent agreement between the observed T_e/T_g of 1.15–1.33 and Hanson's theoretical upper limit of about 1.15.

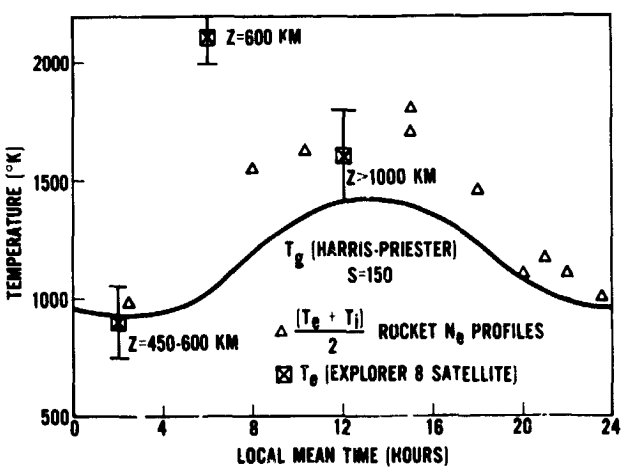


FIGURE 3.—Early charged particle temperature measurements in the upper ionosphere.

In Figure 3, the Explorer VIII data are included with measurements of $(T_e+T_i)/2$ computed from early rocket N_e profiles on the assumption of diffusive equilibrium. Of particular

interest is the ion density profile obtained by Hale (1961) at midday and at about the same time as the Explorer VIII observations. Here a value for $(T_e+_{et})/2$ of 1600°K (Hanson, 1962b) is derived for the region above 1000 km. The data are too sparse for a firm conclusion but there is a suggestion by the inter-comparison that for winter midday mid-solar cycle conditions $T_e \cong T_i$ at altitudes above 1000 km but that both the electron and ion temperatures are moderately higher than the neutral gas temperature. This would be consistent with Hanson's arguments that above 600 km (a) thermal conductivity of the electron gas could support differences between T_e and T_g which are constant with altitude and furthermore, (b) that the electrons rather than the neutral constituents could control the ion temperature so that $T_i > T_g$.

Low midday values of T_e/T_g at high altitudes also can be implied from the measurement in December 1961 of a value for $(T_e+T_i)/2$ of 1235°K (Taylor et al, 1963) from a rocket flight for which an equivalent ion temperature has been inferred (Bauer, 1964). *However*, we emphasize here Equation (1) and the extreme sensitivity of T_e to the electron density. We further emphasize that in the actual case, ratios of Q/N_e^2 which permit only moderate rather than large midday departures from temperature equilibrium as is indicated for December 1960 at high altitudes perhaps represent the exception rather than the rule at middle and high latitudes.

Let us consider now the drastic changes in charged particle temperature characteristics as one moves closer to solar minimum conditions. The average value for the index (S) of solar activity corresponding to the December 1962 profile illustrated in Figure 1 was 85. Taking into account a linear decrease in EUV intensity from the time of the rocket EUV measurements (Hall et al, 1962) and a corresponding decrease in the density of the ionizable constituents and computing Q_c directly from the observed N_e profile, the estimated ratio Q/Q_c is larger than 2. Because of the uncertainties in our knowledge of ionization cross-sections and model atmospheres, the computation of Q/Q_c is suggestive rather than quantitative. Within these uncertainties, it does appear that the EUV effect for low electron densities is sufficient to cause very large electron and possibly runaway electron temperatures.

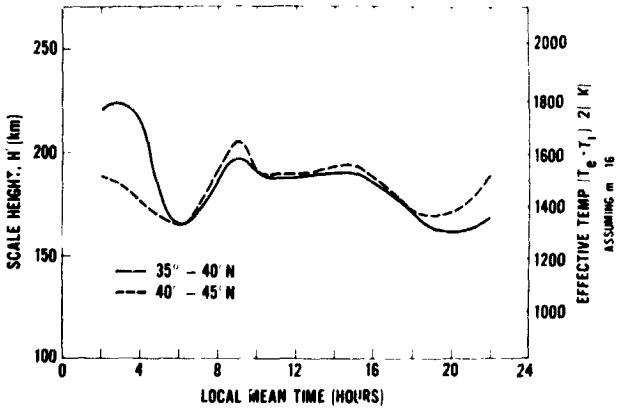


FIGURE 4.—Scale heights at 500 km from ALOUETTE satellite data (Bauer-Blumle, 1964)

In Figure 4 is plotted a mid-latitude diurnal variation of electron scale heights calculated for an altitude of 500 km from electron density profiles obtained by the use of the Alouette satellite during the period October-December 1962 (Bauer and Blumle, 1964). It is seen that on the assumption of diffusive equilibrium and O^+ as the principal ionic constituent, values for $(T_e+T_i)/2$ of 1500°K are indicated at midday. Assuming $T_i = T_g$, a value for T_e/T_g of about 2.0 is indicated (Bauer and Blumle, 1964), which is much in excess of the December 1960 value. This should not be surprising because of the large increase in the ratio Q/N_e^2 at 400 km and above from December 1960 to December 1962.

If the assumed model atmosphere and ionization cross-sections on which the computation of Q/Q_c depends are correct, the fact that runaway electron temperatures are not observed suggests that much of the heat input is not deposited locally. Also it should be emphasized that the ratio of 2.0 for T_e/T_g must be considered an upper limit since it assumes that at 500 km the ion and neutral gas temperatures are the same. Evans (1964) with radar-back-scatter experiments observes at similar latitudes and under similar conditions that $T_i > T_g$ even as low as 400 km. Thus the values for T_e implied from Figure 4 may indeed be somewhat overestimated. That the electron temperature begins to control the ion temperature at an altitude lower than that estimated by Hanson is explainable on the basis that the scale heights illustrated in Figure 4 were measured for a different

model atmosphere and electron density profile than that assumed by Hanson.

In summary, the observational evidence as expected shows a large increase in T_e/T_0 between December 1960 and December 1962 which corresponds to a change in the ratio of the heat input to the square of the electron density. In the 1962 case, there still appears to be a sufficient EUV flux to explain the high daytime electron temperatures observed at 500 km during periods of low electron density. There is indirect evidence that some of the energy is not locally deposited and direct evidence from Evans' results that the ion as well as the electron temperature is raised above the neutral gas temperature.

SEASONAL VARIATION OF T_e

Ionosonde data have shown that the electron density is much higher at the F2 peak in winter than it is in summer. For example, N_{max} measured at Washington, D.C., was on the average more than a factor of two larger in the summer than in the winter of 1962. Corresponding changes in $S_{10.7}$, which reflect changes in the heat input are not observed. These factors are in the direction of making the ratio Q/N_e^2 much higher in the summer than in the winter months. Consequently, it is possible that, at least for mid-latitudes in the Northern Hemisphere, high ratios of T_e/T_0 will persist high into the upper F region throughout a solar cycle. This could explain why some of the early rocket results taken between solar maximum and solar minimum conditions (cf Spencer et al, 1962) show different electron temperature behavior.

DIURNAL VARIATION OF ELECTRON TEMPERATURE

As illustrated in Figure 2, the Explorer VIII satellite results suggest a significant increase in T_e during the early morning hours. High values of T_e in the early morning at the F2 peak are also indicated in ground-based backscatter observations (Bowles et al, 1962) and at higher altitudes by Evans (1964).

Early morning maxima in electron temperature have been confirmed by use of the Explorer XVII satellite from which the maximum value of T_e is placed near 9h local time (Brace et al, 1964). The electron scale height results from Figure 4 also suggest a maximum T_e at approximately the same local time (Bauer and Blumle, 1964). Here the high scale heights for nighttime conditions reflect the importance of light ionic constituents. However, for daytime conditions it would be expected that m_i is relatively constant and thus that the early morning maximum represents a true T_e maximum. It should be pointed out that the nature and existence of an early morning peak in T_e has not been emphasized in the Ariel satellite results (Willmore et al, 1963).

It is possible to show from ionosonde data and Equation (1) on the assumption of no EUV absorption above 300 km that the ratio Q/N_e^2 which controls T_e near the F2 peak maximizes at dawn. However, at high solar zenith angles there could be enough absorption above 300 km to shift the T_e maximum to later in the morning. This reasoning would insert a latitude and altitude dependence on the time of the diurnal T_e maximum. It would be expected that the effect becomes more diffuse at higher altitudes because here the diurnal amplitude of the ionizable constituent has increased relative to the amplitude of the diurnal N_e variation (Bourdeau and Donley, 1963).

LATITUDE VARIATION OF ELECTRON TEMPERATURE

Early ionosonde data showed that the daytime electron density at the F2 maximum increases drastically as one goes from mid-latitudes toward the geomagnetic equator. More recently, results from the satellite Alouette has extended the observation that the geomagnetic field plays an important role in governing the electron density distribution to altitudes well above the F2 peak. In Figure 5 is presented an idealized representation of the latitudinal behavior of N_e prepared by Jackson (private communication) by combining topside sounder (Lockwood and Nelms, 1963) and ionosonde (Wright, 1962) results. Other Alouette data (King et al, 1963) show that the equatorial anomaly builds up earlier in the day at the eastern longitudes. It should be clear from the illustration and Equation (1), that in the daytime T_e should increase with increasing latitude on the basis of the electron density behavior alone, if we assume no EUV absorption above 300 km and no latitude dependence of the neutral gas character-

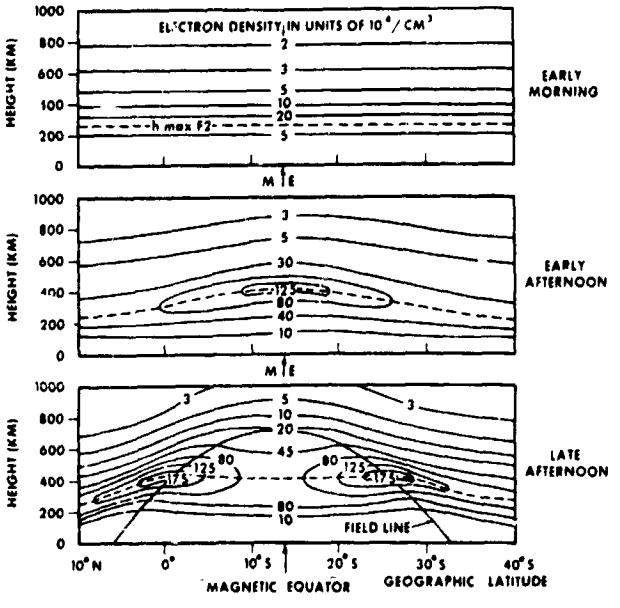


FIGURE 5.—Idealized representation of equatorial anomaly along 75° W meridian based upon data by Lockwood and Nelms (Topside) and J. W. Wright (Bottomside).

istics. An increase of daytime electron temperatures with latitude is indicated by all three methods of charged particle temperature investigation for altitudes below about 800 km. Alouette satellite data suggest constant electron scale heights above 800 km (King et al, 1963).

The ground-based incoherent backscatter results at the geomagnetic equator (Bowles, 1963) show that in the region 200–350 km T_e/T_i is close to 2 during the daytime hours, maximizing at about 275 km. Above about 400 km in the daytime the results show that T_e/T_i is unity. Daytime ion temperatures (Bowles, private communication) for March 1964, in the vicinity of 1000°K are observed Depending on the adequacy of the Harris-Priester reference atmosphere, this would put an upper limit on T_e/T_0 of 1.2 even during solar minimum conditions. Low values for T_e/T_0 at high altitudes would be expected near the geomagnetic equator because of the generally higher values of N_e and because near this location diffusion of photoelectrons vertically tends to be inhibited. This supposes heating only by EUV radiation.

The incoherent backscatter results of Evans (1962, 1964) taken near 50° north magnetic latitude show drastically different behavior of charged particle temperatures. This would be expected because of (a) the generally lower value of N_e than would exist at the equator, and the higher probability of (b) the photoelectron diffusion effect and (c) additional sources of ionization. The earlier results of Evans' taken in March-April 1962, show daytime ratios of T_e/T_i of up to 1.6 in the 300–400 km and that T_e increases with altitude up to 700 km (Evans, 1962). In more recent results taken during July 1963, Evans (1964) offers two possible interpretations for his daytime data obtained for altitudes up to 700 km: (a) if the ionic constituent is all O^+, T_e and T_i continually increase with altitude to values of 2320°K and 2040°K at approximately 700 km or (b) assuming a mixture of 80 percent O^+ and 20 percent He^+ at 700 km, T_e maximizes at about 450 km then decreases to 1960°K at 700 km while T_i increases to a value of 1410°K at 700 km. The trend of ion composition results (Hanson, 1962; Bourdeau et al, 1962; Bowen et al, 1963; Gringauz et al, 1963; Taylor et al, 1963) would imply that the latter alternative is the more likely. If so, it would be consistent with high values of Q/N_e^2 permitting high values of T_e especially below 450 km, and the possibility that cooling to light ionic constituents is becoming effective above 450 km. It should be noted that Equation (1) applies only for O^+ and that the cooling efficiency to ions should be inversely proportional to the ionic mass, m_i.

Direct electron temperature measurements measured with the use of the Ariel satellite also show that T_e significantly increases with latitude (Willmore et al, 1963) the steepest gradient centered at a geomagnetic latitude of about 20°. The midday T_e value given at the geomagnetic equator for an altitude of 400 km is about 900°K which compares favorably with T_0 given by Harris-Priester for the pertinent level of solar activity and with the T_i measurements of Bowles. The trend of the latitude variation at 400 km from the Ariel satellite is generally consistent with the latitude variation N_e and thus both the ground-based results of Bowles and of Evans (Willmore et al, 1963).

However, the situation above 400 km is more complicated. The Ariel results show T_e increasing with altitude up to maximum height of the observations at all latitudes. The continuing increase with altitude of T_e from Ariel at higher latitudes would be consistent with the 1962 results of Evans but not for the likely possibility which Evans offers that T_e decreases above 450 km for

his 1963 results. We suffer here in the comparison from a lack of simultaneity in the observations. An increase of T_e at all altitudes in 1962 is perhaps reconcilable with a possible decrease in T_e above 450 km in 1963 on the base of a lowering with solar activity of the O^+-He^+ transition altitude.

The fate of the photoelectrons which apparently escape from the altitude of their formation is not yet clear. We have made a case that for the December 1962 mid-latitude profile the ratio of Q/Q_e possibly is large enough that the high values of T_e observed up to 500 km at mid-latitudes could be explained on the basis of the EUV effect and heat conduction in the electron gas. In his interpretation of the daytime altitude behavior of T_e from the Ariel satellite, Willmore (1963) finds that for altitudes up to 600 km, the altitude dependence of the heat input computed from the observed T_e and N_e (cf Equation 1) follows the scale height of atomic oxygen and thus also concludes that the main energy input below 600 km is by the photoionization of atomic oxygen. He additionally finds that the increase of T_e above 800 km can be explained by additional energy input from the photoelectrons diffusing from below together with the main energy loss mechanism being thermal conduction in the electron gas rather than collisions with positive ions. His conclusions assume that the photoionization of helium is unimportant and additionally depend on an assumed model atmosphere.

NIGHTTIME CHARGED PARTICLE TEMPERATURE MEASUREMENTS

The charged particle temperature measurements of Figures 2 and 3 are too sparse and the dependency on the assumed reference atmosphere too critical to draw firm conclusions about departures from temperature equilibrium at night for mid-solar cycle conditions. Bowles et al (1962) observe that T_e/T_i is unity at night at the geomagnetic equator with $T_i \cong 600°K$, the latter value being in fair agreement with the Harris-Priester reference atmosphere. Remembering the different altitudes and times of the observations, Willmore et al (1963) report that at 1000 km midnight values of T_e increases from about 800°K at the equator to 1400°K at 60° magnetic latitude. Evans (1964) indicates a small but significant departure from temperature equilibrium in the F region at 50° north magnetic latitude. The definite evidence from the Ariel satellite for quite significant departures from temperature equilibrium at night at medium latitudes have been confirmed by rocket measurements (Brace et al, 1963). The Ariel results show that the nighttime departure from equilibrium becomes more pronounced at the higher latitudes. The nighttime source required to explain the Ariel observations has been estimated to be less than 30 percent of the daytime EUV heat input (Willmore, 1963).

As an example of the sensitivity of nighttime electron temperatures to small sources of heat input, consider Equation (1) and the fact that for solar minimum conditions, N_e varies by a factor of 4 from day to night at 400 km (Bauer and Blumle, 1964). From these considerations, it can be shown that less than 10 percent of the daytime EUV heat input would be required at night to maintain the same temperature difference (T_e-T_p) throughout the day.

SUMMARY AND CONCLUSIONS

For all temporal conditions and latitudes, large departures from temperature equilibrium $(T_e/T_p \geqq 2)$ are observed in the daytime lower F region at the altitude of maximum rate of electron productions. Moderate but significant daytime values for T_e/T_p are maintained to very high altitudes in winter at mid-latitudes in the middle of the solar cycle. Daytime mid-latitude data taken for summer and/or solar minimum conditions when electron densities generally are much lower reveal much larger values of T_e/T_p persisting to altitudes well above the F2 maximum. There is considerable evidence at least for altitudes below 600 km that the diurnal electron temperature maximum occurs in the early morning. Observed increases of daytime electron temperature with latitude follows the observed electron density which is under geomagnetic control. All of these temporal and latitudinal trends in the observed electron temperature are consistent with EUV as the predominant daytime source of electron heating. The uncertainties in calculating the EUV effect make it difficult to infer other possible daytime heat sources at the present time. The charged particle temperature observa-

tions strongly suggest the possibility that not all of the EUV energy is locally deposited, an important factor to be considered in the theories of formation of the ionosphere. There is some evidence principally from backscatter experiments that the ion temperature is controlled by electrons at very high altitudes.

Significant departures from temperature equilibrium at night have been observed especially for conditions close to solar minimum. The estimated intensity of this additional heat source increases with latitude but at all latitudes is only a fraction of the daytime EUV effect.

ACKNOWLEDGEMENT

The author is grateful to S. J. Bauer, L. J. Blumle, J. V. Evans and K. Bowles, for early access to the results of their charged particle temperature observations.

REFERENCES

Aono, Y., K. Hirao and S. Miyazaki, Journ. Rad. Res. Labs., Japan, **9**, 407, 1962.

Aono, Y., K. Hirao and S. Miyazaki, Journ. Rad. Res. Labs., Japan, **8**, 453, 1961.

Bauer, S. J. and Bourdeau, R. E., Journ. Atmos. Sci., **19**, 218, 1962.

Bauer, S. J., Journ. Geophys. Res., **69**, 553, 1964.

Bauer, S. J., and L. J. Blumle, JGR, submitted for publication, 1964.

Bourdeau, R. E., Space Research II, p. 554, 1961.

Bourdeau, R. E., Space Sciences Reviews, **1**, 683, 1963.

Bourdeau, R. E., and Donley, J. L., Proc. Royal Soc. A., in press, presented May 1963.

Bourdeau, R. E., J. L. Donley, E. C. Whipple and S. J. Bauer, Journ. Geophys. Res., **67**, 467, 1962.

Bowen, P. J., R. L. F. Boyd, W. J. Raitt and A. P. Willmore, Proc. Royal Soc. A., in press, presented May 1963.

Bowles, K. L., Ochs, E. R., and Green, J. L., Journ. Natl. Bureau Standards, **66**,, 395 1962.

Bowles, K., XIV General Assembly of URSI, Tokyo, September 1963.

Brace, L. H., N. W. Spencer, and G. R. Carignan, Journ. Geophys. Res., **68**, 5397, 1963.

Brace, L. H., N. W. Spencer and A. Dalgarno, AGU, Washington, April 1964.

Dalgarno A., McElroy, M. B. and Moffet, R. J., Plan. & Space Sci., **11**, 463, 1963.

Evans, J. V. and M. Loewenthal, URSI, Washington, April 1964.

Evans, J. V., Journ. Geophys. **67**, 4994, 1962.

Gringauz, K. I., B. N. Gerozhankin, N. M. Shutti. and G. G. Gdalevich, USSR Geoficicheskaya Sekt., **151**,, 560, 1963.

Hale, L. C., Journ. Geophys. Res., **66**, 1554, 1961.

Hall, L. A., K. R. Damon and H. E. Hinteregger, Space Research III, 1963.

Hanson, W. B., Space Research III, p. 282, 1962.

Hanson, W. B. and F. S. Johnson, Les Congress et Colloques de L'Université de Liège, **20**, 390, 1961.

Hanson. W. B., Journ. Geophys. Res., **67**, 183, 1962b.

Hariss, I., and W. Priester, Space Research III, p. 58, 1962.

King, J. W., P. A. Smith, D. Eccles and F. Helm, Radio Res. Station, I. M. 94, July 1963.

Lockwood, G. K. and G. L. Nemms, Journ. Atmos. Terr. Phys. 1964, in press.

Smith, L. G., Journ. Geophys. Res., **66**, 2562, 1961.

Spencer, N. W., Brace, L. H. and Carignan G. R., Journ Geophys. Res., **67**, 157 1962.

Taylor, H. A., L. H. Brace, H. C. Brinton and C. R. Smith, Journ. Geophys. Res., **68**, 539, 1963.

Willmore, A. P., C. L. Henderson, R. L. F. Boyd and P. J. Bowen, Proc. Royal Soc. A., in press, presented May 1963.

Willmore, A. P., Proc. Royal Soc. A., in press.

Wright, J. W., NBS Technical Note, 138, 1962.

POSITIVE ION TEMPERATURES ABOVE THE F-LAYER MAXIMUM*

N66 37958

R. L. F. BOYD AND W. J. RAITT†

The ion mass analyser on the UK–US satellite Ariel was a spherically symmetric energy spectrometer. The spectral lines may thus be interpreted not only in terms of ionic mass and concentration but also in terms of ion temperature.

Analysis of the data obtained has made available a picture of ion temperature variation with latitude above 400 km. The results show considerable scatter from day to day together with a marked rise in oxygen ion temperature with latitude.

INTRODUCTION

The determination of ion energy (and hence mass) spectrum by the ion probe on the satellite Ariel I (1962 –) provided good data on ion temperatures from 26 April, 1962 to 5 June, 1962. Some data of reduced quality is available until 26 September, 1962. Temperature is obtained from the width of the energy spectrum line for particles arriving at the probe. In principle the lines of hydrogen, helium or oxygen may all yield temperatures. The data presented in this paper all come from a study of the oxygen line which requires a concentration of oxygen greater than about 10^5 cm^{-3}. This implies that for the present study only altitudes below 600 Km. were useful and orbit orientation was such that in the period of this study (Days 117 to 140) this altitude range corresponds to values of local solar time within two hours of noon.

OUTLINE OF THE EXPERIMENT

The probe consisted of a 9 cm. diameter sphere mounted on the spin axis of the satellite. A fine grid 10 cm. in diameter was biassed negatively so as to keep ambient electrons from reaching the sphere, which was subjected to a positive voltage sweep. By constructing the system so that the angular momentum distribution of ions approaching the inner sphere was negligibly distorted by the presence of the grid it becomes valid to obtain the energy distribution of the ions by Druyvesteyn's[1] analysis. The required second derivative of the current voltage curve was obtained electronically and so telemetered.

If R is the kinetic energy of an ion moving with a velocity equal to that of the satellite and K is the energy corresponding to the most probable ion thermal velocity then for R≫K the form of the energy spectrum line is

$$f(E) = [2\pi^{\frac{1}{2}}(\text{K.R.})^{\frac{1}{2}}]^{-1} \exp \frac{-(E-R)^2}{4\,\text{K.R.}}$$

Thus the breadth of the energy spectrum line depends not only on the thermal energy of the ions but also on the velocity of the satellite. It is roughly Gaussian in form with a width at the $\frac{1}{e}$ points of $4(\text{K.R.})^{\frac{1}{2}}$. For monatomic oxygen R is about 5eV and K about 0.1eV so the line width is over 2 volts. While this broadening reduces the mass resolution it greatly eases the problem of determining the ion temperature.

Medicus[2] has obtained a more exact expression for the energy distribution of ions with a Maxwellian distribution arriving at a moving vehicle shown plotted in Fig. 1.

Analysis of the Ariel data has been carried out by fitting values computed from the approximate expression to the observed data.

*This paper was contributed to *Goddard Space Flight Center* under the joint United Kingdom–United States program which developed and launched the satellite Ariel I.

†Both authors are affiliated with the Department of Physics, University College, London, England.

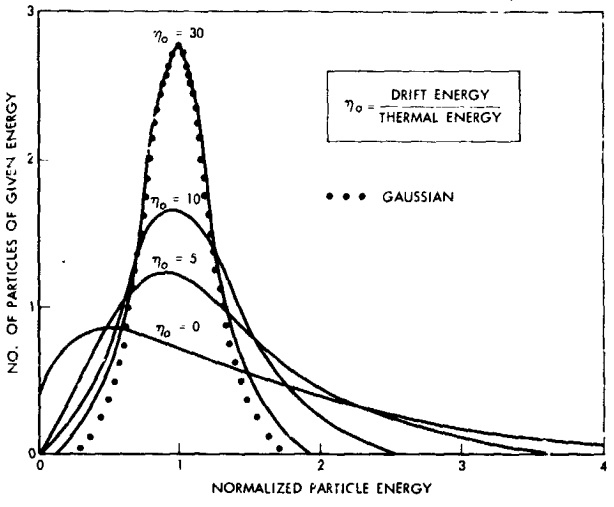


FIGURE 1.—Energy distribution of particles with Maxwellian distribution arriving at a moving vehicle.

EXPERIMENTAL RESULTS

The oxygen ion temperatures between 10 and 12 hours Local Solar Time measured in the longitude range −50° to 150° during the early northern summer of 1962 are plotted as a function of latitude in Fig. 2. The quantity of data is inadequate to make possible a complete decoupling of the effect of altitude from the effect of latitude. During the period selected, however, perigee was a little south of the equator so that the range of altitudes covered is not large. The approximate altitude corresponding to a given latitude is shown by the scale at the top of the figure.

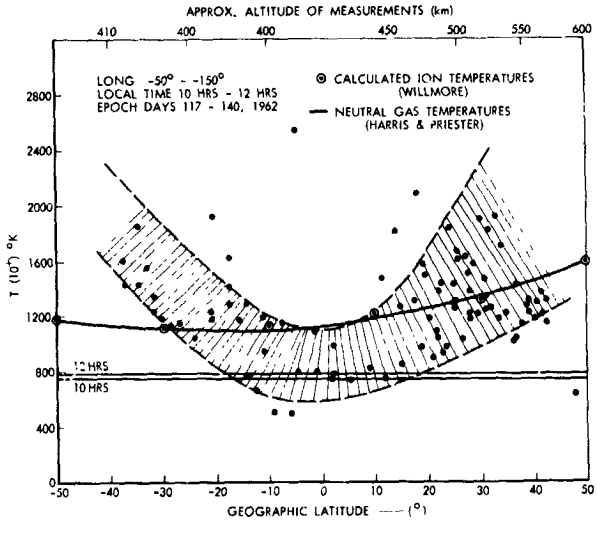


FIGURE 2.—Oxygen ion temperatures as a function of latitude.

Comparison of successive values of ion temperature obtained on a single pass suggest that instrumental consistency is at least as good as ±200 deg. K and possibly a good deal better. Most of the scatter of the points must therefore be taken to reflect real differences in ion temperature from one occasion to another.

Part of the scatter may be attributable to the range of local solar times, to the seasonal change and to the motion of the line of apsides. There may also be some dependence on magnetic activity and on longitude. However, it is not possible to account for most of the scatter in any of the above ways. The main conclusions are therefore that there is a variation in oxygen ion temperatures from one day to another amounting to several hundred degrees and that there is also a marked tendency of the oxygen ion temperatures to increase with latitude.

DISCUSSION OF RESULTS

Fig. 2 shows the neutral gas temperature based on the data of Harris and Priester[3] and a curve of ion temperature computed by A. P. Willmore[4] from the electron temperature data obtained on the same satellite (Willmore et al)[5] on the consideration that the heat flow from the electrons reaches the neutral gas by way of coulomb interaction with the ions. Willmore finds a rise of ion temperature with both latitude and altitude and appropriate weight has been given to both factors in plotting the curve from his data.

Except near the geomagnetic equator where the small dip angle would inhibit upward diffusion of photoelectrons, the energy input to the ionospheric electrons and hence to the ions due to solar radiation will vary as the cosine of the zenith angle.

Since the average fractional energy loss per collision of ions of mass M_i with gas atoms of mass M_g is

$$\frac{\Delta E}{3/2kT_i}=\frac{8}{3}\frac{M_iM_g}{(M_i+M_g)^2}\left(1-\frac{T_g}{T_i}\right)$$ where T_i and T_g refer to the ion and gas temperature.

We can write the energy flux

$$Q=4k\cdot n_i\cdot n_g\cdot S\cdot\left(\frac{3k}{2}\right)^{\frac{1}{2}}\left(\frac{T_i}{M_i}+\frac{T_g}{M_g}\right)^{\frac{1}{2}}\frac{M_iM_g}{(M_i+M_g)^2}(T_i-T_g)$$

where k is Boltzman's constant, n_i and n_g refer to the number densities of ions and gas atoms and S is the appropriate ion-atom energy exchange cross-section.

To a first approximation Q and n_i both vary as the cosine of the zenith $<$, n_g and T_g are constant and $(T_i - T_g) \ll T_g$. Thus writing $Q = Q \cos \chi$ and $n_i = n_0 \cos \chi$ we have

$$T_i - T_g \doteq Q_0 \left[4(3/2kT_g)^{\frac{1}{2}} \cdot n_0 \cdot n_g \cdot S \cdot k \cdot \left(\frac{1}{M_i} + \frac{1}{M_g} \right) \frac{M_i M_g}{(M_i + M_g)^2} \right]^{-1}$$

which shows that the factors most likely to influence the ion temperature in the topside ionosphere are ionic and neutral composition—affecting M_i, M_g and S—and ionic concentrations reduced to zero zenith angle, quantities which are also interrelated by the hydrostatic equation.

N_0 at the layer peak does in fact, fall with increasing latitude so the difference between the gas and ion temperature is to be expected to increase with latitude. The wide scatter in the data may well be related to variations not only in n_g but in ionic and possibly even in neutral particle composition. A preliminary consideration of the rate of energy exchange suggests that for the same flux the temperature difference between oxygen ions and helium atoms must be several times larger than when the ions are helium. The small amount of helium ion temperature data analysed so far supports this view. The possibility of the electric fields having an appreciable effect on the ion temperature would seem to be ruled out by the far larger effect they would have on the electrons.

One important factor must not be overlooked. It is the fact that by selecting only passes giving a high enough density of oxygen to enable oxygen ion temperature to be obtained at the higher altitudes (which correspond to higher latitudes) a premium is put on passes for which the ion temperature is high enough to give an adequate oxygen ion concentration.

REFERENCES

1. M. J. Druyvesteyn. Z. Phys. **64** (1930) 790.
2. G. Medicus. J. Appl. Phys. **33** (1962) 3094.
3. I. Harris and W. Priester. NASA Tech. Note. D–1444 (1962).
4. A. P. Willmore. Proc. Roy. Soc. (In press).
5. A. P. Willmore, R. L. F. Boyd, C. L. Henderson and P. J. Bowen. Proc. Roy. Soc. (1963) (In press).

GEOMAGNETIC CONTROL OF DIFFUSION IN THE UPPER ATMOSPHERE*

S. CHANDRA AND R. A. GOLDBERG

N64-28071

In many recent papers concerned with providing an explanation for the geomagnetic anomaly, agreement with measured data has been obtained from the equations of motion for electrons and ions when used with an empirical boundary condition, whereas poor agreement has resulted from attempts to numerically integrate the commonly employed form of the continuity equation. We have been able to explain this discrepancy by demonstrating that the assumptions used to derive this form of the continuity equations do not agree with observation.

Since the equations of motion do provide a favorable description for the geomagnetic anomaly, we have studied the possible physical models leading to the form of the equations used, and found that although field aligned diffusive equilibrium provides the correct form, a more reasonable assumption concerning electron and ion collisions with neutrals also leads to the same result. We have then been able to provide a more realistic theoretical description of the geomagnetic anomaly by employing an analytic form for the boundary condition which is in better agreement with measurement than those previously used.

Finally, by combining the equations of motion for neutrals, electrons and ions, we have been able to indicate geomagnetic control for the neutral atmosphere in the lower F region of the ionosphere, although the exact shape of this distribution is unknown.

INTRODUCTION

In recent months, it has become increasingly evident that some confusion exists in the understanding of the basic physical mechanisms governing diffusion and the existence of the geomagnetic anomaly in the ionosphere. This apparent confusion arises by comparison of the work of Chandra (1964), (to be referred to as C–1), in which it is shown that the assumption of ambipolar diffusion along a field line cannot lead to geomagnetic control of the charged particles, and such papers as Goldberg and Schmerling (1962, 1963), (to be referred to as GS–I and GS–II), and Goldberg, Kendall, and Schmerling (1964), (to be referred to as GKS), in which this process does appear to produce geomagnetic control of the charged particle density in the ionosphere.

The purpose of this paper is to describe and resolve the confusion which exists in the field at this moment, and then to point out the new problems with which we must contend in order to derive and apply the diffusion equation to ionospheric problems correctly. In addition, a section will be devoted to an improved theoretical description of the geomagnetic anomaly by using an analytic expression for the vertical electron density distribution at the equator which is more in accordance with measurement than the simple Chapman type distribution employed in GKS.

FUNDAMENTAL EQUATIONS AND DEFINITIONS

The major cause of confusion appears to lie in the application of two phrases, viz. *ambipolar diffusion* and *diffusive equilibrium*. Let us investigate and discuss each of these terms to determine how loose usage of them has led to some of the current problems of misunderstanding.

In the normal sense, ambipolar diffusion refers to a plasma in which the negative (electrons) and positive (ions) charges do not move independently due to the influence of the electric field caused by their Coulomb interactions. In this medium, the electrons and ions drift in pairs and this motion of electron-ion pairs is referred to as ambipolar diffusion. The condition for ambipolar diffusion in a neutral plasma is thereby

$$\vec{v}_e = \vec{v}_i = \vec{v} \qquad (1)$$

*Published as *Goddard Space Flight Center Document X-615-64-121*, May 1964.

where $\vec{v}$ is macroscopic velocity and the subscripts e and i refer to electrons and ions respectively. When

$$\vec{v}=0 \tag{2}$$

the condition for diffusive equilibrium is satisfied.

The implications of (1) are quite straightforward, as shown in C–I. In an isothermal atmosphere and in the presence of a magnetic field, this requires $\nabla x(\vec{v}x\vec{B})=0$. In particular, the assumption of field aligned plasma diffusion ($\vec{v}x\vec{B}=0$) can only be satisfied for the trivial case, $\vec{v}=0$, resulting in a hydrostatic distribution of electron density independent of geomagnetic latitude.

On the other hand, favorable comparison between Alouette topside sounder measurements and theoretical calculations of the geomagnetic anomaly has been obtained in GKS by assuming conditions of ambipolar diffusion and diffusive equilibrium along field lines, thereby indicating a possible conflict with the results in C–I. The problem resolves itself once one investigates the meaning of ambipolar diffusion and diffusive equilibrium in the GKS sense.

Let us first write the general equations of motion for neutrals, electrons and ions, respectively, where the subscript n refers to neutral. Following C–I:

$$\frac{n_e m_e m_n}{m_e+m_n}\nu_{en}(\vec{v}_n-\vec{v}_e)+\frac{n_i m_i m_n}{m_i+m_n}\nu_{in}(\vec{v}_n-\vec{v}_i) = -\nabla p_n+n_n m_n\vec{g} \tag{3}$$

$$\frac{n_i m_e m_i}{m_e+m_i}\nu_{ei}(\vec{v}_e-\vec{v}_i)+\frac{n_e m_e m_n}{m_e+m_n}\nu_{en}(\vec{v}_e-\vec{v}_n) = -\nabla p_e+n_e m_e\vec{g}-e\,n_e(\vec{E}+\vec{v}_e\times\vec{B}) \tag{4}$$

$$\frac{n_i m_i m_e}{m_e+m_i}\nu_{ei}(\vec{v}_i-\vec{v}_e)+\frac{n_i m_i m_n}{m_i+m_n}\nu_{in}(\vec{v}_i-\vec{v}_n) = -\nabla p_i+n_i m_i\vec{g}+e\,n_i(\vec{E}+\vec{v}_i\times\vec{B}) \tag{5}$$

where n is number density, ν_{kl} is the collision frequency between the k^{th} and l^{th} particle, m is mass, p is pressure, $\vec{g}$ is gravitational acceleration, e is the absolute value of electron charge, $\vec{E}$ is electric field, and $\vec{B}$ is magnetic field. In writing equations (3)—(5) it is assumed that $\frac{\nu_{lk}}{n_k}=\frac{\nu_{kl}}{n_l}$

In the following we assume that the plasma is in a quasineutral state

$$n_e\approx n_i=N \tag{6}$$

and that the electrons, ions and neutrals obey the ideal gas law in the ionosphere,

$$p_j=n_j kT_j \tag{7}$$

when k is Boltzmann's constant and T is temperature. Furthermore, we assume thermal equilibrium, i.e.

$$T_e=T_i=T \tag{8}$$

Then,

$$p_e=p_i=p \tag{9}$$

In addition, we assume for simplicity that

$$\vec{v}_n\approx 0 \tag{10}$$

Then summation of (4) and (5) provides

$$\frac{m_n m_e N}{m_e+m_n}\nu_{en}\vec{v}_e+\frac{m_n m_i N}{m_i+m_n}\nu_{in}\vec{v}_i=-2\nabla p +N(m_e+m_i)\vec{g}+\vec{J}\times\vec{B} \tag{11}$$

where $\vec{J}$ is current density, defined as

$$\vec{J}=Ne(\vec{v}_i-\vec{v}_e) \tag{12}$$

Since we are investigating ambipolar diffusion and diffusive equilibrium in the GKS sense, it is desirable to write this equation in component form along a field line as

$$\frac{m_n m_e}{m_e+m_n}\nu_{en}\vec{v}_e\cdot\vec{h}+\frac{m_n m_i}{m_i+m_n}\nu_{in}\vec{v}_i\cdot\vec{h} =\left[\frac{-2kT\nabla N}{N}+(m_e+m_i)\vec{g}\right]\cdot\vec{h} \tag{13}$$

where $\vec{h}$ is a unit vector in the direction of the magnetic field. Let us write (13) in more familiar form by using

$$m_e\ll m_i,\ m_n \tag{14}$$

and defining the scale height of the ionizable constituent as H_i, where

$$H_i=\frac{kT}{m_i g} \tag{15}$$

Also, for convenience, we make the approximation

$$m_n\approx m_i \tag{16}$$

Then

$$\frac{m_e\nu_{en}}{2}\vec{v}_e\cdot\vec{h}+\frac{m_i}{4}\nu_{in}\vec{v}_i\cdot\vec{h}=-kT\left(-\frac{\nabla N}{N}+\frac{\vec{i}_r}{2H_i}\right)\cdot\vec{h} \tag{17}$$

Finally, we write

$$\vec{h} = -(\vec{i}_r \sin I + \vec{i}_\theta \cos I) \tag{18}$$

where $\vec{i}_r$ and $\vec{i}_\theta$ are unit vectors in the r and θ directions and I is the magnetic dip angle, reckoned positive when the north seeking pole of the needle points downward. Now, if we treat ambipolar diffusion in the GKS sense, we simply imply that the electron and ion velocity components in the field direction are equal, i.e.

$$\vec{v}_e \cdot \vec{h} = \vec{v}_i \cdot \vec{h} = v_\parallel \tag{19}$$

Applying (18) and (19) in (17), we obtain

$$v_\parallel = \frac{-kT}{\mu\nu}\left[\sin I\left(\frac{1}{N}\frac{\partial N}{\partial r} + \frac{1}{2H_i}\right) + \frac{\cos I}{Nr}\frac{\partial N}{\partial \theta}\right] \tag{20}$$

where

$$\mu\nu = \frac{m_e\nu_{en}}{2} + \frac{m_i\nu_{in}}{4} \tag{21}$$

Assuming that

$$m_e\nu_{en} \ll m_i\nu_{in} \tag{22}$$

because of (14), we may write

$$\mu\nu \approx \frac{m_i\nu_{in}}{4} \tag{23}$$

Equation (20) is a familiar result derived in such papers as Kendall (1962) and GS–II. However, it is clearly not the result of ambipolar diffusion, which is given by (1), but instead, the result of a statement concerning the field line components of electron and ion velocities given by (19).

If we now demand

$$v_\parallel = 0 \tag{24}$$

which is the statement implying diffusive equilibrium along a field line in the GKS sense, we obtain the familiar equation

$$\sin I\left(\frac{1}{N}\frac{\partial N}{\partial r} + \frac{1}{2H_i}\right) + \frac{\cos I}{Nr}\frac{\partial N}{\partial \theta} = 0 \tag{25}$$

which can also be written as

$$\frac{1}{N}\frac{dN}{dr} + \frac{1}{2H_i} = 0 \tag{26}$$

provided we recognize that r and θ are not independent in (25) but related by the dipole field conditions

$$r = r_0 \sin^2\theta \tag{27}$$

and

$$\tan I = 2\cot\theta \tag{28}$$

It is evident that (26) can only be treated in total derivative form if the integration is carried out along the field line.

Statements concerning the components of vectors in a particular direction, such as (19), do not imply any conditions on the total vector. As a result, (25) has not required the assumption of any restrictions on the behavior of the velocity components normal to the field lines.

Equation (26) has been the basis of describing geomagnetic control in the upper F-region in GKS paper. Although (26) has been derived assuming diffusive equilibrium along a field line, it is undesirable to apply this concept because it is of purely hypothetical nature. We now investigate other assumptions to find a more realistic justification for (26).

Let us rewrite (17) as

$$\frac{m_e\nu_{en}}{2}v_{e_\parallel} + \frac{m_i\nu_{in}}{4}v_{i_\parallel} = -kT\left(-\frac{\nabla N}{N} + \frac{\vec{i}_r}{2H_i}\right)\cdot\vec{h} \tag{29}$$

We find two ways for the right hand side of (29) to approach zero. The first imposes a new condition on the velocities, viz:

$$\vec{v}_e = -\frac{m_i\nu_{in}\vec{v}_i}{2m_e\nu_{en}}$$

or

$$v_{e_\parallel} = -\frac{m_i\nu_{in}v_{i_\parallel}}{2m_e\nu_{en}} \tag{30}$$

a result which, although possible, would require a very special condition that the electron velocity be of the order of 10^3 times greater than and in the opposite direction of the ion velocity.

However, if we can demand that the collision frequencies between electrons and neutrals and between ions and neutrals be sufficiently small so that the drag forces arising due to collision be negligible as compared to the pressure gradient, gravity and Lorentz forces it is possible to derive equation (26) without imposing any restriction

on the velocities of electrons and ions. We believe that this assumption is more realistic in the upper F-region where the gyro-frequencies of electrons and ions are much greater than their corresponding collision frequencies.

Although the collision frequency assumption is physically more desirable, it prevents us from obtaining a simple expression for $v_{e\parallel}$ or $v_{i\parallel}$. Instead, we must return to the original equations of motion, (4) and (5), and solve for $\vec{v}_e$ and $\vec{v}_i$ explicitly, as has been carried out in the appendix in C-I. Unfortunately this introduces a very serious complication in the work because of the difficulty in eliminating electric field from the expressions of $\vec{v}_e$ and $\vec{v}_i$ without making specific assumptions about the relationship between $\vec{v}_e$ and $\vec{v}_i$. The implications of these assumptions will be discussed in the latter part of this paper. In the following section we proceed to discuss the physical implications of equation (26).

THE ELECTRON DENSITY DISTRIBUTION WITH A VARIABLE SCALE HEIGHT

Equation (26) can be integrated along a field line to provide the general solution

$$N(r,\theta)=f(r_0,\pi/2)e^{\frac{r\cot^2\theta}{2H_i}} \tag{31}$$

However, if we treat T and m_i constant but recognize that g is proportional to $1/r^2$, H_i is then proportional to r^2, and we obtain

$$N(r,\theta)=f(r_0,\pi/2)e^{\frac{r\cos^2\theta}{2H_i(r)}} \tag{32}$$

In both cases, $f(r_0, \pi/2)$ is an arbitrary function of height at the equator which cannot be determined by the equations of motion from which (31) or (32) are derived. The function $f(r_0, \pi/2)$ must therefore be given as a boundary condition in this problem and can only be determined empirically or by use of additional equations governing the physics of the problem.

Since (31) or (32) depend exclusively upon the equations of motion, it appears that an additional equation, such as the continuity equation, should lead to the desired boundary condition. Unfortunately, as we will show in the next section, the derivation and solution of the continuity equation depend upon a knowledge of $\vec{E}$. Thus, the complexity of the problem becomes quite formidable and it is difficult to anticipate a simple method of solution at this time.

Instead we depend upon an empirical type boundary condition, which may very well be the solution of the correct continuity equation, to derive the explicit form of the electron density distribution.

The incorporation of a Chapman distribution for the boundary condition in (31) leads to the results obtained in CKS. Since such a boundary condition can only be considered as a rough approximation to the shape of the actual vertical electron density distribution at the equator, it is desirable to employ an analytic boundary condition which more closely resembles the true height profiles. Chandra (1963) has proposed a modified form of the Chapman function which includes the effect of variable scale height and which is found to fit the measured vertical distribution for electron density at mid-latitudes far more accurately than the simple Chapman form. We assume here that such a function also describes the vertical electron density distribution at the equator. We can then write

$$f(r_0,\pi/2)=N_{r_{m0}}\exp\frac{1}{2}\left\{1-\frac{r_0-r_{m0}}{H_0\left[1-\alpha\exp\left(-\alpha\frac{(r_0-r_{m0})}{2H_0}\right)\right]}-\exp\left[-\frac{r_0-r_{m0}}{2H_0\left[1-\alpha\exp\left(-\frac{\alpha(r_0-r_{m0})}{2H_0}\right)\right]}\right]\right\} \tag{33}$$

where H_0 is the scale height of the ionic constituent and $N_{r_{m0}}$ is the value of electron density at the equatorial height r_{m0}. The parameter α, which is a measure of departure from the simple Chapman function, and thereby a *shape factor*, is defined as

$$\alpha=\frac{H_0-H(r_{m0})}{H_0} \tag{34}$$

where $H(r_{m0})$ is that value of (Hr_0) at $r_0=r_{m0}$. Also, r_0 is understood to be the radial height specifically at $\theta=\pi/2$.

Although it will not be shown here, (31) and (32) produce nearly identical results in the equatorial region because the small variation of r in the height region of our interest. Furthermore,

the simplified form given by (31) is more convenient for comparison with the results of the GKS paper. We therefore substitute (33) into (31) and obtain

$$N(r,\theta)=N_{r_{m0}}\exp\frac{1}{2}\left\{1-\frac{r\csc^2\theta-r_{m0}}{H_0\left[1-\alpha\exp\left(-\frac{\alpha(r\csc^2\theta-r_{m0})}{2H_0}\right)\right]}+\frac{r\cot^2\theta}{H_t}-\exp\left[-\frac{r\csc^2\theta-r_{m0}}{2H_0\left[1-\alpha\exp\left(-\frac{\alpha(r\csc^2\theta-r_{m0})}{2H_0}\right)\right]}\right]\right\} \tag{35}$$

Equation (35) then provides a general expression for the electron density at all heights and colatitudes provided that we are in a region where the effects of collision can be neglected.

The variation of $N(h,\theta)N_{h_{m0}}$ with colatitude at constant height is shown in Figures 1–4 for various values of α, h_{m0} and H_0. In these figures, we have converted radial height r to altitude h by taking the earth's radius as 6370 km. We have also assumed that $H_0=H_t$ because the effective scale height in (33) approaches H_0 at high altitudes and Chandra (1963) has indicated that H_0 becomes equal to H_t at these altitudes.

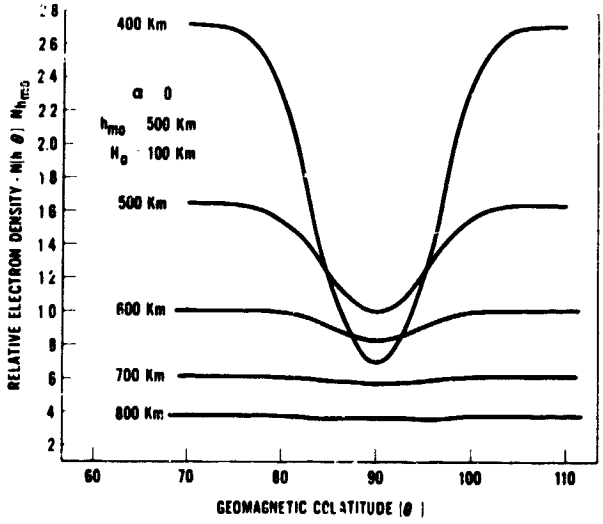


FIGURE 1.—Constant height profiles of relative electron density vs. colatitude for $\alpha=0$, $h_{m0}=500$ km, $H_0=100$ km.

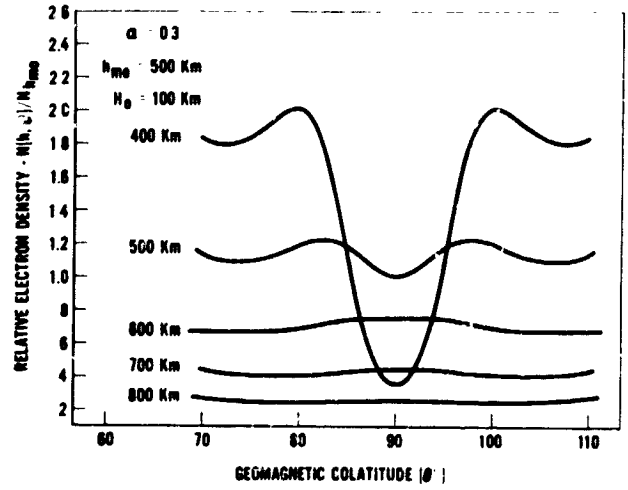


FIGURE 3.—Constant height profiles of relative electron density vs. colatitude for $\alpha=0.3$, $h_{m0}=500$ km, $H_0=100$ km.

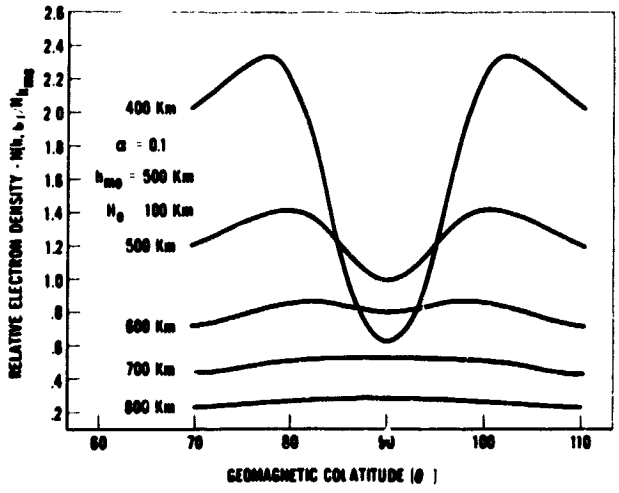


FIGURE 2.—Constant height profiles of relative electron density vs. colatitude for $\alpha=0.1$, $h_{m0}=500$ km, $H_0=100$ km.

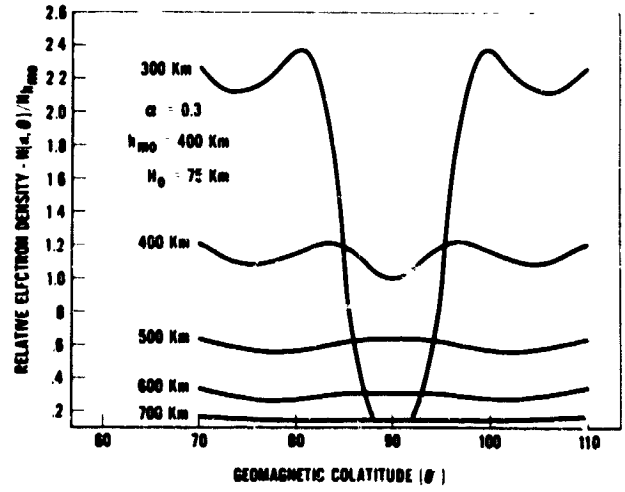


FIGURE 4.—Constant height profiles of relative electron density vs. colatitude for $\alpha=0.3$, $h_{m0}=400$ km, $H_0=75$ km.

We first note that the basic features of the theoretical electron density distribution are unaltered from those first obtained in GKS to describe the geomagnetic anomaly in the vicinity of the equator. Once again the theoretical description breaks down in the bottomside but this is precisely the region where the neglect of momentum transfer terms becomes in-valid. Furthermore, comparison of Figures 3 and 4 clearly shows the insensitivity of the topside results to the parameters h_{m0} and H_0 (except for shifting the constant height profile scale vertically). We therefore conclude the principal properties of the curves can be studied quite extensively by simply altering the shape factor α.

The changes due to variations in α are shown by comparison of Figures 1, 2 and 3. We have also provided a more detailed comparison for one particular height profile in Figure 5. Although

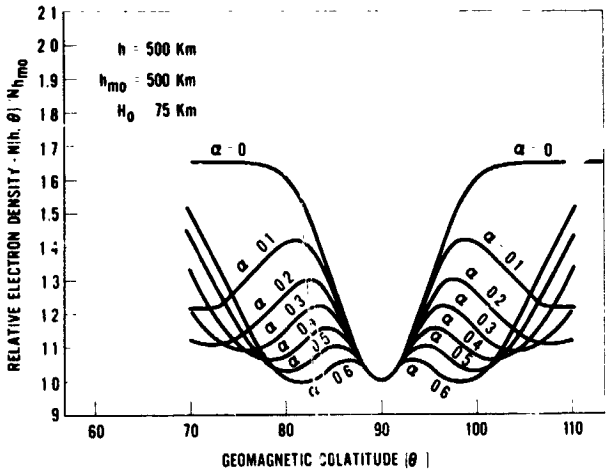


FIGURE 5.—The α dependent behavior of the 500 km profile of relative electron density vs. colatitude for $h_{m0}=500$ km, $H_0=75$ km.

we have included values up to $\alpha=0.6$ to demonstrate the trend of the curves, the highest values are extreme and not likely to be representative of ionospheric conditions. On the other hand, $\alpha=0.1$ to $\alpha=0.4$ are very reasonable values for us to expect under normal conditions representing diurnal and solar cycle variations.

Finally, in equation (35), if we identify the term $H_0[1-\alpha \exp[-\alpha(r \csc^2\theta - r_{m0})/2H_0]]$ with 1/k of the GKS paper, we see that since α is positive, $kH_t>1$ is true for all heights. In particular, if $\alpha=0$, we generate curves which are identical to those in GKS for $kH_t=1$. This explains why $kH_t>1$ provides the closest fit with experimental data in that paper.

PROBLEMS INVOLVED IN THE DERIVATION OF THE DIFFUSION EQUATION

In the previous sections we have seen how the equations of motion for electrons and ions are sufficient to obtain a theoretical description of the electron density distribution in the topside equatorial region of the ionosphere under equinox conditions. This has required us to make certain assumptions concerning collision frequencies or velocity components along field lines and also forced the application of an empirical boundary condition at the equator. In order to produce the empirical boundary condition theoretically and also obtain a solution which is valid in both the topside and bottomside equatorial F region, it is necessary to turn to the continuity equation for additional information. Using the explicit expressions for velocity which are derivable from the equations of motion, it is then possible to derive the diffusion equations associated with the ionosphere.

If we simply require total ambipolar diffusion (equation 1) to occur in the ionosphere so that $\vec{v}$ is independent of the electric field explicitly, and also demand that, in all regions concerned, the motion along field lines are much larger than the drifts normal to field lines, we must then invoke $\vec{v}\times\vec{B}=0$ which, using (4), (5) and (28), gives the constraint equation

$$\frac{1}{N}\frac{\partial N}{\partial r}+\frac{1}{2H}=\frac{2}{N}\frac{\partial N}{r\partial\theta}\cot\theta \qquad (36)$$

This leads to a hydrostatic distribution of electron density which does not agree with measured results, as has been demonstrated in C–I. A second approach (Kendall, 1962, and GS–II) is the assumption that ambipolar diffusion exists only along field lines (see equation (19)). Thus, if we assume that the parallel velocity components of electron and ion velocity are equal and much greater than either of the unequal perpendicular velocity components, we can write (20) as a good approximation for the entire velocity. In mathematical notation, we have

$$\vec{v}\cdot\vec{h}=|\vec{v}_\parallel|=v_\parallel\gg v_{e_\perp},\ v_{i_\perp} \qquad (37)$$

where $v_{e_\perp}$ and $v_{i_\perp}$ are the perpendicular components of electron and ion macroscopic velocities respectively. This implies that

$$\vec{v}_\parallel \approx \vec{v} \tag{38}$$

and provides us with a velocity expression independent of electric field. The general contention has been that (38) allows us to write the steady-state continuity equation in the following form:

$$Q-L=\nabla\cdot N\vec{v}\approx\nabla\cdot N\vec{v}_\parallel \tag{39}$$

where Q and L are production and loss respectively. The procedure has been to substitute (20) into (39) and obtain the well known form of the two dimensional diffusion equation without invoking the equation of constraint (36).

We wish to discuss this approach by first questioning the validity of (37), and then demonstrating that even if it were true, (38) cannot in general imply (39) without the additional inclusion of the constraint equation. This will demonstrate that the field line ambipolar diffusion approach with neglect of the perpendicular velocity components is identical to the total ambipolar diffusion case in which velocities are assumed to lie along field lines. Thus, the results of the two approaches are identical, leading to the conclusion that ambipolar diffusion in which the macroscopic velocity lies along a field line, cannot be the correct physical model to describe the equatorial electron density distribution in the F region of the ionosphere.

Let us first consider (37). We have already seen that the assumption of diffusive equilibrium along a field line ($v_\parallel=0$) leads to a correct description of the electron density in at least the topside region of the ionosphere. If this is the true model of the physical situation, then it is inconsistent with (37) and we cannot expect any results obtained using (37) to provide us with correct results concerning this region. If, on the other hand, the neglect of momentum transfer terms can be attributed to small collision frequencies instead of diffusive equilibrium, (37) need not be violated. This might be a further justification for validating the collision frequency assumption instead of the diffusive equilibrium model. Unfortunately, as we approach the equator, we see from (20) that $v_\parallel$ approaches zero since both sin I and $\partial N/\partial\theta$ approach zero. The latter condition is based strictly upon the empirical condition of symmetry about the equator. We therefore find that no matter how small $v_{e_\perp}$ and $v_{i_\perp}$ may be, there will always be a region about the equator in which (37) does not apply unless

$$v_{e_\perp}=v_{i_\perp}=0 \tag{40}$$

which is identical to the equation of constraint, (36).

We now return to the second question. That is, even if the parallel components of electron and ion velocities are much greater than the perpendicular component, which could still be possible provided $v_{e_\parallel}\neq v_{i_\parallel}$, is it possible to describe the electron density distributions in the entire region of the ionosphere by (39)? We note that

$$\nabla\cdot N\vec{v}=\vec{v}\cdot\nabla N+N\nabla\cdot\vec{v}=(\vec{v}_\parallel+\vec{v}_\perp)\cdot\nabla N+N\nabla\cdot(\vec{v}_\perp+\vec{v}_\parallel) \tag{41}$$

where $\vec{v}$ is now either the electron or ion velocity and

$$\vec{v}=\vec{v}_\parallel+\vec{v}_\perp \tag{42}$$

Now, in order to write (39), we must demand that

$$\vec{v}_\parallel\cdot\nabla N+N\nabla\cdot\vec{v}_\parallel\gg\vec{v}_\perp\cdot\nabla N+N\nabla\cdot\vec{v}_\perp \tag{43}$$

Although (43) could be true for certain special cases, there is not a priori guarantee that (43) will be implied by (37) in general without the additional condition that $v_\perp=0$. Thus, if we are to write (39) as a direct and general implication of (37), we are once again forced to employ the constraint equation.

We cannot state that (40) holds in a very small region about the equator so that its effect outside this region can be neglected. The geomagnetic anomaly itself is a second order effect and we cannot expect to reproduce it by neglecting second order terms which are responsible for its existence.

We therefore find that if ambipolar diffusion exists in the ionosphere, and if it is restricted to the field line direction, we cannot assume (37) without imposing an additional constraint equation. Furthermore, (37) does not generally imply (39) in the ionosphere with or without ambipolar diffusion unless the constraint equation is also employed. However, since (37), (39), and the assumption of ambipolar diffusion along a field line do not provide to the correct description of equatorial electron density, we must conclude that these

assumptions are not valid in a theory leading to a description of the electron density distribution in the equatorial ionosphere.

Kendall (1962), and Rishbeth, Lyon and Peart (1963), have attempted to numerically integrate (39) derived from (20) and (37), without invoking the equation of constraint. They have been unable to obtain the correct description of the geomagnetic anomaly and have therefore concluded that diffusion may not be a very important physical process governing the measured distribution of electron density. However, on the basis of the discussion presented in this section, it now appears that the physical assumptions used in deriving the form of the continuity equation used in their work may not be valid, which simply implies that the diffusion equation is far more complicated than originally believed. Since (19) and (37) are not longer valid, we can no longer equate electron and ion velocities to eliminate electric field. Instead, we must write separate continuity equations for electrons and ions and describe the behavior of electric field before it is possible to obtain the correct theoretical description of the geomagnetic anomaly.

It may appear that the results presented in GS–II are also not valid for the reasons discussed above. However, a closer inspection of GS–II shows that no new information was obtained from the solution of continuity equation than that already available from the equations of motion. The equation discussed in GS–II was simply

$$\nabla\cdot N\vec{v}_{\parallel}=0 \tag{44}$$

where the explicit production and loss terms were neglected in obtaining the series solution. Furthermore, as shown in GKS, the equation of motion leading to (26), whether derived assuming $\vec{v}_{\parallel}=0$, or by making assumptions concerning collision terms, has the identical solution to that obtained from (44) in GS–II. For the case $\vec{v}_{\parallel}=0$, (44) obviously cannot give any new information. This explains why the empirical boundary condition was necessary to obtain a non-arbitrary solution from (44) in GS–II. We should point out, however, that solutions of (39) making use of explicit production and loss terms should not give correct results in the equatorial regions of the ionosphere for the reasons discussed in this section.

THE DISTRIBUTION OF THE NEUTRAL ATMOSPHERE

In this section we will show that when the drag forces are not negligible, as might be the case in the lower F-region and E-region, it is possible to study the behavior of the neutral atmosphere without imposing any restrictions on the velocities of the various constitutents. To obtain the necessary starting equation, we first sum (3), (4) and (5):

$$-\nabla(p_e+p_i+p_n)+(n_nm_n+Nm_i)\vec{g}+\vec{J}\times\vec{B}=0 \tag{45}$$

where we have once again used (14). The component of (45) along the direction of magnetic field is then

$$[-\nabla(p_e+p_i+p_n)+(m_nn_n+m_iN)\vec{g}]\cdot\vec{h}=0 \tag{46}$$

Comparison of (45) and (46) shows that the net force due to pressure gradient and gravity of all particles is perpendicular to the magnetic field and balanced by a current flow force. Next, using (7), (8), (9), (18) and (28), we have

$$\frac{\partial(n_n+2N)}{\partial r}+\frac{\tan\theta}{r}\frac{\partial(n_n+2N)}{\partial\theta}+\frac{n_n}{H_n}+\frac{N}{H_i}=0 \tag{47}$$

where H_n is the scale height of the neutral atmosphere.

Since $N\ll n_n$ and $H_i\approx H_n$, we can write

$$\frac{n}{H_n}+\frac{N}{H_i}\approx\frac{n}{H_n}+\frac{N}{H_n}\approx\frac{n+2N}{H_n} \tag{48}$$

where the suffix on n has been dropped for simplicity. Then (47) becomes, in total derivative form,

$$\frac{d(n+2N)}{dr}+\frac{n+2N}{H_n}=0 \tag{49}$$

Integration of (49) along the field line gives

$$n(r,\theta)\approx n+2N=\vec{g}(r_0,\pi/2)e-\int_{r_0}^{r}\frac{dr}{H_n} \tag{50}$$

where $g(r_0,\pi/2)$ is an arbitrary function of height at the equator and r_0 is defined in (27). If we now demand that the radial distribution of the neutrals obey the normal hydrostatic law at the equator, so that

$$g(r_0,\pi/2)=n_{00}\,e-\int_{r_{00}}^{r_0}\frac{dr}{H_n} \tag{51}$$

where n_{00} is the neutral number density at height r_{00} on the equator, then

$$n(r) = n_{00}\, e - \int_{r_{00}}^{r} \frac{dr}{Hn}, \qquad (52)$$

a result which is entirely independent of θ. If, on the other hand, $g(r_0, \pi/2)$ is perturbed *in any manner* from the exact hydrostatic equilibrium case, we will obtain a distribution for n which does depend on θ. The origin of this angular dependence on the neutrals may seem somewhat surprising until we realize that in selecting a functional form for $g(r_0, \pi/2)$, any deviation in the equatorial neutral distribution from hydrostatic equilibrium must arise due to collisions between neutrals and geomagnetically controlled charged particles. Thus, if the collisions between neutrals and charged particles are sufficiently large to make the momentum transfer forces between charged and neutral particles important, the neutrals will begin to tend toward the angular distribution of the geomagnetically controlled particles. This can also be seen from (3), where it is obvious that we will not obtain the exact hydrostatic distribution in a region when the terms on the left hand side become important. On this basis, we might expect to observe angular variations of the neutral distribution in the bottomside regions of the ionosphere where charged-neutral particle interactions become important.

CONCLUSIONS

From the discussion and results of this paper, we have shown the following:

1. From the equations of motion, it is possible to derive an expression for the electron-density distribution along a field line either by assuming diffusive equilibrium along the direction of the magnetic field or by neglecting the drag forces arising from collisions. The latter assumption appears to be more realistic in the topside of the ionosphere. In either case, it is necessary to assume a radial distribution at the equator to obtain the electron density distribution.

2. We have provided a more accurate formula for the representation of the equinox geomagnetic anomaly than that produced in GKS. Since the empirical boundary condition equation used herein has been shown by Chandra (1963) to fit nearly all vertical profiles of electron density measured to date, we can safely assume that the proper selection of parameters in this formula will lead to a reasonable reproduction of the anomaly in any equatorial region of the ionosphere where interactions of neutrals with charged particles are small because of infrequent collisions.

3. The theory discussed above is semiphenomenological; i.e., it is based on effect and not cause. It does not require a knowledge of the complicated array of physical effects and mechanisms which combine to form the geomagnetic anomaly but, instead, uses an empirical boundary condition which is the accumulated effect of all these causes.

Naturally, if we are to increase our knowledge of the basis mechanisms causing the anomaly and thereby replace the empirical boundary condition by one based on more fundamental considerations than measurement, we must turn to the equations of continuity. Unfortunately, the derivation of the correct continuity equations requires knowledge concerning the electron and ion velocities and/or the electric fields acting on these particles. Currently, most derivations of the continuity equations employ simplifying assumptions, such as $v_{e\parallel} = v_{i\parallel} = v_\parallel$; $v_\parallel \gg v_{e_\perp}, v_{i_\perp}$. The equation derived in the literature under the above assumptions has been numerically integrated by several workers to obtain a theoretical electron density distribution near the equator under steady state conditions. The results obtained by these workers have been unable to account for the gross features of the geomagnetic anomaly, at least to the correct order of magnitude. This has led them to believe that diffusion is of minor importance in governing the geomagnetic anomaly.

We have been able to demonstrate that the velocity assumptions described above do not lead to the proper description of the geomagnetic anomaly. We therefore feel that the assumption about velocities used in the continuity equation rather than the ineffectiveness of motions are responsible for the unsatisfactory description of the geomagnetic anomaly obtained by others.

4. A study of the neutral atmosphere distribution has led us to the conclusion that geomagnetic control of neutrals occurs in any region of the ionosphere where interactions of neutrals with charged particles become important. Since this is most likely to occur in the lower F region of

the ionosphere we suggest that such geomagnetic control of the neutrals might be observable in this region.

ACKNOWLEDGEMENTS

We wish to thank Miss Partica A. Egan for programming the numerical computations used in this work. We also wish to thank Drs. Frank C. Jones and Arnold P. Stokes for many valuable suggestions and critical comments concerning this paper.

Both authors have carried out this research as NAS–NRC Postdoctoral Resident Research Associates.

BIBLIOGRAPHY

CHANDRA, S., *Electron Density Distribution in the Upper F Region*, J. Geophys. Res. **68**, 1937–1942, 1963.

CHANDRA, S., *Plasma Diffusion in the Ionosphere*, J. Atmospheric Terrest. Phys. **26**, 113–122, 1964.

GOLDBERG, R. A. and E. R. SCHMERLING, *The Distribution of Electrons Near the Magnetic Equator*, J. Geophys. Res. **67**, 3813–3815, 1962.

GOLDBERG, R. A. and E. R. SCHMERLING, *The Effect of Diffusion on the Equilibrium Electron Density Distribution in the F Region Near the Magnetic Equator*, J. Geophys. Res. **68**, 1927–1936, 1963.

GOLDBERG, R. A., P. C. KENDALL, and E. R. SCHMERLING, *Geomagnetic Control of the Electron Density in the F Region of the Ionosphere*, J. Geophys. Res. **69**, 417–427, 1964.

KENDALL, P. C., *Geomagnetic Control of Diffusion in the F2 Region of the Ionosphere. I. The Form of the Diffusion Operator*, J. Atmospheric Terrest. Phys. **24**, 805–811, 1962.

KENDALL, P. C., *Geomagnetic Control of Diffusion in the F2 Region of the Ionosphere. II. Numerical Results*, J. Atmospheric Terrest. Phys. **25**, 87–91, 1963.

RISHBETH, H., A. J. LYON and M. PEART, *Diffusion in in the Equatorial F Layer*, J. Geophys. Res. **68**, 2559–2569, 1963.

THE EFFECT OF A VARIABLE ELECTRON TEMPERATURE ON THE EQUATORIAL ELECTRON DENSITY DISTRIBUTION IN THE UPPER IONOSPHERE

RICHARD A. GOLDBERG

By incorporating a model for the measured electron temperature distribution at the magnetic equator in the isothermal and temperature equilibrium theory of Goldberg, Kendall and Schmerling [J. Geophys. Res., *69*, 417–427, 1964], it has been possible to gain further insight into the behavior of the equatorial geomagnetic anomaly under steady state and equinoctial conditions. In particular, it is shown that the measured deviation from thermal equilibrium in the bottomside ionosphere is very influential in allowing extension of the previous theoretical description of the geomagnetic anomaly well into the bottomside ionosphere and to higher latitudes than originally applicable. For completeness, the effect of gravitational variation is now included, but it is shown that this alone contributes only minor improvements to the original results. Finally, several less common properties in the behavior of the geomagnetic anomaly are investigated, and it is shown under what conditions these secondary effects will occur.

INTRODUCTION

The theory describing the equatorial electron density distribution under conditions of thermal equilibrium, equinox, and steady state, originally presented in Goldberg and Schmerling [1963, 1964] and improved upon in Goldberg, Kendall and Schmerling [1964] (to be referred to as GKS), has provided reasonably good agreement when comparison is made to the measured results of the Alouette Topside Sounder Satellite. Furthermore, Chandra and Goldberg [1964] (to be referred to as CG) have demonstrated that the rather artificial concept of field line diffusive equilibrium need not be employed to obtain the necessary equations, but instead, the neglect of collisions between charged and neutral particles is sufficient in this aim. This has enabled us to understand why the theoretical results agree best with data in the topside ionosphere, since this is the region where such an assumption is most reasonable.

Recent theoretical considerations (Hanson and Johnson, 1961; Hanson, 1962; Dalgarno et al, 1963) and recent measurements with rocket probes (Spencer et al, 1962; Brace et al, 1963) and radar backscatter techniques (Evans, 1962 and 1964; Bowles, 1964) have now demonstrated that thermal equilibrium (electron temperature T_e = ion temperature T_i) does not occur in the lower F region ionosphere during the day including that time when the electron density is experiencing nearly steady conditions. It is the purpose of this paper to provide a simple analytical approach for including a temperature model, based on the theoretical or measured T_e distribution, in the theory discussed in GKS and CG and to demonstrate the possible effects of deviation from conditions of thermal equilibrium on the topside electron density distribution. Furthermore, it will be shown how inclusion of this T_e distribution has allowed extension of the theory into part of the bottomside F layer. Comparison with data is also made to provide a possible explanation of several features of the geomagnetic anomaly heretofore unexplained. For completeness, the effects of variable gravity are also included and the slight modifications in the results due to this effect are demonstrated.

*Published as *Goddard Space Flight Center Document X-615-64-279*, October 1964.

SYMBOLS

$\vec{B}$	magnetic field of earth
e	absolute value of electron charge
$\vec{g}$	gravity
H_e	electron density scale height
H_i	ionizable constituent scale height; cf. (16)
H_p	value of H_i at r_p
H_T	scale factor of temperature
H_τ	cf. (7)
$\vec{h}$	unit vector in direction of $\vec{B}$
h_mF2	height of $F2$ layer electron density peak
I	magnetic dip angle; cf. (10)
$\vec{i}_r$, $\vec{i}_\theta$	polar coordinate unit vectors
$\vec{J}$	current density; cf. (3)
K	value of T_e/T_i at r_p
k	Boltzmann's constant
m	mass
N, n	number density; cf. (2)
N_{m0}	equatorial peak electron density
N_mF2	$F2$ layer peak electron density
p	pressure
R	mean radius of earth (6370 km)
r	radial height ($r=z+R$)
r_{m0}	height of equatorial peak electron density
r_0	equatorial radial height
r_p	radial height of temperature peak
T	temperature
$\vec{v}$	velocity
z	altitude
θ	colatitude
ν_{jk}	collision frequency between j^{th} and k^{th} particles
τ	average of electron and ion temperatures; cf. (6)

SUBSCRIPTS

e, i, n	electrons, ions, and neutral particles respectively
m	peak value
mo	equatorial peak value

THE FUNDAMENTAL EQUATION

From CG, combination of the equations of motion for electrons and ions provides

$$Nm_e\nu_{en}\vec{v}_e+\frac{Nm_im_n}{m_i+m_n}\nu_{in}\vec{v}_i= -\nabla p_e-\nabla p_i+Nm_i\vec{g}+ir\,\vec{J}X\vec{B} \quad (1)$$

where we have assumed

$$\begin{aligned} v_n &\ll v_e,\ v_i \\ m_e &\ll m_i,\ m_n \\ n_e &= n_i = N \end{aligned} \quad (2)$$

In (1) and (2), the subscripts e, i, and n refer to electrons, ions, and neutrals respectively; p is pressure; n is number density; ν_{jk} is collision frequency between the j^{th} and k^{th} particles; m is mass; v is velocity; g is gravitational acceleration; B is the earth's magnetic field; and J is current density defined as

$$\vec{J}=Ne(\vec{v}_i-\vec{v}_e) \quad (3)$$

where e is the absolute value of electron charge.

If now, we assume that ν_{en} and ν_{in} are sufficiently small to allow neglect of the drag forces, then

$$-\nabla p_e-\nabla p_i+N\,m_i\vec{g}+\vec{J}X\vec{B}=0 \quad (4)$$

Assuming the electrons and ions behave as ideal gases and taking the field line component of (4), we obtain

$$[-Nk\nabla(T_e+T_i)-(T_e+T_i)k\nabla N+Nm_i\vec{g}]\cdot\vec{h}=0 \quad (5)$$

where T_j is the temperature of the j^{th} type particle, k is Boltzmann's constant, and $\vec{h}$ is a unit vector in the direction of the earth's magnetic field.

Let us define

$$\tau=\frac{T_e+T_i}{2} \quad (6)$$

$$H_\tau=\frac{k\tau}{m_ig} \quad (7)$$

Then

$$\left(\frac{\nabla\tau}{\tau}+\frac{\nabla N}{N}+\frac{\vec{i}_r}{2H_\tau}\right)\cdot\vec{h}=0 \quad (8)$$

where $\vec{i}_r$ is a unit vector in the radial direction. If we assume that the earth's magnetic field is a dipole lying in the r, θ plane, we obtain

$$\frac{1}{\tau}\left(\frac{\partial\tau}{\partial r}\sin I+\frac{1}{r}\frac{\partial\tau}{\partial\theta}\cos I\right)+\frac{1}{N}\left(\frac{\partial N}{\partial r}\sin I+\frac{1}{r}\frac{\partial N}{\partial\theta}\cos I\right)+\frac{\sin I}{2H_\tau}=0 \quad (9)$$

where I is the magnetic dip angle defined by

$$\tan I = 2\cot\theta \tag{10}$$

and θ is geomagnetic colatitude. (In the following discussion, all coordinates given are geomagnetic.) As pointed out in CG, (9) can be written in total derivative form as

$$\frac{d\tau}{\tau}+\frac{dN}{N}+\frac{dr}{2H_\tau}=0 \tag{11}$$

provided the integration is carried out along a field line. The solution of (11) is

$$N(r,\theta)=N(r_0)\frac{\tau(r_0)}{\tau(r,\theta)}e^{-\int_{r_0}^{r}\frac{dr}{2H_\tau}}$$

where r_0 is the equatorial height of the field line of integration, i.e.

$$r_0 = r\csc^2\theta \tag{13}$$

and $N(r_0)$, $\tau(r_0)$ represent the vertical distributions of N and τ at the equator.

THE BOUNDARY CONDITIONS

From (12), it is clear that the complete temperature distribution, given by $\tau(r,\theta)$, must be known in addition to the radial distribution of N at the equator, $N(r_0)$, before $N(r,\theta)$ can be determined uniquely. In previous work, $\tau(r,\theta)$ has simply been specified as a constant everywhere. We now investigate the properties of the theoretical electron density distribution when we include a simple radial model for τ based on measurement and theory.

Recent rocket probe measurements (Spencer, et al, 1962 and Brace, et al, 1963) at Wallops Island, Virginia, show a vertical electron temperature profile under quiet day conditions which departs from the ion temperature at 150 km, peaks between 200 and 250 km with a value of $T_e \approx 2T_i$, and then returns to the $T_i = T_e$ value at about 350 km. Evans [1962] obtains a similar type behavior and later [1964] shows improved results which indicate a peak at 300 km and a return to $T_e/T_i = 1$ above 700 km for radar backscatter measurements in the midlatitude region of Boston, Mass. The only measurements available in the equatorial region are those of Bowles [1964] using the backscatter technique and he reports a T_e/T_i peak value of 2 occurring at about 275 km. Furthermore, he never finds $T_e/T_i > 1$ above 400 km.

Hanson and Johnson [1961], Hanson [1962] and Dalgarno et al [1963] have presented theoretical models for T_e/T_i based on local EUV heating which result in similar type profiles to those measured, viz. $T_e/T_i = 2$ at the peak with the peak occurring between 200 and 250 km. The theoretical models are probably most representative of the equatorial regions where EUV heating is most likely to predominate. The theoretical peak height is somewhat lower than that reported by Bowles [1964] but Dalgarno et al [1963] explain that their lower value for the peak can be attributed to an underestimate of the cooling rate by neglect of the contribution from vibrational excitations. They also state that one should expect a more rapid cooling rate in a warmer atmosphere and this appears to indicate a higher altitude for the T_e/T_i peak at times of higher sunspot number. These are important considerations in selecting the numerical values for the height of the electron temperature peak (r_p) in the next section.

The results discussed above clearly demonstrate the absence of thermal equilibrium in the lower F region of the ionosphere. Unfortunately, the data and theoretical models available are rather limited and the exact behavior and distributional shape is currently rather uncertain. As a reasonable first guess, we therefore choose a simple analytic distribution for T_e which is representative of any of the above profiles but which does not fit any of them in an exact sense. We also consider the temperature behavior to be independent of latitude in the equatorial region under consideration (i.e. from 20°N to 20°S).

An expression which possesses these qualifications is

$$T_e = T_i\left[1+\frac{K-1}{1+\frac{(r-r_p)^2}{H_T^2}}\right] \tag{14}$$

where K is the ratio T_e/T_i at the peak height (r_p) of the T_e distribution and H_T is a scaling factor which governs the thickness of the T_e distribution and which will be referred to as the thermal scale height. A typical plot of (14) is shown in Figure

1 for various values of H_T and for $r_p = 6650$ km. From (14), we obtain

$$\tau(r, \theta) = \tau(r) = \frac{T_i}{2}\left[2 + \frac{K-1}{1+\frac{(r-r_p)^2}{H_T^2}}\right] \quad (15)$$

Although the effect of the gravitational height dependence is small, we include it for completeness in the work which follows. The scale height of the ionizable constituent is given by

$$H_i = \frac{kT_i}{m_i g} \quad (16)$$

If we define H_p as that value of H_i at the height r_p, i.e. the matching point, and we treat T_i as constant in the region under consideration, then

$$H_i = \frac{r^2}{r_p^2} H_p \quad (17)$$

With these definitions, we have

$$H_\tau = \frac{H_p}{2}\frac{r^2}{r_p^2}\left[2 + \frac{K-1}{1+\frac{(r-r_p)^2}{H_T^2}}\right] \quad (18)$$

The form of H_τ given by (18) allows explicit integration of the integral in (12). We obtain

$$-\int_{r_0}^{r} \frac{dr}{2H_\tau} = \frac{r_p^2[F_1+F_2+F_3+F_4]}{2H_p\left[r_p^2 + \frac{H_T^2(K+1)}{2}\right]^2} \quad (19a)$$

where

$$F_1 = \left[r_p^2 - H_T^2\frac{(K+1)}{2}\right]\left[\frac{H_T(K-1)}{2\left(\frac{K+1}{2}\right)^{\frac{1}{2}}}\right]\left[\tan^{-1}\left(\frac{r-r_p}{H_T\left(\frac{K+1}{2}\right)^{\frac{1}{2}}}\right) - \tan^{-1}\left(\frac{r_0-r_p}{H_T\left(\frac{K+1}{2}\right)^{\frac{1}{2}}}\right)\right] \quad (19b)$$

$$F_2 = r_p H_T^2\left(\frac{K-1}{2}\right)\ln\left[\frac{(r_0-r_p)^2 + H_T^2\left(\frac{K+1}{2}\right)}{(r-r_p)^2 + H_T^2\left(\frac{K+1}{2}\right)}\right] \quad (19c)$$

$$F_3 = 2r_p\left[\frac{K-1}{2}H_T^2 \ln r/r_0\right] \quad (19d)$$

and

$$F_4 = \frac{r_0 - r}{r\, r_0}\left[\left(r_p^2 + \frac{H_T^2(K+1)}{2}\right)(H_T^2 + r_p^2)\right] \quad (19e)$$

Finally, we must apply an analytic form for $N(r_0)$. Although Chandra (1962) formulated an improved functional form for $N(r)$ than that described by the simple Chapman function and this has been incorporated in the results of CG, it is necessary to return to the simple Chapman function in the work that follows in order to make direct comparison with Goldberg, Kendall and Schmerling (1964). Hence, we let

$$N(r_0) = N(r_{m0})e^{\frac{1}{2}\left[1 - \frac{r_0 - r_{m0}}{H_e} - e^{-\frac{r_0 - r_{m0}}{H_e}}\right]} \quad (20)$$

where H_e is the electron density scale height at the equator and N_{m0} is the value of electron density at the vertical peak height r_{m0}.

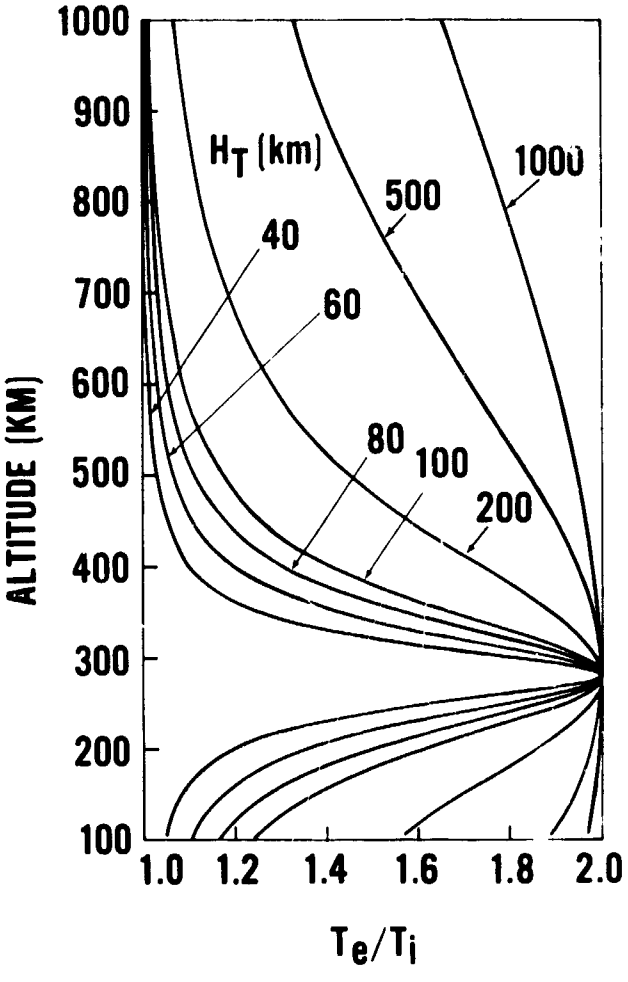


FIGURE 1.—The vertical electron temperature model.

NUMERICAL ANALYSIS AND DISCUSSION

A. The Effect of Gravity

The expressions given by (15), (19) and (20) have been incorporated in (12) to provide an analytic expression for the electron density near the geomagnetic equator above and below the electron density peak under conditions of steady state and equinox. This expression has then been programmed on an IBM 7094 to allow extensive study of its behavior under wide variations in the temperature parameters. The discussion which follows demonstrates how adjustment of the parameters governing the temperature distribution leads to variations in the properties of the geomagnetic anomaly.

Before investigating the actual effect of a variable T_e, let us first study the change induced by simply adding the gravity variation. This is obtained by replacing H_r with H_i and treating $\tau(r,\theta)$ as a constant in (12). Then

$$N(r,\theta)=N(r_0)e^{\frac{r\cos^2\theta}{2H_i}} \tag{21}$$

where H_i possesses the functional dependence on height given by (17). This can be compared to the constant H_i result employed in GKS, viz.

$$N(r,\theta)=N(r_0)e^{\frac{r\cot^2\theta}{2H_p}} \tag{22}$$

where H_2 of GKS is now H_p. (In the discussion and results which follow, the scale height of the electron density distribution H_e has also replaced $1/k$ in GKS notation).

Comparison of (21) and (22) shows that very little change in results should be expected in that region of latitude where $\cot\theta$ and $\cos\theta$ are comparable in magnitude. This encompasses nearly the entire region of our interest. This conclusion is borne out by numerical comparison of (21) and (22) for cases in which $H_p/H_e>1$. A more important result is obtained for cases presented in GKS for $H_p/H_e\leq 1$. In that paper it was shown that angular peaks could not be obtained for this range of selection in parameters, although the initial rise with latitude behaved in equivalent manner to the geomagnetic anomaly. The simple inclusion of gravity now rectifies this situation by allowing angular peaks to form for these cases.

A comparison of the 380 km and 480 km electron density profiles with and without gravity variation for $H_p/H_e=1$ is shown in Figure 2 using the numerical parameters given in Table 1 (The altitude z is related to radial distance r by $r=z+6370$ km). Curve a represents the constant g case given by (22). Curves b and c represent variable $g(r)$ cases in which the matching of H_i with H_p is taken to be at two different heights, viz. 6650 and 6850 km, respectively. We first note that the effect of varying the matching point is rather small and has little if any influence on the features we wish to discuss. We also find that regardless of matching point height, the curves stay relatively close even at high latitudes. The generation of an angular peak for $r_p=6650$ km is not surprising since $H_i/H_p>1$ everywhere above r_p. However, it is surprising to find this result for the 380 km (6750 km) profile when $r_p=6850$, since $H_i/H_e<1$ at this height. This result occurs for all cases investigated and we must conclude that although gravity is a small effect in altering the slope of the curves, as comparison of curves 2b and 2c demonstrate, it is a strong effect in providing a more realistic description of the geomagnetic anomaly should it occur under conditions where $H_p/H_e\leq 1$.

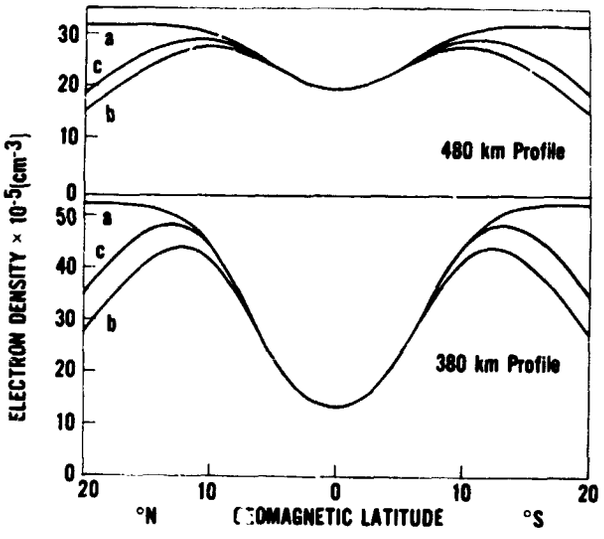


FIGURE 2.—The influence of the gravity variation on a horizontal electron density profile for $H_p/H_e=1$; a) constant g, b) and c) variable g.

B. The Effect of Temperature

Although inclusion of height dependent gravity, $g(r)$, appears to resolve the problem of theoretically describing the geomagnetic anomaly

when $H_i/H_e \leq 1$, it has been shown in CG that this condition is a rather unlikely situation to occur on both empirical and physical grounds. We therefore restrict ourselves in the discussion which follows to $H_i/H_e > 1$.

There are seven parameters to be varied in this problem on the basis of empirical conditions, viz. N_{m0}, r_{m0}, H_e, H_p, H_T, r_p and K. In GKS, it was shown that variation of the first four parameters leads to a description of the equinoctial noontime geomagnetic anomaly during various phases of the solar cycle. Furthermore, although differences in the magnitude and height range of the anomaly occur between high and low sunspot number, no changes in the basic features ascribed to this effect are expected. We will, therefore, limit ourselves to an analysis of the high sunspot case with the knowledge that the results obtained are similar but less pronounced for the intermediate and low sunspot cases. The choice of values for the first four parameters listed above have been made to coincide with GKS by selecting

$$N_{m0} = 19.25 \times 10^5 \text{ electrons/cm}^3,$$

$$H_e = 100 \text{ km}, \; r_{m0} = 6850 \text{ km},$$

and $$H_p = 112.5 \text{ km}.$$

Selection of H_p at the given value insures $H_i/H_e > 1$ in the entire region of interest. The values of the other numerical parameters used are given in Table 1. (These values are selected to coincide as accurately as possible with the temperature profile discussed earlier).

TABLE 1.

Figure	Curve	r_p (km)	H_T (km)	K
1	----------	6650	Variable	2
2	a	----------	0	1
	b	6650	0	1
	c	6850	0	1
3	----------	6650	40	2
4	----------	6650	40	2
5	----------	6650	40	2
6A	----------	Variable	40	2
6B	----------	6850	40	Variable
6C	----------	6850	Variable	2
8	----------	6850	40	2
9	----------	6850	40	2
10	----------	6775	40	2

We can therefore reduce the problem to a three parameter study; H_T and K which determine the thickness and magnitude of the T_e peak respectively; and r_p, which determines the relative distance between the electron temperature and density maxima.

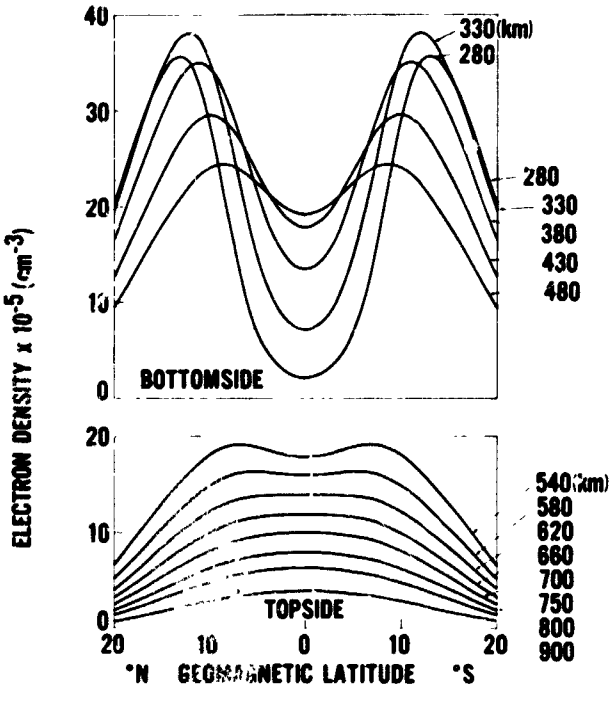


FIGURE 3.—Horizontal profiles of electron density under conditions of variable electron temperature.

Figure 3 represents a sequence of electron density constant height profiles under the effects of T_e variation. Figure 4 represents an equivalent set of vertical profiles. The temperature profile used is illustrated in Figure 1 with parametric values selected for reasons discussed in a previous section.

Comparison of Figure 3 to the thermal equilibrium, constant gravity results given in GKS shows very little modification of the curves out to the angular peak. On the high latitude side of the peak, however, we find a more rapid decrease of electron density in better agreement with measured profiles. This effect is due primarily to the inclusion of the gravitational variation and not the variation in T_e.

A more important result, and one which is entirely due to a variable T_e, is evident in Figure

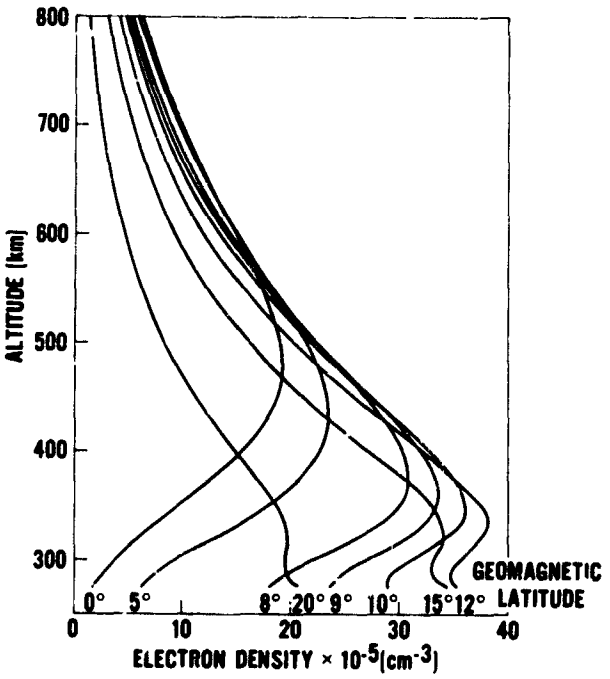


FIGURE 4.—Vertical profiles of electron density under conditions of variable electron temperature.

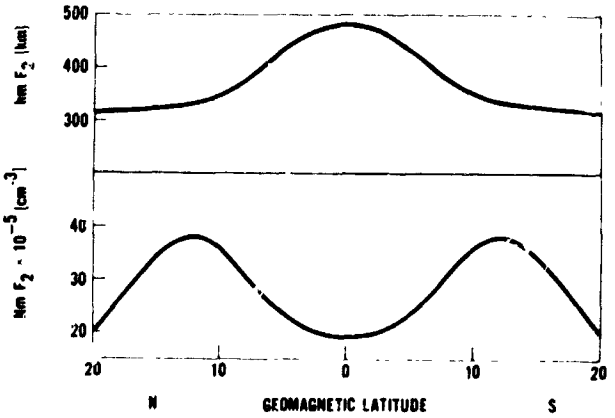


FIGURE 5.—Behavior of N_mF2 and h_mF2 with latitude under conditions of variable electron temperature.

4. We now find qualitative results which match measured data to a height nearly two electron scale heights (H_e) below the equatorial $F2$ peak before discrepancies begin to occur. Since the vertical profiles show a departure from measurement below this height, it appears that the neglect of friction terms in the equations of motion is no longer valid below this region. Nevertheless, it is a valuable extension of the theory to find that the effect of collisions can be neglected to heights well below the equatorial $F2$ peak, especially since it appeared by the results of GKS that collisions were important everywhere below this peak, i.e. in GKS results obtained below the peak do not agree with data and the vertical peaks shown in Figure 4 of this work could not be produced.

The variation of the peak electron density N_mF2 with latitude is given in Figure 5. We observe that a variable T_e produces angular peaks in N_mF2 in accord with measurement, (cf. Croom et al, 1959) whereas in GKS, it was shown that thermal equilibrium is incapable of producing the angular "turnover." Furthermore, this result must be attributed to a variable T_e, since inclusion of variable gravity alone does not alter the results for this parameter given in GKS. The variation of the height of the $F2$ peak, h_mF2, is also given in Figure 5. We observe a leveling out of this height at higher latitudes in accord with measurement (cf. Thomas, 1962), a result which also is not available under thermal equilibrium considerations. A study of these quantities show that their magnitude and shape are very stable to wide variations in K, H_T and $r_{m0}-r_p$, except when K and H_T become very small, i.e. $H_T \leq 29$, $K \leq 1.2$. At these and smaller values we find a rapid transition into the forms published in GKS for thermal equilibrium, in which h_mF2 drops steadily with latitude and N_mF2 rises with latitude without showing any angular peaks.

Figure 6A demonstrates the variation of a typical constant height profile with r_{m0}, viz. 380 km, holding K and H_T fixed. We note that as r_p approaches r_{m0}, the original angular peak converts into a sharper peak adjacent to a relatively flat "ledge." Further study has shown that this effect is not dependent on the absolute value of r_{m0} or the relative distance between the actual constant height profile and r_{m0}. Instead, this behavior is exclusively dependent on $r_{m0}-r_p$.

Because the results which follow depend only on the relative separation $r_{m0}-r_p$ and not on the absolute values of r_p and r_{m0}, and because of the discussion in an earlier section explaining how r_p can actually be larger in a warmer (high sunspot number) ionosphere, we have selected the relatively high value of $r_p \approx r_{m0} = 6850$ km. If the condition $r_p = r_{m0}$ occurs at a lower height, the same results will occur simply shifted this distance in altitude. From Figure 6A we find that the new effect is not present until r_p is very nearly

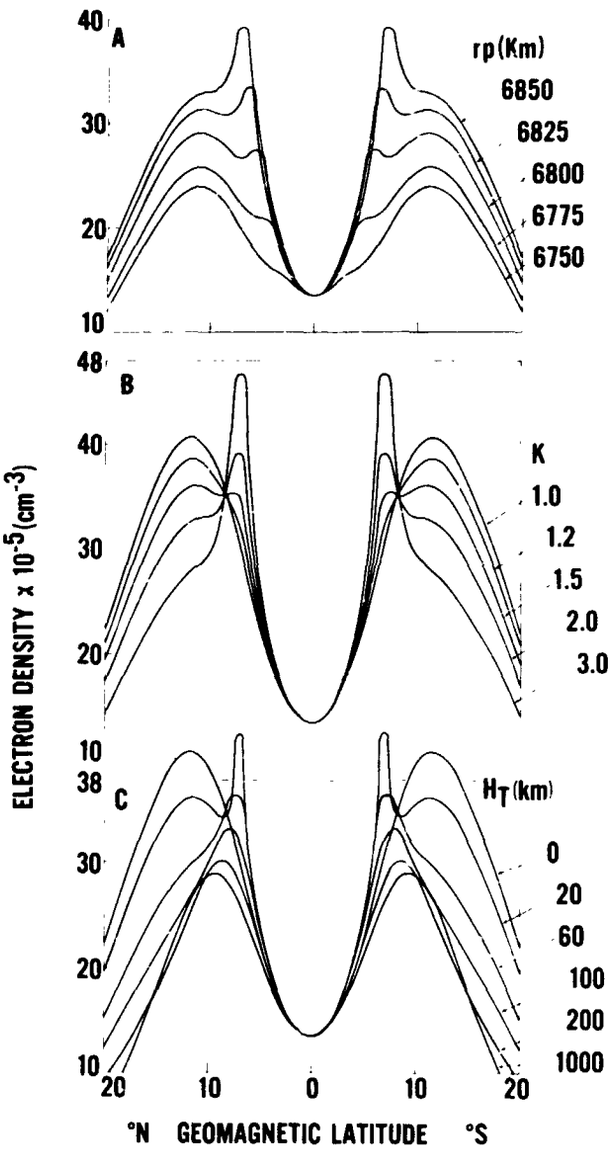


FIGURE 6.—Behavior of a constant height electron density profile (6750 km) with *A*) relative distance between electron temperature and density peaks, *B*) magnitude of temperature peak, *C*) thickness of temperature peak.

equal to r_{m0}, and becomes more pronounced as r_p increases (or r_{m0} decreases).

Although the above effect is not always present, it has appeared in measurements of King et al (1963), an example of which is shown in Figure 7. No attempt has been made to accurately fit this particular profile, however, since there are many combinations of the seven parameters available for such a fit and the profiles showing this property best in the published literature do not always represent noon equinox conditions. Nev-

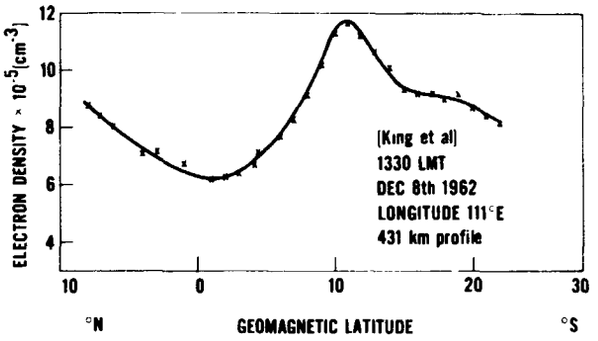


FIGURE 7.—The 431 km electron density profile (King et al, 1963) for December 8, 1962, 1330 LMT, using ALOUETTE data.

ertheless, the effect does occur, and this indicates that the electron temperature and density peaks do lie relatively close together at certain times. During such an occurrence, we would expect the geomagnetic anomaly to appear in a form similar to the result given in Figure 8, where we notice a

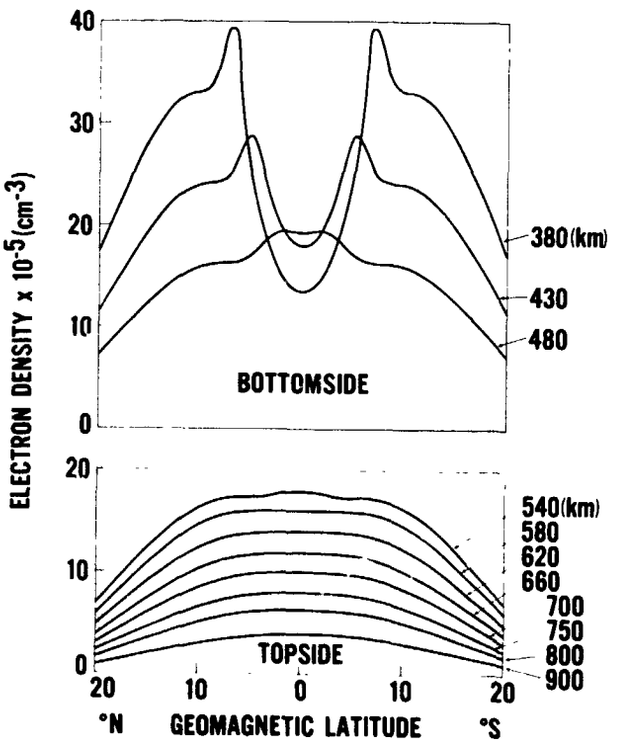


FIGURE 8.—Horizontal profiles of electron density under conditions of variable electron temperature for $r_p = r_{m0}$.

low latitude "ledge" replacing the low latitude "trough" at higher altitudes. We also note a reduction in the height above which the anomaly disappears.

The sequence of vertical profiles corresponding to Figure 8 ($r_p = r_{m0}$) are shown in Figure 9. We

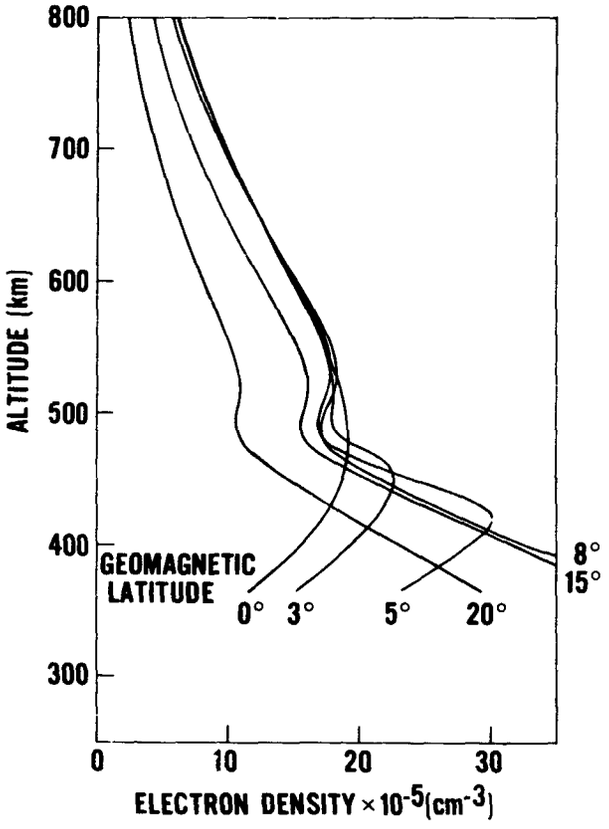


FIGURE 9.—Vertical profiles of electron density under conditions of variable electron temperature for $r_p = r_{m0}$.

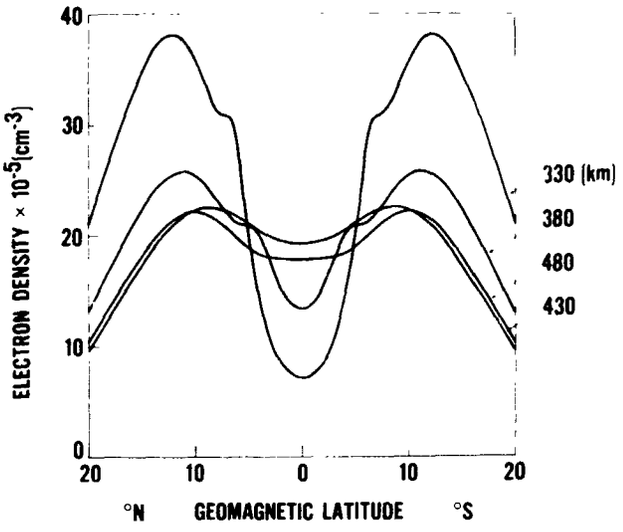


FIGURE 10.—Horizontal profiles of electron density showing the "inner ledge" effect.

find that this extreme lowering of the electron density peak (or raising of the temperature peak) leads to an additional small "bump" in the topside region.

Other smaller effects are also seen upon more detailed study of the results. For example, if $r_{m0} - r_p$ is of the order $.75H_e$ (see Figure 6A), we find the sequence of horizontal profiles illustrated in Figure 10, demonstrating a much smaller ledge occurring on the low latitude side of the angular peaks with no ledge present on the high latitude side. The topside results in this case are identical to those of Figure 3. The determination of this behavior from measurement is difficult, however, because of the small magnitude of the effect.

As partially shown in Figures 6B and 6C, a study of the variations of both K (temperature peak magnitude) and H_T (temperature peak thickness) do not lead to any new conclusions but form a consistent picture, i.e. as K decreases, the ledge gradually disappears leaving the normal horizontal profile for the variable gravity case. As H_T decreases, the effect first sharpens before disappearing for $H_T \leq 15$ km. Similarly, as K and H_T decrease, the topside "bump" in the vertical profile (cf. Figure 9) gradually lowers and blends with the single peak obtained under variable gravity conditions alone. Increasing H_T and K improves the results on the bottomside by reducing the magnitude of the electron density profiles to more reasonable values in this region. Increasing H_T also sharpens the angular peak slightly. Finally, if we consider cases of very large H_T, such as $H_T = 1000$ km, we find a behavior very similar to the $H_T = 0$ situation. This demonstrates that it is the gradient of T_e/T_i, and not its magnitude, which is mostly responsible for the results involving the effect illustrated in Figure 7.

SUMMARY AND CONCLUSIONS

The theory presented in GKS has been extended to include the effects of the variation in gravity and an electron temperature profile based upon empirical and theoretical results. This extension in the theory has led to distinct improvements in the theoretical description of the geomagnetic anomaly when comparison with data is made.

The inclusion of gravitational variation alone has led to an increased gradient on the high latitude side of the theoretical constant height profiles representing the geomagnetic anomaly, this being in better accord with measurement. Furthermore, for the cases in which $H_i/H_e \leq 1$, we now

find angular peaks occurring in the constant height profiles, something which was not available under constant gravity considerations.

The inclusion of an electron temperature vertical profile creates results which are far more remarkable, however. We now obtain a description of the geomagnetic anomaly to heights as low as $2H_e$ below the F2 peak at the equator, thereby indicating that the neglect of friction or collision terms in the equation of motion is allowable down to these heights. The theoretical behavior of N_mF2 and h_mF2 are also found to agree with measurement out to midlatitudes and an angular peaking of N_mF2 is seen to occur. These are results which could not be obtained in GKS under isothermal conditions and thermal equilibrium.

Although T_i has been treated as a constant in the derivation of the expression used to obtain these results, this is not considered to be a serious limitation in the theory since the major gradients in T_i currently appear to be considerably smaller than those in T_e and occur in the very lowest sector of the region under consideration.

Further study with a variable T_e shows that as the height of the electron temperature and density peaks approach one another, we can expect to find an interesting change in the shape of the horizontal electron profiles describing the geomagnetic anomaly. Experimental evidence is presented to show that this behavior does occur at certain times. We also find that the slope of the electron temperature peak rather than its magnitude is more responsible for this effect. In addition, the vertical profiles undergo a slight modification with the appearance of a small "bump" in the topside above h_mF2 during its occurrence. It should be noted, however, that the evidence given in Figure 7 is a topside result and the theoretical results of Figure 8 do not show this effect extending into the topside. Assuming that 431 km was slightly above r_{m0} in Figure 7, we must conclude that either r_p was slightly higher than r_{m0} or that $r_p = r_{m0}$ with the temperature peak exhibiting a sharp gradient (large K, small H_T) on that day, since these are the only possible methods for theoretically obtaining this effect in the topside with the temperature profile assumed. (Naturally, a different electron temperature profile, such as one which possesses a secondary peak in the topside, should not be ruled out as a possible cause of observing this effect in the topside). Since r_{m0} decreases rapidly with sunspot number, we would expect the above behavior to be most frequent during the low sunspot number period of the solar cycle.

Finally, several secondary features originating from the theory due to the magnitude of the separation between N and T_e peaks ($r_{m0} - r_p$) are discussed, but this type of "fine structure" is considered too small to be seen at this time.

ACKNOWLEDGEMENTS

I am grateful for the valuable assistance given to me in the programming and numerical analysis of this work by Christopher E. Bock and William M. Litman. I also offer my deepest thanks to Dr. S. J. Bauer for critically reading the manuscript and making several helpful suggestions.

REFERENCES

Bowles, K. L., Private Communication, 1964.

Brace, L. H., N. W. Spencer and G. R. Carignan, *Ionosphere Electron Temperature Measurements and their Implications*, J. Geophys. Res. **68**, 5397–5412, 1963.

Chandra, S., *Electron Density Distribution in the Upper F Region*, J. Geophys. Res., 68, 1937–1942, 1963.

Chandra, S. and R. A. Goldberg, *Geomagnetic Control of Diffusion in the Upper Atmosphere*, J. Geophys. Res. 69, 3187–3197, 1964.

Croom, S. A. Robbins, and J. O. Thomas, *Two Anomalies in the Behavior of the F2 Layer of the Ionosphere*, Nature **184**, 2003, 1959.

Dalgarno, A., M. F. McElroy and R. J. Moffett, *Electron Temperatures in the Ionosphere*, Planet. Space Sci., **11**, 463–484, 1963.

Evans, J. V., *Diurnal Variation of the Temperature of the F Region*, J. Geophys. Res., **67**, 4914–4920, 1962.

Evans, J. V., *Ionospheric Temperatures during the Launch of NASA Rocket* **8.14** *on July* **2, 1963**, J. Geophys. Res., **69**, 1436–1444, 1964.

Goldberg, R. A. and E. R. Schmerling, *The Distribution of Electrons Near the Magnetic Equator*, J. Geophys. Res., **67**, 3813–3815, 1962.

Goldberg, R. A. and E. R. Schmerling, *The Effect of Diffusion on the Equilibrium Electron Density Distribution in the F Region near the Geomagnetic Equator*, J. Geophys. Res., **68**, 1927–1936, 1963.

Goldberg, R. A., P. C. Kendall and E. R. Schmerling, *Geomagnetic Control of the Electron Density in F the Region of the Ionosphere*, J Geophys. Res. **69**, 417–427, 1964.

Hanson, W. B., *Electron Temperatures in the Upper Atmosphere*, Space Research, Proc. Intern. Space Sci. Symp., 3rd, Washington, 1962, edited by W. Priester

pp. 282–302, North-Holland Publishing Company, Amsterdam, 1962.

HANSON, W. B. and F. S. JOHNSON, *Electron Temperatures in the Ionosphere*, paper presented at Tenth International Astrophysical Colloquium, Liege, Belgium, 1961.

KING, J. W., D. ECCLES, P. A. SMITH, P. DANNAHY, A. LEGG, E. O. OLATUNJI, K. RICE, G. WEBB amd M. WILLIAMS, *Further Studies of the Topside Ionosphere Based on the Topside Sounder Satellite Data*, DSIR Radio Research Station Report No. RRS/I.M. 112 December 1963.

SPENDER, N. W., L. H. BRACE and G. R. CARIGNAN, *Electron Temperature Evidence for Nonthermal Equilibrium in the Ionosphere*, J. Geophys. Res. **67**, 157–175, 1962.

THOMAS, J. O., "The Electron Density Distribution in the F Region of the Ionosphere," in *Electron Density Profiles* edited by B. Maehlum, Pergamon Press, 1962.

A SYNOPTIC VIEW OF IONIC CONSTITUTION ABOVE THE F-LAYER MAXIMUM*

N66 37959

W. J. RAITT, SUSAN LAFLIN AND R. L. F. BOYD†

The UK-US satellite Ariel carried an ion mass spectrometer operating on the principle of analysis of arrival energy of the ions. A very large amount of data has been gathered and it is now possible to obtain a picture of the variation of the ionic composition above 400 km with latitude and longitude.

A strong diurnal effect is evident together with marked geomagnetic control. There is evidence also for a seasonal variation.

INTRODUCTION

The international satellite Ariel I (1962 –) provided an opportunity for a synoptic study of the composition of the topside ionosphere over a prolonged period.[1] Data on ionic constitution was received at stations of the Minitrack network from 26 April, 1962 until 26 September, 1962—a period of 151 days during summer in the northern hemisphere. The relative concentration of oxygen in the topside ionosphere is found to increase with latitude during the day. A marked diurnal variation in composition is also evident. This paper presents some of these data and draws special attention to the evidence for a geomagnetic control of the composition from a set of satellite passes with apogee over the equator and another set with apogee at northern apex.

The apparatus consisted of a gridded probe 10 cm. in diameter. The probe was mounted on the axis of the satellite and being spherically symmetrical was insensitive to the spin of the vehicle.

An ion retarding (positive sawtooth) voltage enabled a Dryvesteyn analysis of the energy distribution of the ions arriving to be obtained, the necessary double differentiation of the current-voltage characteristic being carried out electronically.

The energy of arrival of an ion depends principally on its mass so that the energy spectrum consists of a number of peaks corresponding to ions of mass 16 (oxygen), 4 (helium) or 1 (hydrogen) a.m.u. The integration of these peaks leads to the concentration of each type of ion. Since the first derivative of the current voltage curve which corresponds to this integral of the energy distribution was one of the telemetered quantities, typical raw data of the kind shown in figure 1 was obtained. The sizes of the steps in these

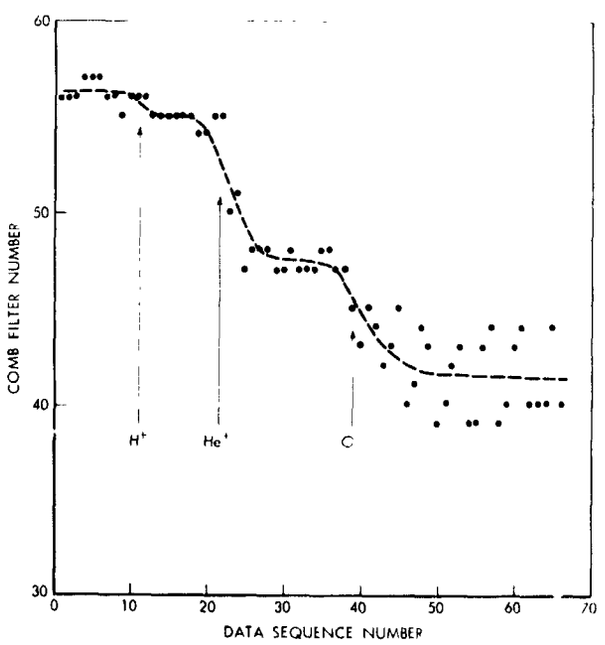


FIGURE 1.—Example of raw telemetered data showing presence of 3 ion types.

*This paper was contributed to *Goddard Space Flight Center* under the joint United Kingdom–United States program which developed and launched the satellite Ariel I.

†The authors are affiliated with the Department of Physics, University College, London, England.

curves lead directly to the concentrations of the various ion species.

THE DAILY VARIATION OF IONIC CONSTITUTION

A number of satellite passes over the equatorial region 10°S. to 10°N. and over the latitude band 35°N. to 45°S. covering a period of 125 days in the northern summer have been analysed, to obtain the diurnal variation in the ion transition altitudes (i.e. where O^+ and He^+ concentrations are equal and where He^+ and H^+ concentrations are equal). Inevitably some seasonal and secular variations are also involved. To reduce the former the altitudes are plotted as a function of solar zenith angle before and after mid-day. In addition, since the basic data gives composition at the actual altitude and local solar time given by the orbital parameters some processing of the data is necessary in order to obtain a single parameter (such as transition altitude) to characterise the state of ionic composition. For this purpose it has been assumed that ion temperature is equal to gas temperature and that the composition at one height may be obtained from that at another height from the hydrostatic equation. The gas temperature function employed was based on the data of Harris and Priester[2] corrected to the mean solar 10 cm. emission for the period covered by the data.

It is certain that this operation introduces some error in scale but it cannot seriously affect the general form of the results.

The diurnal variation of transition altitudes is plotted for the two latitude ranges in figure 2. (The latitude variation will be considered in more detail in the next section). It shows a flat maximum at solar zenith angles corresponding to about 17.00 hours.

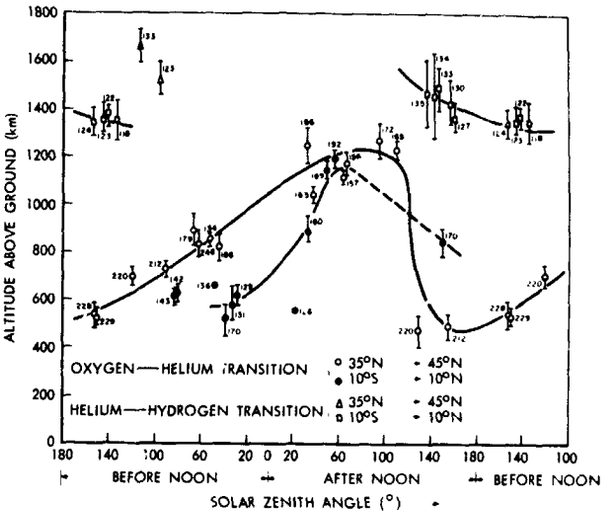


FIGURE 2.—Diurnal variation of transition altitudes.

LATITUDE VARIATION OF IONIC CONSTITUTION

In seeking to establish the variation of composition with latitude it is difficult to eliminate the effects of changing altitude and local solar time. By selecting a short period (four days) during which apogee was nearly over the equator and the descending equatorial passage occurred at about 1500 hours local solar time it is possible to study the region between 45°N. and 45°S. during the period 12.00 to 18.00 hours over which as the previous section shows, the diurnal variation is slowest. At the same time, the altitude range does not change by more than 200 Km. Figure 3 shows contours of percentage oxygen ion concentration betweeen days 189 to 193 in 1962.

There is evidence for a pronounced latitude variation in composition showing higher concentrations of oxygen ions at high latitudes. (The change of composition from 10 percent

$$O^+\left(\frac{[O^+]}{[He^+]}=\frac{1}{9}\right) \text{ to 80 percent } O^+\left(\frac{[O^+]}{[He^+]}=4\right)$$

exceeds the composition change that might be attributed to the 100 Km change in altitude by a factor of 10). The altitude and local solar time are plotted at the side of the figure.

In addition to the evident geomagnetic control of the composition there seems also to be a displacement of the centre of symmetry towards the winter (southern) hemisphere.

LONGITUDE VARIATION OF IONIC CONSTITUTION

A variation of composition with longitude is likely to be due to the asymmetry of the geomagnetic field. This fact is confirmed by the way the contours on figure 3 follow the geomagnetic equator.

Figure 4 shows similar data obtained over the period from day 150 to day 155 in 1962. In this

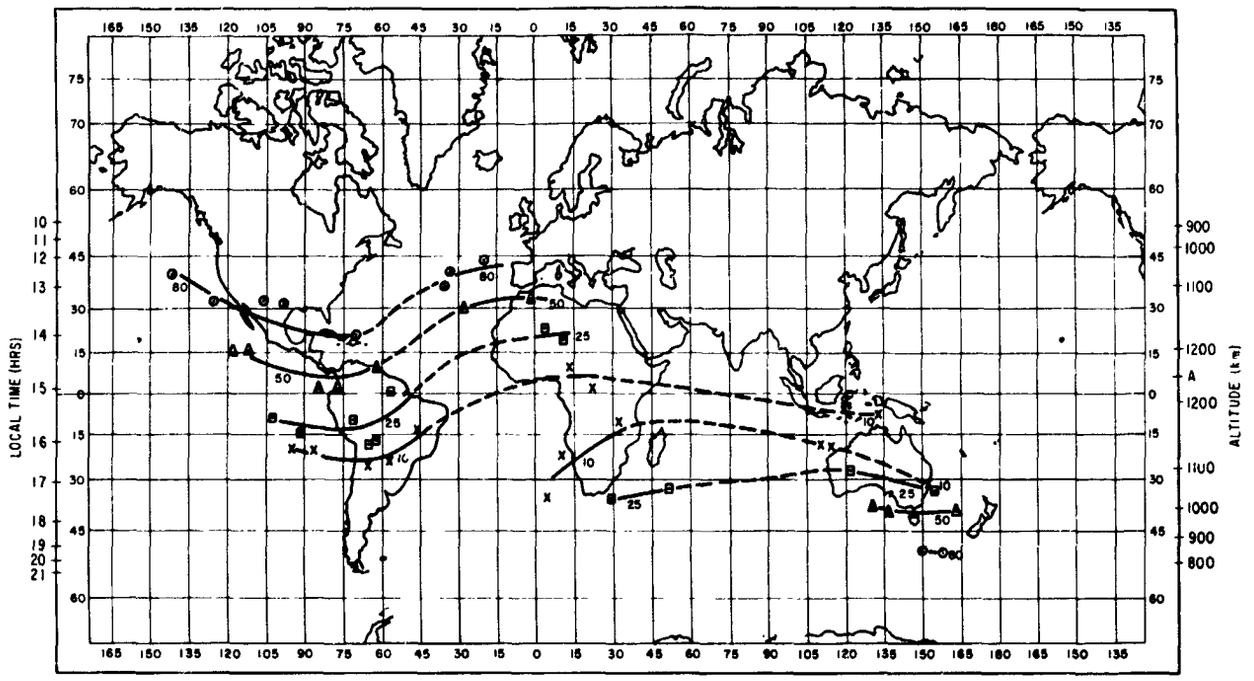


FIGURE 3.—Ion composition contours days 189–193, 1962.

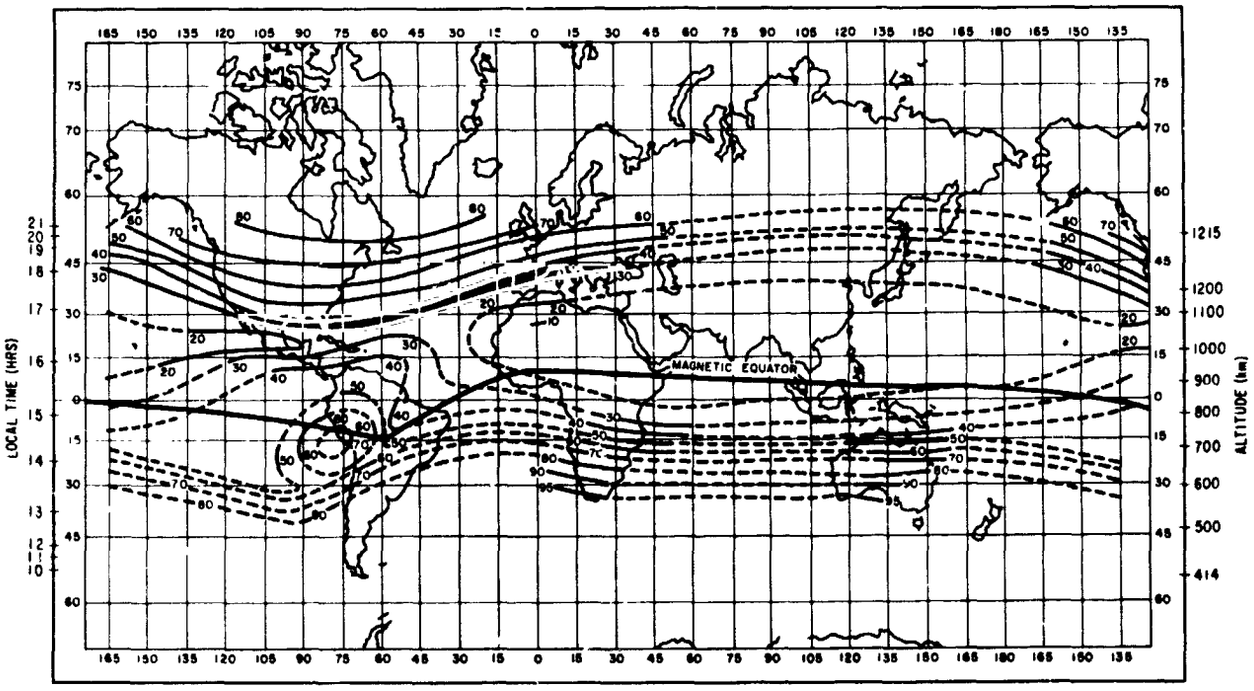


FIGURE 4.—Ion composition contours days 151–155, 1962.

case, the effects of altitude and local solar time on the latitude variation are too great to enable useful conclusions about latitude variation to be drawn. However, variation along any single line of latitude does not suffer from this source of uncertainty and it is again evident that there is a strong geomagnetic control with an extremely complex situation over the South American anomaly. This area of increased oxygen was not detected on the earlier results (days 189–193)

possibly because less data was analysed then and also the altitude and local time is this region were such that the percentage oxygen was only of the order of 10 percent which is near the limit of detection of the probe. On the other hand, there may be a real seasonal or secular difference.

It seems likely that many of the features may be attributed to the dependence of ion temperature on electron concentration[3] while near the equator the horizontal magnetic field (and possibly electric fields) probably make the hydrostatic equation inapplicable. It is worth noting that where the ionic composition ratio is determined by the hydrostatic equation it is the ion temperatures, not the electron temperatures that are relevant. Effects of the magnitude of those illustrated could be produced by an ion temperature change of about 20 percent.

REFERENCES

1. P. J. Bowen, R. L. F. Boyd, W. J. Raitt and A. P. Willmore, Proc. Roy. Soc., (1964) (In press).
2. I. Harris and W. Priester. NASA Tech. Note D-1444 (1962).
3. R. L. F. Boyd and W. J. Raitt. Space Research V.

N66 37960

MAGNETIC SHELL ENHANCEMENTS DURING MAGNETIC DISTURBANCES*

A. P. WILLMORE AND C. L. HENDERSON†

Langmuir probe observations of electron density on the satellite Ariel I for May–July 1962 show that even on magnetically quiet days longitudinal irregularities appear, particularly near the South Atlantic anomaly. The region of locally depressed density near the shell L=4 shows a diurnal variation of intensity, and in the southern hemisphere it inverts to an increase in density during the late morning. A minimum has been observed at the magnetic equator in the evening hours, about 500 km above the peak of the equatorial anomaly.

For disturbed periods enhancements have been observed with a structure resembling magnetic shells, which tend to merge and to approach lower L values as the disturbance increases. For one disturbed period strong enhancements were noted at night near the South Atlantic anomaly. These phenomena imply the influence of energetic particles on topside irregularities.

INTRODUCTION

Langmuir probe measurements of electron density recorded on the satellite Ariel I during the period May to July 1962 (Willmore et al.) have been used to study the quiet-day distribution, and variations due to magnetic disturbances. No major disturbance occurred in this period, but there were several isolated, moderate storms for which definite enhancements could be identified.

Table 1 lists the magnetic data for these disturbances. Unfortunately there were no probe data for most of the peak disturbance period on May 6 and 31. We have therefore presented results for the three other periods.

METHOD OF ANALYSIS

Since the observations were magnetically recorded on the satellite, data for complete orbits were available.

These were separated into northbound and southbound passes, each covering the full latitude range of 54°N. and S. From 14 successive passes a complete scan of latitude and longitude could be mapped. Because of the slow motion of perigee in the orbit, and the slow rotation of the orbital plane, a line of constant latitude on such a

TABLE 1.

Date	3-hourly Kp values	Kp sum
May 4	1−0o 1+0o	0o 0o 0+1o	3+
5	0o 0o 1−0+	0+1−1+3o	6+
6	2−2o 1+3+	4+5o 5−4+	27−
7	3+2+1+1+	2−1−1+2o	14o
8	2+2o 1o 0+	1o 1o 1−3−	11o
May 25	0o 0o 0o 0+	0+0+0+1o	2+
26	0+1−1−1−	1o 1−1+2+	8−
27*	2o 3o 3−2+	3o 4−2o 3−	21−
28	1−1o 1−2−	2o 2o 3−2+	13o
29	1+2o 2+2−	1+1o 1+1o	12o
30	0+1o 1o 0+	1o 0o 1−1o	5+
31	0+2o 3+5−	5o 4−3−4o	26−
June 1	3−2o 3o 3−	2o 1o 2o 2o	17+
2	1+1+1−2−	1−1−2−2−	10−
3	1−0o 0+1−	2−2−1+3−	9o
June 8	0o 1−1+1−	1−1−1o 1o	6o
9	3−4o 3+4−	4−3−4o 5−	29−
10	5−4−3+4o	3o 2−2o 1o	23+
11	1o 2o 2o 1+	2−1−1+2o	12o
12	2+1−2+2+	3−2o 2o 1o	15+
13	1+1+2+2−	1−1−1o 1−	10−
June 19	1−1o 1o 1−	2o 2−1+1o	9+
20	1+1o 1+1o	1−1−1o 1o	8o
21	1o 1o 0+3o	3o 4−4o 2+	18+
22	3− 4−3−2−	1+1+3−2o	18o

*s.c. at 0414 hrs.

*This paper was contributed to *Goddard Space Flight Center* under the joint United Kingdom-United States program which developed and launched the satellite Ariel I.

†Both authors are affiliated with the Department of Physics, University College, London, England.

map is also a line of constant height to within ± 10 km, and of constant local time to ± 10 minutes. This form of presentation was used to show the longitude and magnetic latitude variations of electron density on quiet days. (figs. 1–4).

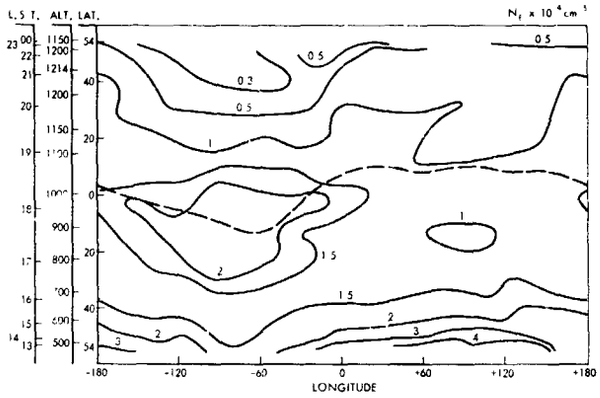


FIGURE 1.—Contours of electron density for northbound passes on May 24, 1962. The dashed line is the magnetic equator.

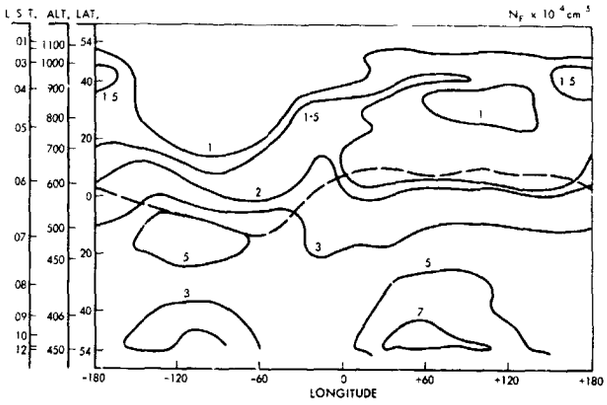


FIGURE 2.—Contours of electron density for southbound passes on May 24, 1962.

For determining disturbance effects, plots of electron density versus latitude for "corresponding" passes on successive days were superimposed. The days included, but were not confined to, those listed in Table 1. The passes were at intervals of 14 orbits, and for corresponding latitudes there was a daily shift of 2° longitude eastwards, a change of less than 20 km in height, and a decrease of 20 minutes in local time. Variations of electron density at a given latitude due to these orbital shifts are in general linear, and can therefore be distinguished from variations which correlate with Kp. A region was classed as showing a disturbance effect only when the electron density for a given latitude differed significantly on the disturbance day from that on the neighboring quieter days.

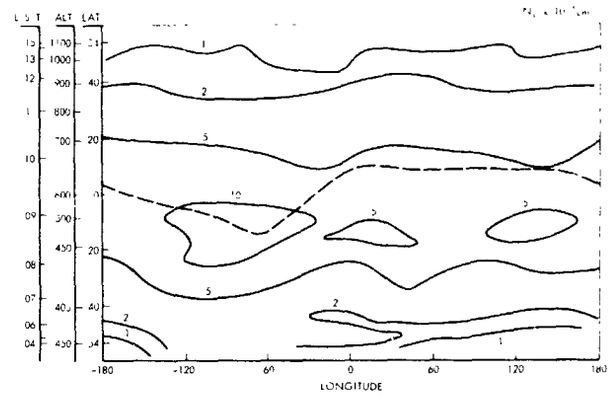


FIGURE 3.—Contours of electron density for northbound passes on June 20, 1962.

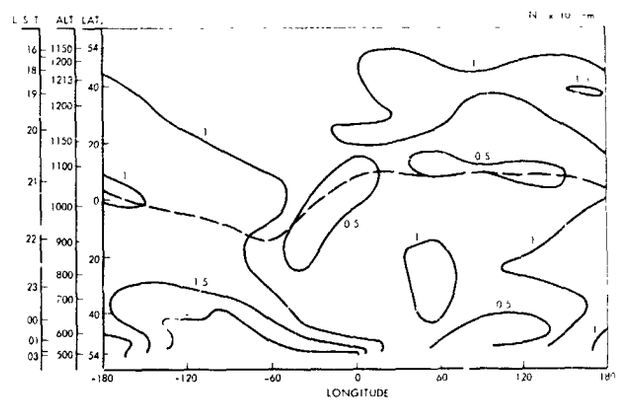


FIGURE 4.—Contours of electron density for southbound passes on June 20, 1962.

No decrease in electron density was unambiguously associated with magnetic activity. Nelms and Warren (this conference) observed a decrease over a wide range of latitudes when Kp reached very high values, up to 9–. However, this referred to altitudes near the maximum of the F layer and so was lower than our observations.

Several regions of increased electron density were observed, and are depicted in figs. 5–8. Bold lines indicated enhancements, fine lines no significant disturbance effect, and blank spaces an absence of data. In some places where there

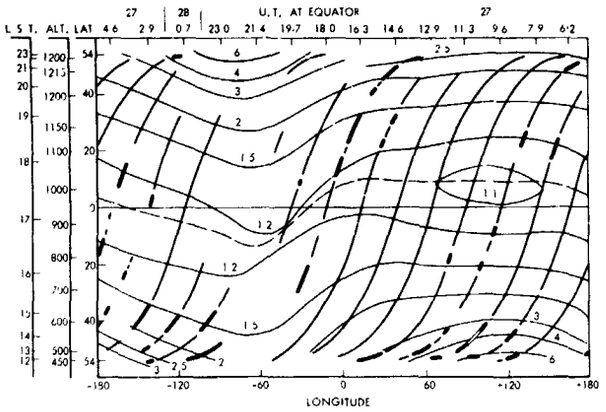


FIGURE 5.—Enhancements for May 27–28, northbound passes. The contours are intersections of L shells with the satellite orbit.

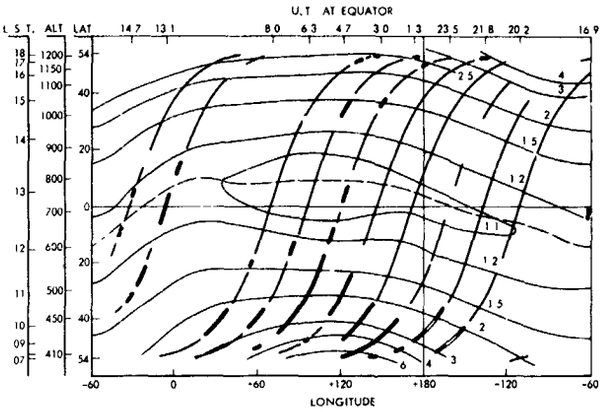


FIGURE 7.—Enhancements for June 9–10, northbound passes.

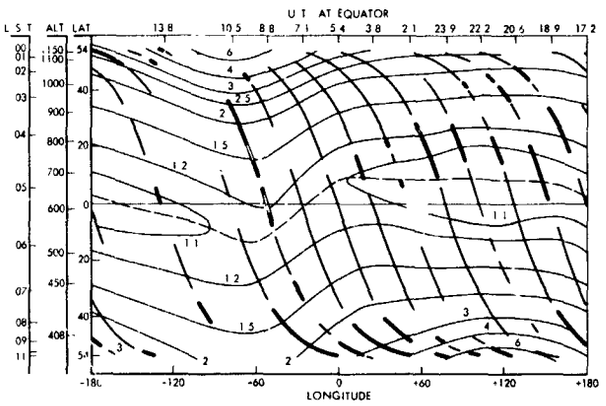


FIGURE 6.—Enhancements for May 27, southbound passes.

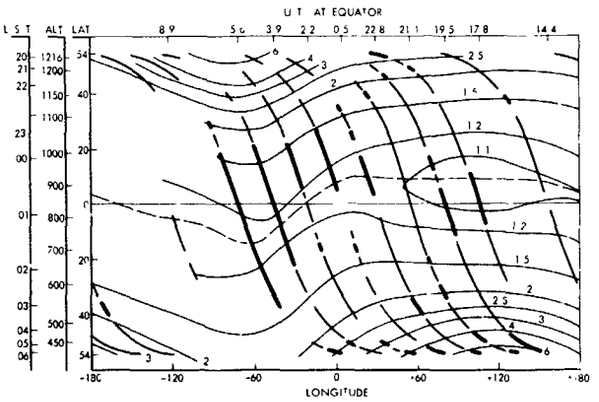


FIGURE 8.—Enhancements for June 9–10, southbound passes.

was a small gap but the sourrounding data suggested a high probability of enhancement, the gap has been bridged by double fine lines. The magnetic equator has been shown as a dashed line, and contours of L values have been drawn. Each pass has been identified by the universal time when the geographic equator was crossed.

QUIET DAY VARIATIONS

Figures 1 and 2 show the contours of electron density for northbound and southbound passes on May 24th, a very quiet day (K sum = 3+). Figures 3 and 4 give the contours for June 20th, when K sum = 8o.

Several interesting regions can be seen:

1. There is a strong maximum on the magnetic equator, centered at longitude —90° This is well-defined on figures 1, 2 and 3, but is absent on figure 4. It is near the South Atlantic anomaly, but displaced about 45° westward. This maximum persists for se eral days at a time, and on other occasions disappears completely. The presence of the maximum is not inconsistent with the model of the diurnal variation of the equatorial anomaly described by King et al (1963a), since the anomaly would not have formed at the local times of figs. 2 and 3, and for fig. 1 the peak would be below the altitude of Ariel. The absence of a maximum in fig. 4 is more puzzling.
2. The regions of high magnetic latitude centered in the northern hemisphere at longitude −90°, and in the southern at

+90°. Muldrew (1963) and others have noted a minimum in electron density extending for about 3° in latitude in the vicinity of the magnetic shell $L=3$ to 4. Figures 1, 2, and 4 show this minimum in the northern hemisphere. It is not present on figure 3. The southern hemisphere region shows a minimum only on figure 4. There is no evidence of it on 1 and 3, and there is a strong maximum on 2. From these and other contour maps it is evident that there is a diurnal variation of this feature. For the southern (winter) hemisphere the electron density (relative to that of neighbouring latitudes) is a minimum during the local night hours, and a maximum in the late morning. For the northern (summer) hemisphere, the local time variations are similar, except that the minimum is deeper and more persistent, and just disappears near noon. It is interesting to note in the high southern latitudes near longitued −100° that the diurnal variation of relative electron density is opposite to that at longitude +90°.

3. On figure 4 a pronounced minimum is evident along the magnetic equator. In the original data this minimum could be for every pass over the equator. It appears to be a feature of the evening ionosphere. It also deviates south of the magnetic equator into the region of the South Atlantic anomaly. The altitude and local time for the minimum as seen in figure 4 are such as to place it some 500 km above the peak of the equatorial anomaly. It is therefore a new and apparently independent phenomenon.

DISTURBANCE EFFECTS

For details of the magnetic conditions for the different disturbance periods, refer back to Table 1. The general impression from the curves of figs. 5–8, and other similar plots, is a tendency for enhancements to occur along magnetic shells, which merge as the disturbance increases. The patterns do not show complete shell formation except in the region of $L>3$. This may be linked with the comparative weakness of the disturbances, and the fluctuations of magnetic activity. Favoured L values would appear to be 1.2 ± 0.1, 1.5 ± 0.1, 2.0 ± 0.2, 2.5, and generally for $L>3$. Other disturbance features seem to vary from storm to storm.

Figures 5 and 6 show enhancements for the minor disturbance of May 27th. From the scale of universal time it will be seen that the passes have been folded to accommodate them all on one scan of longitude. For example on fig. 5 the pass for 0.7 hours on May 28th is next to that for 2.9 hours on May 27th, and the upper portions of passes for 6.2 and 7.9 hours appear on the left of the diagram. A comparison of the enhancements in the high southern latitudes reveals differences in distribution which are due partly to differences in storm time, in that a clear progression to lower L values occurs, particularly in fig. 6. This is probably masked by local time effects to some extent in fig. 5, in that the higher level of undisturbed electron density reduces the relative size of an increase. In fig. 6 it is notable that the enhancements near the magnetic equator are confined almost entirely to the northern hemisphere.

The effects due to a more significant disturbance on June 9th-10th are shown in figs. 7 and 9. Rothwell et al (1963) have reported on the enhancements in the high southern latitudes for this disturbance. Note the shift in the origin of the longitude scale in fig. 7 in order to include the more marked enhancements on the one graph. The major feature of this plot is the broad region of enhancement in the southern hemisphere, which moves toward the equator as the storm intensifies. This storm-time effect receives further support from the coincidence that the disturbance lasted 36 hours, and so a direct comparison of enhancements for the same latitude, longitude, height and local time could be made for successive days. Passes at 1.8, 3.5 and 5.1 hours on June 9th (not shown in the figure) show enhancements in high southern L regions for $L>3$. This compares with L down to 2, 1.5 and 1.3 for the corresponding passes on June 10th. The equatorial enhancements for this figure are similar to those noted by King et al (1963b) in the equatorial anomaly region for the major disturbance of September 22nd, 1963. Figure 8 is

remarkable for the strong night-time enhancement near the magnetic equator, which was most pronounced in the vicinity of the South Atlantic anomaly. Passes for 2.7 and 4.3 hours on June 9th also show this strong equatorial enhancement. The northern hemisphere part of the Sayers shell for $L=1.27$ was very prominent in the longitude region $-70°$ to $+40°$ for the whole of the period June 8th-13th.

The disturbance on June 21st-22nd, the graphs for which are not reproduced, showed enhancements confined almost entirely to the high latitude regions.

CONCLUSIONS

1. Longitudinal irregularities exist even in magnetically very quiet periods. However, the pattern changes to some extent between successive quiet periods.
2. The enhancements show a structure resembling magnetic shell forms, which suggests a particle origin. The tendency of L values to decrease as magnetic activity increases also resembles precipitation zones of the outer Van Allen belt; also the occurrence of a maximum over the South Atlantic anomaly.
3. Therefore it appears that particle effects are at least partly responsible for the ionization irregularities. In this case, the changes in the pattern cannot reflect variations in flux associated with local geomagnetic field changes in the vicinity of the measurements. They must therefore be mapping in some way the geomagnetic field at high L values, where the particles must either be produced or accelerated.

REFERENCES

KING, J. W., SMITH, P. A., ECCLES, D. and HELM, H. 1963a. "The structure of the upper ionosphere as observed by the topside sounder satellite, Alouette." Radio Research Station document No. RRS/I. M.94.

KING, J. W. *et al.* 1963b. "Further studies of the topside ionosphere based on the topside sounder satellite data." Radio Research Station document No. RRS/I.M.112.

MULDREW, D. B. 1963. Canad. J. Phys. **41**, 199.

NELMS, G. L., and WARREN, E. S. "Some irregular variations of the electron density in the topside of the ionosphere." Presented at COSPAR 1964.

ROTHWELL, P., SAYERS, J. and WAGNER, J. H. 1963. "Distribution at high latitudes and during magnetic disturbances." Proc. Roy. Soc. Ariel Discussion Meeting.

WILLMORE, A. P., HENDERSON, C. L., Boyd, R. L. F., and BOWEN, P. J. 1963. "Electron temperature in the upper F-region." Proc. Roy. Soc. Ariel Discussion Meeting.

V. FIELDS AND PARTICLES

MEASUREMENT OF PRIMARY COSMIC RAY CHARGE AND ENERGY SPECTRA DURING 1963*

V. K. BALASUBRAHMANYAN† AND F. B. McDONALD

The differential energy spectra of *H* and *He* nuclei in the interval 100–800 Mev/nuc, the integral *H* and *He* flux >800 Mev/nuc and the cosmic ray charge spectra in the range $Z=1-8$ have been measured using a three element telescope consisting of two thin scintillation counters and a Cerenkov Counter. The measurements were made from Fort Churchill, Canada on a Skyhook balloon flight during June, 1963. The ratio of $Li+Be+B/C+N+O$ appears to be constant in the two regions 400–800 Mev/nuc and >800 Mev/nuc. This ratio extrapolated to the top of the atmosphere is $0.28\pm.08$. A marked difference is observed in the form of the low rigidity differential spectra of *H* and *He*. These measurements are compared with the results available from the last solar cycle.

INTRODUCTION

The objectives of the experiment reported here are two fold:

(a) To develop an improved detector system with adequate charge and energy resolution so that individual charges, their absolute intensities, energy spectra and time variation could be studied with the long range aim of using the detector in a satellite borne experiment.

EXPERIMENTAL SET-UP

A schematic drawing of the detector system is shown in Fig. 1. The system is similar in principle to that reported by McDonald and Webber[1] except for some technical improvements to achieve better charge and energy resolution. For each particle traversing the telescope, the pulse height from all three detectors are recorded. For each detector there is a 512 channel pulse height analyzer and the data is recorded on a balloon borne tape recorder. The total weight of the complete system including batteries and pressure can is 65 lbs.

The results of a balloon flight on 24 June 1963 are reported here. During this flight the balloon remained for 12 hours at a ceiling altitude of 5 gms/cm². The equipment performed reliably throughout the flight and the pre-flight and post-flight calibrations were in agreement with each other. The period during the flights was characterized by a quiet sun, with no abnormal activity.

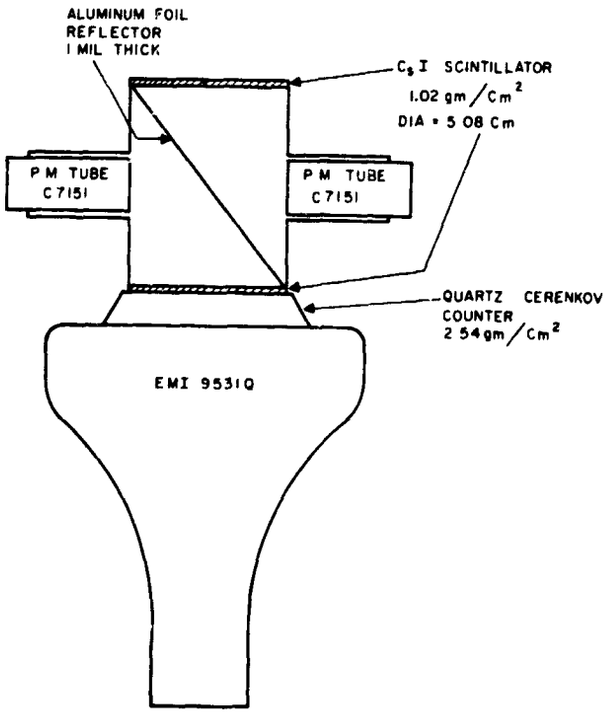


FIGURE 1.—Geometry

*Published as *Goddard Space Flight Center Document X-611-64-63*, 1963.

†National Academy of Science fellow on leave from Tata Institute of Fundamental Research, Bombay, India.

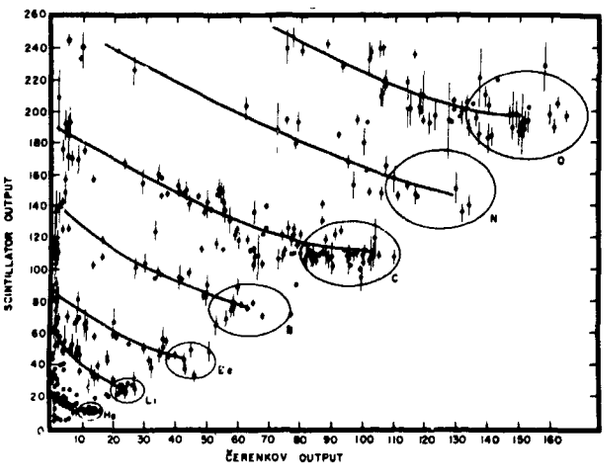


FIGURE 2.—Plot of Cerenkov output vs. ionization loss in telescope.

RESULTS

Fig. 2 shows the charge resolution obtained in the region $Z=2-8$. In the region 350 Mev/nuc (>5 Cerenkov Pulse height) there is excellent resolution between all components. The L/M ratio at the observation level of 5 gm/cm^2 for the energy range 400–800 Mev/nucleon, $=.36\pm.07$ where as for the energy >800 Mev/nucleon the ratio is $.36\pm.09$. The ratio extrapolated to the top of the atmosphere using fragmentation parameters of Friedlander et al.[2] is $0.28\pm.08$ in agreement with the results at Texas 41°N.[3]

A summary of the results on the charge distribution obtained in this experiment is shown in Table 1. Also shown there, are results from previous measurements from other experiments.

MODULATION

Fig. 3 shows the integral intensity of *He* nuclei, the medium and heavy nuclei as a function of neutron count. For rigidity >1.5 BV the agreement between the curve obtained by Webber and McDonald during the last solar minimum, shows that at medium rigidities the sea level neutron count correlates well with the direct measurements. This gives a measure of confidence regarding the accuracy of the measurements.

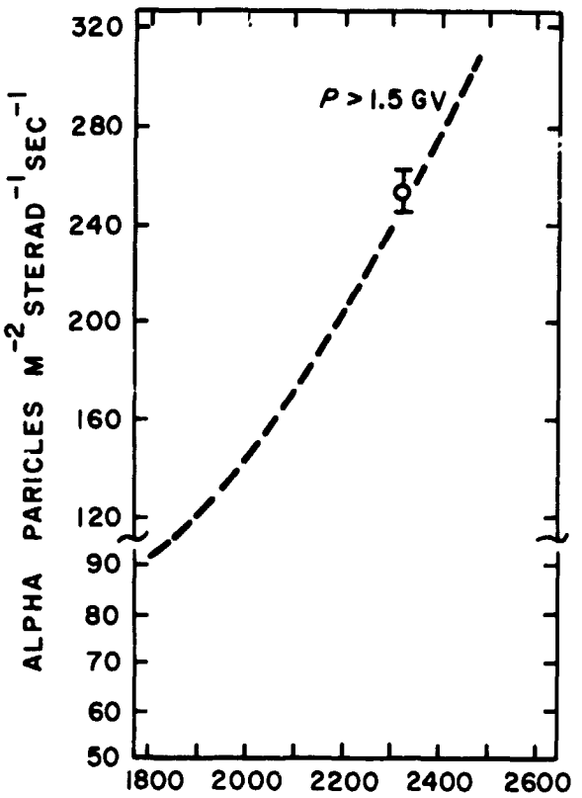


FIGURE 3.—Mt. Washington neutron rate.

Fig. 4 shows the differential rigidity spectra of protons and *He* nuclei. There are some interesting

TABLE 1.—*Comparison of Charge Distributions Obtained at the Top of the Atmosphere Charge Distribution (Percent)*

Charge	Balasubrahmanyan & McDonald	Waddington[3]	Aizu[4] et al	Tamai[5]	Fichtel[6]	O'Dell[7] et al
Li	8.8	3.9	8.8	10.0	7.4	5.3
Be	6.4	1.7	6.0	14.0	5.7	2.3
B	8.1	11.6	10.9	15.7	9.0	7.4
C	31.4	26.0	29.2	18.8	27.1	30.1
N	9.2	12.4	14.8	7.8	15.3	9.7
O	16.6	17.9	14.4	7.3	14.4	19.4
Z 10	19.5	23.9	21.7	20.5	21.7	23.5

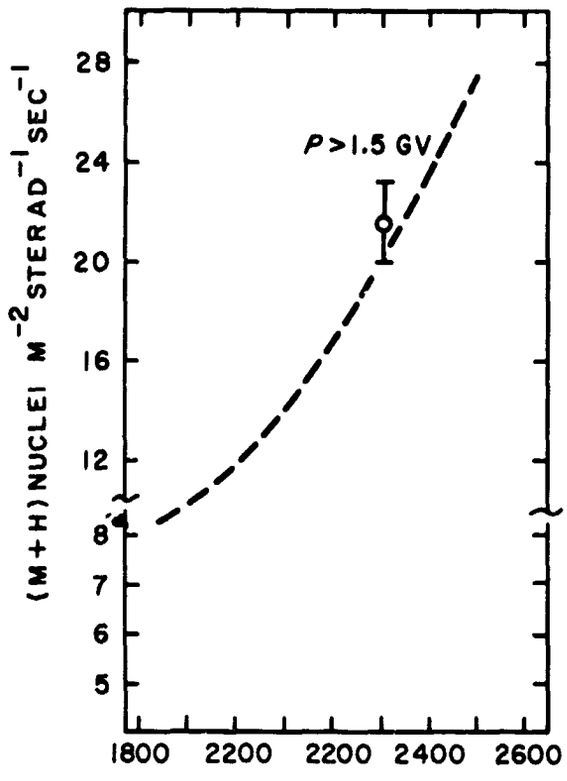


FIGURE 4.—Mt. Washington neutron rate.

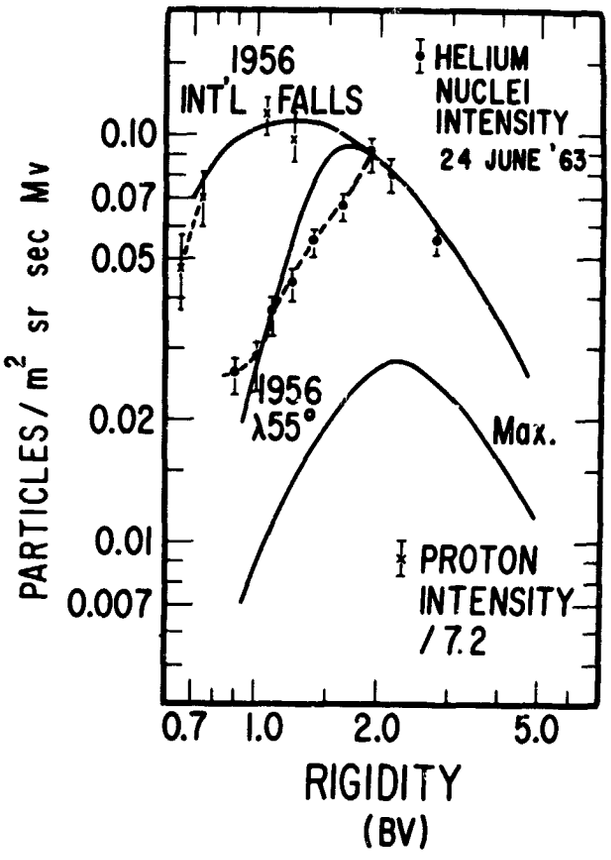
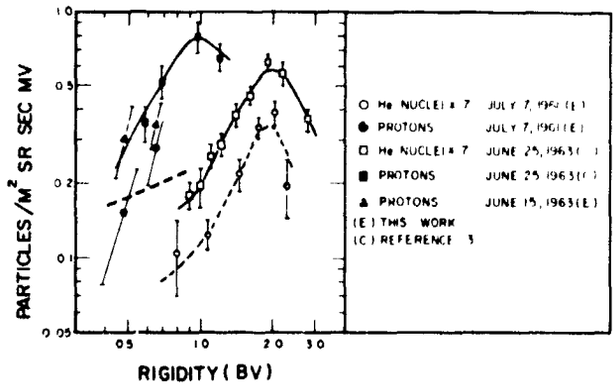


FIGURE 5.—Alpha and Proton Rigidity Spectra

features which this figure reveals. First the differential intensity of Helium nuclei of rigidity >2BV appears to have recovered to the previous solar minimum period. On the low energy side, the previous measurement at International Falls, Minnesota is higher than both the proton and *He* curves of the present experiment. So, on the low rigidity side the recovery is still not complete.

Another interesting feature is the difference in rigidity spectra between the proton and *He* nuclei curves. The recovery for protons has been faster than for *He* nuclei for the same rigidity. In Fig. 5 the results of the counter measurements and that of Nuclear Emulsion measurements by Fichtel et al.[8] are shown. The good agreement between the results of these two entirely different experiments, particularly in the field of intensity measurements of low energy protons, where the secondary corrections etc. are applied on an entirely different and independent basis, gives one a measure of confidence in the interpretation of these results and their bearing on the modulation process.

Parker[9,10] has proposed a diffusion model, wherein the solar wind is responsible for the modulation. In this model, the modulation at a given rigidity is different for particles with different charge to mass ratios because of the velocity dependence of the diffusion mechanism. This model predicts that the differential spectra of the helium nuclei and protons should be different, with the *He* nuclei being suppressed more than



protons. The observations reported here are at least in qualitative agreement with Parker's theory.

REFERENCES

1. F. B. McDonald and W. R. Webber[1], Proceedings of Kyoto Conference II, pp. 429, 1962.
2. M. W. Friedlander, K. A. Neelakantan, S. Tokunaga, G. R. Stevenson and C. J. Waddington, Report of the Cosmic Ray Group, Washington University, St. Louis.
3. C. J. Waddington, Proc. Kyoto Conference, 1961
4. H. Aizu, H. Y. Fujimoto, S. Hasegawa, M. Koshiba, I. Mito, J. Nishimura, K. Yokoi, and M. Schein, Phys. Rev. 121, 1206, 1961.
5. E. Tamai, Phys. Rev. 117(5), 1345, 1960.
6. C. E. Fichtel, Nuovo Cimento, 19(6), 1100, 1961.
7. F. W. O'Dell, M. M. Shapiro, and B. Shiller, Proc. of Kyoto Conference III, pp. 23, 1619.
8. C. E. Fichtel, D. Guss, D. A. Kniffen and K. A. Neelakantan, Proc. Jaipur Conference, 1963.
9. E. N. Parker, Phys. Rev. 109, 1874, (1958).
10. E. N. Parker, Phys. Rev. 110, 1445, (1958).

MAGNETOHYDRODYNAMIC SHOCK STRUCTURE IN A MOVING MAGNETIC FIELD*

DAVID B. BEARD†

The shock structure around a cylinder immersed in a cold collisionless plasma stream containing a weak magnetic field has been investigated in two dimensions. The stream pressure on the boundary of the sheath is proportional to $\cos^2\psi$ where ψ is the angle between the stream velocity and the normal to the sheath. The thickness of the sheath and particle density in the sheath on the upstream side of the cylinder have been found to be nearly constant. A tear drop shaped vacuum is formed downstream from the cylinder. The plasma electrons have essentially all the ion stream kinetic energy in the transition sheath upstream of the cylinder but the ion stream energy is regained downstream from the cylinder. The importance of this shock structure when applied to the earth's magnetosphere and the generation of Type I comet tails is discussed briefly.

1. INTRODUCTION

The solar wind, consisting of a cold plasma flowing constantly at high velocity away from the sun, is now known to contain a weak magnetic field [1,2] oriented somewhat perpendicular to the stream velocity vector. Though the magnetic field carried along in and moving with the stream velocity is so weak that the magnetic field pressure is only one percent of the stream pressure, the field prevents the particles from moving independently of one another. The result is that any object larger than a few ion gyroradii will create a magnetohydrodynamic shock when immersed in the solar wind.[3-9]

The distinctive features of shocks generated in the solar wind are caused by the shape of the object creating the shock and the effect of arresting the velocity of the magnetic field lines which initially must move with the wind velocity since the stream pressure is overwhelmingly greater than the magnetic pressure. The plasma and its field are compressed on the upstream surface of the obstacle and much of the stream kinetic energy is diverted to gyromotion perpendicular to the compressed magnetic field. A simple extension of ordinary M theory of the magnetic pinch effect shows that on entering the compressed field region through the charge separation electric field on the surface, the plasma electrons will acquire essentially all of the kinetic energy of the ion stream in the solar wind. This energy is returned to the ion stream after the plasma in the compressed field region moves around to the downstream side of the obstacle.

In the work presented here a simple model is discussed in which a cylinder of intense magnetic field parallel to its axis presents an obstacle to a cold plasma stream containing a weak magnetic field parallel to the cylinder axis. Besides the inherent interest of the model in understanding magnetohydrodynamic shock structure under such unusual conditions, the model illuminates some gross aspects to be expected in astrophysical problems which occur, for example, in the boundary of the earth's magnetic field and in the tail rays of comets.

2. PHYSICAL CHARACTERISTICS OF THE TRANSITION SHEATH SURROUNDING A CYLINDER IN A MAGNETIZED PLASMA STREAM

The simple idealized model considered here is illustrated in Fig. 1. A proton-electron plasma, collisionless and with no thermal motion, is streaming parallel to the polar axis with velocity v_0, and encounters a cylinder of radius R containing an arbitrarily intense magnetic field sufficient

*Published as *Goddard Space Flight Center Document X-640-64 164*, June 1964.

†Goddard Space Flight Center, University of California, Davis, California, and University of Kansas, Lawrence, Kansas.

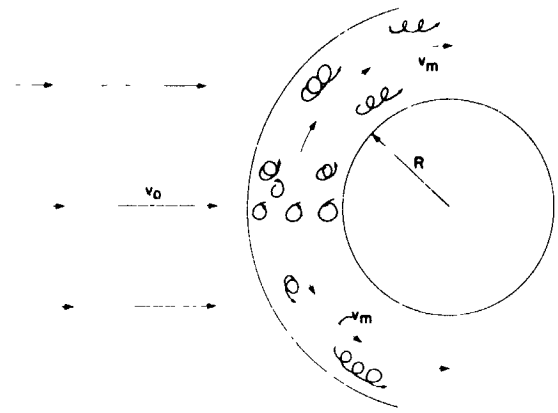


FIGURE 1.—Illustration of solar wind containing a weak magnetic field perpendicular to the plane of the figure incident on a cylinder of radius R. The arrows represent the velocity of the field lines (particle guiding centers) in the free stream v_0, and in the transition sheath, v_m. The circles and loops represent partial trajectories of the particles while in the sheath.

to exclude the incident plasma stream. (In general the obstacle need furnish only a spatially varying magnetic field pressure, $H_g^2/8\pi$, but a cylindrical shape has been introduced at the start to simplify the discussion of the paramount physical features resulting from the analysis.) Finally, the incident plasma stream is assumed to contain a weak magnetic field, parallel to the cylinder axis, which is carried along with the plasma. The stream magnetic field exerts a pressure of only one percent of the stream pressure and the field traveling with the stream velocity v_0 is maintained by the small gyromotion of the plasma in the moving velocity frame of the plasma stream. In this connection it might be noted that the preliminary analysis by Dr. M. Neugebauer of the Mariner R plasma data yielded an effective plasma thermal energy of the order of 10–30 electron volts in a stream energy of the order of a kilovolt. Dr. Neugebauer's analysis was made by examining the energy spread parallel to the stream motion, but the presence of a magnetic field and the gyromotion it causes guarantees that the result is valid for the particle energy component perpendicular to the stream motion as well.

The particle pressure on the outer boundary of the transition sheath is given by

$$p_p = (n_0 v_0 \cos\psi)(m_p v_0 \cos\psi) = n_0 m_p v_0^2 \cos^2\psi \qquad (1)$$

where n_0 is the proton density in the stream m_p is the proton mass and ψ is the angle between the normal to the outer boundary surface of the transition sheath and the stream velocity vector. The second paranthetical factor represents the momentum change of the incident particles at the boundary as they enter the sheath. The external pressure confining the particle and enhanced magnetic field pressure in the sheath acts perpendicularly to the surface. Hence there is no change in particle center of motion momentum component parallel to the boundary which would result in a momentum change factor not linearly dependent on $\cos\psi$.

There is, of course, a very important change in the particle motion on crossing the boundary of the sheath and that is that near the stagnation point, in particular, much of the stream motion is diverted to gyromotion perpendicular to the magnetic field. Although the magnetic field is too weak to contribute effectively to the pressure near the stagnation point, it does cause the charges to rotate about the lines of force. Particle pressure, however, dominates every other consideration and the guiding center of the particles are caused to move towards decreasing pressure gradients "dragging" the magnetic field lines along with them. The particle pressure within the transition sheath is given by

$$p_p = \tfrac{1}{2} n_s m_p v_p^2 = \tfrac{1}{2} n_s m_p (v_0^2 - v_m^2) \qquad (2)$$

where n_s is the ion density within the sheath, v_p is the particle velocity component perpendicular to the field, and v_m is the velocity of the guiding center, that is, the velocity of the magnetic lines of force. The factor 1/2 arises because of the two dimensional motion of the particles within the sheath as opposed to the monodirectional motion of the stream pressure. In three dimensions, where the particle motion is also along the lines of force, the factor might be 1/3.

We thus obtain an expression for the pressure equality on the boundaries of the sheath.

$$n_0 m_p v_0^2 \cos^2\psi + H_0^2/8\pi = \tfrac{1}{2} n_s m_p (v_0^2 - v_m^2) + (n_s^2/n_0^2) H_0^2/8\pi = H_c^2/8\pi \qquad (3)$$

where H_c is the magnetic field intensity in the cylinder at the inner boundary. Since the plasma is collisionless, the magnetic field intensity is directly proportional to the particle density, and

thus the magnetic field intensity anywhere in the sheath is $(n_s/n_0)H_0$. Particle thermal pressure would be easy to include for similar reasons. Since n is proportional to H and the square of the thermal velocity, v_{pt}^2, is proportional to H if the motion is adiabatic, the particle thermal pressure is a constant multiple of the magnetic pressure and may be included by multiplying H_0^2 by an appropriate factor.

The change in the external pressure, P, as a function of the distance along the boundary of the sheath, l, results in a pressure gradient which accelerates the plasma within the sheath.

$$n_s m_p \frac{dv_m}{dt} = n_s m_p v_m \frac{dv_m}{dl} = -\frac{dP}{dl} + \frac{d}{dl}\left(\frac{n_s^2}{n_0^2}\frac{H_0^2}{8\pi}\right) + \frac{m_p}{2}(v_0^2 - v_m^2)\frac{dn_s}{dl} \quad (4)$$

which may also be obtained by differentiating Eq. 3 with respect to l. Throughout most of the upstream side of the cylinder H^2_0 is negligible compared with the particle pressure and Eq. 3 may be approximated by

$$n_s = 2P/m_p(v_0^2 - v_m^2)$$

which satisfies Eq. 4 since the last two terms in Eq. 4 are negligible in these circumstances. Thus, $v_m \sim v_0 \sin\psi$ throughout the transition sheath on the upstream side and n_s is nearly constant and equal to $2n_0$. As ψ approaches $\pi/2$, however, the magnetic pressure becomes significant and n_s decreases with l depending on ψ as

$$n_s \sim n_0[1 + \tfrac{1}{4}(8\pi n_0 m_p v_0^2/H_0^2)\cos^2\psi - (3/32)(8\pi n_0 m_p v_0^2/H_0^2)^2\cos^4\psi] \quad (5)$$

Using the approximations valid on the upstream side of the cylinder the position of the outer boundary to the transition sheath is easily obtained by requiring particle conservation. The number of particles arriving per second per unit cylinder length within a distance, y_2, from the polar axis is $n_0 y_2 v_0$; the number leaving the transition region per second per unit cylinder length is $n_s(r_2 - r_1)v_0 \sin\psi$ (see Fig. 2).

$$r_2 - r_1 \sim r_2 \sin\theta_2/2 \sin\psi \quad (6)$$

For a cylindrical obstacle for which r_1 is constant, $\theta_2 = \psi$, and

$$r_2 = 2r_1 \quad (7)$$

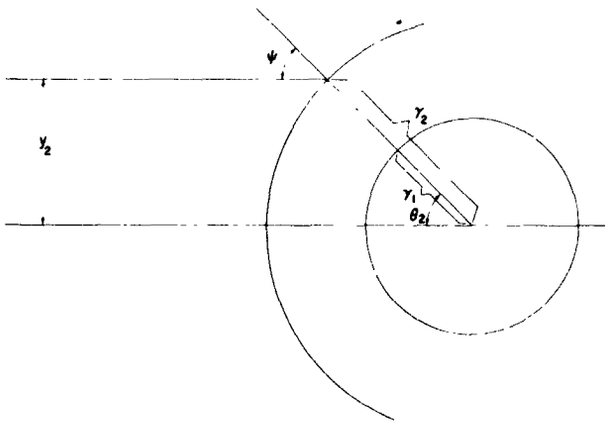


FIGURE 2.—Diagram of use of constant flux condition in determining the sheath thickness, $r_2 - r_1$, by which particles leave a volume of unit height after entering the volume within the length y_2.

If the obstacle were not cylindrical a differential form of the constant flux condition, Eq. 6, would be required

$$n_0 v_0 d\,l \cos\psi = n_s(r_2 - r_1)v_0 d(\sin\psi)$$

$$\frac{d\psi}{dl} = \frac{n_0}{n_s}\frac{1}{r_2 - r_1} = \frac{1}{r_2}\frac{d\psi}{d\theta} \quad (8)$$

which yields Eq. 7 in the region where $n_s = 2n_0$ if r_1 and r_2 are now regarded as functions of ∂. If n_s is a decreasing function of θ, as it is near $\theta \sim \psi \sim \pi/2$, $r_2 - r_1$ increases faster than r_2 and $\frac{d\psi}{d\theta}$ is less than unity. In this region, the parametric equations 5 (or rather 3) and 8 must be solved numerically.

The magnetic pressure cannot be neglected in the region

$$\theta \sim \psi \gtrsim \cos^{-1}[H_0/(8\pi n_0 m_p v_0)^{1/2}] \sim 84° \quad (9)$$

where the stream velocity is

$$v_m \sim v_0 \sin\psi \sim 0.99 v_0 \quad (10)$$

and

$$(v_0^2 - v_m^2)^{1/2} \sim 0.10\, v_0$$

Therefore, on the downstream side of the cylinder, the guiding center of stream velocity in the transition sheath is again approximately v_0, and the velocity perpendicular to the field lines and v_0 is less than $0.10\, v_0$. The stream expands freely into the vacuum behind the cylinder so as to form a diffuse triangle shaped tail, based on

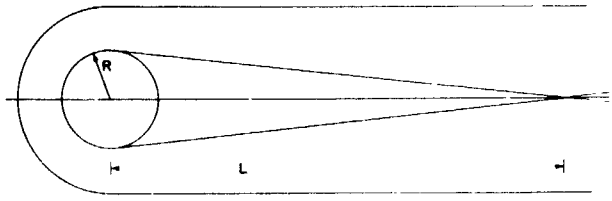


FIGURE 3.—The shock structure behind a cylinder of radius R. The length of the vacuum downstream from the cylinder $L \sim RV_0/(v_0^2 - v_m^2)^{\frac{1}{2}}$.

simple kinematics, which closes more than ten cylinder radii downstream and leaves a turbulent wake as illustrated in Fig. 3.

3. THE SHOCK PRODUCED ENERGETIC ELECTRONS IN THE TRANSITION SHEATH

The preceding section has given us a general description of our simple two dimensional model problem as illustrated in Figs. 1–3. A weakly magnetized plasma moving with velocity v_0 enters a transition sheath where the velocity of the particle's guiding centers, that is, the velocity of the magnetic lines of force, changes abruptly at the outer boundary of the transition sheath to a lower velocity, v_m. Ferraro[10] and especially Dungey[11] and Rosenbluth[12] have shown how a charge separation layer will be created at any boundary where the magnetic field is suddenly increased on the side away from the plasma. Beard,[13] Grad,[14] and Sestero[15] have further developed the theory of the electric field on the boundary which serves to energize the electrons at the expense of the positive ions. The electric field arises due to the difference in momentum between the electrons and the positive ions, and thus there is no field in any region where the momenta of the opposite charges are equal.

In the simple two-dimensional problem discussed in this paper, it might be anticipated that when the plasma enters and passes through the transition sheath the charges will have gained or lost potential energy (depending on the sign of their charge) as though they had passed through permeable condenser plates at the boundary. Hence, if there is no reverse electric field, the electrons will have acquired essentially all of the initial stream kinetic energy of the positive ions.

Following Rosenbluth's [12,13] analysis in which the plasma did not enter the region of increased field intensity, we write, for the changed conditions of this problem, the equations of motion of a charge initially moving parallel to the x axis

$$\ddot{x} = \frac{q}{m}E(x) + \frac{q}{mc}H\dot{y} \tag{11}$$

$$\ddot{y} = -\frac{q}{mc}H(\dot{x} - v_m) \tag{12}$$

where the coordinate system is illustrated in Fig. 4, q and m are the charge and mass of the charged

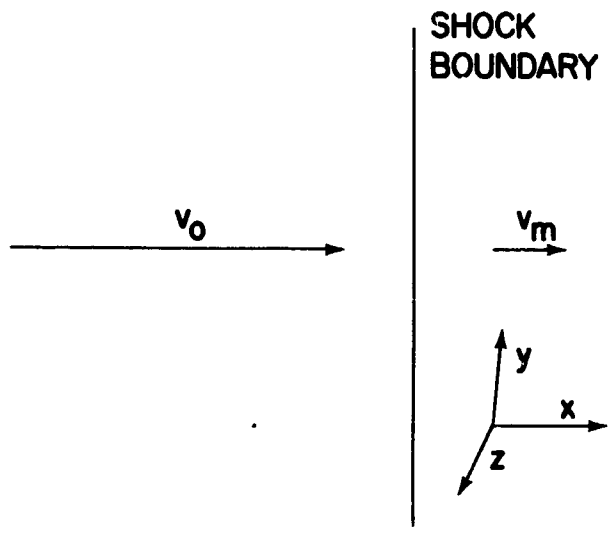


FIGURE 4.—The coordinate system used in analyzing particle motion in a magnetic field parallel to the z axis whose field lines travel with a velocity v_0 to the left of the shock boundary and with velocity v_m to the right of the shock boundary.

particle, E is the electric field perpendicular to the boundary which is tangent to the y, z plane, H is the magnetic field and is parallel to the Z axis, and v_m is the velocity of the guiding centers (i.e. the magnetic lines of force) which are assumed to be moving entirely parallel to the x axis for this part of the analysis. Eq. 12 is easily integrated and yields the y component of the particle velocities.

$$\dot{y} = \frac{-q}{mc}x / \int_{-\infty}^{t} H\,(x' - v_m)dt' \tag{13}$$

Hence the y component of the particle momenta is the same for both electrons and protons.

The particle densities are inversely proportional to the x component of the average velocity of the particles. To avoid the impossibly large electric

fields resulting from excessive charge separation the x component of the average velocities of the particles must be kept equal everywhere. This requires that in being slowed from an initial velocity v_0 to the final guiding center velocity of V_m, the x-component of the deceleration of the electrons must be equal to the deceleration of the ions. Thus Eq. 11 and 13 yield the equality

$$\frac{e}{m_e}E(x)-\left(\frac{e}{m_e c}\right)^2 H\int_{-\infty}^{t} H\,(x'-v_m)dt'=-\frac{e}{m_i}E\,(x)$$

$$-\left(\frac{e}{m_i c}\right)^2\int_{-\infty}^{t} H\,(x'-v_m)dt'$$

where the subscripts e and i refer to electron and positive ion and $q_e=-e$, $q_i=e$. Neglecting the right hand side of this equation as being of order m_e/m_i smaller than the left hand side we obtain the familiar M theory result

$$E(x)=\frac{e}{m_e c^2}H\int_{-\infty}^{t} H\,(x'-v_m)dt' \qquad (14)$$

If Eq. 13 and 14 are inserted into Eq. 11, then the integral of Eq. 11 after multiplication by x is to zero order in m_e/m_i

$$x_i^2=v_0^2-\frac{2e}{m_i}\int_{-\infty}^{x} E\,(x')dx'$$

Let X_M represent the position at which the incident proton first reaches the final guiding center velocity, V_m, (i.e. the position beyond which the compressed field is constant.) then,

$$\tfrac{1}{2}m_i(v_0^2-V_m^2)=e\int_{-\infty}^{X_M} E\,(x')dx' \qquad (15)$$

Eq. 15 has the very obvious physical interpretation that the total potential drop for $-\infty<x<x_M$ is equal to the change in the ion stream kinetic energy divided by the ion charge. For $x>x_M$ charge neutrality does not require $\dot{x}_e/x'_i$; only their averages, V_m, must be equal.

Since the momenta of the electrons and protons are equal, no further charge separation takes place; the gyroradii of the particles are equal. The momentum of an ion in the stream $m_i v_0$ has been reduced to $(m_e/m_i)^{1/2}m_i v_0$ in exerting the pressure on the boundary needed to confine the enhanced magnetic field interior to the boundary. The kinetic energy of the ions has been reduced, however, from $\frac{1}{2}m_i v_0^2$ to $\frac{1}{2}m_e v_0^2+\frac{1}{2}m_i v_M^2$. The considerations of the previous section are still valid since the total particle pressure in the transition layer is not affected by the electrical potential at the boundary since the electron energy in the transition layer is $\frac{1}{2}m_i v_0^2-\frac{1}{2}m_i v_M^2$.

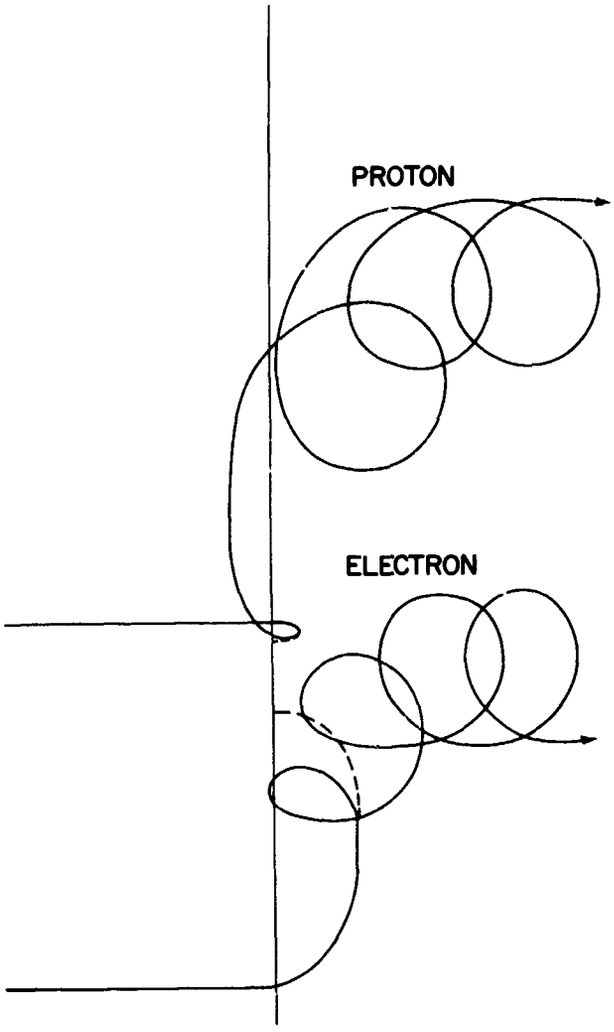


FIGURE 5.—Electron and proton trajectories initially traveling with a moving magnetic field which is slowed down and compressed, downstream from the shock boundary.

A free hand drawing of the particle motion as it crosses the boundary is shown in Fig. 5 for conditions near the stagnation point. The conclusions are essentially unchanged at other points on the boundary as illustrated in Fig. 6 if one considers only the component of the guiding center

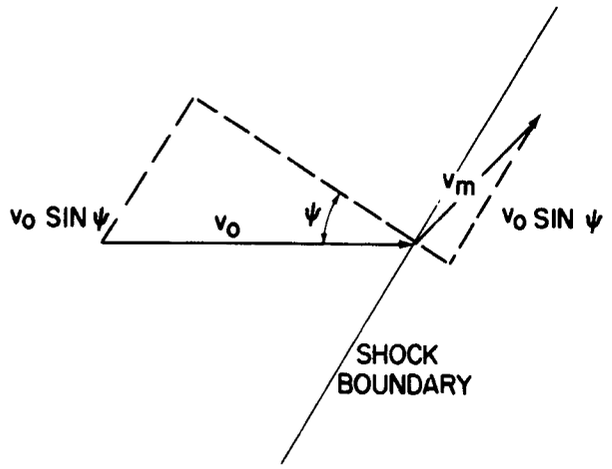


FIGURE 6.—The geometry of the guiding center or field line motion on both sides of the shock boundary for oblique incidence.

motion perpendicular to the boundary. The component of the stream velocity parallel to the boundary is the same on both sides of the boundary and therefore the problem reduces to the stagnation point case in a coordinate frame which moves with the velocity $v_0 \sin\psi$ parallel to the boundary and $v_0 \cos\psi$ supplants v_0 in the stagnation point considerations. In accelerating to nearly the free stream velocity again in its motion away from the stagnation point, the motion of the plasma in the transition sheath causes the initial ion kinetic energy to be recovered at the expense of the electron energy.

CONCLUSIONS

A cold plasma stream containing a weak magnetic field (field pressure much less than particle stream pressure) will be compressed on the upstream surface of a cylindrical obstacle whose axis is parallel to the magnetic field. The particle density and magnetic field intensity within the transition sheath on the upstream side are double their free stream value. The thickness of the sheath is equal to the cylinder radius over almost all of the upstream surface and thickens slightly near the joining of the upstream and downstream surfaces due to the additional magnetic pressure, however small, in the transition sheath. The enhanced magnetic pressure in the transition sheath, however slight, prevents the ions from returning completely to their initial free stream motion on the downstream side of the cylinder. As a result, a plasma stream, even though the particles initially have a unique velocity, will expand into the vacuum behind the cylinder. In the case of a plasma with some thermal velocity $(kT/2\pi m)^{1/2}$ the vacuum downstream would be a diffuse isoceles triangle whose height would be $v_0 (2\pi m/kT)^{1/2} R$ where v_0 is the initial stream velocity and R is the cylinder radius (1/2 the triangle base.)

The enhanced magnetic field in the transition sheath creates a charge separation layer at the boundary of the undisturbed plasma stream. The total change in electrical potential across this layer is equal to half the stream pressure on the transition sheath. The plasma stream moves through the charge separation layer, and, on entering the transition sheath, the electrons acquire all the kinetic energy the ions would have had perpendicular to the magnetic field in the absence of the charge separation layer. At the stagnation point the electron kinetic energy in the transition sheath equals the initial ion kinetic energy in the free stream, but elsewhere in the transition sheath the electron kinetic energy equals the difference between the ion free stream kinetic energy and the ion stream energy in the transition layer.

ASTROPHYSICAL APPLICATIONS

The application of this work to Type I comet tails is straightforward and will be reported elsewhere. The slight ionization of massive cometary molecules through charge transfer with protons in the solar wind cause the interplanetary magnetic field lines to slow down drastically in overcoming the inertia of the massive cometary ions initially at rest. The energetic electrons in the transition sheath which forms on the surface of the "slowed" field lines then rapidly and efficiently ionizes hundreds of cometary molecules per cc. The energetic electrons will further rapidly diffuse out along the field lines (as suggested by Axford in private conversion) and pull the ions out along the field lines so as to form the striking ray structure known as Type I comet tail rays.

The simple model examined in this paper also yields a good qualitative understanding of the phenomena observed at the edge of the earth's magnetic field and enables predictions about the

energy of the charges in the magneto-sheath to be made. The sphericity of the earth's magnetosphere introduces a new physical dimension. In the plane containing the stagnation point and the magnetic field lines the particle pressure both inside and outside the outer boundary of the transition sheath (A. J. Dessler has suggested the name magneto-sheath) is still proportional to $\cos^2\psi$. The particles are decoupled, however, from the magnetic field in the direction of the field, and conservation of magnetic flux requires that the magnetic field intensity be approximately independent of distance from the stagnation point or $\cos\psi$ in the case of a cylinder and to *increase* away from the stagnation point in the case of a sphere. Therefore, in this plane the shock boundary around a circular surface would expand faster than in the plane at right angles to the interplanetary magnetic field. The particle density in this plane away from the stagnation point is less than in the plane perpendicular to the magnetic field.

The particle pressure in this plane is *not* communicated along the field lines and this introduces a complication in the analysis depending on the isotropy of the particle velocity discussed below. Free hand drawings of the field lines are illustrated in Figs. 7 and 8.

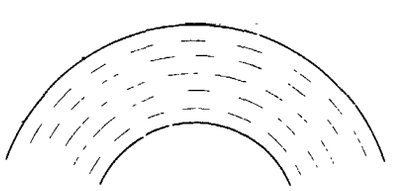


FIGURE 7.—Magnetic field lines indicated by dashed lines in free interplanetary space and in the transition sheath in the plane containing the field direction.

Another aspect of the magnetosphere problem is that turbulence on the surface or the sharp kinks produced in magnetic field lines where the interplanetary field lines are joined to the lines in the magnetosheath will invalidate the simple two dimensional analysis presented in this work. The particle motion in the magnetosheath will be more nearly isotropic in three dimensions

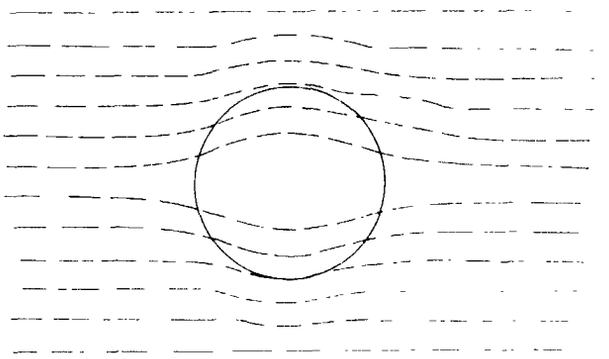


FIGURE 8.—Magnetic field lines, indicated by dashed lines, viewed in projection from the upstream side as they diverge and slide around a spherical obstruction bowing out most pronouncedly in the equatorial plane.

instead of being isotropic in twod imensions. This causes the particle density and field intensity to be three times the interplanetary values and the thickness of the sheath near the stagnation point to be more nearly one-half of the radius of curvature of the magnetosphere.

Independent of these considerations the sudden increase of the magnetic field intensity from the interplanetary field value to the magnetosheath value will cause the magnetosheath electrons to have the same momentum perpendicular to the field lines as the protons do. Thus the particle energy and pressure perpendicular to the field lines will be monopolized by the electrons. The particle energy and pressure parallel to the field lines will be monopolized by the protons, however, since charge separation effects will cause the electron and proton velocities to equalize.

ACKNOWLEDGEMENT

I would like to thank Dr. John Banister of Sandia Corporation for a helpful discussion of this problem. I would also like to thank both the Sandia Corporation and the Goddard Space Flight Center for their hospitality and support of this work.

REFERENCES

1. P J. COLEMAN JR., L. DAVIS JR., E. J. SMITH, and C. P. Sonett Science, **138**, 1099, (1962)
2. N. F. NESS, C. S. SCEARCE, and J. B. SEEK, J. Geophys. Res., in press, (1964).
3. F. DE HOFFMAN and E. TELLER, Phys. Rev. **80**, 692, (1950).
4. P. L. AUER, H. HURWITZ JR., and R. W. KILB, Phys. of Fluids, **5**, 298 (1962).

5. P. J. KELLOGG, J. Geophys. Res., **67**, 3805, (1962).
6. J. R. SPREITER, and W. P. JONES, J. Geophys. Res., **68**, 3555, (1963).
7. W. I. OXFORD, J. Geophys. Res., **67**, 3791, (1962).
8. D. B. BEARD, J. Geophys. Res., **69**, 1159, (1964).
9. L. LEES, Proc. AIAA New York meeting, Jan. 1964 and Cal. Inst. of Tech. preprint.
10. V. C. A. FERRARO, J. Geophys. Res., **57**, 15, (1952).
11. J. W. DUNGEY, Cosmic Electrondynamics, Cambr. Univ. Press, Cambridge, England, 132–152, (1958).
12. M. N. ROSENBLUTH, Dynamics of a Pinched Gas contribution to Magnetohydronamics edited by R. Landshoff, Sanford Univ. Press, Palo Alto, 57–66, (1957).
13. D. B. BEARD, J. Geophys Res., **65**, 3559, (1960).
14. H. Grad, Phys. Fluids, **4**, 1366, (1961).
15. A. SESTERO, Phys. Fluids, **7**, **45**, (1964).

SOLAR CYCLE CHANGES IN INNER ZONE PROTONS

ROBERT C. BLANCHARD† AND WILMOT N. HESS

N65-29503

Time dependent calculations of the inner Van Allen belt proton population show that large changes of the population up to a factor of 50 will take place during the solar cycle. The effect is most pronounced for the region of B–L space corresponding to minimum altitudes of 300–700 km. Because different energy protons respond to the changing solar cycle at different rates the proton energy spectra will change with time also.

The purpose of this paper is to calculate the expected changes in inner zone proton populations with time in the solar cycle.

Freden and White [1959] identified the penetrating component of the inner zone of the Van Allen belt as due to energetic protons and measured the energy spectrum of the protons of $E>75$ Mev. This and subsequent experimental work on Atlas rockets [Freden and White, 1960; Freden and White, 1962; Armstrong, Harrison, Heckman and Rosen, 1961] has confirmed and extended this finding and we now have a well-established proton energy spectrum at $L\sim1.4$, $B\sim.20$, which is about apogee for these flights, for near solar maximum (see Fig. 1). The analysis of this spectrum [Freden and White, 1960] has shown conclusively that the protons are produced by neutron decay.

One of the features of this component of the radiation belt is the time constancy. Measurements after a solar flare [Freden and White, 1962] gave essentially the same fluxes of protons as before the flare. For $L>1.6$ changes in proton spectra are seen [Naugle and Kniffen, 1961] associated with solar proton events [Pizzella, 1961]. These may be explainable in terms of neutrons produced in the polar regions by solar protons [Lenchek, 1962; Lingenfelter, to be published]. For $L<1.6$ only gradual changes in the proton populations have been observed on Explorer VII [Pizzella, 1961]. It was originally suggested [Pizzella, 1961] that such changes implied that neutrons could not produce all inner zone protons but this does not seem to be the case. Hess [1962] showed that slow changes in proton population were expected due to changes in the galactic cosmic ray flux during the solar cycle and more

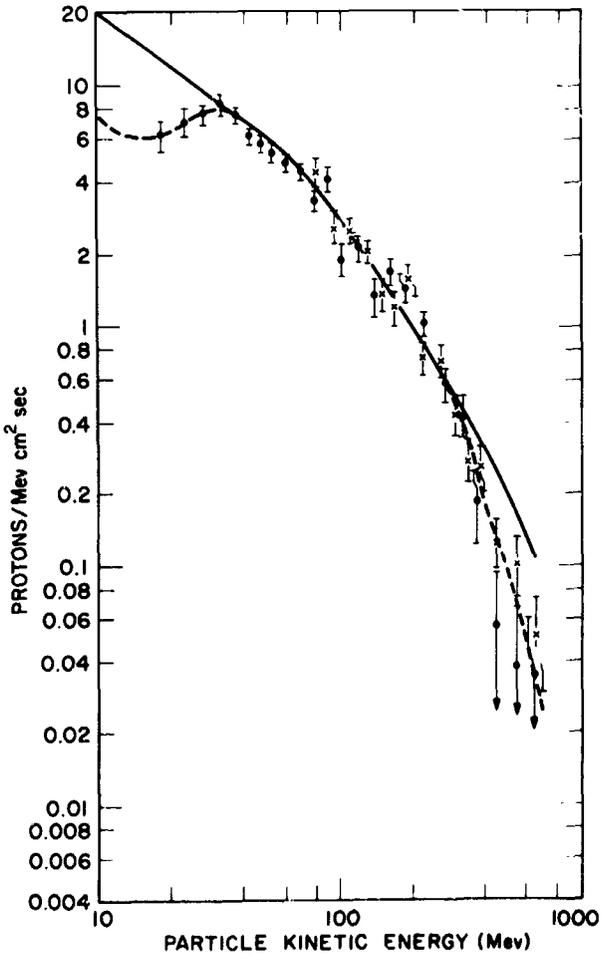


FIGURE 1.—The experimentally measured inner zone proton energy spectrum (Freden and White 1962) for $L\sim1.4B\sim.20$.

†This work was done as part of the requirements for a Master's Degree from the College of William and Mary.

importantly due to changes in the upper atmospheric density during the solar cycle. We will now make a quantitative discussion of the expected changes in proton populations for $L<1.6$.

The continuity equations for protons is usually written [Singer, 1958; Hess, 1959; Freden and White, 1960]

$$\frac{dN(E)}{dt}=+S(E)-L(E)+\frac{d}{dE}\left[N(E)\frac{dE}{dx}v\right]=0 \quad (1)$$

and the steady state population of protons $N(E)$ solved for. We are now interested in time variations so we will write the time dependent form of this equation

$$\frac{dN(E,t)}{dt}=+S(E,t)-L(E,t)+\frac{d}{dE}\left[N(E,t)\frac{dE}{dx}(t)v\right] \quad (2)$$

The source term $S(E, t)$ will use the neutron decay source strength of Hess for solar maximum $S(E)$ multiplied by a function of time $f(t)$ to consider the solar cycle variation in galactic cosmic ray flux as determined by McDonald and Webber [1961]

$$S(Et)=S(E)f(t)=\frac{0.8E^{-2\cdot0}}{\nu\gamma T_n}\frac{re^3}{r}F(t) \quad (3)$$

where $f(t)$ varies from 1.0 at solar max to 1.25 at solar min in a similar way to Fig. 2.

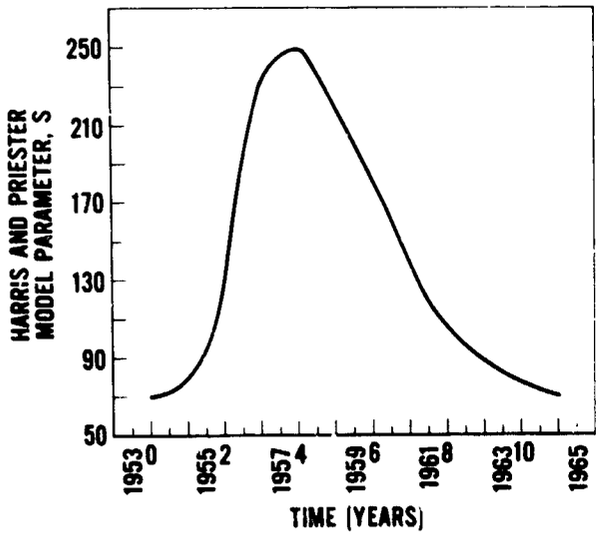


FIGURE 2.—Solar cycle used in these calculations S is the Harris and Priester (1962) model parameter which is related to but not the same as the 10 cm flux intensity.

For the loss term $L(E, t)$ following Freden and White [1960] we consider nuclear collisions

$$L(E,t)=N(E,t)v\Sigma_i\sigma_i\bar{\rho}_i(t) \quad (4)$$

where $\bar{\rho}_i(t')$ is the average atmospheric density of atmospheric component i (O_2, O, N_2 or He) and σ_i is the inelastic cross section for collisions (assumed geometric).

The last term in equation 2, the change in proton population due to slowing down, varies with the solar cycle due to the atmospheric density change

$$\frac{dE}{dx}=\left[\frac{\rho(t)}{2.69\times10^{19}}\right]\times\frac{dE}{dx}\Big|_{NTP} \quad (5)$$

where $\bar{\rho}$ (t) is the average atmospheric density of equivalent oxygen atoms. The variation of $\bar{\rho}_i(t)$ and $\bar{\rho}(t)$ with time in the solar cycle is the major reason for the change in proton population.

AVERAGE ATMOSPHERIC DENSITIES

A major part of this problem was to determine the average atmospheric densities used in calculating the loss rate of trapped protons by coulomb scattering. To do this the theoretical model of the upper atmosphere developed by Harris and Priester [1962] was used. This model gives the time dependences of the atmosphere for both solar cycle and diurnal variations. It agrees well with several measurements of density by satellite drag (King-Hele, 1963; Bryant, to be published) and recently has been checked by preliminary data of densities from the Explorer XVII satellite. Spencer et al [1963] This model atmosphere is the most complete description of the time dependence of atmospheric densities available and agrees well with the current experimental data. We must perform several operations on the H and P model data to get it in form to use for this problem. Harris and Priester give the atom densities n_i of the several atmospheric constituents i in the form of.

$$n_i(h, t, S)$$

where h is altitude above the earth, t is local time, and S is a model parameter related to, but not the same as, the intensity of the average 10 cm solar flux, F, in watts/m^2-cps$\times10^{-22}$. Recent studies of atmospheric densities [Harris and

Priester, 1963] show that the model number S is the same as $\bar{F}$ near solar max but near solar min $S<\bar{F}$. For example, $\bar{F}=70$ corresponds to the model $S=100$. Fig. 2 shows the solar cycle variation of S we have used based on the last solar cycle.

The first operation on the H and P atmosphere is to average over local time. The protons we will consider live long enough so that they will encounter the midday density bulge and the nighttime minimum many times and will average them out.

Secondly, we perform a longitude average. As the particles drift around the earth their mirror points do not stay at constant altitude but rather follow a certain path in B–L space. Several such constant B–L rings are shown in altitude-longitude coordinates in Fig. 3. The particles dip

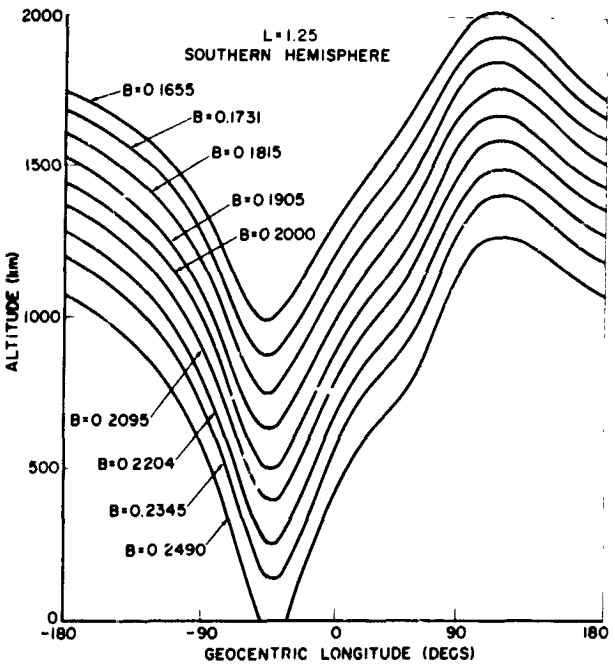


FIGURE 3.—Variation of the altitude of B-L rings with longitude.

lowest in the South Atlantic due to the nature of the earth's magnetic field. A B–L map of minimum altitudes is shown in Fig. 4. Due to the variation in altitude of the protons' mirror point, longitude averages were made using the B–L rings every ten degrees in both northern and southern hemispheres. This step gives the average mirror point density for a proton's motion. Walt has recently shown that the protons do not

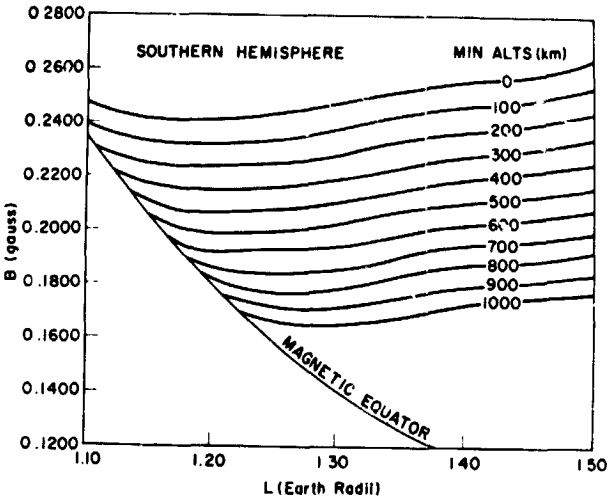


FIGURE 4.—B–L map of minimum altitudes.

drift in longitude at a constant rate because of the variation of the magnetic field gradient and field line curvature. This effect is ignored here. It probably amounts to about a 20 percent correction on average densities. Values of atmospheric densities are not known that well now anyway.

The third operation on the H and P atmosphere is to average over a bounce from one mirror point to the conjugate mirror point. Since the protons live much longer than a bounce period, and since no change in direction of the proton is considered during slowing, it is permissible to average the density in this fashion. A dipole field is assumed in carrying out this step.

Ray [1960] gave the bounce averaging process as

$$\rho=\frac{\int\rho(\lambda)ds}{\int ds} \tag{6}$$

where the element of length along the particle's orbit is

$$ds=\frac{vdt}{\cos\alpha}=\frac{d\ell}{\cos\alpha} \tag{7}$$

where $d\ell$ is the element of length along the field line. Using

$$d\ell^2=dr^2+r^2d\lambda^2 \tag{8}$$

and

$$r=L\cos^2\lambda \tag{9}$$

with the mirror equation and the expression for the magnetic field variation along a field line

combining and substituting in equation 6 gives

$$\bar{\rho}_i = \frac{\int \rho_i(\lambda) A(\lambda) d\lambda}{\int A(\lambda) d\lambda} \tag{10}$$

where

$$A(\lambda) = \frac{\cos^4 \lambda 1 + 3 \sin^2 \lambda}{[\cos^6 \lambda (1+3 \sin^2 \lambda_m)^{\frac{1}{2}} - \cos^6 \lambda_m (1+3 \sin^2 \lambda)^{\frac{1}{2}}]^{\frac{1}{2}}} \tag{11}$$

The subscript m here corresponds to the particles' mirror point. This bounce averaged density has been integrated on a computer.

The last step is to combine the five i constituents N_2, O_2, O, He and H to give the averaged number of equivalent oxygen atoms

$$8\bar{\rho} = 14\bar{\rho}(N_2) + \bar{\rho}(O) + 16\bar{\rho}(O_2) + 2\bar{\rho}(He) + \bar{\rho}(H) \tag{12}$$

Values of $\bar{\rho}$ at different times in the solar cycle for $L = 1.25$ are shown in Figure 5.

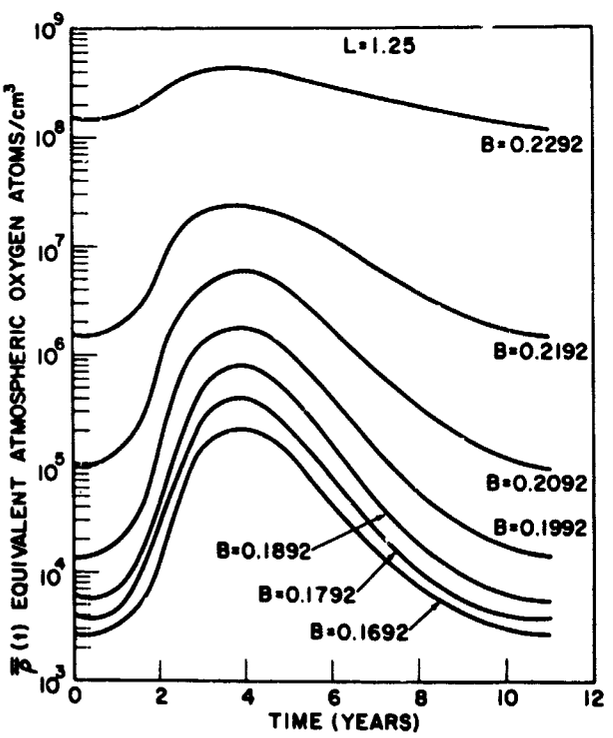


FIGURE 5.—Average atmospheric densities at different times of the solar cycle at different B–L points.

We can use the data on the rate of energy loss dE/dx for oxygen at NTP [Aaron, Hoffman and Williams, 1949] in equation 5 with this averaged density $\bar{\rho}$ to calculate the loss rate.

RESULTS

Equation (2) is integrated stepwise starting with $N(E) = O$ at $t = O$ to build up to an oscillating proton population which is the same from one

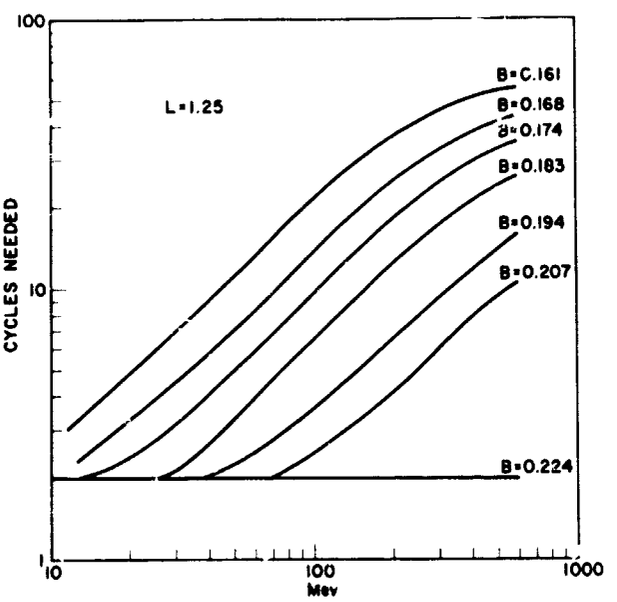


FIGURE 6.—The number of solar cycles needed for a group of protons to come within 1% of its final oscillating population for $L = 1.25$ starting with $N(E) = O$ at $t = O$.

solar cycle to the next. In Figure 6 is shown the number of solar cycles after which the population change per cycle is less than 1 percent for $L = 1.25$. After achieving this final condition the proton energy spectrum varies during one solar cycle as shown in Figures 7–11. The dotted curves top and bottom in these figures are what the proton spectrum would be if steady state conditions were achieved at solar max and solar min. Steady state clearly is not achieved for high energy protons or for high altitudes.

It is also interesting that this calculation predicts a spectral peak above 100 Mev for certain conditions (Figure 7, year 3) due to the time lag

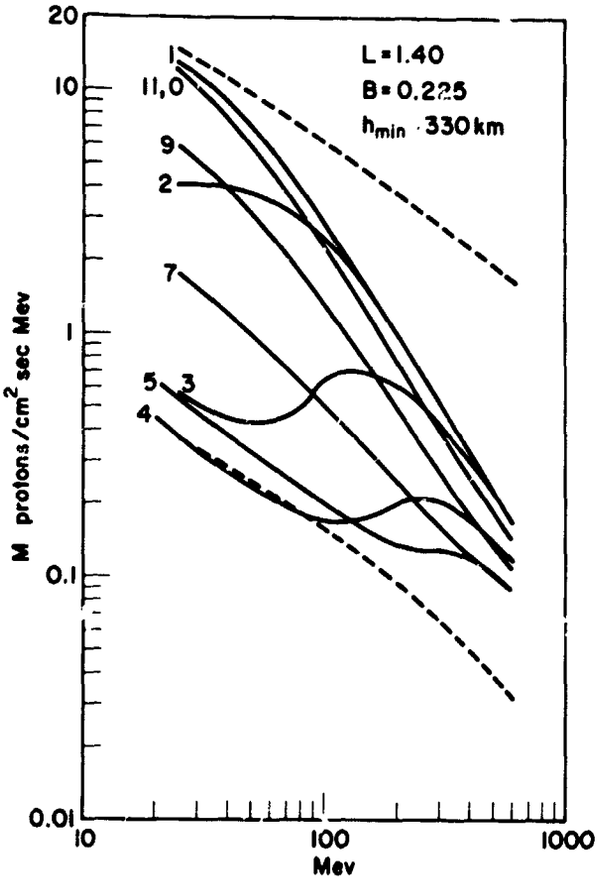


FIGURE 7.—Proton energy spectra at different times in the solar cycle for $L=1.40$, $B=.225$. The curves are labeled by the time in years from solar min.

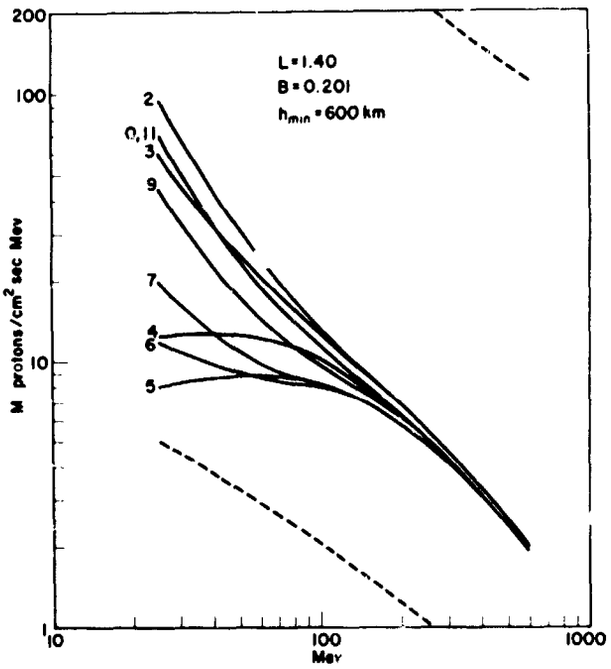


FIGURE 8.—Proton energy spectra at different times in the solar cycle for $L=1.40$, $B=.201$. The curves are labeled by the time in years from solar min.

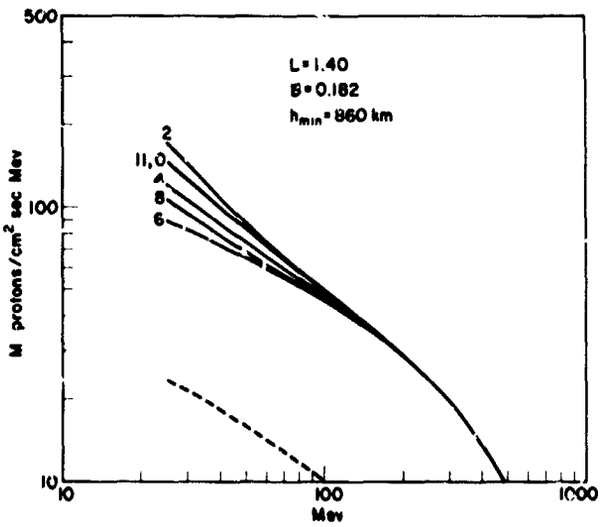


FIGURE 9.—Proton energy spectra at different times in the solar cycle for $L=1.40$ $B=.181$. The curves are labeled by the time in years from solar min.

in different energy protons adjusting to solar cycle changes.

Figures 12 and 13 show time histories for various conditions. Large variations from solar max to solar min populations are seen in these figures. The amplitude of these changes for $L=1.25$ and 1.40 are shown in Figures 14 and 15.

A comparison of past experimental results and the calculations of this paper is shown in Table I. In discussing these experiments we can consider three types of information.

In comparing time changes of the proton fluxes two situations occur:

1. Some individual experiments run long enough (a year or more) so that they should see proton flux changes directly. From Table I, experiments (*c*) and (*d*) are of this time. Although in experiment (*d*) it is not one experiment on one satellite, it is many identical experiments on essentially identical satellites, and therefore this experiment falls in this category.

2. Different experiments on different satellites can be compared to provide data on proton flux changes. Data of this type is given in experiments (*a*), (*e*) and (*f*) in Table I.

The third type of data is given by:

3. Direct measurement of the proton energy

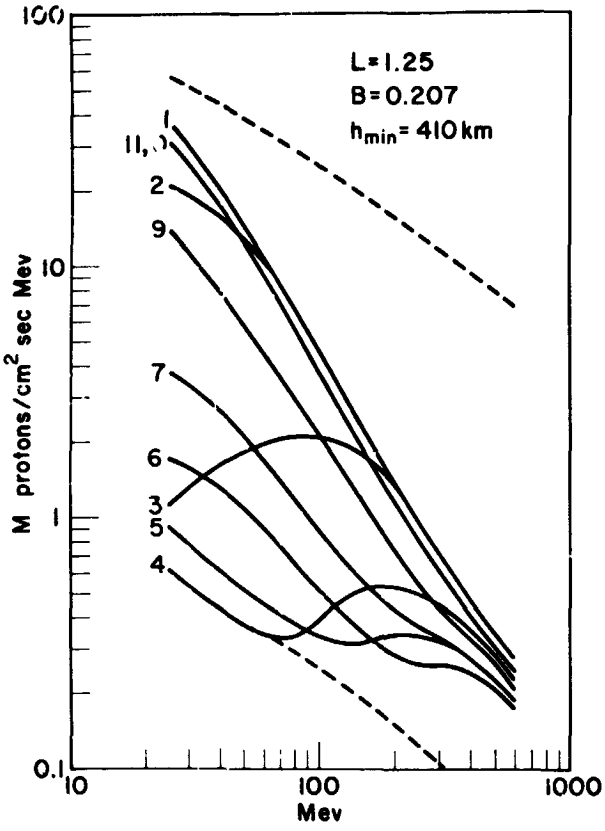


FIGURE 10.—Proton energy spectra at different times in the solar cycle for $L=1.25$, $B=.207$. The curves are labeled by the time in years from solar min.

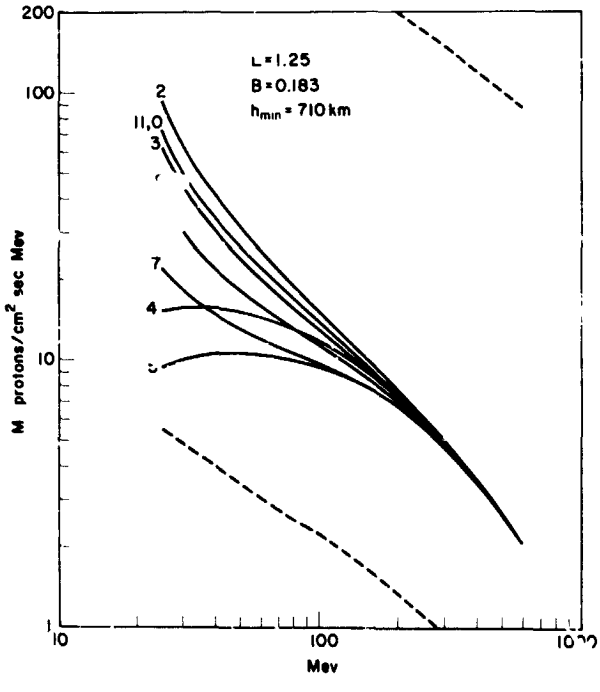


FIGURE 11.—Proton energy spectra at different times in the solar cycle for $L=1.25$, $B=.183$. The curves are labeled by the time in years from solar min.

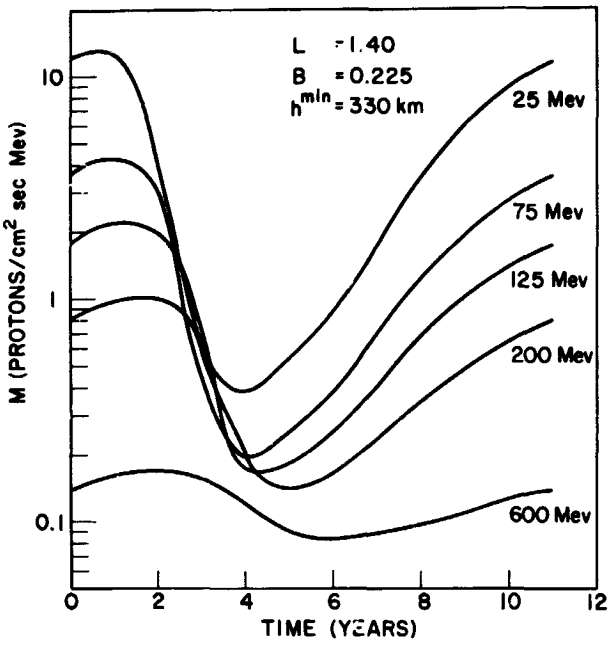


FIGURE 12.—Proton time histories for $L=1.40$, $B=.225$.

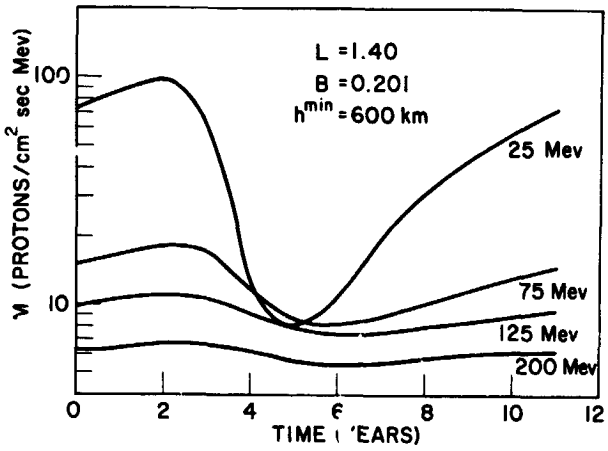


FIGURE 13.—Proton time histories for $L=1.40$, $B=.201$.

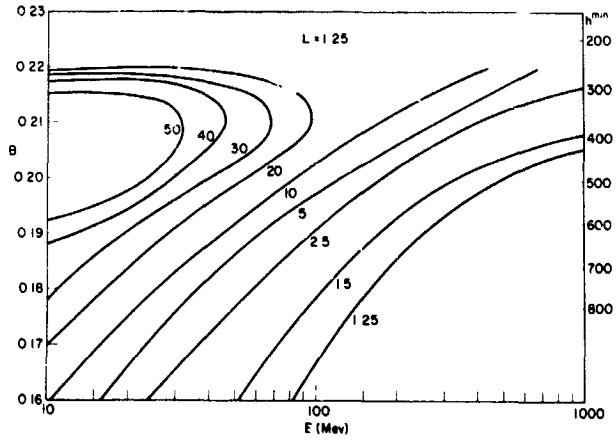


FIGURE 14.—The amplitude of the proton population change during the solar cycle for $L=1.25$.

TABLE 1.

Experiment	Date of Experiment	Time in Solar Angle	L	B	H Min.	Results of Experiment	Predictions from Present Paper
(a) Naugle and Kniffen (1963)—Emulsions on NERV	Sept., 1960	6.7	(a) 1.54	.209	610 km	Spectrum~flat from 20 to 60 MeV $\frac{Jo\ (31)\text{ at }6.7}{Jo\ (31)\text{ at }4.1\text{ (Exp. IV)}} \sim k$	Spectrum should fall about a factor of 3 in this energy range (Fig. 8. curve 7) $\frac{Jo\ (31)\text{ at }6.7}{Jo\ (31)\text{ at }4} = 1.3$ (Fig. 8)
			(b) 1.47	.223	400 km	Spectrum~flat from 10 to 100 MeV $\frac{Jo\ (31)\text{ at }6.7}{Jo\ (31)\text{ at }4.1\text{ (Exp. IV)}} \sim 2.5$	Spectrum should fall about a factor of 4 in this energy range (Fig. 7, curve 7) $\frac{Jo\ (31)\text{ at }6.7}{Jo\ (31)\text{ at }4} = 3.5$ (Fig. 7)
(b) Freden and White (1962)—Emulsions on Atlas	1960	6.5	~1.40	~.202	600 km	Spectrum peaks at ~ 40 MeV $\frac{Jo\ (20)}{Jo\ (40)} \sim .75$	No spectral peak at t=6.5 (Fig. 8) $\frac{Jo\ (20)}{Jo\ (40)} \sim 2.0$ but at t=5 do have spectral peak $\frac{Jo\ (20)}{Jo\ (40)} = 0.9$
(c) Pizzella, McIlwain and Van Allen—GM counter on Exp. VII	Oct. 1959 to Dec. 1960	5.8 to 7.0	1.40	.22	390 km	$\frac{Jo\ (E>18)\text{ at }7.0}{Jo\ (E>18)\text{ at }6.0} \sim 2.3$	$\frac{Jo\ (E=25)\text{ at }7.0}{Jo\ (E=25)\text{ at }6.0} = 2.0$ (Fig. 12)
(d) Filz and Yagoda—Emulsions on Discoverer	Dec. 1960 to June 1962	7.0 to 8.5	vary-ing	vary-ing	400 km	$\frac{Jo\ (E>55)\text{ at }8.5}{Jo\ (E>55)\text{ at }7.0} \sim 2.0$	$\frac{Jo\ (E=75)\text{ at }8.5}{Jo\ (E=75)\text{ at }7.0} \sim 2.3$ (Fig. 12)
(e) Heckman and Nakano—Emulsions on Low altitude polar satellites	Sept. 1962 to Sept. 1963	8.7 to 9.7	vary-ing	vary-ing	400 km	$\frac{Jo\ (65)\text{ at }9.5}{Jo\ (65)\text{ at }9.0} = 1.0 \pm .10$ $\frac{Jo\ (65)\text{ at }9.5}{Jo\ (65)\text{ at }4.1\text{ (Exp. IV)}} \sim 3$	$\frac{Jo\ (65)\text{ at }9.5}{Jo\ (65)\text{ at }9.0} = 1.2$ $\frac{Jo\ (65)\text{ at }9.5}{Jo\ (65)\text{ at }4.1} = 12$
(f) Freden and Paulikos—Solid state counters on Polar satellites	Sept. and Oct., 1962	8.8	1.30	.200	530 km	$\frac{Jo\ (7)}{Jo\ (85)} \sim 13$	
			1.47	.223	400 km	$\frac{Jo\ (7)}{Jo\ (85)} \sim 23$	$\frac{Jo\ (10)}{Jo\ (85)} \sim 11$ (Fig. 7)
			1.54	.209	610 km	$\frac{Jo\ (7)}{Jo\ (85)} \sim 43$	$\frac{Jo\ (10)}{Jo\ (85)} \sim 24$ (Fig. 8)
			low	high		$\frac{J\ (E>31)\text{ at }8.8}{J\ (E>31)\text{ at }4.1\text{ (Exp. IV)}} \sim 3$	$\frac{J\ (E=50)\text{ at }8.8}{J\ (E=50)\text{ at }4.1} = 13$
(g) Rowland Bakke, Imhof, Smith (1963) scintillators on Atlas pods	March, 1963	9.2	1.27	.216	340 km	no maximum at 40 MeV $\frac{Jo\ (20)}{Jo\ (40)} \sim 2.0$	$\frac{Jo\ (20)}{Jo\ (40)} = 2.2$ (Fig. 10)
						$\frac{Jo\ (10)}{Jo\ (60)} \sim 8.0$	$\frac{Jo\ (10)}{Jo\ (60)} = 9.0$

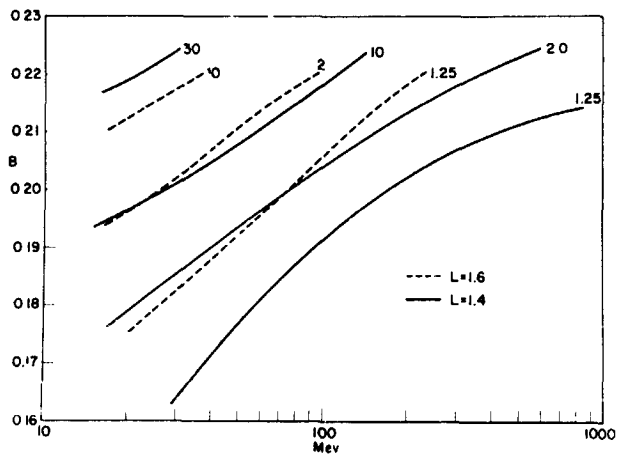


FIGURE 15.—The amplitude of the proton population change during the solar cycle for $L=1.4$ and 1.6.

spectrum in one experiment. Experiments (*a*), (*b*), (*e*), (*f*) and (*g*) are of this type.

In comparing flux time change experiments of type (1) with the calculations, the agreement is good. Two experiments covering the period 1959 to 1962 give results consistent with the calculations.

Comparing with type (2) data does not work as well. The observed changes of flux with time agree qualitatively with those predicted but not quantitatively. The experimental flux ratios are smaller than the calculated ones. Two remarks are in order here. First, type (2) data usually has larger errors attached to it than type (1) because it involves the systematic errors of two experiments while type (1) data involves no systematic errors—only statistical errors are involved in the time variations in the flux. Because of this the experimental flux ratios using type (2) data are probably not as accurate as those using type (1) data. Secondly all the comparisons of type (2) data involve Explorer IV data so there is no independent check of experimental consistency.

We may have over-estimated the amplitude of the change of proton populations in this present calculation by as much as a factor of two because our solar cycle model uses $S=70$ at solar minimum while recent data of Harris and Priester (1963) shows that $S=100$ is probably more appropriate. Changing the shape of the model solar cycle (Figure 2) will also affect the results of the calculation. The solar cycle must rise more rapidly towards solar max. than it falls to solar min. in order to produce the changes in spectrum calculated here. Solar cycle 16 was quite symmetrical and should not give the calculated results.

Of the five experiments of type (3) that help understand the proton energy spectrum the first two (*a*) and (*b*) in 1960 showed a modest sized peak at about $E=40$ Mev. The more recent experiment (*g*) of Rowland et al in 1962 shown in Figure 16 does not show such a peak. The solid curve on Figure 16 compared with the data of Rowland is the results of the present calculations normalized. The agreement is fairly good. Recent experiments (*f*) (*g*) show a large low energy proton population but cannot be compared with early experiments because the early ones did not go down to such low energies. Figures 7–11 show that the calculated spectral shape should not change much during the period of these measurements—from 6 to 9 years solar cycle time The proton fluxes should increase but the shape of the spectrum stays nearly the same. Striking changes in spectral shape should occur on the upswing of the solar cycle. The agreement with spectral shape in Figure 16 is not bad but the peaks in experiments (*a*) and (*b*) should not occur at $t=6$. It may be that a neutron absorption as suggested

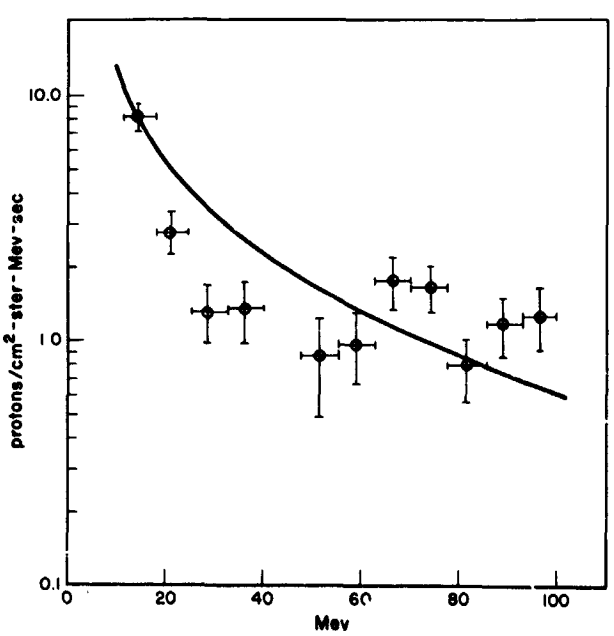


FIGURE 16.—The energy spectrum of inner zone protons measured by Rowland et al, 1963, on an Atlas pod at $L=1.27$, $B=.216$ shown for comparison is the spectrum calculated in this paper for $L=1.27$, $B=.207$, $t=9$.

by Freden and White (1962) is required to explain the peak.

All of the information presented so far on Figures 6 through 15 are mirror point fluxes M, that is, the flux of protons mirroring per unit volume at one particular B and L. For low altitudes where the atmosphere changes rapidly with altitude it is nearly correct to compare this flux with measured omnidirectional flux values. For these low altitudes the omnidirectional flux is very nearly the same as the mirror point flux. To show a more complete picture of the solar cycle proton changes we have converted to omnidirectional fluxes using equations (24) of Hess and

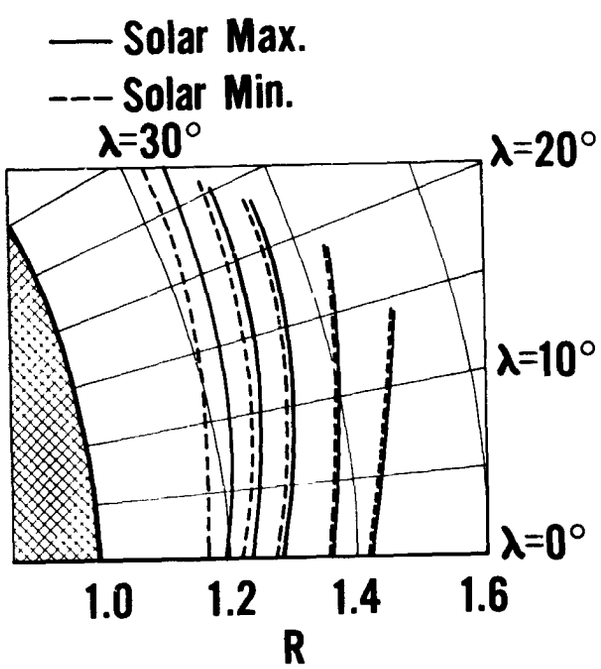


FIGURE 17.—An $R-\lambda$ map of the omnidirectional flux of $E=25$ Mev protons at solar max and solar min.

Killeen (1961). Figure 17 shows a calculated R–λ map of the 25 Mev proton omnidirectional fluxes J at solar max and solar min. An isoflux contour is clearly at lower R for solar minimum.

In conclusion we have shown that large changes in proton fluxes will take place during the solar cycle for those regions of B–L space corresponding roughly to minimum altitudes of about 300–700 km. Changes in proton spectra will occur also. Comparison of the calculations with available experimental information are not conclusive. Some kinds of experimental data agree quantitatively with the calculations. Other data agrees qualitatively but not quantitatively. The crucial tests of the calculations will involve measuring changes in the proton flux and spectrum as we approach solar maximum during 1966–68. We know experimentally (King-Hele 1963) that the atmospheric density changes used in the calculation are reasonable These calculated effects must take place with about the magnitudes shown here unless there are features of the inner belt protons which we do not now understand.

REFERENCES

1. ARMSTRONG, A. H., F. B. HARRISON, H. H. HECKMAN, and L. ROSEN, Charged particles in the inner Van Allen radiation belt, *J. Geophys. Research*, 66 351–8 (1961).
2. ARON, W. A., B. S. HOFFMAN, F. C. WILLIAMS, "Range-Energy Curves" (2nd Rev. 1949) U. S. A.E.C., Univ of Calif. Rad. Lab.
3. BRYANT, R., Densities Obtained from Drag on the Explorer XVII Satellite, to be published.
4. FILZ, R., and H. YAGODA "Observations on Trapped Protons in Emulsions Recovered from Satellite Orbits" to be published in Space Research IV and private communication.
5. FREDEN, S. C., R. S. WHITE, Protons in the Earth's Magnetic Field Phys. Rev. Letters **3** 9–11 (1959).
6. FREDEN, S. C., R. S. White, "Particle Fluxes i the Inner Radiation Belt" *J. of Geophys. Res.*, vol. 65, no. 5, May 1960.
7. FREDEN, S. C., R. S. WHITE, "Trapped Proton and Cosmic-Ray Albedo Neutron Fluxes", J. of Geophys. Res., vol. 67, no. 1, January 1962, p. 25.
8. FREDEN, S. C., and G. A. PAULIKAS, "Trapped Protons at Low Altitudes in the South Atlantic Magnetic Anomoly" Tech. Doc. SSD-TDR-63-342, Aerospace Corporation, Inglewood, Calif.
9. HARRIS, I., and W. PRIESTER, "Theoretical Models for the Solar-Cycle Variation of the Upper Atmosphere" X-640-62-70, June 1962, and J. A. S. **19** 286 (1962), Time-Dependent Structure of the Upper Atmosphere.
10. HARRIS, I., and W. PRIESTER, Relation between Theoretical and Observational Models of the Upper Atmosphere, J. of Geophysical Res. **68** 5891, 1963.
11. HECKMAN, H. and G. H. NAKANO "East-West Assymmetry in the Flux of Mirroring Geomagnetically Trapped Protons" J.G.R. **68** 2117, 1963 and private communications.
12. HESS, W. N., "Discussion of paper by Pizzella, McIlwain and Van Allen" J. of Geophys. Res., **67** 4886 (1962) and "Lifetime and Time Histories of Trapped Radiation Belt Particles" Space Research Vol. III, to be published.
13. HESS, W. N. and J. KILLEEN "Special Distribution of Electrons from Neutron Decay in the Outer Radiation Belt" J. G. R. **66** 3671 (1961).

14. Hess, W. N., "Van Allen Belt Protons from Cosmic-Ray Neutron Leakage" Phys. Rev. Letter **3** 11 (1959).
15. King-Hele, D. G., "Decrease in Upper-Atmosphere Density since the Sunspot Maximum of 1957–58" Nature **198** 832 (1963).
16. Rowland, J. H., J. C. Bakke, W. L. Imhof, and R. V. Smith, "Radiation Environment Experiment" Tech. Report SSD-TDR-63-194, Lockheed Missiles and Space Co., Palo Alto, California.
17. Lencheck, A. M., J. of Geophys. Res., **67** 2145 (1962).
18. Lingenfelter, R. E., to be published.
19. McDonald, F. B., and W. R. Webber, "A study of the rigidity and charge dependence of primary cosmic ray temporal variations" Journal of the Physical Society of Japan, vol. 17, Supplement A-II, Proceedings of the International Conference on Cosmic Rays and the Earth Storm, September 4–15, 1961.
20. Naugle, J. E., and D. A. Kniffen, Flux and Energy Spectra of the Protons in the Inner Van Allen Belt, Phys. Rev. Letter **7** 3 (1961) and "Variations of the Proton Energy Spectrum with Position in the Inner Van Allen Belt" J.G.R. **68** 4065 (1963).
21. Pizzella, G., C. E. McIlwain and J. A. Van Allen, Time Variations of Intensity of the Earth's Inner Radiation Zone, October 1959 through December J.G.R. **67** 1235, 1962.
22. Ray, Ernest C., "On the Theory of Protons Trapped in the Earth's Magnetic Field" J. of Geophys. Res., vol. **67**, no. 4, April 1960, pp. 1125–1133.
23. Singer, S. F., "Trapped Albedo Theory of the Radiation Belt" Phys. Rev. Letter **1** 181 (1958).
24. Spencer, N. W., L. H. Brace, C. R. Corignan, D. R. Taeusch and H. Niemann, "The Concentration and Temperature of Molecular Nitrogen and Electrons in the 120 to 350 km Region," L. H. Brace and N. W. Spencer, "Geophysical Implications of Explorer XVII Electrostatic Probe Measurements," Horowitz, R., and G. P. Newton, "First Direct In-Site Measurements of Atmospheric Density from Explorer XVII," Reber, C. C., "Preliminary Results Regarding Neutral Atmosphere Composition from the Explorer XVII Satellite," abstracts at the Boulder AGU Meeting, Dec. 1963, Transactions of AGU **44** 884 (1963).

N64-27353

A NEUTRAL-POINT EXPANSION OF THE IDEAL MAGNETOSPHERE*

RONALD BLUM†

The idealized model of the geomagnetic field–solar wind interaction yields a singular (neutral) point on the magnetopause at which the magnetic field vanishes. If we expand the fields in a power series around this point, including quadratic terms, we can derive an approximate equation of the magnetopause in a small neighborhood of the neutral point which is consistent with the idealized boundary conditions to fourth order. We then consider the additional pressure due to multiple reflections of particles in this neighborhood and show that less than a 4 per cent correction to the single-reflection pressure condition is necessary.

INTRODUCTION

In the idealized model of the geomagnetic field–solar wind interaction [*Beard*, 1960] we assume the field to be excluded from the plasma and contained in a cavity called the magnetosphere. The boundary of the cavity, or magnetopause, is unknown, but we stipulate the dynamic condition that the magnetic pressure just inside the magnetopause be exactly balanced by the kinetic pressure of solar wind particles elastically reflected from the surface (thermal effects are neglected). Figure 1 is a schematic drawing of the magnetosphere with the dipole located at the origin of coordinates and perpendicular to the direction of the plasma stream.

Mathematically, this situation is described by the field

$$\mathbf{H} = -\text{grad}\ \Omega \qquad \nabla^2\Omega = 0 \qquad (1)$$

inside the magnetosphere, and the boundary conditions

$$\mathbf{H}\cdot\text{grad}\ F = 0 \quad \text{(confinement)} \qquad (2)$$

$$\mathbf{H}\cdot\mathbf{H} = \beta^2 \cos^2\chi \quad \text{(pressure)}$$

which hold on the surface, $F(x', y', z') = 0$, where χ is the angle of incidence of the incoming plasma (Figure 1), and $\beta^2 = 8\pi(2nm_iV^2)$. V, n, and m are the plasma drift velocity, ion number density, and ion mass, respectively.

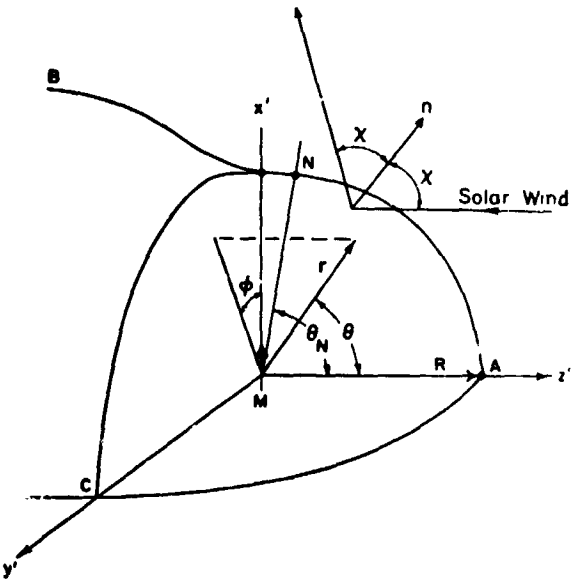


FIGURE 1.—Schematic drawing of the magnetosphere.

A singular point on this unknown surface is N, the neutral point, where the magnetic field lines 'split,' the field vanishes, $\chi = \pi/2$, and the gradient is in the x' direction. Although the problem is unsolved as yet, there are some good approximations to the general shape [*Mead and Beard*, 1963; *Spreiter and Briggs*, 1962; *Midgley and Davis*, 1963]. However, these approximations generally fail in the region of the neutral point, a region of interest since, if it does exist, it is likely to be

*Published in *Journal of Geophysical Research*, 69(9): 1765–1768, May 1, 1964.

†National Academy of Sciences—National Research Council Postdoctoral Research Associate with NASA.

unstable, and a possible point of entry for high-energy particles into the magnetosphere.

The neutral point considered here is an X-type neutral point [*Dungey*, 1958, pp. 39–41, 51–52, 98–102]; however, it lies *on* a bounding surface, which must run parallel to one of the limiting

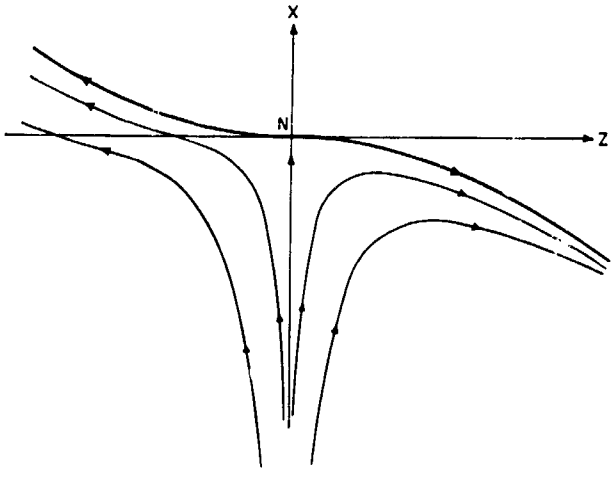


FIGURE 2.—Fields near the neutral point in the noon meridian plane.

field lines (Figure 2). We must assume that surface currents can be made to account for the disappearance of the field outside the surface (magnetopause).

SERIES REPRESENTATION

We can represent the field $\mathbf{H}=\beta\mathbf{h}$ by means of its scalar potential Ω. We first transform to the (x, y, z) coordinate system with origin at N such that

$$R(dx, dy, dz)=(dx', dy', dz') \tag{3}$$

In a region small compared to the apex radius R, we have $x^2+y^2+z^2\ll 1$. The scalar potential is expanded in the form

$$-\Omega/\beta=ax^2+by^2+cz^2+dx^3+ey^3+fz^3 +mxy^2+nxz^2+rzx^2+szy^2 \tag{4}$$

We must include cubic terms in the potential, since we expect the limiting field lines to be curvilinear in the xz plane. There are no linear terms, because the field must vanish at N; terms in y to the first power are deleted, because symmetry requires $H_y=0$ in the xz plane. The dependence on xz has also been omitted, since for $y=0=z$ there should only be an x component of H on the x axis (Figure 2).

The gradient of (4) yields the normalized field:

$$\begin{aligned} h_x&=2ax+3dx^2+my^2+nz^2+2rxz \\ h_y&=2by+3ey^2+2mxy+2szy \\ h_z&=2cz+3fz^2+2nxz+rx^2+sy^2 \end{aligned} \tag{5}$$

and the requirement $\nabla\cdot\mathbf{H}=0$ yields

$$\begin{aligned} &a+b+c=0 \quad 3d+m+n=0 \\ &3f+r+\cdots\ e=0 \end{aligned} \tag{6}$$

Furthermore, as Figure 2 shows, we expect $a<0$, $c>0$.

NOON MERIDIAN CONTOUR

In the xz plane the boundary conditions (equation 2) on the noon meridian contour become $(dx/dz=x')$:

$$h_x/h=-F_z/F_x=x' \tag{7}$$

$$h_x^2+h_z^2=x'^2/(1+x'^2) \tag{8}$$

If we substitute for x' in (7) and (8), then

$$h_x^4+h_x^2(2h_z^2-1)+h_z^4=0 \tag{9}$$

and, since both components must vanish at N,

$$h_x^2=(1-2h_z^2-(1-4h_z^2)^{1/2})/2\to h_z^4\to N \tag{10}$$

From Figure 2 we see that h_x is opposed in sign to the coordinate z on the noon meridian countour. Thus, sufficiently close to N,

$$h_x=\pm h_z^2 \text{ for } z\lesssim 0 \tag{11}$$

If we substitute into (7) and differentiate,

$$x''=\pm dh_z/dz \quad z\lesssim 0 \tag{12}$$

which leaves two alternatives: either x'' is discontinuous at N or $x''=0$ at N. The latter implies $c=0$, which is inconsistent with the geometry of the limiting field line in the xz plane. Thus the curve, $\{x=x(z),\ y=0\}$, must be represented as two separate power series for $z\lesssim 0$. Since (11) must be satisfied on the noon meridian contour, it gives an implicit representation of that contour near N. This must agree with $x=\int dz h_z/h_z$ of

(7) and with Figure 2. The leading term of (11) is

$$2ax+nz^2=\pm 4c^2z^2 \qquad (13)$$

If $n=0$, equation 13 implies that a is of positive sign, which contradicts Figure 2. Therefore we must assume that n has two different values, according to the sign of z. This is permissible as long as $\mathbf{h}$ remains continuous at $z=0$, and $\nabla\cdot\mathbf{h}=0$ everywhere. If we compare (13) with the leading term of (7), we find

$$x'=\pm h_z\lesssim 0 \qquad x=\pm cz^2 \qquad (14)$$

near N. Thus, (13) becomes

$$n=\pm 2c(2c-a) \text{ for } z\lesssim 0 \qquad (15)$$

SURFACE REPRESENTATION

Since the gradient is in the x direction at N, we can represent the surface by two second-order expansions in (y, z):

$$F(x,y,z)=x-A^{\pm}z^2-B^{\pm}y^2-C^{\pm}yz=0 \qquad (16)$$

for $z\lesssim 0$. By symmetry, $B^+=B^-=B$; (14) requires $A^+=-A^-=c$. If we form the dot product $H\cdot\nabla F=0$ and substitute for x form (16) we find the confinement condition satisfied to second order in z^2, y^2, and yz when

$$m=2B(2b-a) \qquad (17)$$

$$C^{\pm}(1-b/a-c/a)=0 \qquad (18)$$

The divergence conditions (equations 6) require that d also be discontinuous, since $3d+m+n=0$, although m is continuous. Therefore, from (17) and (15),

$$3d=\pm 2c(a-2c)+2B(a-2b)z\lesssim 0 \qquad (19)$$

Equation 18 is incompatible with $a+b+c=0$ unless $C^{\pm}=0$. Although (19) forces h_z to be discontinuous at $z=0$ ($a\neq 2c$), the 3 dx^2 term is negligibly small compared to the ax term in (5).

Third-order terms in $\mathbf{H}\cdot\nabla F=0$ are eliminated by simply setting $f=0=s$. Then, by the divergence condition, $r=0$ also. Thus in the power series representation, the confinement condition of (2) is satisfied up to fourth-order errors, on the postulated surface of (16). If we substitute the expansions for $\mathbf{h}$ and $F(x, y, z)$ into the pressure condition of (2), we find

$$h_x^2+h_y^2+h_z^2=\cos^2\chi$$
$$\doteq(\partial F/\partial z)^2(1-\text{2nd-order terms}) \qquad (20)$$

and expanding $\mathbf{b}$ yields

$$4c^2z^2+4b^2y^2+\text{4th-order terms}$$
$$=4c^2z^2(1-\text{2nd-order terms}) \qquad (21)$$

By setting $b=0$ we can reduce the error in the second boundary condition to fourth order. Our final expansion for the field in the neighborhood of N is given by:

$$-\Omega/\beta=cz^2-cx^2-(m\pm 6c^2)x^3/3$$
$$+mxy^2\pm 6c^2xz^2 \qquad (22)$$

and the surface of the magnetosphere near N is given by

$$F(x,y,z)=x-(\pm)cz^2-my^2/2c \qquad (23)$$

for $z\gtrsim 0$.

MULTIPLE REFLECTIONS AT N

Midgley and Davis [1963] have observed that the effect of multiple reflections of particles near N might seriously alter the pressure condition (equation 2) near the neutral point. In the second-order approximation, we can show that the shape of the surface is consistent with the pressure condition to within 4 per cent. This effect would be most pronounced in the xz plane; we shall calculate the added pressure δp at a point (x, z) on the noon meridian contour due to multiple reflections.

Figure 3 represents a particle incident on the point (x_0, z_0) on the contour $x=cz^2$, $z\lesssim 0$, $y=0$, with an angle of incidence χ_0. This particle, on first reflection, strikes the magnetopause again at (x, z) where

$$(x-x_0)/(z-z_0)=c(z+z_0)=\tan 2\chi_0 \qquad (24)$$

and

$$\cot\chi_0=-x'(z_0)=-2cz_0 \qquad (25)$$

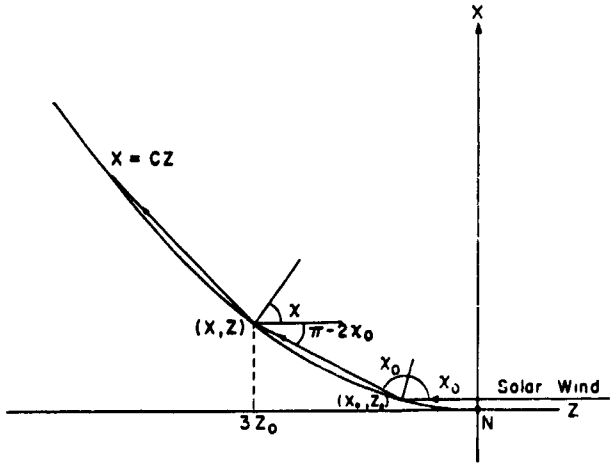


FIGURE 3.—Multiple reflections near the neutral point.

therefore,

$$z = 4cz_0/(1-4c^2z_0^2) - z_0 \doteq 3z_0 \qquad (26)$$

as indicated in Figure 3. We note that the angle of incidence is $\pi - 2\chi_0 + \chi$ at (x, z), and the change in the normal component of momentum will be correspondingly smaller. If we pursue the particle to the third point of reflection we find

$$c(z+3z_0) = \tan 2(\chi - \chi_0) \quad z \doteq 5z_0 \qquad (27)$$

However, the effect of the particle, by this time, is negligible.

We now consider the additional pressure δp at (x, z) due to particles reflected from (x_0, z) and compare it with the pressure $p = 2nmV^2 \cos^2 \chi$ of the incident plasma stream (ignoring the y coordinate). If $nV\, dx_0$ particles/sec are incident on an area of $ds_0 = (dx_0^3 + dz_0^2)^{1/2}$ at (x_0, z_0), they are reflected onto an area $ds = (dx^2 + dz^2)^{1/2}$ at (x, z), where their momentum changes by $2mV \cos(\pi - 2\chi_0 + \chi)$ per particle. Therefore, since $dx = 9dx_0$,

$$\delta p = -2mnV^2 \cos(\chi - 2\chi_0)(dx/ds)/9 \qquad (28)$$

If we substitute for χ, χ_0 from (25) and note that $dx/ds = \cos\chi$, then we find

$$\delta p/p = 1/27 \ll 1 \qquad (29)$$

We can conclude that the effect of multiple reflections near the neutral point is negligible and that the pressure condition is valid over the entire magnetosphere in this idealized representation.

REFERENCES

BEARD, D. B., The interaction of the terrestrial magnetic field with the solar corpuscular radiation, *J. Geophys. Res.*, *65*, 3559–3568, 1960.

DUNGEY, J. W., *Cosmic Electrodynamics*, 184 pp., Cambridge University Press, Cambridge, 1958.

MEAD, G. D., and D. B. BEARD, The shape of the geomagnetic-field solar-wind boundary, Goddard Space Flight Center rept., November 1963.

MIDGLEY, J. E., and L. DAVIS, JR., Calculation by a moment technique of the perturbation of the geomagnetic field by the solar wind, *J. Geophys. Res.*, *68*, 5111–5123, 1963.

SPREITER, J. R., and B. R. BRIGGS, Theoretical determination of the form of the boundary of the solar corpuscular stream produced by interaction with the magnetic dipole of the earth, *J. Geophys. Res.*, *67*, 37, 1962.

POSSIBLE SOURCE MECHANISM FOR LOW-ENERGY GALACTIC ELECTRONS

KARL A. BRUNSTEIN*

A calculation is made of the expected secondary electron flux resulting from the knock-on collisions of the primary nuclear beam with the interstellar gas. The model includes ionization losses and a statistical Fermi-mechanism energy gain. Comparison is made with recent satellite experimental data.

INTRODUCTION

Recent interest in cosmic-ray electrons has been confined largely to higher energies. Specifically, experimental results [1–3] in the energy region of the order of 100 MeV to several BeV have been of interest because of their bearing on the problem of galactic radio emission. The study of lower energy electrons, although probably not of direct importance to the radio emission question, is of importance because of its relationship to the higher energy electron spectrum, and because of its bearing upon the questions of solar modulation and energetic electron production.

Several workers in the field have arrived at the conclusion that the primary cosmic-ray beam must traverse several g/cm^2 of interstellar material prior to being sampled at or near the earth.[4–6] This necessarily implies a flux of low-energy electrons in equilibrium with the primary beam due to the knock-on process in the interstellar gas. This problem has been extensively studied for knock-on electrons due to muons in various substances.[7–9] The equilibrium problem in the interstellar gas is somewhat different from the laboratory experiments described in Refs. 7 and 8 because of the absence of the cascading process in the interstellar gas and the enhanced ionization loss rate in the partially ionized hydrogen.[10] In addition, there is the possibility of further acceleration of the secondary electrons in the inter-stellar material.[11]

It is not clear that these galactic electrons of low rigidity could penetrate into the solar cavity; however, recent work by Palmeira and Balasubrahmanyan [12] suggests that, at least during solar minimum, they can. This question is not considered here. The question of solar modulation is a separate one. By considering the knock-on flux as expected in the absence of solar influence and comparing with experimental data obtained outside the magnetosphere, new information concerning solar influence may be inferred.

PROCEDURE

A model is adopted in which the knock-on electrons, once produced, lose energy due to the ionization effect and gain energy due to a statistical Fermi mechanism. It is further assumed that the electrons tend to remain in the somewhat localized regions in which they are produced and that the losses due to diffusion out of the galaxy are negligible at these low rigidities. In addition, synchrotron losses are neglected at the energies in question here. Then, assuming a source of knock-on electrons and the predominance of the ionization loss and statistical gain mechanisms, a calculation of the low-energy electron spectrum is made.

Assuming a primary proton beam not varying appreciably with time, one can write the equation for the density of knock-on electrons as

$$\partial N(E,t)/\partial t+\alpha N(E,t)-(k-\alpha E)[\partial N(E,t)/\partial E] = Q(E)$$

*National Aeronautics and Space Agency—National Academy of Sciences—National Research Council Regular Post-Doctoral Resident Research Associate.

with $N(E,0)\equiv 0$, where

$N(E,t)$ = electron density at energy E and time t in electrons/m³-MeV,

$(dE/dt)_{\text{Fermi}}=\alpha(E+M_e c^2)$ defines α,

$k=|dE/ds|\rho c-\alpha M_e c^2$, dE/ds being the ionization loss rate, and

$Q(E)$ = production rate in electrons/m³-MeV-sec.

It is possible to solve the differential equation for arbitrary production rate $Q(E)$. The solution is found to be

$$N(E,t)=\sum_{n=0}^{\infty}\frac{(k-\alpha E)^n(1-e^{-\alpha t})^{n+1}}{(n+1)!\alpha^{n+1}}\frac{d^n}{dE^n}Q(E).$$

Adopting the Bhabha[9] cross section for knock-on production and the rigidity spectrum of McDonald and Webber [13] for the galactic proton beam, we may write for the production rate

$$Q(E)=\varphi(E)/E^2-\gamma(E)/E,$$

where

$$\varphi(E)=8\pi CM_e c^2 a\left(\frac{Mc^2}{Ze}\right)^{-1.25}\times\int_{(1-2M_e c^2/E)^{1/2}}^{\beta_{\max}}\beta^{-4.25}(1-\beta^2)^{-0.375}d\beta$$

$$\gamma(E)=4\pi Ca\left(\frac{Mc^2}{Ze}\right)^{-1.25}\times\int_{(1-2M_e c^2/E)^{1/2}}^{\beta_{\max}}\beta^{4.25}(1-\beta^2)^{0.625}d\beta,$$

$C=0.150$ cm²/g,

and

$a(Mc^2/Ze)^{-1.25}=5420$ (m²-sr-sec)$^{-1}$.

The dependence of φ and γ upon E makes this rigorous approach impractical. Instead, by use of the mean value theorem, one finds that $Q(E)$ may be approximated by

$$Q(E)\cong AE^{\delta},$$

where A and δ are readily evaluated. Substitution of $Q(E)=AE^{\delta}$ into the differential equation for $N(E,t)$ enables one to use the method of characteristics to solve the equation to yield

$$N(E,t)=\frac{1}{k-\alpha E}\left(\frac{A}{\delta+1}\right)\left[\frac{1}{\alpha^{\delta+1}}\left(k-\frac{k-\alpha E}{e^{\alpha t}}\right)^{\delta+1}-E^{\delta-1}\right].$$

Taking $\rho=2\times10^{-26}$ g/cm² [14] and letting $t\to\infty$, we get, setting $dJ/dE=(C/4\pi)N(E)$

$$\frac{dJ}{dE}=\frac{2.48\times10^{-13}}{1.625\alpha}\,\frac{E^{-1.625}-(k/\alpha)^{-1.625}}{(k/\alpha)-E}.$$

In addition to this flux calculated for the primary proton on hydrogen interaction, there will be a significant contribution from the heavier nuclei in the cosmic-ray beam. The knock-on production rate at a given primary velocity is very nearly a function of Z^2.[15] We then write the relation for the contribution of nuclei of change Z_i as

$$\left(\frac{dJ}{dE}\right)_i=Z_i{}^2\frac{J_i(>\beta)}{J_p(>\beta)}\left(\frac{dJ}{dE}\right)_p.$$

Using relative fluxes as given in the review by Ginzburg and Syrovatsky,[14] we arrive at the conclusion that the knock-on contribution from primaries of charge $Z\geq 2$ will be approximately 0.75 times the proton contribution. The total expected knock-on flux is then approximately 1.75 times the proton contribution.

The ionization loss rate for electrons of 3 to 15 MeV is nearly independent of energy for materials of low Z. In the interstellar hydrogen gas, however, it is fairly strongly a function of the degree of ionization. A degree of ionization of 10 per cent with a corresponding dE/ds value of 5 MeV/g/cm² has been taken.[10]

The calculated electron fluxes for different values of α are plotted in Figs. 1 and 2.[16] It is seen that the resultant intensity is a strong function of α, the parameter in the statistical acceleration mechanism. Typical electron fluxes as measured with IMP-A are also shown in the figures. It is seen that the range of α values selected, $1/\alpha=3\times10^{14}$ sec to $1/\alpha=3\times10^{15}$ sec, allows a fairly good matching of the theoretical and experimental fluxes. The value $\alpha\sim10^{-15}$ sec^{-1} does not appear unreasonable.[17,18]

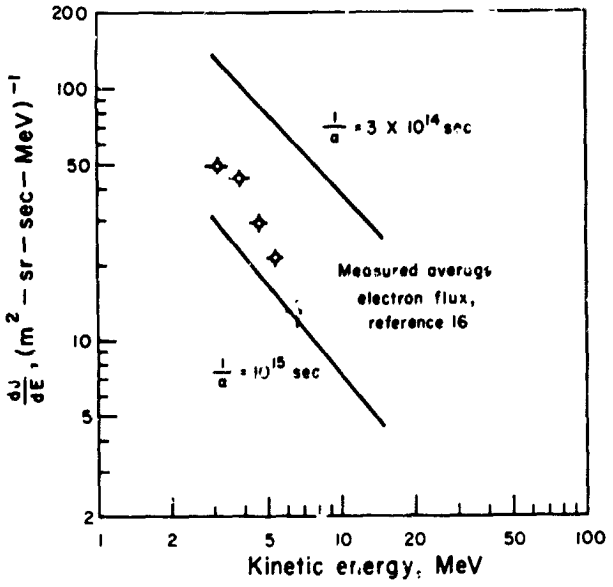


FIGURE 1.—Calculated spectra for values of α indicated. Circles: average electron flux from Ref. 16.

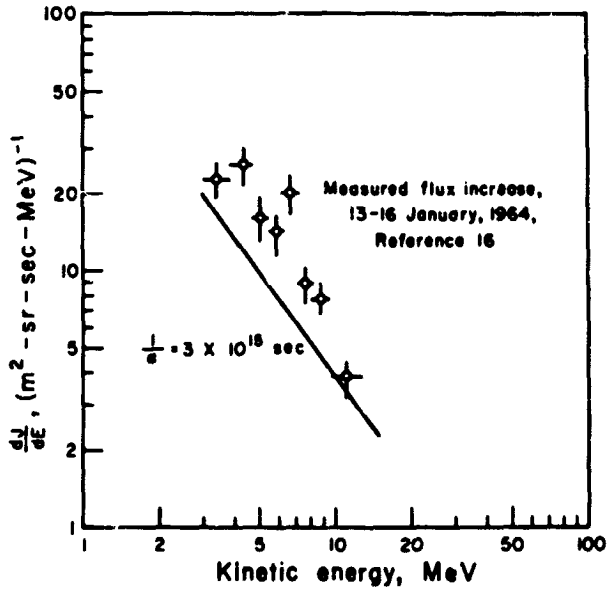


FIGURE 2.—Calculated spectrum for $1/\alpha = 3 \times 10^{15}$ sec. Circles: a typical flux increase, taken from Ref. 16.

It is not clear at this time whether the measured flux increase or the entire measured flux can be attributed to the knock-on process. Both possibilities are suggested by the reasonableness of the α values required.

CONCLUSION

The electron-positron flux resulting from the proton-proton interactions in the interstellar material has been discussed by several authors.[19,20] DeShong, Hildebrand, and Meyer[3] conclude, based on electron-positron ratios, that a substantial portion of the electron flux above 50 MeV must have an origin other than proton-proton collisions. It is speculated that a substantial portion of the lower energy electron flux seen in space may be attributed to knock-on electrons acted upon primarily by ionization losses in the interstellar gas and a Fermi type acceleration process. This, of course, requires that the low-energy electron flux should be composed largely of negative electrons. In addition, any long term solar modulation should be of an inverse activity dependence, similar to the primary nuclear beam. Both of these expectations will be subjected to experimental test by proposed experiments during the next solar half-cycle.

The knock-on process should produce secondary electrons in the BeV energy range also. The theoretical cross section in this case contains spin-dependent terms, and one does not feel as trusting of it as in the low-energy case where the interaction is one of Coulomb force only. In addition, these higher energy electrons may diffuse out of the galactic disk more readily and will also be subject to synchrotron losses. It is nevertheless interesting to plot the low-energy electron

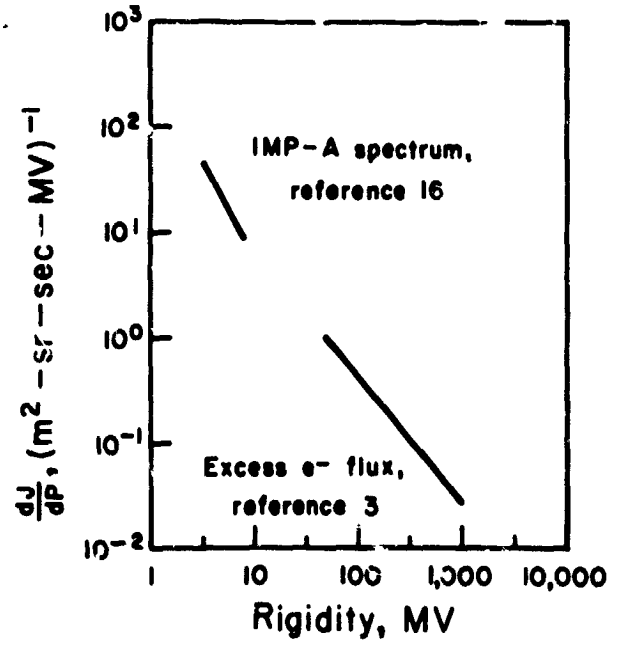


FIGURE 3.—The average electron flux from Ref. 16 shown along with the excess electron flux arrived at in Ref. 3.

flux along with the higher energy flux as has been done in Fig. 3. It is suggested by Fig. 3 that the knock-on process at higher energies may also be of significance.

Adopting for the moment the conclusion that the low-energy electrons as seen by IMP-A are due to the knock-on process, leads to the conclusion that the Fermi mechanism must be moderately effective for these low-energy electrons and that the parameter α has the value $\alpha \sim 10^{-15}$ sec^{-1}.

ACKNOWLEDGEMENT

I would like to express my appreciation for help given by Dr. Phillip Abraham in overcoming mathematical complexities. Dr. Thomas L. Cline kindly made his data available prior to publication.

REFERENCES

[1] J. A. Earl, Phys. Rev. Letters **6**, 125 (1961).

[2] P. Meyer and R. Vogt, Phys. Rev. Letters **6**, 193 (1961).

[3] J. A. DeShong, R. H. Hildebrand, and P. Meyer, Phys. Rev. Letters **12**, 3 (1964).

[4] F. W. O'Dell, M. M. Shapiro, and B. Stiller, International Conference on Cosmic Rays and the Earth Storm, Kyoto, 1961 (unpublished).

[5] S. Hayakawa, K. Ito, and Y. Terashima, Progr. Theoret. Phys. (Osaka) Suppl. **6**, (1958).

[6] H. Aizu, Y. Fujimoto, S. Hasegawa, M. Koshiba, I. Ito, J. Nishimura, and K. Yokoi, Progr. Theoret. Phys. Suppl. (Kyoto) **16**, 54 (1960).

[7] W. W. Brown, A. S. McKay, and E. D. Palmatier, Phys. Rev. **76**, 506 (1949).

[8] W. E. Hazen, Phys. Rev. **64**, 7 (1943).

[9] H. J. Bhabha, Proc. Roy. Soc. **A164**, 257 (1938).

[10] S. Hayakawa and K. Kitao, Progr. Theoret. Phys. (Kyoto) **16**, 139 (1956).

[11] E. Fermi, Phys. Rev. **75**, 1169 (1949).

[12] R. Palmeira and V. K. Balasubrahmanyan, J. Geophys. Res. (to be published).

[13] F. B. McDonald and W. R. Webber, Goddard Space Flight Center Contributions to 1961 Kyoto Conference on Cosmic Rays and the Earth Storm, Greenbelt, Maryland, 1961 (unpublished)

[14] V. L. Ginzburg and S. I. Syrovatsky, Progr. Theoret. Phys. (Kyoto) Suppl. **20**, 1 (1961).

[15] N. F. Mott, Proc. Roy. Soc. (London) **A124**, 425 (1929).

[16] T. L. Cline, F. B. McDonald, G. H. Ludwig, Phys. Rev. Letters (to be published).

[17] V. L. Ginzburg, *Progress in Elementary Particles and Cosmic Ray Physics* (North-Holland Publishing Company, Amsterdam, 1958), Vol. IV, Chap. V, p. 335.

[18] P. Morrison, S. Olbert, and B. Rossi, Phys. Rev. **94**, 440 (1954).

[19] S. Hayakawa and H. Okuda, Progr. Theoret. Phys. (Kyoto) **28**, 517 (1962).

[20] F. C. Jones, Journal Geophys. Res. **68**, 4399 (1963).

STUDIES OF SOLAR PROTONS WITH EXPLORERS XII AND XIV*

D. A. BRYANT†, T. L. CLINE, U. D. DESAI‡ AND F. B. McDONALD

Four solar proton events observed by Explorers XII and XIV in 1961 and 1962 are discussed. These events are directly associated with solar activity, and in three cases, are followed either by secondary events delayed about 2 days or by recurrent events on succeeding solar rotations. It is shown that the rate of propagation of solar protons in some of these primary events is linearly dependent on particle velocity and that this dependence makes it possible to separate the source characteristics from the propagation effects. In each event which shows this velocity dependence the propagation curves of all observed energies agree and reach maximum intensity at a time much greater than the rectilinear travel time. These results lead immediately to the conclusions that propagation involves an important degree of scattering and that the degree of scattering is independent of energy over the observed range of 1.4 to 500 mev. The energy spectrum at the time of escape from the sun, the "source spectrum," is determined in each event which shows this velocity dependence and is described by a power law in kinetic energy from a few mev to several hundreds of mev. A striking feature of all events is the existence of periodic fluctuations in intensity, which are simultaneous at all energies, having a period from 1. to 1.5 hours depending on the event. This result indicates a local origin, and it is suggested that the fluctuations reflect processes occurring in the region between the magnetosphere and the earth's shock front.

I. INTRODUCTION

During the last solar cycle it has become apparent that large solar flares are often accompanied by the acceleration of protons and other nuclei to energies sometimes exceeding tens of bev. Over a five-year period centered around the last solar maximum about sixty of these events were observed. Solar-proton events have been studied by many techniques. Ionization chambers, mu-meson telescopes, and neutron monitors at sea level provide indirect information at energies above a few bev. With high-altitude balloons, measurements down to about 80 mev are possible. Rocket measurements have no inherent low-energy threshold but are limited to brief samples of an event. Studies by radio techniques of the ionization produced in the D layer are sensitive to energies down to about 10 mev but have little energy resolution. Satellite studies can combine the advantages of the other techniques: they can provide almost continuous coverage with good energy and time resolution and with no inherent energy threshold, yielding a higher probability of detection of solar-proton events than can be provided by the other methods. Finally, a satellite with a sufficiently eccentric orbit can provide measurements of solar protons free of interaction with the earth's magnetic field and free of disturbances by the earth's trapped radiation.

The satellites Explorer XII and Explorer XIV carried cosmic-ray instruments with response for solar protons of energy as low as 1.4 mev. They were active from 16 August to 5 December 1961 and from 2 October 1962 to 1 August 1963, respectively. We discuss here several findings resulting from the study of the major solar proton events observed with these satellites during 1961 and 1962. Anomalous features of certain other events will be treated in later papers.

II. THE APPARATUS

Explorers XII and XIV were essentially identical satellites with similar orbits. The apogee of

*Published as *Goddard Space Flight Center Document X-611-64-217*, July 1964.

†Goddard Space Flight Center and D.S.I.R. Radio Research Station, Slough, Bucks, England.

‡Goddard Space Flight Center and Physical Research Laboratory, Navrangpura, Ahmedabad, India.

Explorer XII was about 13 earth radii and that of Explorer XIV about 16 earth radii; thus, for more than half the time in each orbit they were beyond the outer edge of the magnetosphere which varied from about 7 to over 12 earth radii. Gaps about 8 hours wide appear in the data as a result of discarding the observations made during the passes of the satellite through the magnetosphere. Various characteristics of the satellites are listed in Table I.

The detectors used for solar proton measurements in Explorers XII and XIV provided differential energy measurements from 1.4 mev to 500 mev and an integral measurement at about 600 mev. There were three detectors: a scintillation-counter telescope, a geiger-counter telescope and a single scintillation counter. The scintillation-counter telescope with a 32-channel pulse-height analyzer provided the differential-energy measurements for energies above 50 mev and the integral measurement at 600 mev. The geiger-counter telescope used in the single mode and in the coincidence mode gave integral measurements at 30 mev and 100 mev. Low-energy measurements were made with the single scintillation counter and an 8-channel integral pulse-height analyzer. The axis of the geiger-counter telescope was oriented parallel to the satellite's spin axis and the other detectors were normal to it. A summary of the dynamic range and resolution of each of the detectors is given in Table II. Corrections were applied to all the data for particles penetrating the detector shielding and for particles clipping the edges of the scintillators; only when these corrections were small were the data used for analysis. The detectors are described more fully in a previous publication (Bryant *et al.*, 1962).

III. THE EVENTS

In this paper we confine our attention to the solar proton events following the flares of 10 September, 28 September, and 10 November 1961 and the flare of 23 October 1962. Some details of these flares and the plage regions producing them are given in Table III. The table also contains comments on the flares (H. Prince and R. Hedeman, private communication, 1963) and on the accompanying type IV radio emission (A. Maxwell, J. Warwick, private communications, 1963). Some of the flares were followed by delayed effects and these are also listed. Effects delayed by the prompt 2-day sun-to-earth transit of enhanced plasma and effects delayed by the rotation of the sun (27-day recurrent events) are listed separately.

A calendar of all the solar particle events observed during the intervals discussed here is shown in Figure 1, in which the occurrence of

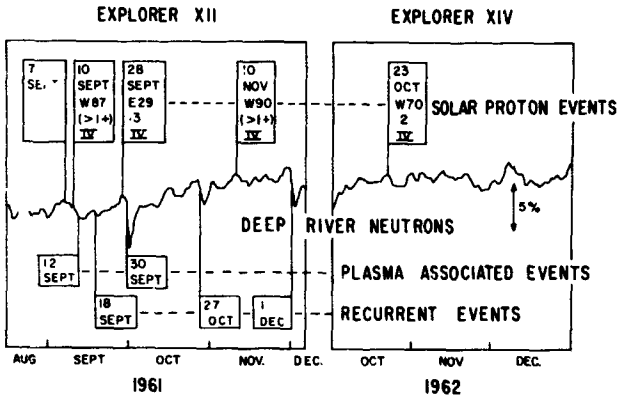


FIGURE 1.—Solar particle events observed with Explorers XII and XIV during 1961 and 1962. The occurrence of each primary solar proton event is indicated above the cosmic-ray neutron intensity plots with a flag which displays the solar longitude and classification of the parent flare and the existence of associated type IV radio bursts. Occurrences of both varieties of delayed proton events are also indicated; a plasma-associated event occurs with a delay of two days and a recurrent event takes place when a long-lived solar streamer passes central meridian on a succeeding solar rotation.

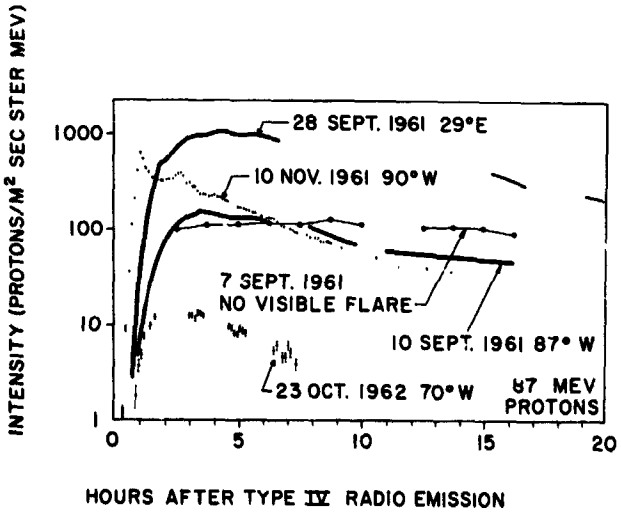


FIGURE 2.—The intensities of 87-mev protons vs time after the type IV emissions during the five primary solar proton events. The shapes of only two events, those of 28 September 1961 and 23 October 1962, are seen to be quantitatively similar, as monitored in this manner.

TABLE I.—*Satellite Characteristics*

Parameter	Explorer XI	Explorer XIV
Dates of Operation	16 Aug. to 5 Dec. 1961	2 Oct. 1962 to 1 Aug. 1963
Orbital Period	26.5 hours	36.5 hours
Apogee, Geocentric	83,600 km.	104,800 km.
Initial G. Perigee	6,700 km.	6,700 km.
Sun-apogee angle from earth	≈0° min. to ≈110° max.	≈70° to 180° and back to ≈0°
Initial Spin Period	2.2 seconds	5. seconds
Direction of Spin Axis	+47° right ascension −27° declination	(Uncertain due to precession)

TABLE II.—*Detector Characteristics*

Detector	Plastic Scintillator Telescope	Geiger Counter Telescope	CsI Crystal
Proton Energy Range	>55 mev	>30 mev	1.4 to 22 mev
Information Recorded	32 differential channels	Coincidence mode: / Single mode:	8 integral levels
Energy Intervals Processed	Energy Interval; Mean Energies 55 to 118 — 87 mev 118 to 150 — 135 mev 150 to 200 — 175 mev 200 to 255 — 228 mev 255 to 335 — 295 mev 335 to 500 — 418 mev (above 600 mev)	Energy Lower Limits ≈30 mev / ≈100 mev (sensitivity is a function of energy)	Mean Energies 2.2 mev 3.8 mev 5.7 mev 7.9 mev 14.5 mev
Geometric Factor (cm^2 ster)	13.4	150. / 12.7 (at 100 mev)	2.85
	(Corrections take into account variation of geometric factor with energy)		
Time Resolution	5 minutes of storage each 7 minutes	1.6 seconds of storage each mode during 5 out of 7 minutes	1.6 seconds of storage each level for 5 out of 7 minutes
Direction of Detector Axis	Normal to spin axis	Parallel to spin axis	Normal to spin axis

TABLE III.—*Details of the Flares*

Event	10 Sept. 1961	28 Sept. 1961	10 Nov. 1961	23 Oct. 1962
Plage				
Number	6212	6235	6264	6581
Rotation Order	2	3	4	2
Earlier, later	6197, 6235	6212, ----	----------	6563, ----
Flares Produced	69	15	9 after 5 Nov.	----------
Flare				
Class	>1+	3	>1+	2
Position	W87, N12	E29, N13	W90, N19	W70, N03
H α (UT):				
Start, Max., Stop	1950, 2020, 2054	2202, 2223, 2530	1434, 1444, 1450	1642, 1708, 1745
Type IV Onset (UT)	1937	2212	1445	1656
Comments	A very large, bright flare.	A very large, bright flare.	Accompanied by unusual loop activity.	Occurs where there are no spots.
Optical and Radio Observations (UT)	Loops at 2017, throughout 2025–2055	Type III bursts earlier on 27 Sept.	Limb prominence at 1430. Loops for many hours with max. at 1555	Type III G at 1649 and later.
	Considerable Type II	Type III G and Type II	Type III G and Type II	Type II
2-Day Delayed Effects				
Cosmic Ray Decreases	11 to 14 Sept.	30 Sept. to 5 Oct.	None	None
Low Energy Particles	12 Sept. max.	30 Sept. max.	None	None
Geomagnetic Activity	11 Sept., 1606 U.T.	30 Sept., 2108 U.T.	None	None
Recurrent Effects				
Cosmic Ray Decreases	None (central meridian passage on 28–30 Sept.)	28 Oct. to 1 Nov.	1 Dec. to 4 Dec.	None
Low Energy Particles	----------	27 Oct. max.	1 Dec. max.	None
Geomagnetic Activity	----------	26 Oct., 1940 U.T. 28 Oct., 0820 U.T.	1 Dec. to 4 Dec.	None

primary events and delayed effects are plotted against the cosmic-ray intensity measured by a sea-level neutron monitor (courtesy of H. Carmichael, private communication, 1962). Examination of the data showed no evidence either for any other smaller discrete events or for any separate, continuous increases over the general background level. The intensity of one energy interval of protons, centered at 87 mev, is plotted for all five primary events in Figure 2, indicating the widely varying behavior of the events. The event of 7 September 1961, which does not lend itself to obvious correlation to any known flare, will be discussed in a later paper.

IV. THE EVENT OF 28 SEPTEMBER 1961

The solar proton event of 28 September 1961 was initiated by a class 3 flare 29° east of central

meridian. Table III shows the pertinent solar and geophysical data. Some aspects of the event were considered in an earlier paper (Bryant *et al.*, 1962).

Intensity vs. time profiles of various energy components of the event are shown in Figure 3. Time is measured in hours from 2208 UT, the time of emission at the sun of a brief x-ray burst observed by Anderson and Winckler (1962) from 2216 to 2217 UT. For the first few hours points are plotted at intervals of about 7 minutes, and later points are hourly averages. The energy parameter is the mean energy of each interval as outlined in Table II. It is clear that the time taken to reach maximum intensity increases with decreasing energy. The data are less complete at the lower energies because at first the correction due to high-energy particles which penetrate the shielding of the low-energy particle detector is large, and because the satellite entered the trapped-radiation zone soon after the first effect became small.

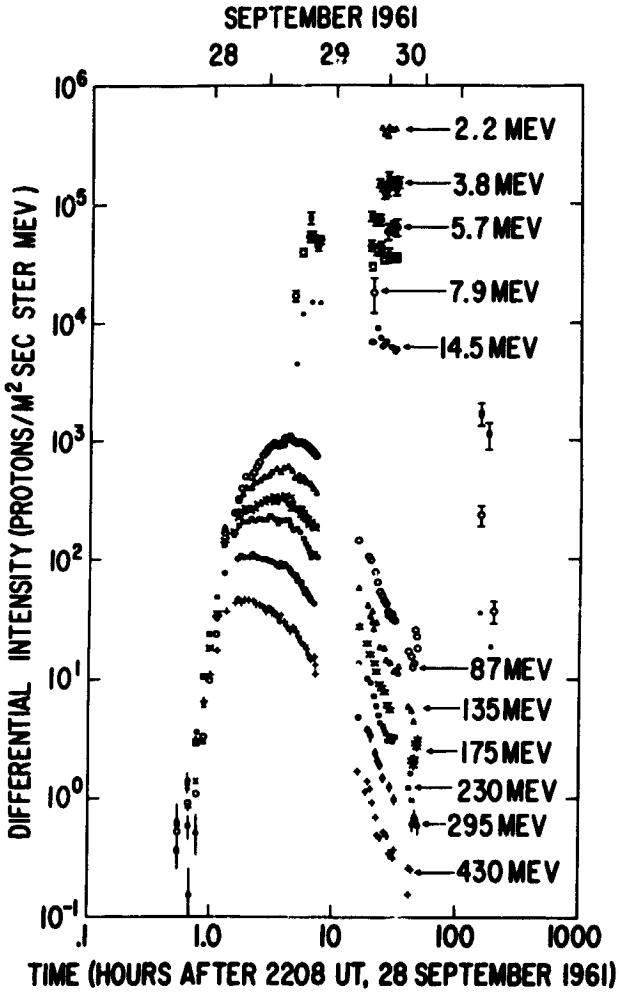


FIGURE 3.—The differential intensities of solar protons during the 28 September 1961 event plotted against time after the x-ray burst at the sun. The data are interrupted when the satellite passed through the magnetosphere and when the delayed increase occurred on 30 September 1961.

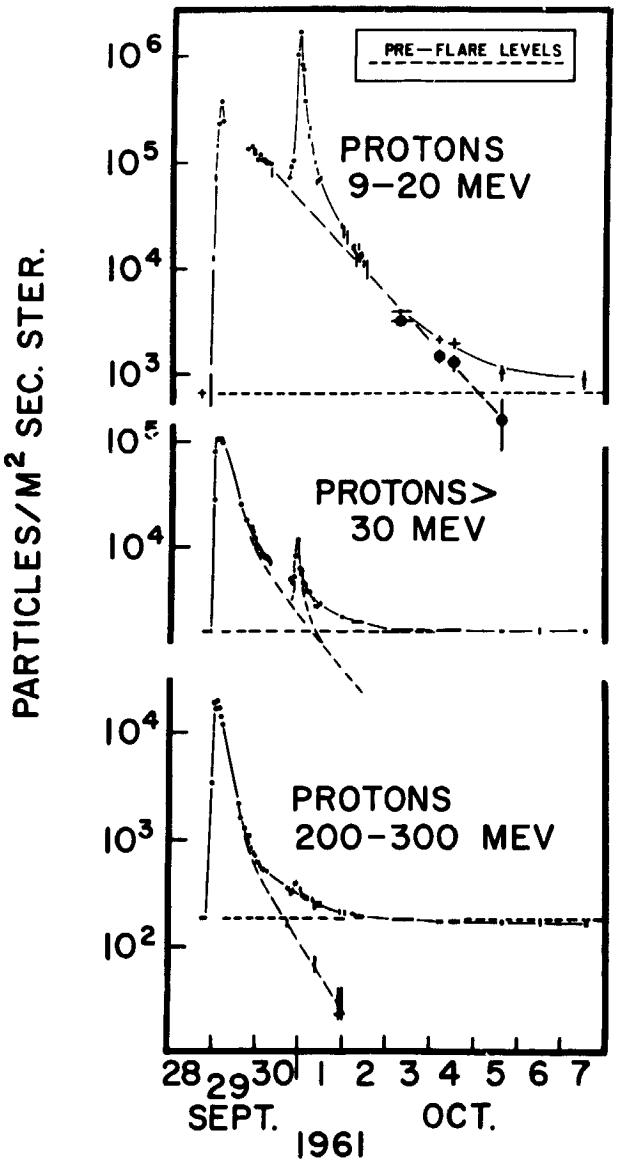


FIGURE 4.—Representative proton intensities between 28 September and 7 October, showing the delayed intensity increase of predominately lower-energy protons on 30 September 1961. The energy spectra of these particles are relatively constant with time, unlike those of the velocity-ordered primary solar proton event, and their arrival times are essentially constant with energy, occurring at the time of arrival of the enhanced solar plasma, two days after the flare.

There is a departure from a smooth decay at about 48 hours. At that time an increase took place in the intensity of low-energy particles which was associated with the arrival of a solar plasma stream that produced a sudden-commencement geomagnetic storm and a Forbush decrease. This occurrence is described in detail in the earlier paper, but for the sake of completeness it is briefly described here. The details of this increase, omitted in Figure 3, are shown for some energy intervals in Figure 4 on a linear time scale. Late on 30 September there is a sudden increase in intensity which is more marked at lower energies. The increase starts at about 1930 UT just before the sudden commencement of a magnetic storm at 2108 UT. The maximum intensity of particles of energy above 3 mev during the increase is more than ten times that attained during the main solar-proton event. Axford and Reid (1962) reported riometer observations of this burst of low-energy protons and had previously observed a similar event on 10 February 1958 (Reid and Axford, 1962). Other events of this kind have been reported by Charakhchyan *et al.* (1963).

After a full rotation of the sun, when the plage region responsible for the flare of 28 September was again close to central meridian, there was another burst of low-energy protons. This time the increase was much smaller. Figure 5, which is a plot of the total intensity above 3 mev on a further-compressed linear time scale, shows the full sequence of events. This recurrent event was also associated with enhanced solar plasma responsible for a recurrent geomagnetic storm and a recurrent Forbush decrease. We have put forward this event and a similar one following the 10 November 1961 flare as new evidence for the existence of long-lived solar streams (Bryant *et al.*, 1963).

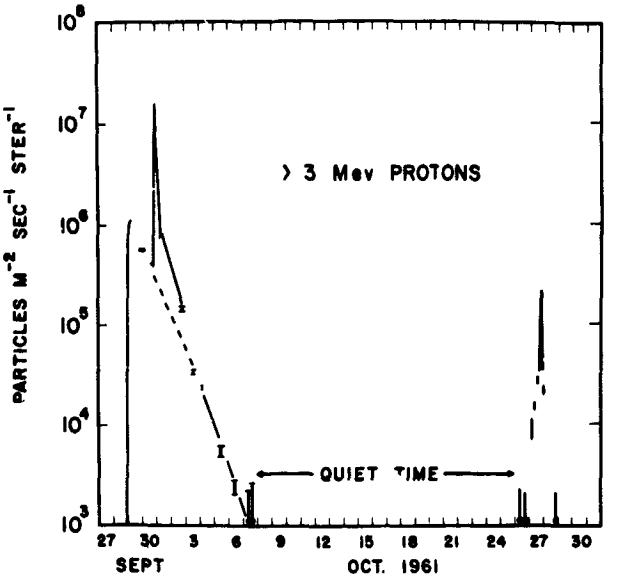


FIGURE 5.—The intensity of protons of energy above 3 mev between 30 September and 28 October 1961. The delayed increase on 30 September is superposed on the primary solar-proton intensity decay and the recurrent event on 27 October follows the completely event-free intervening period.

a) Velocity Dependence

The energy dependence of the rate of rise of intensity suggests that the propagation of the particles is a velocity-dependent process. A striking linear dependence on velocity is revealed by the following analysis.

We assume that all particles were accelerated at the same time or, more strictly, that they were all accelerated within a time interval short compared with the interval between acceleration and observation. We may then determine the distance a particle has traveled between acceleration and observation by taking the product of particle velocity and time from the beginning of the event. The time used here for the beginning of the event is 2208 UT, the time of the x-ray burst at the sun. We then convert the intensity vs. time profile to an intensity vs. distance profile. The intensity vs. distance profile is effectively a distribution in distance traveled. This distribution is a property of the medium through which the particles have traveled. Figure 6 shows the result of this treatment in which the distance traveled is measured in astronomical units. In constructing this figure the intensities of each component have been scaled to give the best fit to a common curve. The physical meaning of this normalization will be examined further below.

We note from Figure 6 that all components lie very closely on a common curve, apart from small-scale deviations to be discussed later as a separate topic. The fact that we have essentially a common curve shows that particles of all energies have traveled a given path length with equal probability; this is true for all path lengths to the extent that the various components of Figure 6 do

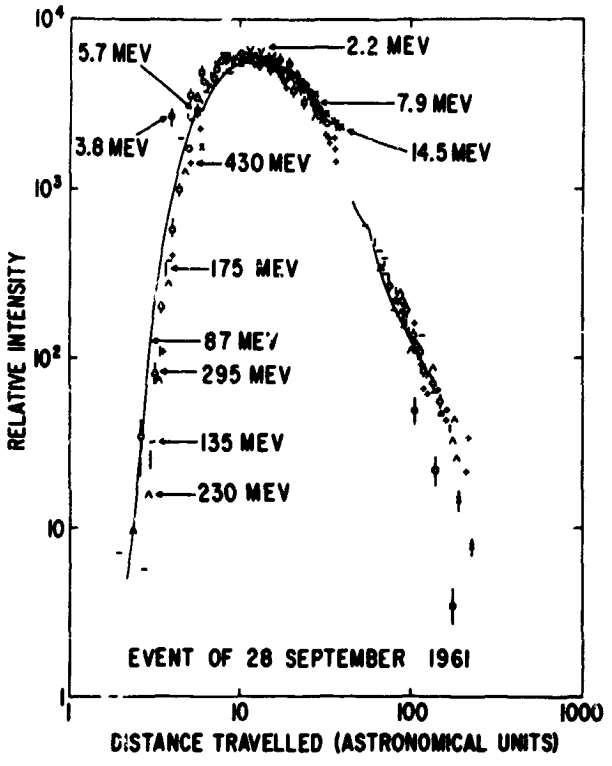


FIGURE 6.—The intensity vs time plots of Figure 3 converted to relative intensity vs distance plots. The distance is computed for each energy component by taking the product of the corresponding particle velocity and time from the event; the intensities are scaled to give the best fit to a common propagation curve. This fit occurs over a dynamic range in energy of a few hundred and a velocity range of 14, and over a time duration of several days.

lie on a common curve. We note that the distance traveled by most particles is many astronomical units, which indicates that propagation involved an important degree of scattering. Further, the degree of scattering is not a function of energy over the observed range. This suggests that the mode of propagation is a diffusion-like process and that energy-dependent processes, such as drift across magnetic field lines, play a minor role. In fact, it has been shown that the equation for simple diffusion describes the propagation curve of this particular event through its maximum (Bryant *et al.*, 1962). It fails to do so, though, during both the early anisotropic phase and the later stages where boundary conditions are important.

b) Source Spectrum

We now discuss the physical meaning of the scaling factors used to construct Figure 6. Consider the relative intensity of any two components of the event. We have recorded the intensities not as a function of time but as a function of distance traveled, and found that the relative intensity is essentially constant over a range from 2 to more than 100 astronomical units. There is nothing to suggest that an extrapolation back to zero distance in invalid. The relative intensity of two components at zero distance is, by definition, a measure of the shape of the source spectrum.

The scaling factors used to produce Figure 6 provide these relative intensities and Figure 7 shows the source spectrum obtained directly from them. The source spectra of two other events analyzed in a similar way are also shown. The ordinate of Figure 7 is arbitrarily chosen to show the maximum intensity reached at the earth. The differential intensities shown are proportional to the absolute differential intensities of protons produced at the sun and retain, therefore, the same spectral form, but the constant of proportionality which depends on the geometry of propagation is unknown. The spectrum is well represented by a power law in kinetic energy with a slope of about -1.7.

We note that, apart from small-scale fluctuations, this event is completely described by two graphs: the source spectrum of Figure 7 and the distribution of path length during propagation of Figure 6.

V. THE EVENT OF 23 OCTOBER 1962

The event of 23 October 1962 is the smallest discussed here and probably the lowest-intensity primary solar-proton event studied to date. It was initiated by a class 2 flare occurring 70° west of central meridian; further details of the flare and associated phenomena are listed in Table III. The event was also observed by the cosmic-ray equipment on Mariner II (H. Anderson and V. Neher, J. Van Allen, private communications, 1963).

Figure 8 shows the intensity vs. time profiles of various components of the event. We see that again the higher-energy components reach a maximum earlier than the lower-energy components. The intensity vs. time profiles, corrected for particle velocity and superimposed in the same way as for the event of 28 September 1961, are shown

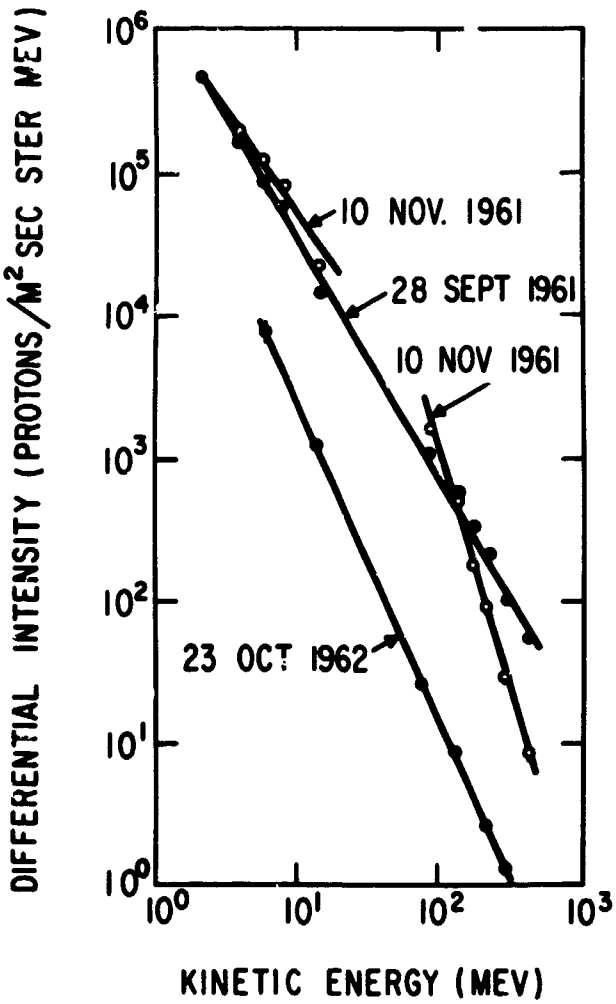


FIGURE 7.—The source spectra of three solar proton events. The intensities plotted are arbitrarily chosen to show the maximum intensities reached at the earth; as explained in the text the relative scaling of the two sections of the spectrum of 10 November is not necessarily meaningful. In the case of the solar proton events which totally conform to a velocity-dependent behavior, such as those of 28 September 1961 and 23 October 1962, the source spectrum is the unique differential energy spectrum of the protons at the time of their escape from the sun; in each event the source spectrum is proportional to that shown here with a constant which depends in an unknown way on the geometry of propagation.

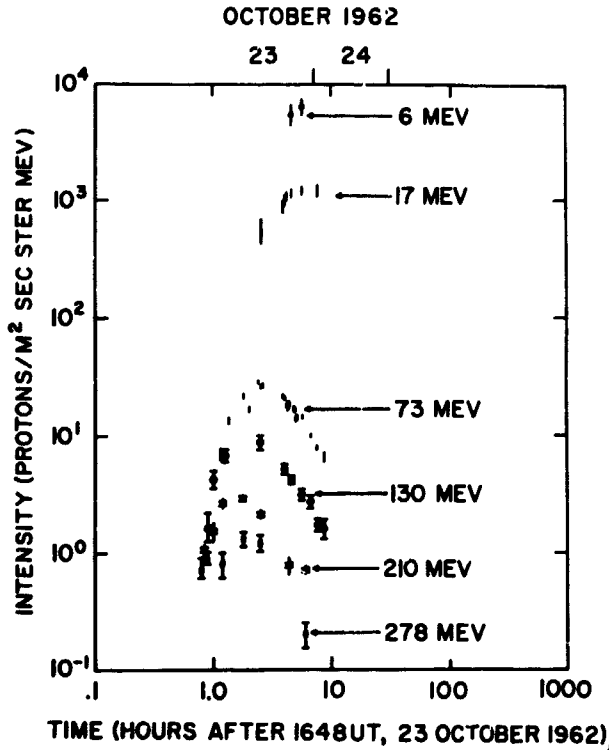


FIGURE 8.—The differential intensities of solar protons during the 23 October 1962 event plotted against time after the onset at the sun of type IV radiation. The behavior of this event is qualitatively similar to that of the early portion of the 28 September 1961 event. Due to the relatively low intensity of the event, the differential intensities of protons of energy above 300 mev could not be measured, the protons of energy below 5 mev did not arrive until after the satellite entered the magnetosphere.

in Figure 9. We find a good fit to a common curve so again the observation that higher-energy particles arrive earlier is explained quantitatively as a dispersion effect. The source spectrum obtained from the relative normalization used to construct Figure 9 is shown in Figure 7. We note that again it is well represented by a power law in kinetic energy. The slope of the spectrum in this case is about -2.3.

VI. THE EVENT OF 10 NOVEMBER 1961

The solar proton event of 10 November 1961 was one of two events observed by Explorer XII resulting from flares on the West limb of the sun (Table III). A long-lived solar stream emanating from the flare region produced a recurrent event 21 days later on 1 December (Bryant *et al.*, 1963).

Intensity vs. time profiles for this event are shown in Figure 10. A feature of these curves is a sudden drop in intensity at 1.3 hours (1546 UT). It occurs in all components that show a measurable intensity at this time. A second drop occurs at 3 hours (1730 UT). As we shall mention again in Section VII, these fluctuations are very likely unusually large cases of a feature common to all

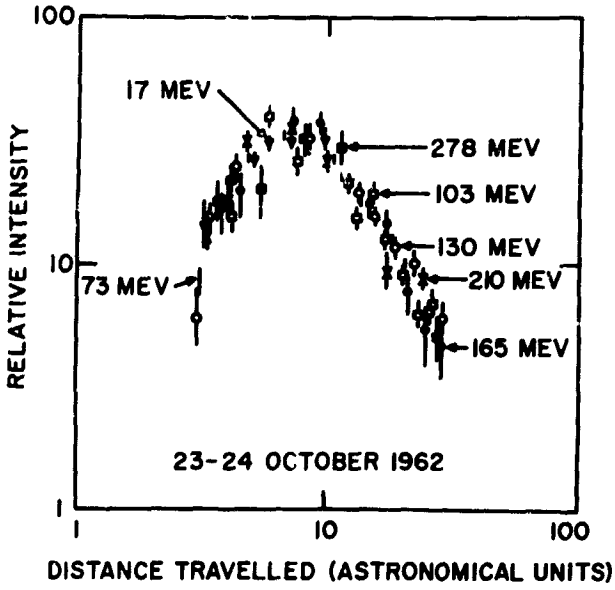


FIGURE 9.—The intensity vs time plots of Figure 8 converted to a relative intensity vs distance plots, scaled to give the best fit to a common propagation curve.

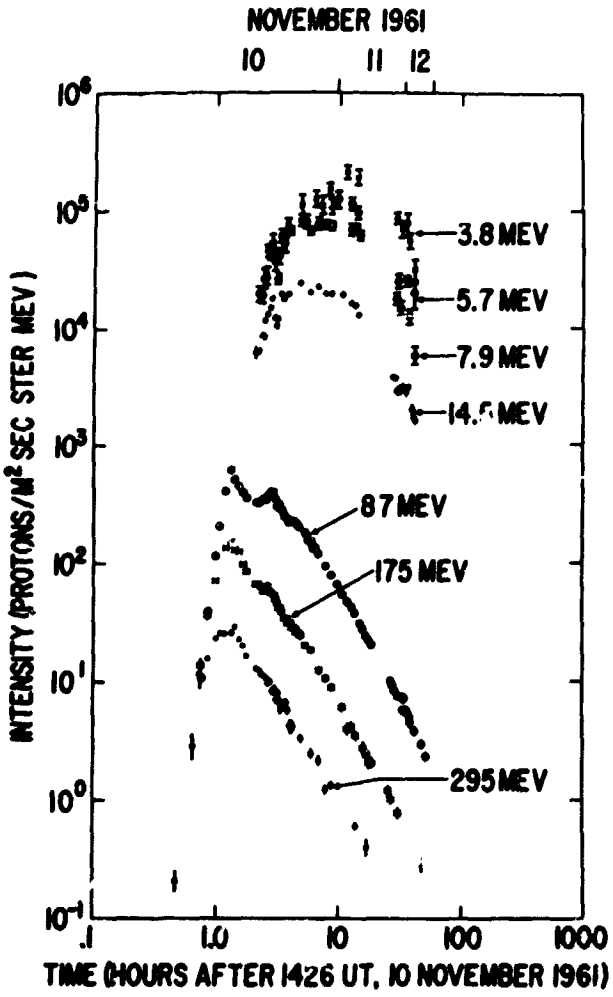


FIGURE 10.—The differential intensities of solar protons during the 10 November 1961 event plotted against time after the flare. A sudden drop in the intensities of the higher-energy protons occurs at 1.3 hours; this discontinuity in the slopes of the propagation curves occurs before the arrival of most of the lower-energy protons.

four primary events discussed here, namely a periodic intensity fluctuation with a period of about one hour. However, since these particular changes are so unusually large and sharp they may be due to some other cause. The flare producing this event was accompanied by unusual loop activity, so it may not be by chance that the solar-proton intensity exhibits unusual behavior. We are unable to link any specific happening at the sun with the changes taking place near the earth but we are able to deduce when it should occur. Figure 11 shows the integral intensities above 60 mev and above 200 mev plotted in a linear scale to show how drastically the behavior changed. The data are consistent with the breakdown having occurred simultaneously at all energies, but the time resolution of the measurements would permit a 15 minute dispersion and so is consistent with the dispersion to be expected from the velocity difference acting over a distance of 1 astronomical unit. The breakdown occurs in the 87 mev component somewhere between 1546 and 1556 hours. Since 87 mev particles take 20 minutes to travel 1 a.u., the first ones to bring information of a change would have left the sun between 1526 and 1536 if they traveled in a straight line, but (more likely) between 1521 and 1531 if they traveled along a curved path defined by the interplanetary magnetic field. Optical or radio information would reach the earth 8 minutes later, somewhere between 1529 and 1539. Though the flare showed great activity (see Table III for comments) we have found no evidence for any drastic change of conditions in this period.

a) Velocity Dependence and Source Spectra

The complicated structure of this event defies a description in terms of velocity dependence alone. As will be shown below, there are occasions on which a velocity dependence is revealed. Figures 12a and 12b show velocity-compensated intensity-distance profiles for the higher and lower energies taken separately. The higher-energy

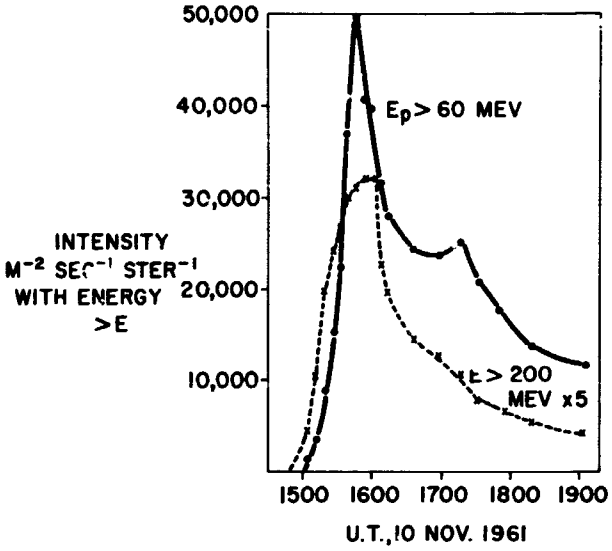


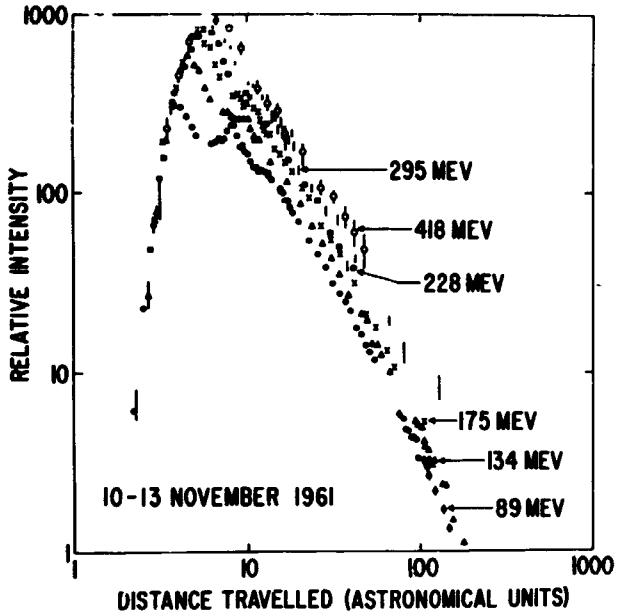
FIGURE 11.—The integral intensities of protons above 60 mev and above 200 mev during the first few hours of the 10 November 1961 event, plotted on a linear scale. These components are chosen so as to display the maximum dispersion of the times of peak intensity; since measurements were made every 7 minutes, the data are consistent with the peak intensities having occurred simultaneously.

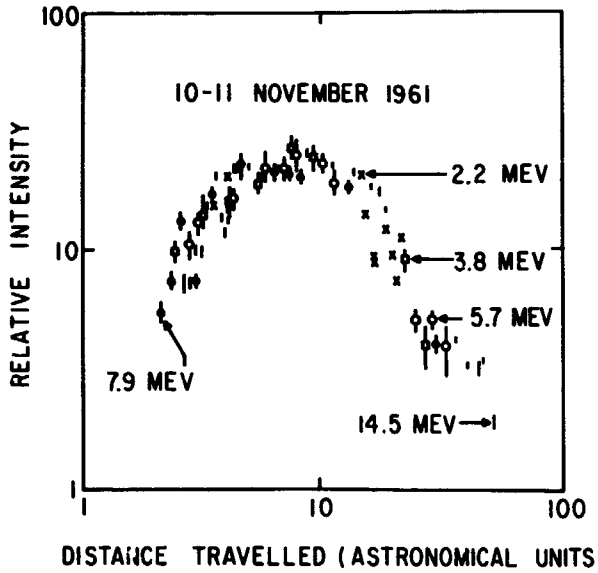
components are velocity dependent before the sudden drop in intensity at 1.3 hours, and the lower-energy components are velocity dependent throughout the event. The fact that the lower-energy components do not show a measurable intensity until after the breakdown of velocity dependence at higher energy introduces a further uncertainty in interpretation.

Two source spectra are shown for this event in Figure 7, one for the higher and one for the lower energies. Since the velocity-compensated intensity vs. time profiles are different, the relative normalization of the two sections of the spectrum is not meaningful. The normalization for each section of the spectrum is arbitrarily chosen to be the maximum intensity reached at the earth. It may be that the common fit of the low-energy intensity curves is not significant, since the deviation from a common propagation envelope increases as energy decreases among the high-energy groups.

b) Time Dependence

Although this event shows a velocity dependence that holds very well at certain times, it cannot be completely described by a single source





FIGURES 12a and 12b.—The intensity vs time plots of Figure 10 converted to relative intensity vs distance plots. Two groups are separately scaled to give the best fit to a common high-energy curve before the 1.3-hour peak, and to a common low-energy curve after it. A unique high-energy propagation curve does not exist throughout the event, but appears to be approached asymptotically with increasing energy. A common low-energy curve is apparent, but it may relate more directly to a local modulation than to interplanetary propagation.

spectrum and single propagation curve. Further, there are two sudden intensity drops that occur over a wide energy range with no dispersion. They almost certainly result from sudden changes

in the propagation medium which probably take place near the earth but, from the above discussion, which may also take place at the sun. Other effects of this kind are described in the next section and discussed more fully in section VIII.

VII. THE EVENT OF 10 SEPTEMBER 1961

The event of 10 September 1961 was initiated by a flare on the west limb of the sun. The intensity vs. time profiles of Figure 13 show that the event was dominated by intensity changes occurring simultaneously at all energies. Only gross features, such as the fact that the lower-energy components reach maximum intensity later than the higher-energy components, can be attributed to dispersion. There are no systematic changes of behavior with energy and consequently no quantitative fit to velocity dependence. A linear plot of three sample components (Figure 14) typifies the irregular behavior of the event. The high intensity reached by this and lower-energy components early on 12 September, about 1 day after the flare, is probably due to the arrival of enhanced solar plasma that produced a small cosmic-ray decrease at about that time following geomagnetic activity at 1106 UT on 11 September. This increase is like the delayed arrival of low-energy particles seen about 48 hours after the 28 September 1961 event, but in this case there is a much slower onset no doubt having to do with the west limb location of the flare.

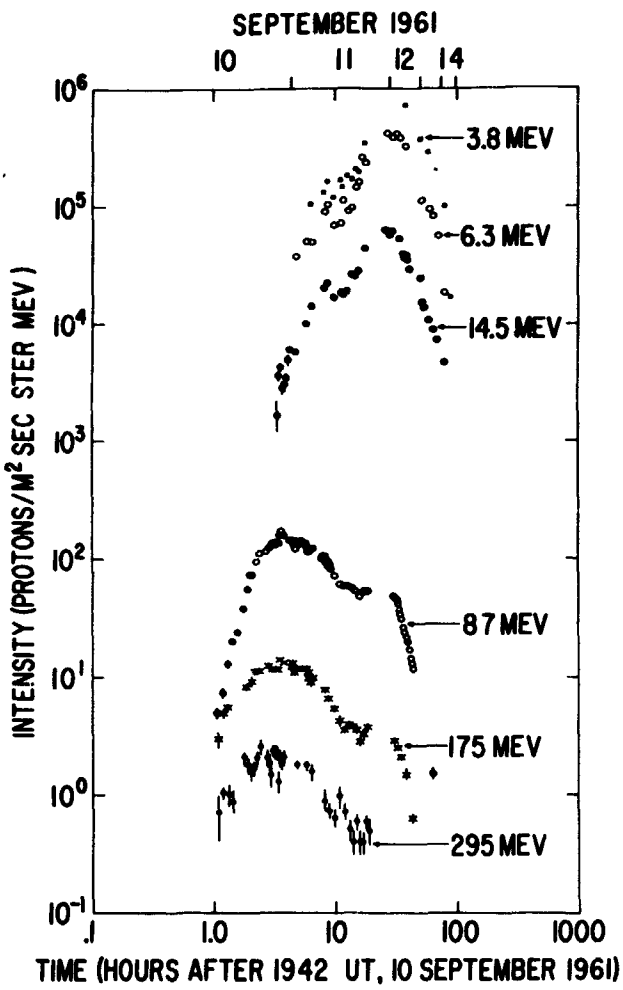


FIGURE 13.—The differential intensities of solar protons during the 10 September 1961 event plotted against time after the flare. There is a qualitative velocity ordering for the first 10 hours; during this time the higher-energy protons tend to increase in intensity before the lower-energy protons do, but there is no quantitative fit to a velocity dependence. Later in this event the intensities of protons of all energies increase together in a manner qualitatively similar to that of 30 September; this occurrence is also accompanied by cosmic-ray and geomagnetic fluctuations associated with the arrival of enhanced solar plasma.

a) Fluctuations

Superimposed on the large-scale features of the intensity vs. time profiles are a series of fluctuations which are nearly periodic, with the same frequency and phase at all energies. They take place in all these events, but are most clearly marked in the event of 10 September 1961. Figure 15, which shows the integral intensity vs. time profiles on a linear scale for the 5.7-mev and the 30-mev components of this event, illustrates these

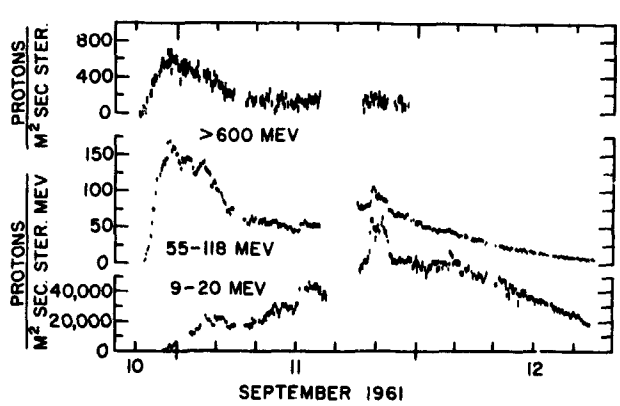


FIGURE 14.—A linear plot of the intensity of minimum-ionizing protons and of two lower-energy groups during the 10 September event, displaying both the irregular behavior of the intensity early in the event and the delayed increased on 11 September. These intensity variations dominate the pattern of this event.

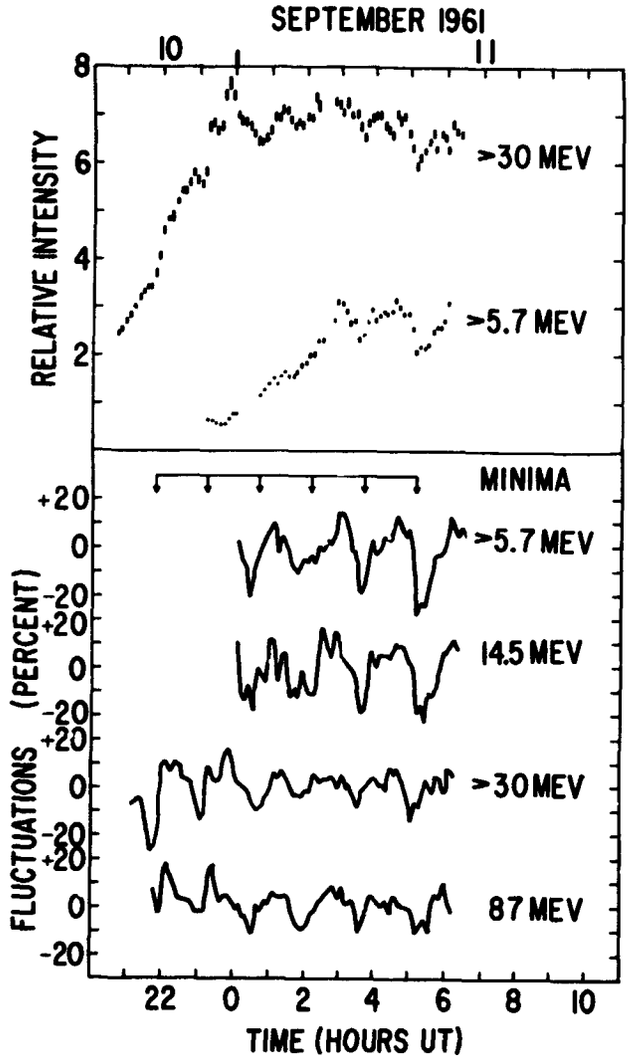


FIGURE 15.—Linear plots vs time of two integral proton intensities during the 10 September event; the deviations from the running means of intensity of these and two differential intensities are also shown in order to illustrate the relative fluctuations. A regular modulation appears having period and phase that are constant with energy; the times of minimum relative intensity are indicated to illustrate the approximate periodicity. Since there is no discernible velocity dispersion of this modulation within an event, these fluctuations are believed to have a local origin.

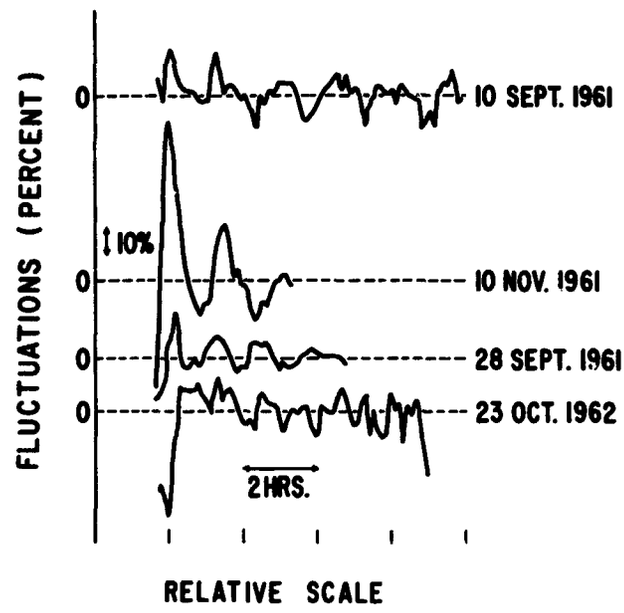


FIGURE 16.—The fluctuations of the 87-mev component of the solar proton intensities of four solar proton events. The period of the fluctuation is constant within an event but varies from 1. to 1.5 hours with the event.

fluctuations. To show the periodic fluctuations more clearly, the lower-frequency components have been removed by subtracting the running mean of one period length; the result for several energy components is shown in the lower half of the figures. Fluctuations of this kind occur to some extent in all events: Figure 16, in which the relative time scale is in units of hours, shows the fluctuation of the 87-mev component for all events. The sudden drops in intensity near the beginning of the 10 November 1961 event have the same frequency and phase as the other fluctuations, so it appears that they may not be isolated happenings caused directly by activity at the sun but the first two members of a series of fluctuations, although they were unusually large and very sharp. The amplitude of the fluctuations depends little on energy in any of the events. Table IV summarizes the periods and amplitudes of the fluctuations.

VIII. DISCUSSION

The following general statements can be made summarizing the results presented above.

(1) Solar proton events observed in the 2- to 600-mev energy region fall into at least three categories: the primary events have maximum intensity soon after the parent flare; these are sometimes followed about two days later by secondary events that occur with the arrival of the solar plasma; they also are occasionally followed by recurrent events that do not immediately follow solar activity but occur when the parent plage region passes central meridian on successive 27-day intervals.

TABLE IV.—*Solar-Proton Event Characteristics*

Event	10 Sept. 1961	28 Sept. 1961	10 Nov. 1961 Low Energy	10 Nov. 1961 High Energy	23 Oct. 1962
Energy Range of Velocity Dependence (mev)	(none detected)	1.4 to 500	1.4 to 22	55 to 500	4 to 330
Exponent of Source Spectrum		−1.7	−1.5	−3.5	−2.3
Scale of Source Spectrum at Max. Intens. $(cm^2\ sec\ ster\ mev)^{-1}$		1.6×10^6	1.4×10^6	$7.\times10^9$	$4.\times10^5$
Most Probable Distance (a.u.)		12	8	9	8
M.F.P. (a.u.) from Diffsuion Theory		0.04	0.06	0.055	0.06
Fluctuation Period (hours)	1.4	1.3	(none detected)	1.5	1.0

(2) The intensity vs. time profiles often show a linear velocity dependence. In two of the events under discussion (28 September 1961 and 23 October 1962) this velocity dependence lasts throughout the events. Exceptions occur in the event of 10 November 1961, which shows some early departure from this behavior and in the event of 10 September 1961, which shows no quantitative agreement with velocity dependence.

(3) All events show a series of periodic intensity fluctuations with a period of from 1. to 1.5 hours. These occur without dispersion.

If we convert those intensity vs. time profiles which show a quantitative, linear velocity dependence to probability vs. distance profiles, we are led directly to the following conclusions:

(4) Properties of the propagation medium and properties of the source can be studied separately and, in particular, the shape of the energy spectrum at the source can be obtained.

(5) Propagation involves an important degree of scattering (since the most likely distance traveled is about 10 astronomical units).

(6) The degree of scattering is independent of particle rigidity in the region below 1.5 bv (since all the propagation curves are the same).

a) Velocity Dependence

The analysis described in section IV, showing the velocity dependence of these events, has constituted a necessary but insufficient test. For example, we might have found an even closer fit had we compensated the intensity vs. time profiles by velocity raised to the power n, where n is close to but not equal to 1. We now show that the best value of n is indeed 1. To find the best value of n, the various intensity vs. time profiles were shifted horizontally and vertically to give the best fit to a chosen reference curve. No account was taken of particle velocity at this stage. The factors by which the curves were horizontally shifted were then plotted against particle velocity. The result is shown in Figure 17a in which, for comparison, lines of slope unity have been drawn through the points. The normalizations are arbitrary and depend on the curves chosen for reference. We find that the slopes of all lines are close to unity, the mean value being 1.0 ± 0.1. The energy intervals over which this test could be performed are implied by the figure and are indicated in Table IV. The various parts of Figure 17a have been superimposed in Figure 17b, where, as in 17a, a line of slope unity has been drawn through the points for comparison. (The energy range studied here does not extend sufficiently far into the relativistic region to permit a test of rigidity dependence as opposed to simple velocity dependence.)

Further evidence of velocity-dependent propagation in solar-proton events was provided by a comparison of alpha-particle and proton intensities during the events of 12 November 1960 and

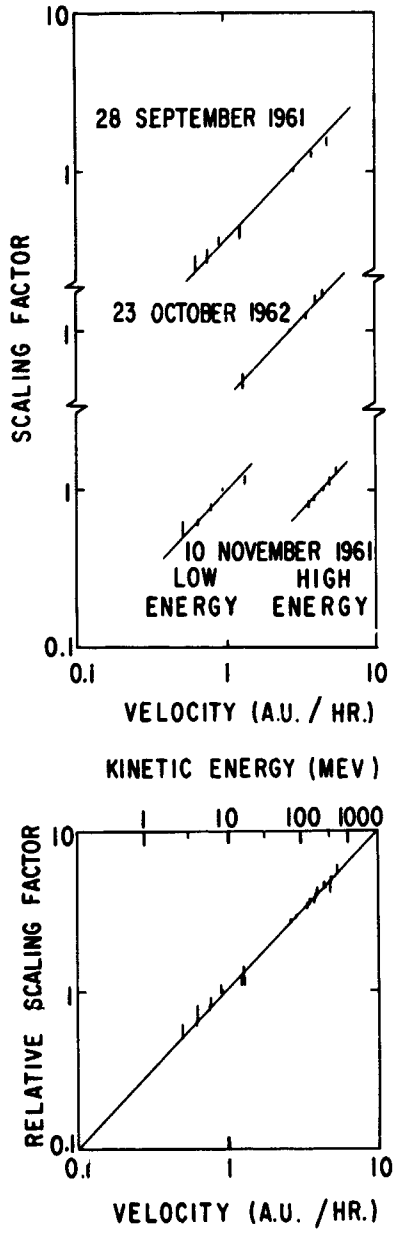


FIGURE 17a and 17b.—The scaling factors by which the differential intensity vs time plots of each event were shifted horizontally to give the optimum fit to a common curve. This effort was made to determine the nature of the dependency on velocity; comparison lines of slope 1 are used to indicate the closeness of fit to a linear velocity dependence.

15 November 1960 by Biswas *et al.* (1962, 1963). Several rockets carrying nuclear emulsions were fired into these events, and the relative intensity of alpha particles and protons having the same velocity was found to be the same at certain times even though the energy spectra of protons and alpha particles were changing with time.

b) The Source Spectrum

We note from Figure 7 that a given source spectrum is well represented by a power law in kinetic energy, although the event of 10 November 1961 is an exception in that two power-law spectra are required. The 28 September 1961 spectrum is remarkable in that it fits a power law over a dynamic range of nearly three decades in energy. The fact that there is no deviation from this smooth, simple spectral representation, even at the low-energy portion of this source spectrum, prompts us to put forward the argument, based purely on aesthetic grounds, that the amount of matter traversed by the solar protons after acceleration was less than the range of 1 mev proton, that is, about 1 milligram cm^{-2}. If we consider a beam of particles penetrating an absorber, we find the emerging spectrum very much depleted in particles with energies near that required to just penetrate the absorber. There is consequently a rapid change of slope at low energies. It seems unlikely, therefore, that an excess production of lower-energy protons would occur in such a manner as to exactly compensate their absorption in an amount of material greater than their range resulting in so simple a form of source spectrum.

We can draw no conclusions here about the solar-proton acceleration process from the shape of the source spectra, since there probably are a number of processes that would produce the observed spectra. Table IV gives the slopes of the spectra observed in the events under discussion.

c) The Propagation Medium

For the purpose of this discussion we define the propagation medium to be the medium through which solar protons travel after acceleration and escape from the acceleration mechanism. Let us now consider where the scattering takes place.

Meyer, Parker and Simpson (1956) suggested that during the 23 February 1956 event scattering took place in interplanetary space beyond the earth's orbit. Parker (1963) has discussed a mechanism of simple diffusion throughout interplanetary space; E. Roelof (private communication, 1963) has found that scattering by magnetic fields in interplanetary space that are irregular in space and/or time would produce a degree of scattering that is a function of particle rigidity.

McCracken (1962) introduced the idea of small-angle scattering caused by irregularities in an otherwise quasi-radial interplanetary field to account for the onset of isotropy in several solar-proton events. Anderson *et al.* (1959) and Gold (1962) have suggested that the processes of scattering and drift in the strong magnetic fields near the sun must play an important part in solar-proton propagation. Bryant *et al.* (1962) and Hoffman and Winckler (1963) found that the intensity vs. time profiles of several solar-proton events are well represented by those to be expected from a process of simple diffusion in interplanetary space.

The propagation curves of the three velocity-dependent events are compared in Figure 18.

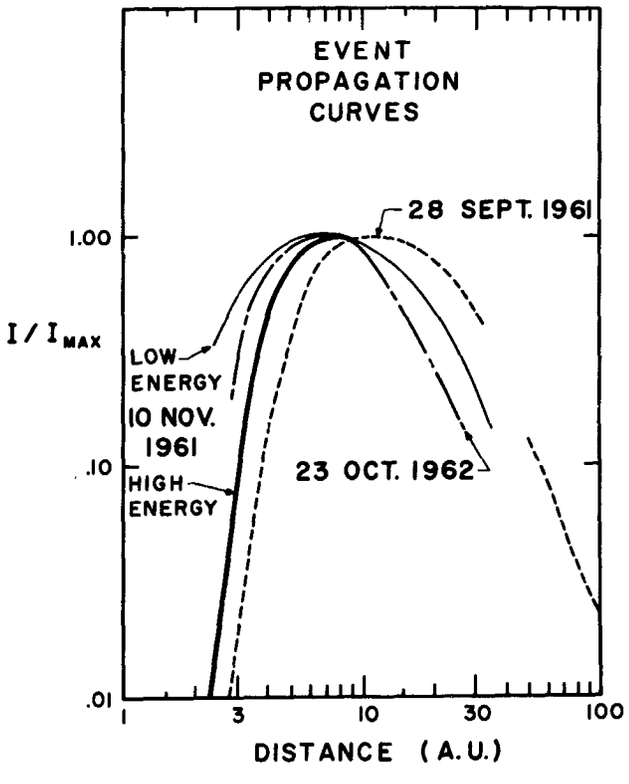


FIGURE 18.—The propagation curves of the three velocity-dependent events. These curves represent distributions in the distance travelled by solar protons from release into the propagation medium to the point of observation.

Apart from that of the low-energy component of 10 November 1961, the curves are geometrically similar to a remarkable degree. It is meaningful, therefore, to associate with each of the curves a parameter describing the rate of propagation of the event. A convenient parameter, though somewhat poorly defined, is the most probable distance traveled—that is, the distance traveled by the particles arriving at maximum intensity. These distances are listed in Table IV. We notice that the ordering of the propagation geometry in the three events correlates either with the phase of the solar cycle or with the displacement in longitude of the flare from about 70 degrees west, the origin of the earth-intercepting garden-hose line. For example, the slowest event is that from a flare at 29 degrees east; it is also the first in time. No significance is claimed for either correlation on the basis of such a small number of events.

We have mentioned that the intensity curves which fit a common propagation curve in a given event also fit the equation for simple diffusion. The resulting effective mean free paths and other parameters are listed in Table IV. In spite of this fit, it is difficult to see how solar protons could undergo a process of simple diffusion in interplanetary space especially since the garden-hose magnetic field has been observed to play an important part in guiding solar protons away from the sun. Another difficulty is that the observed velocity independence of the scattering implies that the scattering takes place at discrete scattering centers, rather than as a continuous process, as would otherwise seem more reasonable. The finding that the degree of scattering is independent of rigidity, combined with Roelof's result for continuous scattering in irregular magnetic fields, rules out such a mechanism at least for the events under discussion.

In order to reconcile a diffusive propagation with the previously observed guiding by the garden-hose interplanetary field, we consider a model in which scattering occurs near the sun and the particles escape to the earth after diffusing from the flare to the foot of a line of force providing direct access to the earth (Gold, 1959). We meet at once with two difficulties. Firstly, particle drift produced by gradients and curvatures in any general field near the sun would lead to a rigidity-dependent propagation. Secondly, there must be a certain amount of scattering in interplanetary space to account for the onset of isotropy (McCracken, 1962). Since any pitch-angle scattering would in general be accompanied by a change of guiding center there would be some

motion across the lines of force. The question then arises of the relative importance of motion along the lines of force and motion across them. We have then reached a process of anisotropic diffusion in interplanetary space instead of one of simple isotropic diffusion. Anisotropic diffusion conceivably could account for the observations, but there is no present theoretical treatment against which to test the data.

d) The Stability of the Propagation Medium

One important consequence of a velocity-dependent propagation is that low-energy protons travel through the propagation medium later than the high-energy protons and yet suffer the same degree of scattering. The average properties of the propagation medium can, therefore, remain constant for at least two days. It is important to remember, though, that changes must have been occurring in the magnetic field structure during the period of propagation. During the event of 28 September 1961, for example, there was an enhanced solar plasma moving out from the sun. Where the solar protons are isotropic in interplanetary space, even gross changes in magnetic field structure can produce only small effects if they take place many mean free paths from the point of observation. Only changes in magnetic field structure occurring near the satellite could strongly influence the measured intensity. We suggest that the character of the 10 September 1961 event was largely determined by such effects, dominating the intensity vs. time character more than the dispersion that otherwise should have been present. One can only speculate at this stage on the details of the mechanism producing time-dependent effects.

e) Local Disturbances of the Propagation Medium

As was mentioned above, the periodic intensity fluctuations are almost certainly a result of magnetic field structure in the region of interplanetary space near the earth. The fluctuations could result from, for example, adjacent regions of strong field and weak field constituting a trapping region or "magnetic bottle." Such regions would exclude some incident particles, thereby lowering the particle density in the region of the weaker field. Such regions might eventually become filled as a result of scattering from magnetic irregularities and, in fact, the magnitude of the fluctuation does decrease with time. This mechanism does not account for the regularity of the fluctuation unless there is postulated a characteristic wave motion of interplanetary plasma. The wave motion is not a property of all interplanetary plasma but is confined to the region between the magnetosphere and the quasi-stationary bow wave postulated by Axford (1963) and others. They pointed out that interaction between the solar wind and the magnetosphere should create a bow wave standing several radii from the magnetosphere on the sunlit side of the earth; this phenomenon was recently observed with a plasma detector (Bridge *et al.*, 1964) and a magnetometer (Ness *et al.*, 1964) on Explorer XVIII. All solar-proton measurements reported here were made outside the magnetosphere but behind this bow wave. It is possible, therefore, that the modulation of solar-proton intensity was confined to this region alone: regular structure in this region may be due directly to regular structure in the solar wind or it may be a natural frequency of the region.

Fluctuations with a period of approximately one hour were noted in the Explorer X magnetic field and plasma data (Heppner *et al.*, 1963; Bonetti *et al.*, 1963) where they were interpreted as a result of the passage of the space probe in and out of the pulsating boundary of the magnetosphere. Mathews *et al.* (1961) noted fluctuations with a period of 1.25 hours in neutron monitor records of the event of 12 November 1960. They suggested that these fluctuations were a direct result of pulsating decreases in the horizontal component of the earth's field measured at the equator. Winckler *et al.* (1961) also observed similar fluctuations in balloon measurements of solar-proton intensity. All of these forms of fluctuation may have been due to the same cause, namely, either a regular structure in the interplanetary plasma density or a natural frequency of the region between the bow wave and the magnetosphere.

REFERENCES

ANDERSON, K. A., ARNOLDY, R., HOFFMAN, R., PETERSON, L., and WINCKLER, J. R. 1959, *J. Geophys. Res.*, **64**, 1133.

ANDERSON, K. A., and WINCKLER, J. R. 1962, *J. Geophys. Res.*, **67**, 4103.

AXFORD, W. I., and REID, G. C. 1962, *J. Geophys. Res.*, **67**, 1962.

A*xford*, W. I. 1962, *J. Geophys. Res.*, **67**, 3791.

BISWAS, S., FICHTEL, C. E., and GUSS, D. E. 1962, *Phys. Rev.*, 128, 2756.

BISWAS, S., FICHTEL, C. E. GUSS, D. E., and WADDINGTON, C. J. 1963, *J. Geophys. Res.*, **68**, 3109.

BONETTI, A., BRIDGE, H. S., LAZARUS, A. J., ROSSI, B., and SCHERB, F. 1963, *J. Geophys. Res.*, **68**, 4017.

BRIDGE, H., EGIDI, A., LAZARUS, A., and LYON, E. *Explorer XVIII Symposium*, Gaddard, 1964.

BRYANT, D. A., CLINE, T. L., DESAI, U. D., and MCDONALD, F. B. 1962, *J. Geophys. Res.*, **67**, 4983.

———. 1963, *Phys. Rev. Letters*, **11**, 144.

CHARAKHCHYAN, A. N., TULINOV, V. E., and CHARKHCHYAN, T. N. 1961, *Proc. Intern. Conf. Cosmic Rays and the Earth Storm*, Kyoto, 2, 365.

GOLD, T. 1959, *J. Geophys. Res.*, **64**, 1665.

———. 1962, *J. Phys. Soc. Japan*, **17** (A–2), 600.

HEPPER, J. P., NESS, N. F., SCEARCE, C. S., and SKILLMAN, T. L. 1963, *J. Geophys. Res.*, 68, 1.

HOFFMAN, D. J., and WINCKLER, J. R. 1963, *J. Geophys. Res.*, 68, 2067.

MATHEWS, T., TAMBYAHPILLAI, T., and WEBBER, W. R. 1961, *Monthly Notices of the Royal Astron. Soc.*, **123**, 97.

MCCRACKEN, K. G. 1962, *J. Geophys. Res.*, **67**, 423, 435 and 447.

MEYER, P., PARKER, E. N., and SIMPSON, J. A. 1956, *Phys. Rev.*, **104**, 768.

NESS, N., SCEARCE, C., and SKILLMAN, T. *Explorer XVIII Symposium*, Goddard, 1964.

PARKER, E. N. 1963, *Interplanetary Dynamical Processes*, (New York: Interscience Publishers).

REID, G. C., and AXFORD, W. I. 1962, *J. Geophys. Res.*, **67**, 3590.

WINCKLER, J. R., BHAVSAR, P. D., and PETERSON, L. 1961, *J. Geophys. Res.*, **66**, 995.

cont'd on enclosure

N65-10683

VELOCITY DEPENDENCE AND SOURCE SPECTRA OF SOLAR PROTON EVENTS*

D. A. BRYANT†, T. L. CLINE, U. D. DESAI‡ AND F. B. McDONALD

It has become clear from recent satellite measurements[1,2] that solar protons can be transported from the sun to the earth in three distinct ways in addition to the direct way sometimes observed with cosmic-ray monitors at sea level.[3] We wish to describe these four modes of propagation and then show how the velocity dependence of one of these modes makes it possible to determine the energy spectrum of the particles at the time of release from the sun.

We classify solar particle events consisting of radiation with energy above 1 Mev into four distinct classes. The distinction between the classes depends upon the way in which the particles are transported from the sun to the region of the earth. The first two classes consist of particles that arrive at the earth in a manner determined by particle velocity; the first class consists of predominantly higher-energy anisotropic particles that arrive after nearly a direct transit and the second class consists of those that arrive after a diffusive propagation; the third class consists of those that arrive in a manner determined by the motion of enhanced solar plasma; the fourth class consists of those that depend upon the rotation of the sun. Typical delays between the occurrence of a solar flare and the arrival at the earth of particles in these four classes are different. The delay of the first class event is close to the rectilinear transit time for highly relativistic particles and is therefore only a few minutes. The arrival-time delay of the second class depends upon the rate at which particles propagate through interplanetary space and so varies with energy up to a number of hours. The typical delay for particles in the third class is not a function of energy since particles of all energies arrive with the enhanced solar plasma responsible for Forbush decreases and geomagnetic storms; it is the transit time of solar plasma across one astronomical unit, that is, about two days. The fourth class event takes place near the time of central-meridian passage of a plage region responsible for a flare and solar particle events of the other classes during the previous solar rotation. This fourth class is closely associated with long-lived solar streams. The delay between the parent flare and the arrival of these particles is not a function of energy since particles of all energy arrive with the plasma stream. The delay depends upon the solar longitude of the parent flare and may therefore be as long as one solar rotation, or 27 days. Figure 1 shows the times of occurrence of events of the latter three classes seen by Explorers XII and XIV during parts of 1961 and 1962. Four sequences of events are evident.

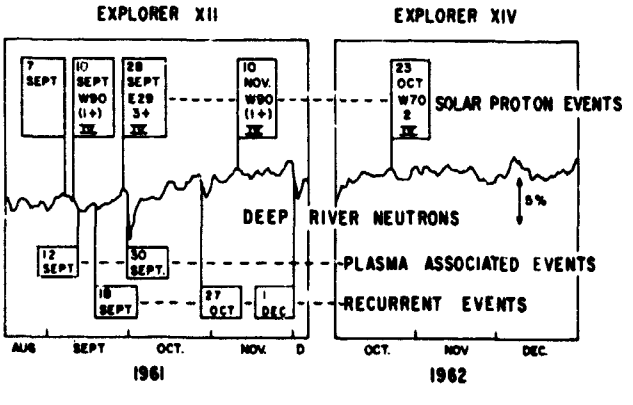


FIGURE 1.

*Published as *Goddard Space Flight Center Document X-611-64-63*, and also published in the *Proceedings* of the International Conference on Cosmic Rays, Jaipur, India, 1:70–82, April 1964.

†Goddard Space Flight Center and D.S.I.R. Radio Research Station, Slough, Bucks, England.

‡Goddard Space Flight Center and Physical Research Laboratory, Narrangpura, Ahmedabad, India.

No events of the first class were seen with sea-level monitors during these time intervals. Figure 2 shows the variation in intensity of interplanetary protons of energy greater than 3 Mev during a sequence involving all the three classes of events seen. Superimposed on the intensity decay of the velocity-ordered event of 28 September is the plasma-associated event of 30 September, followed in turn after 27 event-free days by a recurrent event on 27 October.

We now confine our attention to several velocity-ordered events observed on Explorers XII and XIV. We show how it has been possible in these events to determine separately the influence of the propagation medium and the form of the energy spectrum of particles released from the sun, that is, the source spectrum. These deductions were made possible by the fact that differential energy measurements could be made outside the magnetosphere with the equipment carried by these satellites.

A striking velocity dependence was shown by the solar proton event of 28 September 1961. This is shown by Figures 3 and 4. Figure 3 shows intensity vs. time profiles for various differential energy components. The abscissa is in units of hours from the time of the flare. Figure 4 shows the behavior of the intensities of the same differential components of the event but this time plotted not as a function of time but as a function of distance traveled. The distance traveled is simply the product of particle velocity and time from the flare. The intensity curves of the various components have been vertically scaled to give the best fit to a common curve. The physical meaning of this normalization will be examined further below. We note from Figure 4 that all components lie very closely on a common curve. We may interpret Figure 4 as a measure of the probability that a particle should travel a given distance before reaching the earth from the sun. The fact that we have essentially a common curve shows that particles of all energies traveled a given path length with equal probability. This is true for all path lengths to the extent that the various components of Figure 4 lie on a common curve. The statistical distribution of path length traveled is clearly a property of the propagation medium of interplanetary space. We note that the distance traveled by most particles is an order of

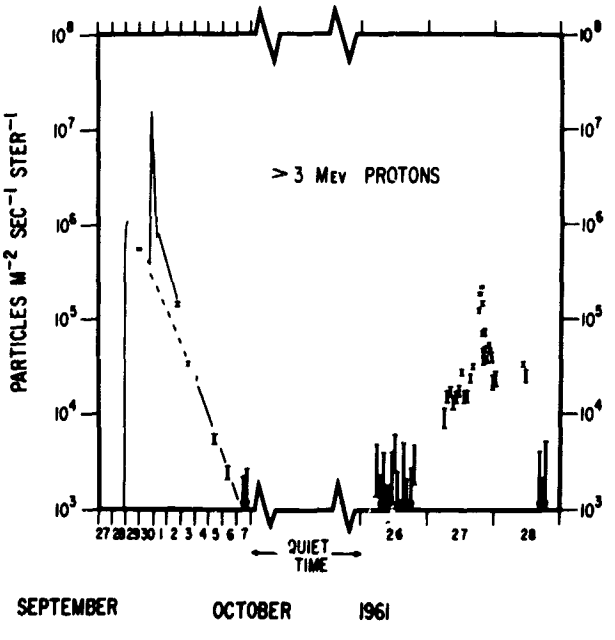


FIGURE 2.

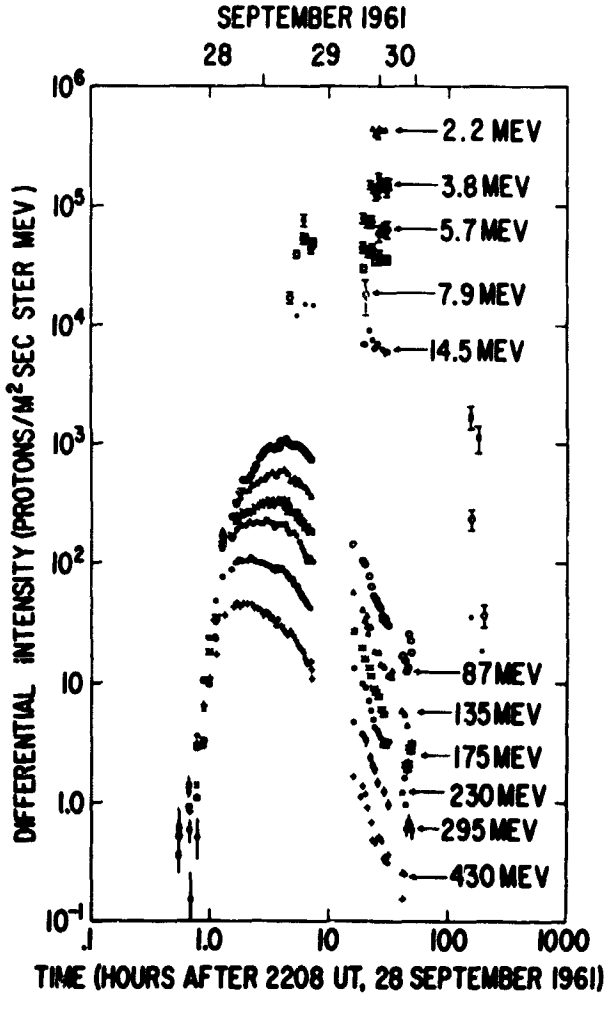


FIGURE 3.

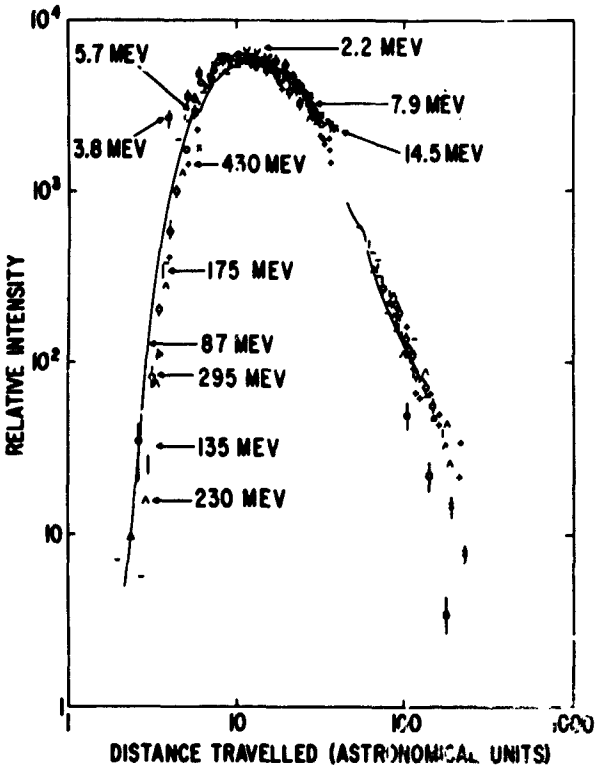


FIGURE 4.

magnitude larger than one astronomical unit. This indicates that propagation involved an important degree of scattering. Further, the degree of scattering is not a function of energy over the range examined. This suggests that the mode of propagation is a diffusion-like process and that energy-dependent processes, such as drift across magnetic field lines, do not play a dominant role. In fact, the equation for simple diffusion fits the propagation curve of this particular event through its maximum. It does not fit, though, at the beginning where anisotropy is dominant and at the end where boundary conditions must be taken into account.

Some of the other solar proton events we have observed with Explorers XII and XIV show the same degree of good fit to common, velocity-compensated, intensity vs. distance curves, but a few contrast by not fitting at all. We believe that these exceptions do not weaken our argument for velocity dependence but strengthen it by illustrating that there are times when the properties of the propagation medium cannot, by this technique, be sorted from the source characteristics because, for example, the medium could be changing as the particles are propagating through it. In fact, some solar events may thereby be combinations of events of our first three categories. Although our observations indicate that solar proton intensities in some cases depend very closely upon the first power of velocity, a choice between velocity and rigidity dependence cannot be made from these data alone. There are indications from earlier emulsion measurements of solar proton and alpha intensities,[4] however, that velocity dependence is preferable.

We discuss now the physical meaning of the scaling factors used to construct Figure 4. Let us consider the relative intensity of two components of the event. We have recorded the intensities not as a function of time but as a function of distance traveled and found that the ratio of intensities is essentially constant over a range from 2 astronomical units to more than 100 astronomical units. There is nothing to suggest that an extrapolation back to zero distance is invalid.

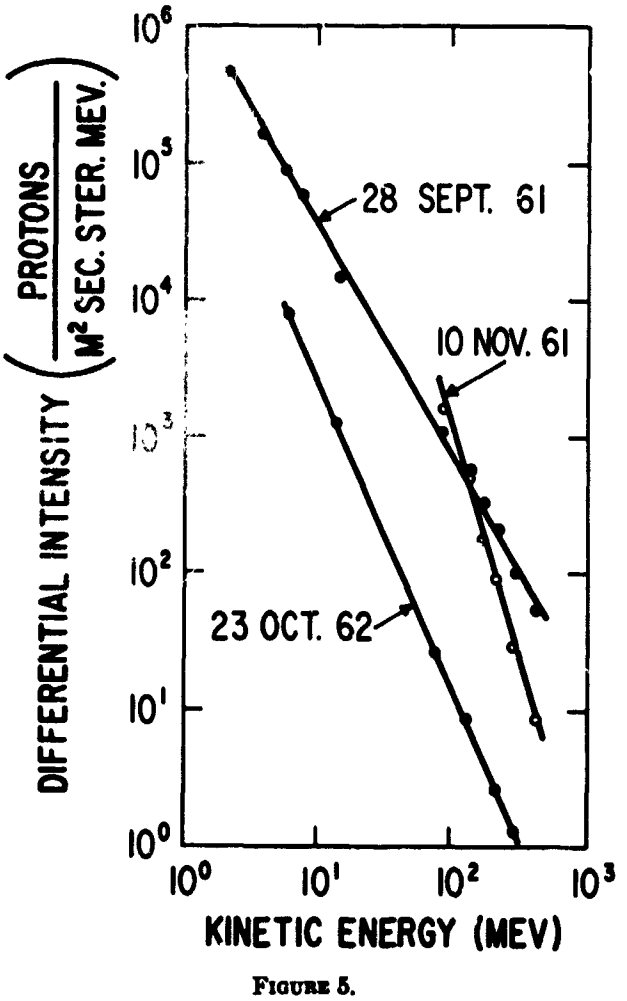


FIGURE 5.

The ratio of the intensities of two components at zero distance is, by definition, a measure of the shape of the source spectrum. Figure 5 shows the source spectrum obtained directly from the scaling factors used to produce Figure 4. The source spectra of two other events analyzed in a similar way are also shown. The ordinate of Figure 5 is arbitrarily chosen to be the maximum intensity reached at the earth. The differential intensities shown are proportional to the absolute differential intensities of protons produced at the sun and retain, therefore, the same spectral form, but the constant of proportionality depends on the geometry of propagation, which is unknown. (For example, since simple diffusion theory fits this propagation curve through maximum intensity, the numerical solution it gives for the source intensity is of the same spectral form, but that solution is for diffusion in an infinite, isotropic sphere, and is probably not a meaningful one.)

We note from Figure 5 that the source spectra are commonly very well represented by power laws in kinetic energy. This fact prompts us to put forward the argument, based purely on aesthetic grounds, that the amount of matter traversed by the solar protons after acceleration was less than the range of 1 Mev proton, that is, about 1 milligram cm^{-2}. It seems highly unlikely that an excess production of lower-energy protons would so exactly compensate their absorption in an amount of material greater than their range such as to produce so simple a form of source spectrum.

An interesting feature of these events is the existence of small-scale deviations from a common curve. Superimposed on the generally velocity-dependent intensity-time profiles are fluctuations which are nearly periodic with the same frequency and phase over the entire energy range studied. These fluctuations are evident in the velocity-ordered events discussed above, but are more striking in the 10 September 1961 event which showed no velocity dependence and was no doubt influenced by greater interplanetary disorder. Figure 6 shows plots of some sample intensities and indicates the periodic fluctuations. In this event the period is about 1.5 hours; in other events the period is slightly different. Since the transit-time dispersion over the energy range studied is significant, the fact that, in any given

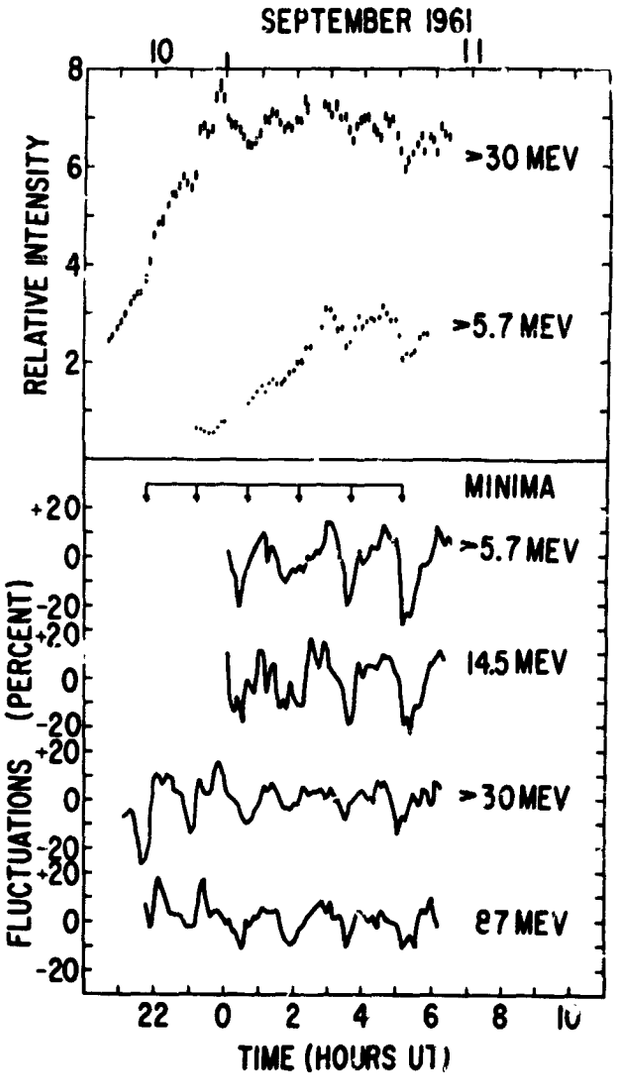


FIGURE 6.

event, the fluctuations have the same period and are in phase at all energies shows that their origin is local. We suggest, therefore, that they reflect the magnetic field structure in local interplanetary space, but we as yet have no explanation for their periodicity.

REFERENCES

1. New Evidence for Long-Lived Solar Streams in Interplanetary Space, D. A. Bryant, T. L. Cline, U. D. Desai and F. B. McDonald, *Phys. Rev. Letters*, 11, No. 144 (15 August 1963)
2. Explorer 12 Observations of Solar Cosmic Rays and Energetic Storm Particles after the Solar Flare of September 28, 1961, D. A. Bryant, T. L. Cline, U. D. Desai and F. B. McDonald, *J. Geophys. Res.*, 67, No. 13, 4983, (December 1962)

3. Observation of a Short-Lived Cosmic-Ray Solar Flare Increase with a High-Counting-Rate Meson Detector, R. A. R. Palmeira and K. G. McCracken, *Phys. Rev. Letters*, 5, No. 1, 15, (July 1, 1960).

4. Hydrogen, Helium and Heavy Nuclei from the Solar Event on November 15, 1960, S. Biswas, C. E. Fichtel, D. E. Guss and C. J. Waddington, *J. Geophys. Res.*, 68, No. 10, 3190, (May 15, 1963).

See A65-15-23

COMMENTS ON THE VANGUARD 3 MAGNETIC FIELD DATA AND ANALYSIS*

JOSEPH C. CAIN AND SHIRLEY HENDRICKS

The principal results of the Vanguard 3 magnetic field experiment are summarized in a report and a paper [*Cain et al.*, 1962 *a*, *b*]. Since some of the material presented in these documents has been used by several authors, it is worth while to clarify a few points and to report a modification indicated by a later analysis.

The Vanguard 3 report gives a series of 63 spherical harmonic coefficients fitted to the data themselves by the iterative procedure of *Jensen and Cain* [1960] (reviewed by *Heppner* [1964]). This field, designated Vanga, used g_n^m and h_n^m to $n=m=7$ and gave an rms deviation from the data of 21 γ. The data selected for this fit were the 2797 observations given by *Cain et al.* [1962*a*] as $L=0$(=best data), excluding four such observations with residuals (measured F−computed F) over 100 γ. These observations were those recorded at Blossom Point, Maryland, on October 4, 1959, at 0h 47m 23s, 0h 47m 37s, 19h 50m 11s and 19h 50m 29s. In the Vanguard 3 report we said 'The Vanga field should be construed only as a smooth curve through these data and not necessarily as a physically meaningful description of the geomagnetic field in regions remote from the Vanguard measurements.' We should note here that there is also no claim that the Vanga coefficients provide a description of other than the *total scalar field* $|\bar{F}|$ within the spacetime volume over which the data are taken. The validity of these statements is amply illustrated by *Leonard* [1963] in connection with the Johnson Island tests. In his paper he shows that the observed auroral emission seen at a point conjugate to the bomb explosions could best be predicted by the set of 48 coefficients derived by *Jensen and Cain* [1962] and that the worst agreement resulted from using the Vanga coefficients. This disagreement is not surprising, since the field line tracing requires a knowledge of the vector field and since the nearest Vanguard 3 observation was made more than 3000 miles away!

One feature of the Vanguard 3 data relative to the Vanga fit was the systematic variation of the residuals with altitude. This fact was illustrated in Figure 2 of *Cain et al.* [1962*b*] and led to the conclusion that 'A smaller systematic variation of the field was noted that appeared to require external sources. . . .' Subsequent analysis of the data using a slightly modified technique has changed this conjecture. This improvement was to solve the whole set of equations simultaneously in the linear expansion of potential rather than using the previously adopted technique of solving for small sets (up to 16) of coefficients at once, holding the rest constant, and iterating so as to correct all coefficients for a given n at a time. These methods were used, since difficulties were originally found in finding stable solutions to the Vanguard 3 data, and it was thought that the problem was one of losing accuracy in the solution involving large determinants.

The results of using the improved techniques are shown in Table 1, where we give the smallest

TABLE 1.—*rms Residuals between Vanguard 3 Data and Sets of Spherical Harmonics using Different Numbers of Coefficients*

Maximum $n=m$	No. of Coefficients	rms Residuals, γ
3	15	83.2
4	24	25.6
5	35	17.6
6	48	16.9 Vangb
7	63	16.7

*Published in *Journal of Geophysical Research*, 69(19):4187–4188, October 1, 11964.

rms residual to the Vanguard 3 data that could be determined with the specified number of coefficients. As can be seen in this table, it was possible to obtain a fit to a rms of 17.6 γ with only 35 coefficients, whereas the previous best fit of 21 γ required 63 coefficients. Only spatial terms are used here owing to the short period of 84 days over which the data are taken. However, an inspection of the isoporic charts in F over the regions of measurement show changes in the earth's surface up to 15 γ in a quarter year. It is thus likely that a somewhat better fit to the Vanguard 3 data will be made possible by the inclusion of secular change terms even over this short an interval.

The 48-term set designated 'Vangb' in Table 1 was used to re-evaluate the conclusions drawn in the Vanguard 3 paper. The only change is illustrated in Figure 1, which gives the average deviation with altitude of the observations from this new field. The systematic shift with altitude illustrated in Figure 2 of the original paper [*Cain et al.*, 1962*b*] vanishes. Thus there are no longer any detectable local sources in the volume of the Vanguard 3 measurements that contribute systematically to the field at the level $\sim 6\gamma$. The major conclusion of the Vanguard 3 paper that the sources of the D_{st} field must lie above altitudes of the order of 2500 km remains unaltered. The positive correlations previously obtained between D_{st} at the earth's surface and the residuals to the new fit are not lessened. Instead, the correlations shown in Figure 9 of the Vanguard 3 paper are slightly more positive.

Although the Vangb field gives comparable results in serving as a tool for analysis of the Vanguard 3 magnetic field data, its validity outside the regions of the data is even less than for Vanga. This fact was well illustrated when we attempted to find the Vangb conjugate point for Johnson Island to compare with those computed and observed by *Leonard* [1963]. The result is that for Vangb the magnetic equator passes nearly *through* Johnson Island and thus the conjugate point starting with an altitude of zero falls within 100 km!

Further analyses of the main geomagnetic field have been done which include the use of the Vanguard 3 data as a valuable addition to the total data set. The results of this work were reported recently [*Daniels and Cain*, 1964] and will form the basis of a publication in which a realistic field will be presented. It is hoped that workers in the field of geomagnetism will meanwhile confine the use of the Vanga set of coefficients to interpolations of the scalar field $|\bar{F}|$ in the spacetime volumes enclosed by the Vanguard 3 data and not attempt extrapolations to other regions where the result is likely to be fictitious.

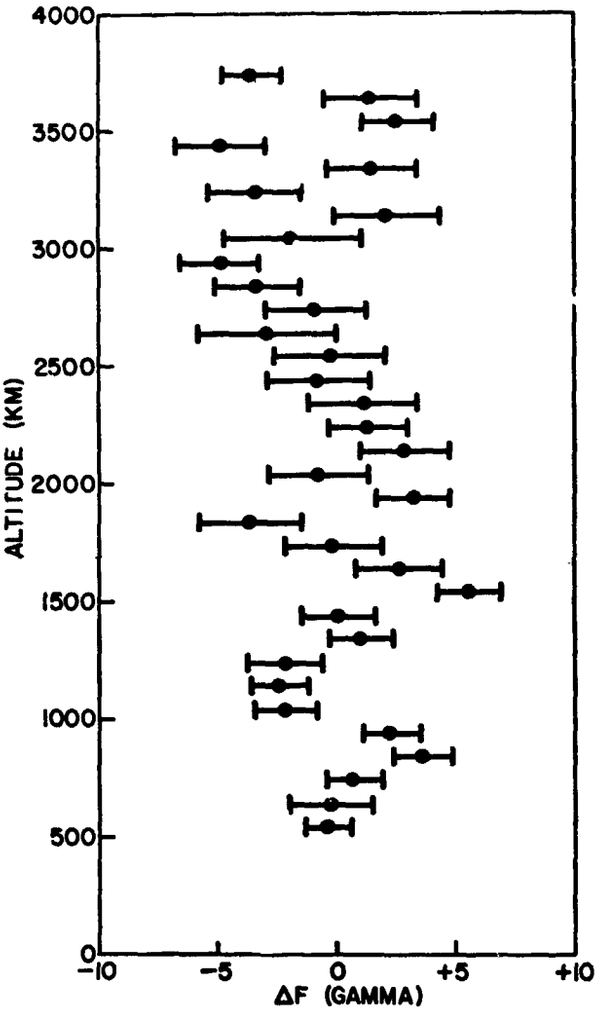


FIGURE 1.—Average deviation of the Vanguard 3 magnetic data from the 48 coefficient Vangb field plotted as a function of altitude. These average residuals are given with error bars signifying the standard error of each mean.

REFERENCES

CAIN, JOSEPH C., IVAN R. SHAPIRO, JOHN D. STOLARIK, and JAMES P. HEPPNER, Measurements of the geomagnetic field by Vanguard 3 satellite, *NASA Tech. Note D-1418*, 1962*a*.

CAIN, JOSEPH C., I. R. SHAPIRO, J. D. STOLARIK, and J. P. HEPPNER, Vanguard 3 magnetic-field observations, *J. Geophys. Res.*, **67**, 5055–5069, 1962*b*.

DANIELS, W. E., and J. C. CAIN, An evaluation of the confidence levels of the spherical harmonic coefficients for the geomagnetic field (abstract), *Trans. Am. Geophys. Union*, **45**, 39, 1964.

FINCH, H. F., and B. R. LEATON, The earth's main magnetic field—epoch 1955.0, *Monthly Notices Roy. Astron. Soc., Geophys. Suppl.*, **7**(6), 314–317, 1957.

HEPPNER, J. P., The world magnetic survey, *Space Sci. Rev.*, **2**, 315–354, 1963.

JENSEN, D. C., and J. C. CAIN, An interim geomagnetic field (abstract), *J. Geophys. Res.*, **65**, 2500, 1960.

LEONARD, Robert S., Selection of a model of the earth's magnetic field, *J. Geophys. Res.*, **68**, 6437–6440, 1963.

N65-35933

PARTICLE ACCELERATION IN CISLUNAR SPACE*

A. G. W. CAMERON

The recent discovery[1] of a characteristic red luminescence when enstatite achondrites are bombarded by 40-keV protons led Kopal and Rackham[2, 3] to observe the Moon in order to determine whether a similar luminescence might be excited in lunar materials by solar wind bombardment. On the night of November 1–2, 1963, Kopal and Rackham observed an enhanced red emission from the vicinity of the crater Kepler. The duration of the red glow was of the order of 10 min or longer, and it appeared twice within a 2-h period. No repetition of this phenomenon was observed on the following night or at the next lunation.

Among the circumstances which may be relevant to an explanation of this phenomenon are the following: the enhanced emission occurred within one day of Full Moon. There had been a class 3 solar flare on October 28 and two class 1 flares on November 1, and the Earth was in a magnetically disturbed region of space. Cosmic-ray neutron counts indicated that recovery from a Forbush decrease was occurring at the time of the luminescence.

Kopal and Rackham [2] noted that the red glow from the Kepler region approximately doubled the brightness of the lunar surface there, so that the amount of energy emitted in the red was approximately 10^4 ergs/cm^2 sec. They further concluded [2] that since the highest luminescent efficiency (in the Bustee meteorite) had been found to be 20 percent,[1] it was likely that the incoming energy responsible for the luminescence was at least 5×10^4 ergs/cm^2 sec. The other two enstatite achondrites examined by Derham and Geake [1] had a luminous efficiency lower by a factor of three, so it is not unlikely that the incident energy exceeded 10^5 ergs/cm^2 sec. Because approximately the same luminous efficiency is found during bombardment of the meteorite powder by photons of several MeV as by protons of 40 keV[4], it appears likely that the above conclusions ear not sensitive to the energy of the particles bombarding the lunar surface.

Kopal and Rackham [2] have discussed a possible interpretation that the above events resulted from bombardment of the lunar surface by an enhanced flow of solar plasma initiated by the preceding solar activity. They noted that the energy flux required was five orders of magnitude greater than that normally available in the solar wind, which must be considered a difficulty. Kopal [4] has suggested that an alternative mechanism might be required in which the lunar bombardment would involve particles, of energy greater than those of the solar wind, which had been accelerated in the preceding class 3 flare and trapped by the chaotic magnetic field configuration then existing in interplanetary space. It is the purpose of this communication to suggest a further alternative.

The region of interaction between the solar wind and the Earth's magnetosphere is very complex; but the following general picture has emerged (see, for example, Hines [5]). In the solar direction the pressure of the solar wind compresses the magnetosphere, but the solar wind must then flow around the magnetosphere, so that in the antisolar direction a long cavity is formed into which the terrestrial magnetic field can expand. Quite large energetic particle fluxes are observed near the magnetosphere boundary in the antisolar direction, and it has been suggested that the region of trapped radiation may extend to a distance comparable with that of the Moon.[6] A recent analysis of terrestrial magnetic activity has shown that the general level of activity is slightly changed for several days on either side of Full Moon, suggesting that the antisolar magnetic cavity may extend to the lunar distance.[7] Since the solar wind

*Published in *Nature* 202(4934):785, May 23, 1964.

exhibits supersonic flow with respect to the Earth, it is necessary that a standing shock wave must be formed beyond the magnetopause. The highly turbulent magnetic fields between the shock front and the magnetopause have been observed.[8]

It seems to be the case that wherever turbulent fluid motions and chaotic magnetic fields exist in Nature, the acceleration of charged particles to energies much exceeding thermal energies takes place. This is inferred to be the case in strong extragalactic radio sources, in supernova remnants, in solar flares, and in the magnetosphere itself. One acceleration mechanism which can be very efficient in certain circumstances is the second-order acceleration of charged particles traversing time-varying fluctuations in a magnetic field system.[9] This is probably responsible for the acceleration of particles in the outer part of the magnetosphere which are then dumped directly into the atmosphere to produce aurorae. The process will be enhanced at times of greater magnetic activity when the solar wind is exerting varying pressure on the magnetopause.

It is a logical extension of these considerations to expect that extensive particle acceleration will take place in the shock zone beyond the magnetopause. The particles thus accelerated will be discharged down the magnetic field lines in the antisolar direction. By analogy with the auroral discharges, it is possible that the particle beams in the antisolar direction may carry an energy flux large compared to that in the solar wind. However, since the energy going into the charged particle beams must be derived from that of the solar wind, it is evident that such particle streams will be isolated in local regions of space, and that the integrated energy flux of plasma and particles together cannot exceed the initial energy flux of the plasma flow.

The luminescence near the crater Kepler observed by Kopal and Rackham may have been produced by these energetic particle streams. It may also have been produced by trapped radiation in the distant tail of the magnetosphere. Many more observations will be needed to evaluate the plausibility of these two possibilities. In either case, such luminescence can be expected to occur principally at times near Full Moon and during unusually large magnetic activity on the Earth. Because of the localized nature of the incident particle streams, lunar luminescence excited in this way is likely to be observed only in small regions of the Moon at any one time and to exhibit variations with periods of a few minutes.

Because of the complex character of the phenomena suggested here, it is not possible to suggest the range of energies to which particles are likely to be accelerated in the shock zone, and hence it is not clear whether such particle streams will constitute a radiation hazard to a manned lunar landing or to manned spaceflight in the antisolar direction.

I wish to thank Prof. Z. Kopal for a stimulating discussion of his observations.

REFERENCES

[1] Derham, C. J., and Geake, J. R., *Nature*, **201**, 62 (1964).

[2] Kopal, Z., and Rackham, T. W., *Nature*, **201**, 239 (1964).

[3] Kopal, Z. and Rackham, T. W., *Sky and Telescope*, **27**, 140 (1964).

[4] Kopal, Z. (private communication).

[5] Hines, C. O., *Science*, **141**, 130 (1963).

[6] Frank, L. A., Van Allen, J. A., and Macagno, E., *J. Geophys. Res*, **68**, 3543 (1963).

[7] Stolov, H., and Cameron, A. G. W. (to be published).

[8] Ness, N. F., *IMP Satellite Symposium at Goddard Space Flight Center* (March 1964).

[9] Shen, C. S. (to be published).

N65-19722

DETECTION OF INTERPLANETARY 3- TO 12-MEV ELECTRONS*

T. L. CLINE, G. H. LUDWIG, AND F. B. McDONALD

In this Letter we report the direct observation of interplanetary electrons of energy above 3 MeV with the IMP-1 satellite (Explorer 18).

Electrons observed in the primary interplanetary radiation in the BeV energy region by Earl[1] and in the 200-MeV energy region by Meyer and Vogt[2] are believed to be of galactic origin because their energies are as high as those assumed to be necessary for their penetration into the inner solar system and because their measured intensity agrees with that which was anticipated to account for galactic radio emission. Support was lent to this hypothesis when the modulation characteristics of these particles were observed[3] to be similar to those of cosmic-ray protons and their positron-to-electron ratio was found[4] to be compatible with an origin of at least half of them in meson-producing cosmic-ray interactions in the interstellar medium. We feel that the existence of an interplanetary flux of electrons lower in energy by orders of magnitude is interesting because of the possibility that these too may have a cosmic origin. If so, their study should yield entirely new information about the galactic electron sources and modulation characteristics. If they are of solar origin, there are analogous implications. We wish to demonstrate here that the flux of lower-energy electrons we observe is indeed a primary component of the interplanetary radiation, and to discuss its properties in terms of its possible origin, either galactic or solar.

The observations reported here were made with a scintillator telescope on Explorer 18, a satellite placed in an elliptic orbit with an apogee height of 193,000 kilometers. Data were taken from the launch, on 27 November 1963, until 6 May 1964 when the satellite passed into a long period in the earth's shadow, causing failure of the detector. During this time interval the apogee moved from the sunlit side of the earth beyond the magnetosphere (terminating at about 70,000 km) and beyond the earth's shock front (observed with a magnetometer[5] and plasma sensor[6] at about 100,000 km) to the region behind the earth and inside the shock front. Electron data taken only when the satellite was beyond 125,000 kilometers are reported here; throughout the life of the instrument these data continued to be free from effects due to the trapped radiation.

The detector was developed[7] to study low-energy cosmic-ray protons, electrons and light nuclei. It is composed of three scintillators: two in coincidence, measuring energy loss and total energy, and a guard counter in anti-coincidence. When a table of intensity vs. measured energy loss vs. measured total energy is constructed from data taken at apogee, there is seen a distinct counting rate component of minimum-ionizing energy loss and of low apparent energy. An analysis of the topology of distributions in energy loss and in total energy through this minimum-ionizing component indicates that indeed it is composed of three distinct particle groups: On group with total energies corresponding to electrons that stop within the detector, a much smaller group with a high apparent total energy equal to or exceeding the energy loss of a minimum-ionizing cosmic ray traversing the detector, and a third group with very low total energies. We believe that the latter two components are surely secondary radiations composed of, respectively, cosmic rays that avoid detection by the guard counter (for example, by turning into neutrals through reactions within the detector) and gamma rays made locally in the spacecraft, producing random and coherent coincidences between the energy loss and total energy detectors.

*Published as *Goddard Space Flight Center Document X-611-64-368*, November 1964

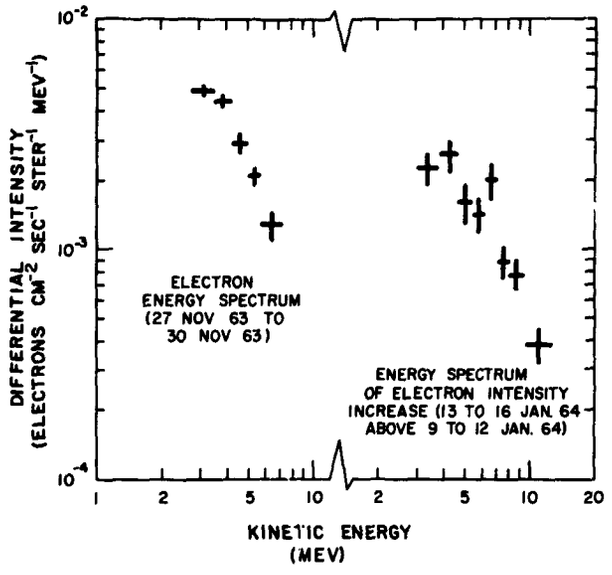


FIGURE 1.—Differential energy spectra of electrons observed beyond 125,000 kilometers from the earth. The first spectrum is from the apogee of the first orbit; the second is the difference between measurements from the 13th and 12th orbits and indicates the first significant increase in intensity.

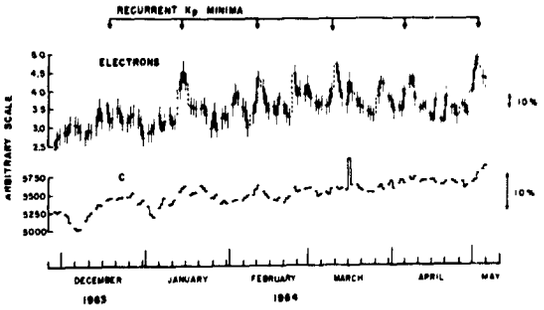


FIGURE 2.—Integral counting rate of electrons throughout the active life of the instrument plotted in quarter-orbit averages. The counting rate "C" of cosmic rays into a thinly shielded scintillator and the times of one recurrent minimum of the interplanetary index K_p are also shown. Recurrent Forbush decreases are seen in the cosmic rays in early December and January and a small solar proton event occurs in March; other increases can be largely attributed to the electron mixture in the cosmic rays.

These secondary effects were eliminated to produce Figure 1a, which shows the energy spectrum of electrons obtained during the first orbit (27 to 30 November 1963) at a time when the observed electron intensity was at a typical minimum and when there were no measurable time variations. Figure 1b shows, for comparison, a spectrum of the difference between the first statistically significant intensity increase (13 to 16 January 1964) and the immediately preceding intensity (9 to 12 January). No background corrections were necessary to produce the latter distribution since the electron intensity increase was unaccompanied by an increase of either secondary gamma rays or spurious cosmic rays; it was therefore possible to determine the intensity to higher energies. The nearly identical shapes of the two corrected spectra suggest that the electrons seen daily may have the same origin as the extra ones seen on days of increased electron flux. The integral intensity of electrons of energy between 2.7 and 7.5 MeV is $210. \pm 10.$ electrons/m²sec.ster., and that of the increase between 3. and 12.5 MeV an additional $100. \pm 10.$ electrons/m²sec.ster.

To demonstrate that most of the observed electrons are not of local or secondary origin at the satellite (e.g., such as knock-on or cascade-shower electrons produced in or near the detector) we consider their time variations. Figure 2 shows the counting rate of these electrons, partially corrected for slow gain drifts in the detector, plotted in the form of one-quarter-orbit averages throughout the active life of the instrument. (The gaps in the data occur at times when the satellite is within 125,000 kilometers; the other three points per orbit are plotted so that each center one represents data taken from beyond 185,000 kilometers.) Also shown are a comparison plot "C" of the integrated cosmic-ray flux into a scintillator with about 0.3 gram cm^{-2} shielding, and the times of a recurrent minimum in the interplanetary magnetic activity index K_p with a period of one solar rotation.

A dominant feature of the electron rate is the appearance of many statistically meaningful intensity increases, including one series apparently coincident with the recurrent K_p minimum. These electron intensity increases were not accompanied by comparable increases in the integral cosmic-ray intensity above 15 MeV: the magnitude of the electron modulation is 50 per cent on occasions, while the cosmic rays undergo modulations of less than 5 per cent. Further, following the flare of 16 March 1964 there was a solar-proton event, accompanied by Type IV solar radio emission,[8] during which the flux of protons of energy between 15 and 75 MeV briefly increased by

several orders of magnitude, while the 3- to 8-MeV electron flux rose less than 50 ± 25 per cent. (Figure 2 shows the quarter-orbit average of the total integrated cosmic ray flux increasing at that time by about 10 per cent.) These comparisons demonstrate that, at most, an insignificant fraction of the electron modulation results from modulations of cosmic rays of energy above 15 MeV.

Modulations of protons with energies below 15 MeV, such as 27-day recurrent solar proton events similar to those observed[9] with Explorer 12, were not monitored with our apparatus; but these would be expected at the times of recurrent Forbush decreases and geomagnetic activity, rather than at the time of our repeating electron increases. Several such 1- to 10-MeV proton intensity increases were observed early in the life of the satellite by Fan, Gloeckler and Simpson[10] but these were about two weeks out of phase with our electron enhancements and appear to be accompanied by, if anything, decreases in the electron intensity and in the galactic cosmic rays.

Finally, a study of 3-hour averages of the observed intensity of these electrons indicates no variations with distance from the earth, either during orbits of minimum intensity or during times of increased intensity; the electron rate is constant through the shock front to a distance of up to 50 per cent beyond it. Further, the satellite's passage through the wake of the moon[5] was unaccompanied by an electron intensity variation. Thus, these electrons are not secondary to cosmic rays or solar protons or due to geophysical processes.

We feel that the question of whether these primary electrons originate at the sun or in the galaxy cannot be difinitely answered on the basis of the available data; however, the following properties of these electrons are consistent with their being galactic. First, the differential energy spectrum of this 3- to 12-MeV component fits smoothly onto a spectral plot of the cosmic-ray electron intensities [1, 2, 4] at much higher energies. Second, the time variations of the electrons can be compared to those of cosmic rays in that there is a strong correlation between the electronint ensity increases and quiet interplanetary conditions, evidenced by K_p minima and very small sea-level cosmic-ray intensity increases. Third, there appears to be a long-term increase of electron intensity after a correction of the same order is applied for a slow, monitored drift in detector gain; if this increase is real, it is similar to the 11-year modulation of cosmic rays as solar minimum is approached. However, the fact that the differential cosmic-ray proton intensity is peaked at about 1 BV/c rigidity and negligible below 150 MV/c markedly contrasts with the fact that electrons of rigidity $\gtrsim 3.5$ MV/c are more abundant than those of greater rigidity. Parker has recently pointed out[11] that particles with gyroradius close to the idealized irregularity scale of the modulating medium should be deflected more than those of either extreme; thus these electrons of low rigidity might originate in the galaxy and penetrate the solar system as easily as those of great rigidity.

In spite of the foregoing arguments for galactic origin, it is not impossible that the electrons came instead from the sun. Several possibilities present themselves. For example, relativistic electrons might be generated over most of the upper surface of the solar atmosphere, in which case regions of enhanced and expanded plasma (which contain recurrent proton fluxes) would tend to contain fewer electrons while regions of quiet-time streaming would contain more, as we have observed. Further, the deceleration of the electrons in the enhanced plasma might be much greater than that in the quiet-time streaming. Alternatively, the electrons might be associated with the development of new sunspot regions, which is a characteristic of this phase of the solar cycle and appears to correlate weakly with the observed pattern of intensity increases.[12] We have not, however, found a correlation with any solar radio or optical activity.

The results we quote here are preliminary: an evaluation of the detector response, providing a more exact spectrum, and a detailed investigation of the time variations will be given elsewhere. We are happy to acknowledge the efforts of the many people who made the IMP-1 a success.

REFERENCES

1. J. A. Earl, Phys. Rev. Letters 6, 125 (1961).
2. P. Meyer and R. Vogt, Phys. Rev. Letters 6, 193 (1961).

3. P. MEYER and R. VOGT, J. Geophys. Res. **66**, 3950 (1961).
4. J. A. DESHONG, JR., R. H. HILDEBRAND and P. MEYER, Phys. Rev. Letters **12**, 3 (1964).
5. N. F. NESS, C. S. SCEARCE and J. C. SEEK, Space Research **5**, N. Holland Publ. (COSPAR Proceedings, Florence, 1964).
6. H. S. BRIDGE, A. EGIDI, A. LAZARUS, E. F. LYON and L. JACOBSON, Space Research **5**, N. Holland Publ. COSPAR Proceedings, Florence, 1964).
7. D. A. BRYANT, G. H. LUDWIG and F. B. MCDONALD, I.R.E. Transactions on Nuclear Science *NS*-**9**, 376 (1962).
8. A. MAXWELL, private communication.
9. D. A. BRYANT, T. L. CLINE, U. D. DESAI and F. B. MCDONALD, Phys. Rev. Letters **11**, 144 (1963).
10. C. Y. FAN, G GLOECKLER and J. A. SIMPSON, Goddard IMP Symposium, March 1964.
11. E. N. PARKER, J. Geophys. Res. **69**, 1755 (1964).
12. H. DODSON PRINCE and R. HEDEMAN, private communication.

7164-2601/1

TELEVISION CINEMAPHOTOGRAPHY OF AURORAS AND PRELIMINARY MEASUREMENTS OF AURORAL VELOCITIES

T. N. DAVIS† AND G. T. HICKS‡

During October 1963 a series of tests were performed at Fort Churchill, Manitoba on the usefulness of an image orthicon closed-circuit television system for indirect auroral photography. The system used was similar to one described in a paper by Hicks (1963). It was equipped with an f/0.75 lens providing an effective field of view of 16°. Recording was accomplished by photographing the television monitor with a 16 mm cinema camera.

Operated at maximum sensitivity, the system enabled direct viewing of 8th and 9th magnitude stars on the television monitor. At this sensitivity level, the system was somewhat better than the unaided, dark-adapted human eye in the detection of weak auroral structures. No difficulty was found in obtaining: auroral photographs of even the weakest structures, with exposures of 1/60 sec at the normal sound cinema rate of 24 frames per second. For brighter forms it was necessary to electronically reduce the system sensitivity. Examples of some of the photographs are presented in Figure 1.

Most of the photographs were recorded on Eastman Plus-X black and white film. In addition, color photographs on Eastman High-Speed Ektachrome film were obtained by sequential insertion of broad-band color filters into the optical system. A color cycle consisting of the placement of red, green and blue filters into the system, was completed in 1/8 sec. The result was a cinema film strip showing a frame with a red image, then a frame with green image, and finally a frame of blue image; the cycle was then repeated. Projected at normal sound speed, the color film shows



FIGURE 1.—From top to bottom, four consecutive photographs of a rayed arc with evenly-spaced and relatively stable ray structures. The exposure of each photograph is 1/60 second; the interval between adjacent photographs is 1/24 sec. The spacing between the ray structures is near 1.8 Km. Note the subtle temporal changes in the aurora that are evident by comparing adjacent photographs.

†Goddard Space Flight Center. NASA-National Academy of Sciences—National Research Council Post-Doctoral Resident Associate on leave of absence from Geophysical Institute, University of Alaska, College, Alaska.

‡E. O. Hulburt Center for Space Research, U.S. Naval Research Laboratory, Washington, D.C.

much flicker but the eye integrates sufficiently so that the colors are additive. Since the spectral response of the system is broader in wavelength than the eye, the color films show more red (due to OIλ6300–6364 and N_2 First Positive bands) and blue (due mainly to N_2^+ First Negative bands near λ4278 and λ4652) than one normally sees by direct visual observation. The filters used in the preliminary tests are very wide, near 800Å at half peak transmission. In the future much narrower filters will be employed in order to isolate specific auroral emissions.

Both the color and black and white television cinema-photography techniques provide data for the study of a variety of morphological aspects of auroral structures. These aspects include auroral motions, auroral pulsations, and the growth, decay and configurations of individual structures. The technique may allow the examination of the relative spatial distributions of the principal auroral emissions in more detail than has been possible previously.

Preliminary analysis of the data has been aimed at the measurement of apparent horizontal velocities with emphasis on the faster motions. These fast motions are of folds within auroral structures, of ray bundles or of well defined regions of enhanced luminosity (streaming). In some cases it is not clear which type of motion is being observed. A summary of 49 velocity measurements is presented in Figure 2. The values presented there are obtained by assuming an auroral lower border height of 100 Km. This assumption and the limited precision of angular calibration of the television system as used in October 1963 lead to uncertainties in the velocity magnitudes which are estimated at 10 percent. The measurements were performed on moving structures visible on consecutive film frames ranging in number from 6 to 125 (time duration 0.2 to 5 secs.).

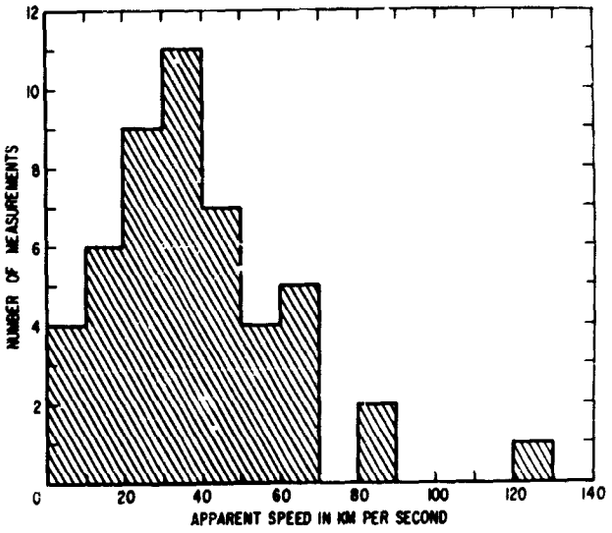


FIGURE 2.—Range of 49 measurements of horizontal speeds of auroral rays and regions of enhanced luminosity. The speed spectrum between 0 and 130 Km per/sec is divided into 13 equal blocks and the number of measurements falling within each block is shown.

REFERENCES

HICKS, G. T., Image orthicon techniques and satellite cinemaphotography; ANNA 1-B, *Photographic Science and Engineering*, 7, 328-330, 1963.

N65-19714

AN EXPERIMENTAL EXAMINATION OF LOW-ENERGY COSMIC RAY HEAVY NUCLEI*

C. E. FICHTEL, D. E. GUSS, AND K. A. NEELAKANTAN†

In a sounding rocket experiment flown from Fort Churchill on September 4, 1963, a finite flux of cosmic ray nuclei with charges greater than that of helium was detected at energies below the experimental energy cutoff of balloon-borne experiments. The particles were examined by extending large sheets of nuclear emulsions from the sides of the rocket during its period outside the atmosphere. The flux of medium nuclei ($6 \leq Z \leq 9$; Z = nuclear charge) in the energy range from 30 to 150 MeV/nucleon was measured as $0.67 \pm .13$ particles/($m^2 sr$ sec), and the flux of ($10 \leq Z \leq 19$) nuclei in the energy range from 40 to 190 MeV/nucleon was $0.31 \pm .09$ particles/($m^2 sr$ sec). A finite flux of light nuclei ($3 \leq Z \leq 5$) was also seen in the 30 to 110 MeV/nucleon region. The abundances of medium and ($10 \leq Z \leq 19$) nuclei relative to helium nuclei in the same energy intervals were found to be significantly less than the relative abundances previously determined at high energies. However, within the statistical uncertainty, the relative differential flux values observed in the vicintiy of the earth are consistent with the helium, medium, and ($10 \leq Z \leq 19$) nuclei having the same source spectrum at least above about 0.2 BeV/nucleon for a wide range of source spectral shapes and an interstellar path length of the order of those normally assumed (i.e., 2.5 g/cm^2 or slightly larger). Other possible interpretations also are considered.

INTRODUCTION

Over the last decade the study of cosmic radiation has progressed considerably and has reached the point where the fundamental properties are becoming reasonably well established.[1–3] The proton energy spectrum now is being measured from 0.01 BeV/nucleon to approximately 10^{10} BeV/nucleon, and the composition of the cosmic radiation has been measured numerous times in the energy region above approximately 0.2 BeV/nucleon. In this region, the composition appears to be independent of energy, at least up to about 10^6 BeV/nucleon, with the possible exception of a small relative increase of the light nuclei in the 0.2 to 0.5 BeV/nucleon interval.[4–8] In particular, the helium to medium nuclei and the helium to heavy nuclei ratios are known to be the same to within about a 15 percent experimental uncertainty in tne region from 0.3 to 7.5 BeV/nucleon. The data also have revealed that there is a strong modulation of the cosmic radiation which is fairly certainly associated with the solar cycle. In the measured energy interval, the variation of cosmic ray intensity is greatest at the lowest energies.[9]

In an effort to increase our general experimental knowledge of the cosmic radiation and at the same time obtain some new insight into some of these questions, an experiment was undertaken to examine the intensity of the heavier nuclei in the cosmic radiation in the region below about 0.2 BeV/nucleon. The first question to be answered was whether or not there are any medium or heavy nuclei in this low-energy region. A cosmic ray source with a very high energy threshold for acceleration of the high charges could create a near absence of these particles, since only some particles which have been degraded in energy in their interstellar travels would be present. However, if particles are present, the measurement of their properties can provide some restraints on the combined questions of the interstellar travel of cosmic rays and the spectra at the origin.

The local solar modulation has the same effect on all the particles of primary concern in this experiment—namely, *He*, *C*, *N*, *O*, *Ne*, etc.—

*Published as *Goddard Space Flight Center document X-611-64-345*, November 1964

†NAS-NASA Research Fellow, on leave from Tata Institute of Fundamental Research, Bombay, India.

because these nuclei all have the same charge to mass ratio and hence the same velocity for a given charge.

To measure the intensity of low-energy heavy nuclei, nuclear emulsions, which are detectors particularly suited for this purpose, were exposed to the cosmic radiation above the earth's atmosphere on a sounding rocket at Fort Churchill, Canada. The use of a sounding rocket rather than a polar orbiting recoverable satellite has several advantages. First, several sounding rocket shots can be spaced at desired intervals, whereas—at least until now—it has not been possible to obtain a single emulsion exposure on a recoverable satellite under less than about 2 g/cm^2 of material in a region of space where low-energy particles are not excluded by the earth's magnetic field. Also, satellite exposures are very expensive because of the cost of the necessary modifications to an existing system and the cost of testing to meet the rigid design specifications. The design of a satellite system is complicated by the requirement that the emulsions must be protected from high temperatures and still be exposed under very little matter. Further, the fact that the geomagnetic cutoffs are uncertain demands that a time resolution device be included in a satellite experiment to obtain absolute fluxes, since emulsions themselves integrate over time. Finally, the high Van Allen belt radiation background arising from the South Atlantic anomaly is an additional disadvantage associated with a satellite exposure.

To overcome the principal difficulty associated with the sounding rocket exposure—namely, having only a short exposure time available, a large area of emulsion was extended from the side of the payload during the portion of the flight when the rocket was above the atmosphere. The rocket was fired from Fort Churchill, Canada, so that the particles of interest could reach the extended detector at full intensity without having been excluded by the earth's magnetic field.

EXPERIMENTAL PROCEDURE

The nuclear emulsion detectors used in this experiment were 600-micron-thick Ilford G–5 emulsions which were assembled in eighteen packs consisting of eighteen emulsions, 6.5 cm by 9.9 cm, placed on top of each other and covered with Mylar and black electrical tape to give a water- and light-tight package. Six of these packs were placed in each tray and held in place by means of a small metal lip. A photograph of the emulsion section of the payload is shown in Fig. 1. The trays in turn were kept inside of the payload until 61 seconds after launch of the sounding rocket, at which time extension of the trays was begun. Before reentry and 411 seconds after launch, the retraction of the trays was begun. Extension and retraction each took about 7 seconds. These events and others are indicated in the time-altitude curve in Fig. 2.

The rocket itself was an Aerobee 150, which had the capability of carrying the 171-pound payload of this experiment to an altitude of 151 miles. In addition to the standard Aerobee nose cone, there were two extension sections; the one closest to the rocket was a recovery section, and the other was

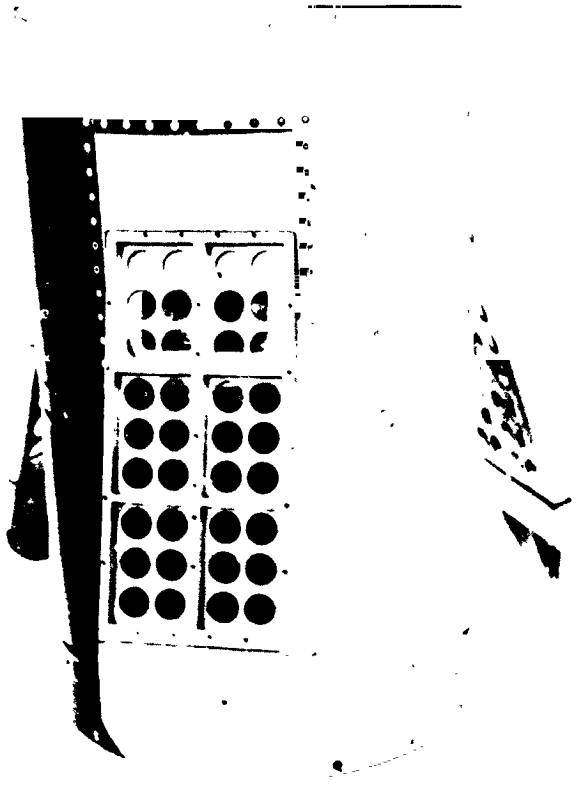


FIGURE 1.—Photograph of the rocket payload section with the emulsions removed and the emulsion trays extended. Six emulsion packs were placed in each of three symmetrically positioned trays. The angle of the trays with respect to the rocket axis was 17.5 degrees.

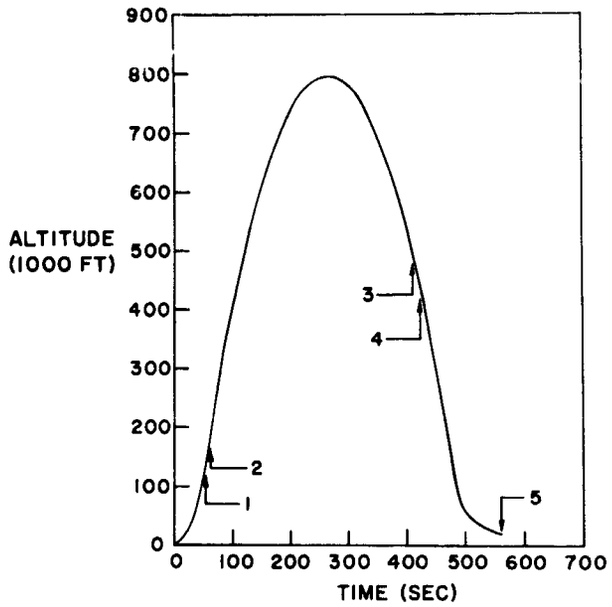


FIGURE 2.—Rocket time-altitude trajectory, showing the sequence of events: (1) rocket burnout (2) emulsion tray extension (3) emulsion tray retraction (4) payload separation (5) parachute deployment and SARAH beacon activation.

the section which housed the extension mechanisms for the emulsion trays described in the preceding paragraph. The water-tight housing for the emulsion in the retracted position extended up into the standard nose cone section. In addition, the nose cone itself contained a radar beacon, battery power for the extension and retraction motors, and a telemetry section. Magnetometer data for rocket aspect, data on the extension and retraction of the trays, and rocket parameters including acceleration and chamber pressure were telemetered.

The rocket was launched successfully at 1311 U.T. on September 4, 1963, and all parts of the payload functioned properly. It was a quiet day geophysically[10] and the Mt. Washington neutron monitor counting rate was 2318.[11] The nuclear emulsions were recovered, processed, and found to be in generally good condition. A total emulsion surface area of 821 cm^2 was available for analysis after elimination of the area held under the tray lips and the immediately adjacent area, where the clear solid angle would have been greatly reduced.

A complete area scan was made under a microscope of the top emulsion of each of the packs, as well as a complete rescan to check scanning efficiency. In the original scans, all tracks which were dark and which either had delta rays or were wider than a track formed by a single line of grains were accepted, regardless of the angle the track made in the emulsion. These tracks were then analyzed by a scientist to separate the slow proton and helium tracks from the partiles with charges greater than 2. At this point, all tracks which were within 20 degrees of the perpendicular to the emulsion were rejected. The method of charge and energy analysis is essentially the same as that used previously in other work and described in detail in a paper by Biswas, Fichtel, and Guss [12] and, therefore, will not be repeated here. Only particles which ended in the emulsion were analyzed. When this work was completed, it was found that the sample of particles, although small, was sufficiently large to permit an actual flux determination rather than simply to set an upper limit.

The calculation of the solid angle of collection involves a number of features which are indicated in Fig. 3. First, there is the restriction on the angle with respect to the plane of the emulsion (called "dip angle") mentioned above, corresponding to section A in the figure. Second, there is a very small solid angle, consisting of tracks of small dips and azimuthal angles in the emulsion which are closest to the rock t axis, which is excluded because particles would have had to pass through the rocket to enter the emulsion within this solid angle. This solid angle segment varies with position on the surface of the emulsion, but a typical segment labeled "B" is shown in Fig. 3. The final and largest portion of the solid angle which was excluded or partially excluded is made up of those angles which are forbidden to the particles by a combination of their paths along the magnetic field and the earth's atmosphere.

The method of calculating the angle of a particle with respect to the vertical for a given altitude above the earth after the particle has mirrored at a lower altitude has been discussed in an earlier paper by Biswas, Fichtel, and Guss.[12] Although the details are complicated and will not be repeated, the net effect is to increase the angle with respect to the vertical in which particles may arrive from 90 degrees to a larger angle which increases with altitude above the earth. There

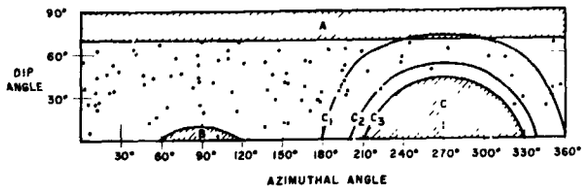


FIGURE 3.—Distribution of ending $Z \geq 6$ nuclei (indicated by black circles) in the solid angle of collection. The "dip angle" is the angle with respect to the plane of the emulsion; the "azimuthal angle" is that with respect to the perpendicular to one of the edges of the emulsion measured in the emulsion plane. The shaded area "A" was excluded from analysis because tracks in this segment had a dip too great to be analyzed. The shaded area "B" was excluded from analysis because particles at these angles had passed through the rocket material before entering the emulsion; this area varied with position in the emulsion and a typical segment is shown. The curves C_1, C_2, and C_3 are the curves for particles with space angles greater than 90°, 70°, and 60° with respect to the magnetic line of force. These curves varied somewhat as the coning angle of the rocket varied, and the ones shown are typical. The shaded area "C" is forbidden to the low-energy heavy nuclei under consideration because they would be stopped by ionization energy loss in the atmosphere between the rocket and their mirror point.

is only a very small solid angle in which particles pass through a significant amount of atmosphere, but not enough to effectively remove them from consideration. Further, the rocket was not quite vertical but had a small coning angle in addition to its spin about the principal axis; therefore, there was a small variation in the position of the set of angles in the emulsion which are at a given angle with respect to the vertical during the flight. This effect increases somewhat the solid angle in the emulsion in which primary tracks may appear but has essentially no effect on the total solid angle for collecting primary particles. Some of the curves related to this problem are shown in Fig. 3 near the area marked "C" and are labeled C_1, C_2, and C_3. The explanation of these curves is given in the figure caption. No tracks should have been observed in the shaded area C of the solid angle diagram, and none were.

The effective solid angle is then calculated from the equation

$$\Omega = \int\int (f \cdot \cos\theta) \sin\theta \, d\theta \, d\varphi,$$

where θ is the angle with respect to the perpendicular to the emulsion and φ the azimuthal angle in the emulsion: f is a weighting factor which is 0 in the shaded areas of Fig. 3 where no tracks were accepted and 1 in most of the rest of the area except around C, where it varied between 0 and 1 with position, depending on the fraction of the extended time that the primary particles could reach these angles. The cos θ factor occurs because the emulsion collecting area is a flat surface. Since the area marked "B" in Fig. 3 varies with position on the emulsion surface, Ω varies with position, but only very slightly because of the small contribution of segment B. The collection time also varies slightly with position because of the small, but finite, time required to extend the tray. With all the above factors taken into consideration, the effective primary area—collection time—solid angle factor, called "$A\,\Omega T$," was $60.7 m^2$ sr sec.

The major correction to the raw data is the one which accounts for the background tracks formed during ascent before the trays are extended and during descent after the trays are retracted. In a more recent version of this experiment flown in July 1964, this correction was eliminated by including a sliding plate mechanism which permitted separation of the tracks formed during the period that the emulsion trays were extended from those tracks made at other times. The results of this experiment will be reported later when the data reduction and analysis is complete.

The background consisted of particles of relatively large ambient energies because they had to pass through several g/cm^2 of material before reaching the emulsion, the exact amount of material depending on the angle and the height of the rocket in the atmosphere. An estimate of this background correction therefore could be made from balloon flight data because the particles composing the background had initial energies which were sufficiently great to reach balloon altitudes. For this purpose, the flux and the energy spectrum of low-energy heavy nuclei were measured in the nuclear emulsion plates carried on a balloon flight made from Fort Churchill on July 15, 1963, when the cosmic ray flux level was known to be essentially the same as on September 4, 1963, because the neutron monitor counting rates for these days differed by only 0.2 percent. The excellent

correlation between low-energy cosmic ray intensities and neutron monitor counting rates has been shown previously by McDonald and Webber.[13] In principle, the background correction could be very complex; in practice, a good approximation is obtained by simple summations because the degraded spectrum changes slowly in shape with increasing amounts of material present and, secondly, because the relative times of moving through variable amounts of residual atmosphere and remaining under a constant amount of a few g/cm² of material were about the same for the emulsions in the rocket and those on the balloon.

One final correction which had to be made to the raw data was the increase in the number of collected particles by an amount which took into account the probability that the particles might interact before ending or, in the cases of some of the higher energy particles included in the analysis, leave the stack if it entered the edge at an unfavorable angle. The upper limit to the energy was kept small in order to keep this correction small.

RESULTS AND DISCUSSION

After completion of the analysis outlined in the previous section, the results shown in Table 1

TABLE 1.—*Differential fluxes for medium nuclei and* ($10 \leq Z \leq 19$) *nuclei on September 4, 1964.*

Kinetic energy (MeV/nucleon)	$6 \leq Z \leq 9$ [p/(m² sr sec MeV/nucleon)]	$10 \leq Z \leq 19$ [p/(m² sr sec MeV/nucleon)]
50	0.0040 ± 0.0015	----------
55	----------	0.0005 ± 0.0006
90	0.0055 ± 0.0018	0.0005 ± 0.0006
130	0.0072 ± 0.0024	0.0040 ± 0.0015
170	Not measured	0.0028 ± 0.0014

were obtained. The first pointo to be made is obvious: namely, that these fluxes are quite clearly significantly different from zero. Hence, a finite flux of cosmic ray medium nuclei in an energy range as low as 30 to 70 MeV/nucleon has been observed in the vicinity of the earth. Further, there is a finite flux of heavy nuclei with energies at least as low as 110 MeV/nucleon.

In the rest of the paper, the heavy nuclei will be divided into two groups—nuclei with nuclear charges from 10 to 19, hereafter called ($10 \leq Z \leq 19$) nuclei, and very heavy nuclei ($Z \geq 20$)—because it is advantageous to keep the variation in energy loss within a charge group to within tolerable limits and because there were very few nuclei with charges of 20 or more. Comments on light nuclei ($3 \leq Z \leq 5$) will be confined to the last paragraph of this section, since they are a separate subject.

The next subject of interest is the comparison of the differential energy spectra obtained in this experiment with the differential spectra at higher energies and with the spectrum of helium nuclei. In this comparison, it is important to remember that the cosmic ray differential energy spectrum varies with the period in the solar cycle. Figure 4

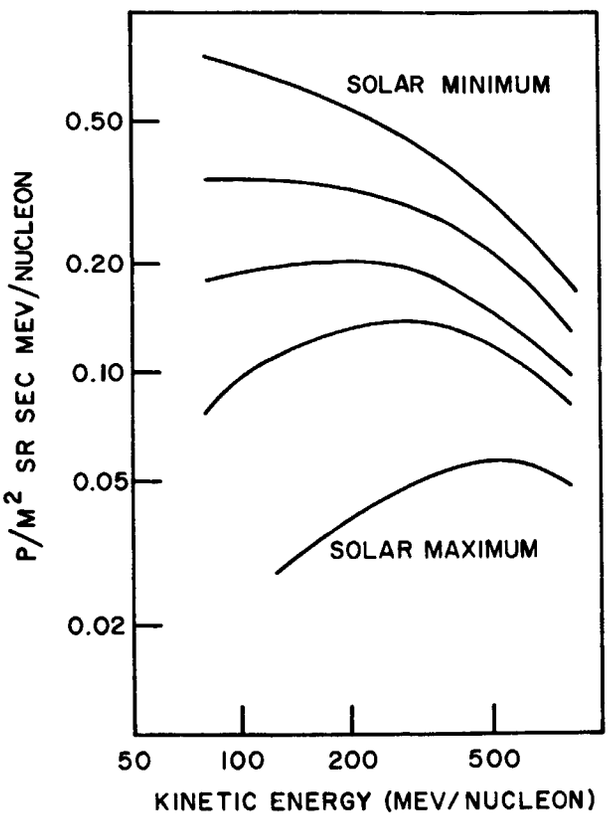


FIGURE 4.—Differential energy spectrum for helium nuclei at various times during the solar cycle. These curves were calculated from data published by Webber (see Ref. 1).

shows the differential energy spectrum for helium for various periods in the solar cycle.[1] The existing data indicate that the medium and heavy nuclei have energy spectra similar to that of the helium nuclei, only reduced in intensity.[1] The medium nuclei group, for example, is 0.063 times the helium particle intensity; and the charge group from $Z = 10$ to 19 is 0.021 times the helium particle intensity. The possibility of small differences of the order of 15 percent or less in the region from about 200 to 400 MeV/nucleon cannot be excluded; above 1.5 BeV/nucleon, they probably are correct to within 10 percent.

At the time of the firing of the sounding rocket from which the data under discussion were obtained, the cosmic ray flux had passed through the minimum value of the cycle a few years earlier and was slowly increasing. The curve in Fig. 5 shows the approximate shape of the helium particle differential spectrum at the time of the rocket shot. It has been shown[1] that the particle spectrum is a smoothly varying function of energy and that for this period in the solar cycle a 1 percent variation in the Deep River neutron monitor rate, which gives an estimate of the higher energy particle intensity, corresponds to about a 10 percent variation in the helium particle differential flux at 200 MeV/nucleon. Therefore, for purposes of correlating the measurements made here, an uncertainty of less than that will be introduced by comparing the results obtained here with helium spectra obtained when the neutron monitor was within 1 percent of the reading at the time the rocket was in the air. The authors know of four helium spectra in the energy region from about 80 to 600 MeV/nucleon that were obtained during the summer of 1963 which satisfy this condition.[14-17] An average of these spectra, which are in close agreement, was used as the basis for the curve in Fig. 5.

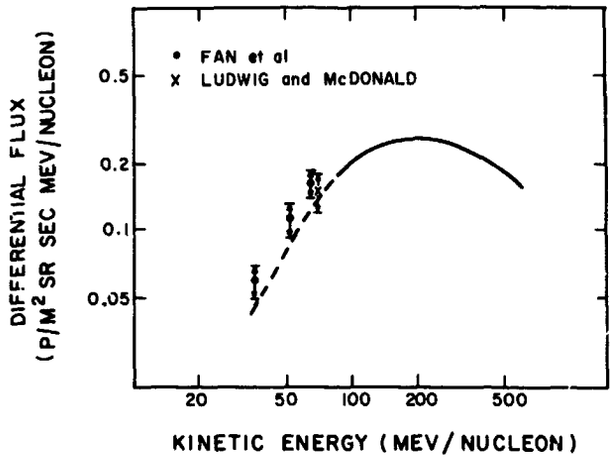


FIGURE 5.—Helium nuclei differential energy spectrum obtained as explained in the text. The low-energy points, from which the low-energy portion of the curve was deduced, are those of Fan et al, (Ref. 18) and Ludwig and McDonald (Ref. 19).

In addition, Fan et al[18] and Ludwig and McDonald[19] have obtained a helium energy spectrum in the 30 to 80 MeV/nucleon region on Explorer XVIII during the period January through March 1964, when the neutron monitor rate was typically from 1 to 2 percent higher than the rate at the time of the measurement of this experiment. To compensate for a systematic change in intensity in this region, the curve in Fig. 5 was extended below 80 MeV/nucleon by a smooth connection to data above 80 MeV/nucleon, by keeping the shape of the 30 to 80 MeV/nucleon data but reducing the intensity appropriately. This procedure could, at most, introduce a noticeable error only in the lowest medium nuclei energy interval.

Since no comparable data were available for the heavier particles at the time of writing of this paper, the helium particle curve was multiplied by the ratios mentioned above to obtain the best possible approximation of the curves for the heavier particles at higher energies.

The results displayed in Figs. 6 and 7 indicate that the low-energy spectrum of the medium nuclei observed in the vicinity of the earth falls below that of the helium nuclei multiplied by 0.063. Similarly, the $(10 \leq Z \leq 19)$ nuclei curve falls below the corresponding one for helium nuclei multiplied by 0.021. Hence, the abundance of helium nuclei relative to these higher charge groups apparently increases in the low-energy region until it is well above the fairly constant value which it has from about 400 MeV/nucleon to very high energies.

The spectra which are observed at the earth represent the source spectra after they have passed through interstellar matter and have been modulated within the solar system. Whereas

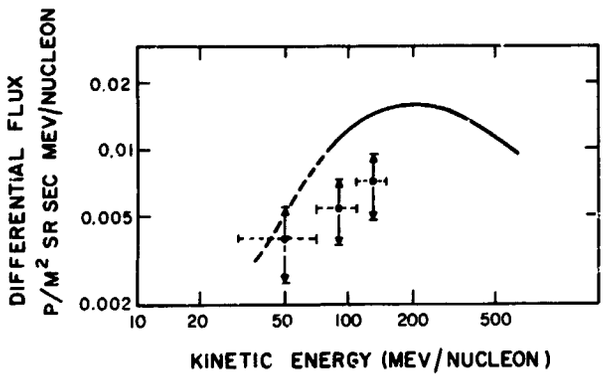


FIGURE 6.—Differential energy spectrum for medium nuclei. The curve is that for helium nuclei (Fig. 5) multiplied by 0.063.

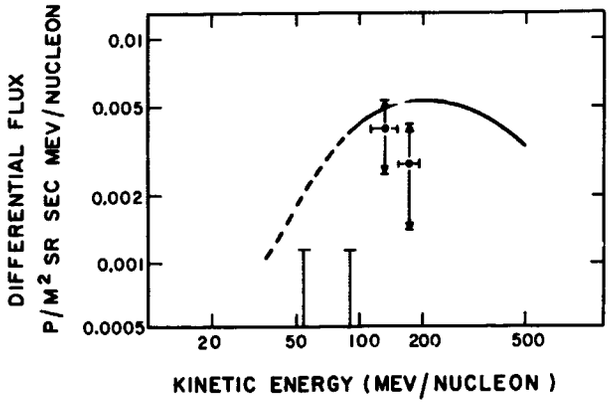


FIGURE 7.—Differential energy spectrum for $(10 \geq Z \geq 19)$ nuclei. The curve is that for helium nuclei (Fig. 5) multiplied by 0.021. For the two lowest energy points, only the upper limit corresponding to one standard deviation is shown.

the solar system modulation affects only the intensity of the increment of flux in a given energy interval, interstellar space contains enough material along the path of the particle to change appreciably the particle energy as well as the intensity. In the latter case, it normally is assumed that the intensity is changed significantly only by fragmentation in interactions and not by the complicated time-dependent magnetic effects which cause the intensity variation in the solar system. Therefore, the differential energy flux at the source for the i^{th} type of particle $j_s(i)$ is related to the differential energy flux observed in the vicinity of the earth but outside the region where the earth's magnetic field excludes, or partially excludes, particles $j_o(i)$ by the relation

$$j_o(i)[_{E=E_0} = j_s(i)]_{E=E_s(i)} \left[\frac{\Delta E_s(i)}{\Delta E_0}\right] h(v_0,R) f(i,v,\rho), \tag{1}$$

where ΔE_0 and ΔE_s represent the small energy increments containing the set of particles at the observation point and the source respectively, i is the type of particle, h the solar modulation function, f the interstellar intensity variation factor, v the particle velocity, R the rigidity (momentum/unit charge), and ρ the amount of interstellar matter traversed between the source and the earth.

The expression $f(i, v, \rho)$ is well known and is given in detail in a paper by Hayakawa.[20] However, the parameters to be substituted into the equations are not known exactly. The parameters include the mean free paths of the different elements in space, the probability of one type of particle emerging from an interaction caused by another particle and the amount of material traversed. Table 2 gives the values of the parameters used in the calculation and the references from which the parameters were obtained.[21–24] When more than one reference is given, the parameter listed in the Table is a weighted average. In Table 2, P_{ij} gives the average number of secondaries of type "j" formed in an interaction of a particle of type "i" in an interaction with a hydrogen nucleus. λ_i' is the absorption mean free path, which is given by the equation

$$\frac{1}{\lambda_i'} = \frac{1}{\lambda_i}(1 - P_{ii}), \tag{2}$$

where λ_i is the interaction mean free path.

TABLE 2.—*Parameters used for extrapolation through interstellar matter.*

	Reference		References
$\lambda_a = 14.6$ g/cm²	20	$P_{aa} = 0.07 \pm 0.07$	21–24
$\lambda_M = 6.0$ g/cm²	20	$P_{(L+M+H)a} = 1.3 \pm 0.5$	21–24
$\lambda_H = 4.0$ g/cm²	20	$P_{MM} = 0.14 \pm 0.04$	21–24
		$P_{HM} = 0.21 \pm 0.10$	21–24
		$P_{HH} = 0.40 \pm 0.15$	21–24

The above parameters are known to vary with energy; however, their exact dependence is not well known. Above approximately 100 MeV/nucleon they are thought to be nearly constant and not to vary appreciably until the energy/nucleon is below about 30 or 60 MeV.[23 25] To reach the earth at the observed energies after passing through several g/cm² of interstellar hydrogen, the particles under consideration must spend either all or almost all of their time in interstellar space at energies above 100 MeV/nucleon. Therefore, it seems reasonable to use the parameters obtained at higher energies in view of the above considerations and the lack of sufficient information to calculate more exact values. Including the generally accepted partially tested hypothesis that the energy per nucleon does not vary significantly in an interaction leads, then, to the conclusion that f is not a function of velocity in the region of interest if ρ is not a function of energy.

The true modulation function $h(v, R)$ is not known. Many models have been presented, and each seems to have some advantages and some weaknesses.[26]: For the purposes of the present discussion, it is sufficient to note that, since all the proposed modulation mechanisms involve only magnetic or electric fields and the amount of material traversed within the solar system is negligible, particles with the same charge to mass ratio and hence the same rigidity for a given velocity will be depressed by the same amount for a given velocity.

The remaining term on the right side of Eq. (1), which multiplies $j_s(i, E)$, is $(\Delta E_s(i)/\Delta E_0)$. This term arises from the change in the width of the energy interval in which particles are contained as they lose energy. Since the rate of energy loss per nucleon varies with the nuclear species, this term will affect the different nucleon groups in different ways. Further, because of the different rates of energy loss for a given E_0, $E_s(i)$ will be different for different nuclear species; and hence $j_s(i, E)]_{E=E_{s(i)}}$ can vary for a fixed E_0 even if $j_s(i, E)$ is the same for all nuclear groups.

Consider now the ratio of the differential flux values of two different nuclear types with the same charge to mass ratio. From Eq. (1), this ratio is

$$\frac{j_0(i,E)]_{E=E_0}}{j_0(k,E)]_{E=E_0}} = \frac{j_s(i,E)]_{E=E_s(i)}}{j_s(k,E)]_{E=E_s(k)}} \frac{[\Delta E_s(i)/\Delta E_0]}{[\Delta E_s(k)/\Delta E_0]} \frac{f(i,v,\rho)}{f(k,v,\rho)} \tag{3}$$

If ρ, the fragmentation parameters, and the mean free paths are independent of velocity, $f(i, v, \rho)/f(k, v, \rho)$ becomes a constant independent of velocity. Hence,

$$\frac{j_0(i,E)]_{E=E_0}}{j_0(k,E)]_{E=E_0}} = \frac{j_s(i,E)]_{E=E_s(i)}}{j_s(k,E)]_{E=E_s(k)}} \frac{[\Delta E_s(i)/\Delta E_0]}{[\Delta E_s(k)/\Delta E_0]} [K(\rho,i,k)],\ [K \neq g(V)] \tag{4}$$

Notice also that

$$\Delta E_s(i) \lambda \Delta E_0 \xrightarrow[v \to c]{} 1 \tag{5}$$

and, for reasonable spectra,

$$\{j_s(i, E)]_{E=E_{s(i)}}\}/\{j_s(k, E)]_{E=E_{s(k)}}\} \xrightarrow[v \to c]{} \text{(Const.), for a fixed } E \tag{6}$$

The implications of the experimental results now will be examined by comparing the data with the predictions based on several suggested source spectra and interstellar mean free paths. If it is assumed first that Eq. (4) is valid, the expected ratio for $j_0(i, E)]_{E=E_0}$ and $j_0(k, E)]_{E=E_0}$ can be calculated for different values of ρ and the experimental knowledge of the limiting value for the ratio at high energies, where Eqs. (5) and (6) apply.

The most commonly assumed source spectrum is a power law in the total energy with an exponent of 2.5 as given by

$$j_1(i, W) = K_1(i)/W_N^{2.5}, \tag{7}$$

where W_N is the total energy/nucleon. In Figs. 8 and 9 the observed ratios of the differential

medium nuclei and the ($10 \leq Z \leq 19$) nuclei to the differential helium nuclei spectral points obtained in the manner described earlier are plotted and compared with the ratio expected on the basis of the observed high-energy ratio and an assumed spectral shape of the form of Eq. (7). The best recent value for the estimate of the amount of interstellar matter traversed—at least for high-energy particles—is 2.5 g/cm², which is based on a calculation by Badhwar et al[23] and Badhwar and Daniel[25] with heavy emphasis on the high-energy (>1.5 BeV/nucleon) composition data of O'Dell et al.[27] The expected ratio has been calculated for both 2.5 and 5.0 g/cm² of interstellar hydrogen. Notice that this ratio is relatively insensitive to a change of a factor of 2 in interstellar matter in this range of values. Notice also that the agreement between the experimental points and the curves for these assumed conditions is satisfactory.

As a second example, a source spectrum of the form of Eq. (7) above 300 MeV/nucleon and of the form of Eq. (8) below 300 MeV/nucleon is chosen:

$$j_2(i) = K_2(i)/(E_s)^{0.67} \qquad (8)$$

The expected ratios for this source spectrum, the observed high-energy ratios, and 2.5 and 5.0 g/cm² of interstellar matter also have been calculated and are shown in Figs. 8 and 9. Here again there is a relatively small difference in the curves for 2.5 and 5.0 g/cm², and there is satisfactory agreement with the experimental data. Smooth changes in the source spectrum in general have little effect on the resultant ratios unless the variations are very large, that is, appreciably greater than those selected here.

Hence, the following conclusion can be drawn: The experimentally observed helium to medium and helium to ($10 \leq Z \leq 19$) nuclei ratios are consistent with the assumption that the source spectra are the same and that the particles have passed through the same amount of material, which is in the range of 2.5 to 5 g/cm². This conclusion is relatively independent of the exact shape of the source spectrum. On the other hand, differences in the source spectra between helium and medium nuclei would appear relatively quickly in the form of a disagreement

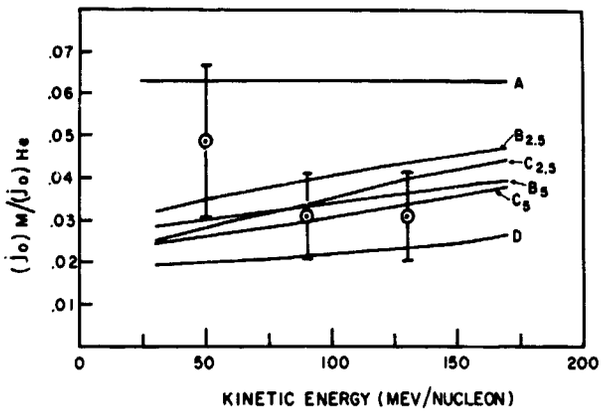


FIGURE 8. -Ratio of the differential flux of medium nuclei to helium nuclei expected after passage of the particles through various amounts of interstellar gas. Curve A: 0 g/cm² for all spectral shapes; curve $B_{2.5}$ and curve B_s: 2.5 g/cm² and 5.0 g/cm², respectively, assuming the source spectrum of Eq. (7); curve $C_{2.5}$ and curve C_s: 2.5 g/cm² and 5.0 g/cm², respectively, assuming the source spectrum of Eq. (7) for particles with kinetic energy greater than 300 Mev/nucleon and the source spectrum of Eq. (8) for particles with kinetic energy less than 300 Mev/nucleon; curve D: passage through 2.5 g/cm² for relativistic particles, 6 g/cm² for particles with kinetic energy less than 400 Mev/nucleon, assuming the source spectrum of Eq. (7).

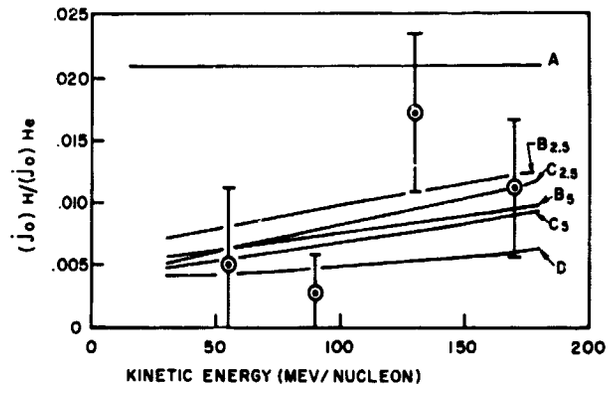


FIGURE 9.—Ratio of the differential flux of ($10 \geq Z \geq 19$) nuclei to helium nuclei expected after passage of the particles through various amounts of interstellar gas. Curve A: 0 g/cm² for all spectral shapes; curve $B_{2.5}$ and curve B_s: 2.5 g/cm² and 5.0 g/cm², respectively, assuming the source spectrum of Eq. (7); curve $C_{2.5}$ and curve C_s: 2.5 g/cm², and 5.0 g/cm², respectively, assuming the source spectrum of Eq. (7) for particles with kinetic energy greater than 300 Mev/nucleon and the source spectrum of Eq. (8) for particles with kinetic energy less than 300 Mev/nucleon; curve D: passage through 2.5 g/cm² for relativistic particles, 6 g/cm² for particles with kinetic energy less than 400 Mev/nucleon, assuming the source spectrum of Eq. (7).

between the calculated and observed helium to medium nuclei ratio as a function of energy.

The results obtained here, then, suggest there is good reason to think that above about 0.2 BeV/nucleon the source spectra of all components are at least similar. Information below this energy is not forthcoming from this approach because nuclei of the higher charges being considered must have at least this energy initially to reach the earth.

Next, assume that the particles have gone through different amounts of interstellar material depending on their energy. Dahanayake et al,[28] for example, suggest that the lower energy particles, below about 400 MeV/nucleon, have passed through more material than the high-energy ones, which are assumed to go through 2.5 g/cm^2 on the basis of the work mentioned earlier. In this case, $f(i, v, \rho)/f(j, v, \rho)$ of Eq. (3) is not a constant independent of energy. However, if the high-energy ratio of the observed fluxes is known and if the parameters of Table 2 are used, the additional expected suppression resulting from low-energy particles going through more material than high-energy ones can be calculated. The results obtained, assuming that all particles have a source spectrum of the type given by Eq. (7) and that the particles below 400 MeV/nucleon have passed through 6 g/cm^2 (as suggested by Dahanayake et al,[28]) while the high-energy ones (kinetic energy$>$1.5 BeV/nucleon) have passed through 2.5 g/cm^2, are also shown in Figs. 8 and 9. Here the agreement is poorer; further, these assumptions lead to a ratio in the 300 to 400 MeV/nucleon region which is appreciably more than one standard deviation below several measurements of the medium to heavy nuclei ratio in that region. In addition, Hildebrand and Silberberg [29] and Webber [30] now have shown that the existing experimental data on the He_3/He_4 ratio probably can be reconciled with a mean free path in the low-energy region which is the same or only slightly larger than 2.5 g/cm^2, by taking into account the properties of secondaries from interactions and the effects of the solar modulation mechanism on particles with different charge to mass ratios.

Light nuclei also were observed in the low-energy region from 30 to 110 MeV/nucleon, but no quantitative value for the differential flux will be quoted because a high scanning detection efficiency was not achieved for these particles. Since these light nuclei generally are assumed to arise from heavier nuclei, the type of theoretical analysis outlined above does not apply. There is, however, the alternate problem of interest: namely that, if the modulation effect is rigidity dependent, light nuclei might be expected to have slightly different energy spectra from the medium or heavy nuclei. If the rigidity dependence of the modulation can be determined by other means, the relative abundance of light and medium nuclei provide an independent estimate of the amount of interstellar matter traversed by the cosmic radiation at these low energies. The analysis was not pressed beyond the determination of the existence of a light nuclei flux because there was not an adequate number of particles to determine the light to medium ratio with sufficient accuracy to see a deviation from the high-energy ratio; and the additional work involved is tremendous. With an improved technique and a higher flux rate, we hope to be able to measure the relative abundance of light nuclei in the 1964 flight to be reported later.

CONCLUSIONS

The answer to the basic question of whether or not there are low-energy heavy nuclei below the energy cutoff set by material above detectors flown on balloons has been seen to be "Yes." There definitely is a finite flux of medium nuclei in an energy range as low as 30 to 70 MeV/nucleon and ($10 \leq Z \leq 19$) nuclei with energies at least as low as 110 MeV/nucleon. The abundance of medium and ($10 \leq Z \leq 19$) nuclei relative to helium nuclei in the energy regions from 30 to 150 and 30 to 190 MeV/nucleon, respectively, are less than the relative abundances at higher energies. This difference can be explained quantitatively by the higher rate of energy loss of the particles of higher charge in the interstellar matter. It also was shown that within the statistical uncertainty the resulting differential flux measurements are consistent with the helium, medium, and ($10 \leq Z \leq 19$) nuclei having the same source spectrum at least above about 0.2 BeV/nucleon for a wide range of source spectral shapes, including ones normally assumed. For a mean free path independent of energy this result also is

fairly insensitive to the values of cross sections and fragmentation parameters assumed, principally because relative values of ratios are being considered. The conclusion is independent of the solar modulation mechanism because, since all particles considered have the same charge to mass ratio, their relative abundances at a given velocity will be unaffected by the local solar modulation. No information can be obtained about the source spectrum below about 0.2 BeV/nucleon because the particles of high charge must have approximately this energy to reach the earth after passing through interstellar matter.

It has been established that there is a finite flux and that the differential flux is consistent with similar source spectra for helium, medium, and heavier nuclei and an interstellar path in hydrogen of 2.5 g/cm^2 within the relatively large uncertainties of these initial measurements. It now seems worthwhile to measure these spectra again to look for a possible variation with the period in the solar cycle and to examine the matters discussed in this paper in greater detail. This problem is being pursued.

REFERENCES

[1] W. R. Webber, Progress in Elementary Particle and Cosmic Ray Physics VI (North Holland Publishing Co., Amsterdam) p. 77 (1962).

[2] Proceedings of the International Conference on Cosmic Rays, Jaipur, India, Commercial Printing Press Limited, Bombay (1963).

[3] Proceedings of the International Conference on Cosmic Rays and the Earth Storm, Kyoto, 1961, J. Phys. Soc. Japan **17**, Suppl. A-III (1962).

[4] H. Aizu, Y. Fujimoto, S. Hasegawa, M. Koshiba, I. Mito, J. Nishimura, and K. Yokoi, J. Phys. Soc. Japan **17**, Suppl. A-III, 38 (1962).

[5] D. E. Evans, Nuovo Cimento **27**, 394 (1963).

[6] F. Foster and A. Debenedetti, Nuovo Cimento **28**, 1190 (1963).

[7] C. Fichtel, Nuovo Cimento **19**, 1100 (1961).

[8] F. B. McDonald and W. R. Webber, J. Geophys. Res. **67**, 2119 (1962).

[9] See, for example, F. B. McDonald and W. R. Webber, J. Geophys. Res. **69**, 3097 (1964).

[10] J. Virginia Lincoln, J. Geophys. Res. **69**, 525 (1964).

[11] Courtesy of the Enrico Fermi Institute for Nuclear Studies, Chicago.

[12] S. Biswas, C. E. Fichtel, and D. E. Guss, Phys. Rev. **128**, 2756 (1962).

[13] F. B. McDonald and W. R. Webber, J. Phys. Soc. Japan **17**, Supp A2, 428 (1962).

[14] V. K. Balasubrahmanyan and F. B. McDonald, J. Geophys. Res. **69**, 3289 (1964).

[15] C. E. Fichtel, D. E. Guss, D. A. Kniffen, and K. A. Neelakantan, work to be submitted to Phys. Rev.

[16] J. Ormes and W. R. Webber, Phys. Rev. Letters **13**, 106 (1964).

[17] P. S. Freier and C. J. Waddington, Phys. Rev. Letters **13**, 108 (1964).

[18] C. Y. Fan, G. Gloeckler, and J. A. Simpson, Trans. Am. Geophys. Union, Vol. **45**, 8 Sept. 1964.

[19] G. H. Ludwig and F. B. McDonald, to be submitted for publication in Phys. Rev. Letters.

[20] S. Hayakawa, Progr. Theoret. Phys. (Kyoto) **15**, 111 (1956).

[21] M. W. Friedlander, K. A. Neelakantan, S. Tokunaga, G. R. Stevenson, and C. J. Waddington, Phil Mag **8**, 1691 (1963).

[22] H. Aizu, Y. Fujimoto, S. Hasegawa, M. Koshiba, T. Mitto, J. Nishimura, and K. Yokoi, Progr. Theoret. Phys. (Kyoto), Suppl. 16 (1960).

[23] G. D. Badhwar, R. R. Daniel, and B. Vijayalakshmi, Progr. of Theoret. Phys. (Kyoto) **28**, 607) (1962).

[24] S. Hayakawa, K. Ito and Y. Terashima, Progr. Phys. (Kyoto), Suppl. **6**, 1 (1958).

[25] G. D. Badhwar and R. R. Daniel, Progr. Theoret. Phys. (Kyoto) **30**, 613 (1963).

[26] For a discussion of this problem see, for example, Ref. 9 or C. E. Fichtel, D. E. Guss, G. R. Stevenson and C. J. Waddington, Phys. Rev. **133**, B818 (1964).

[27] F. W. O'Dell, M. M. Shapiro, B. Stiller, J. Phys. Soc. Japan **17**, Suppl. A-III, 23 (1962).

[28] C. Dahanayake, M. F. Kaplon, and P. J. Lavakare, J. Geophys. Res. **69**, 3681 (1964).

[29] B. Hildebrand and R. Silberg, Regional IQSY Symposium, Buenos Aires, Aug. 3-8, 1964.

[30] W. R. Webber, Regional IQSY Symposium, Buenos Aires, Aug. 3-8, 1964.

MODULATION OF LOW ENERGY GALACTIC COSMIC RAYS*

C. E. FICHTEL, D. E. GUSS, D. A. KNIFFEN, AND K. A. NEELAKANTAN†

The exact way in which the cosmic ray energy spectrum changes with solar activity has been of particular interest because the experimental determination of this feature provides one of the most severe tests of any theory related to cosmic ray modulation. The helium and proton components in addition to being the most abundant also have the additional advantage of having charge to mass ratios which differ by a factor to two. Therefore, having a markedly different velocity for a given rigidity, these particles provide a means for separating rigidity and velocity effects.

In order to continue the study of the problem of the variation of the galactic cosmic ray proton and helium nuclei spectra, a series of balloon flights was made from Fort Churchill, Canada by Goddard beginning in 1961. It is the aim of this paper to present the results of the nuclear emulsion studies of the hydrogen and helium nuclei obtained in 1962 and 1963 and compare them to those obtained in 1961[1] and the earlier results of many experimentalists from the period of increasing solar activity.

One method of comparing the period of rising cosmic ray flux with the declining phase is to plot the integral helium nuclei flux above 250 MeV/n as a function of the neutron monitor rate which gives some measure to the higher energy cosmic rays of all charges. Figure 1 shows the smooth curve determined from many points for this period by McDonald and Webber.[2] Also shown on this figure are the points obtained by the Goddard emulsion group in 1961, 1962, and 1963. As can be seen, the points agree well with the smooth curve.

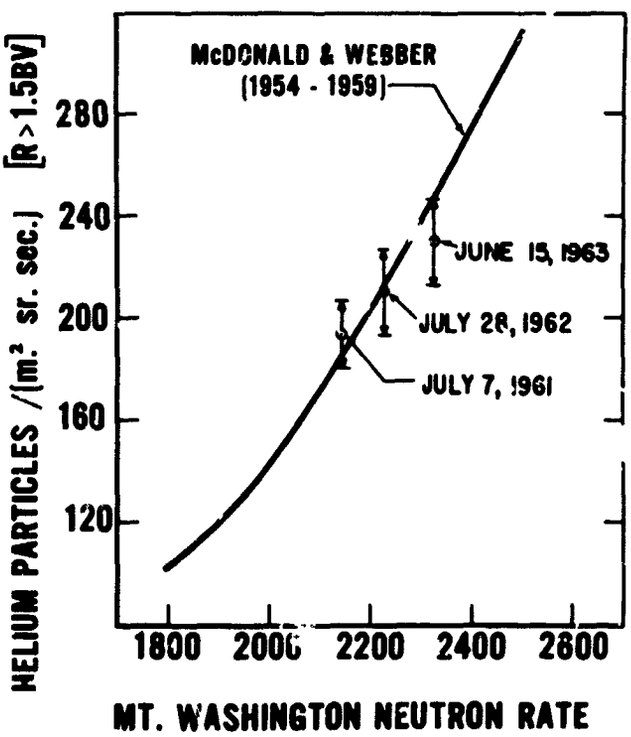


FIGURE 1.—Integral helium particle flux as a function of the Mt. Washington neutron rate.

A much more detailed and stringent comparison between the two periods is the comparison of the differential energy spectra of the helium nuclei in the present experiment with those obtained for the same neutron monitor rate during the declining phase of cosmic ray flux. Figure 2 shows the spectra for various neutron monitor counting rates during the declining phase of cosmic ray activity as given in Webber's review article.[2] These curves fit both the proton data and the helium nuclei data multiplied by seven. In general, the helium nuclei data extended down to about 1.2 BV/Nucleon and the proton data

*Published as *Goddard Space Flight Center Document X-611-64-63.*

†Goddard Space Flight Center, on leave of absence from Tata Institute of Fundamental Research, Bombay, India

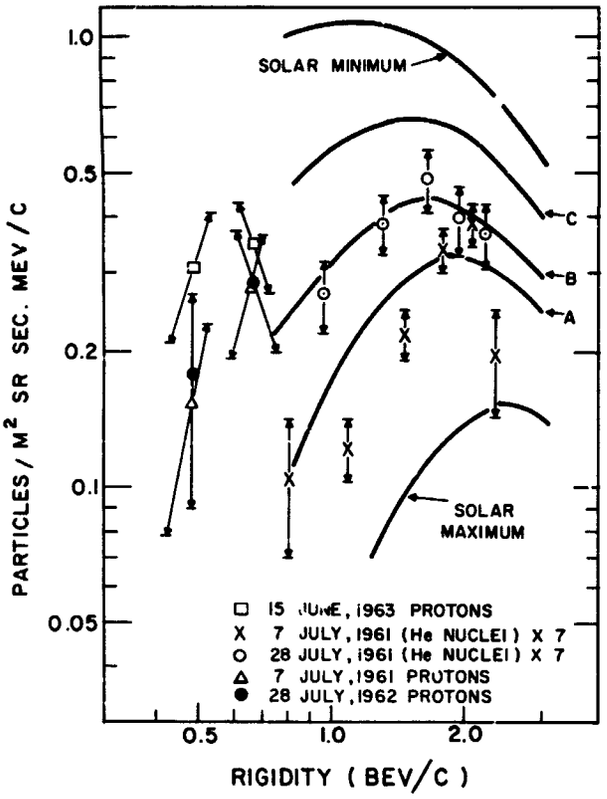


FIGURE 2.—Differential proton and helium fluxes as a function of rigidity. The solid curves are taken from figure 31a of Webber's review article and represent either the differential proton flux or the helium nuclei flux multiplied by seven for solar maximum. Curves A, B, and C correspond to times when the neutron monitor rates were the same as on July 7, 1961, July 28, 1962, and June 15, 1963, respectively.

extended from 0.8 to 1.4 BV. As can be seen, there is reasonable agreement between the helium nuclei data and the curves for the declining cosmic ray phase at the same neutron monitor rates although there is a suspicion that the helium points from this work may fall below the curve at the lowest rigidities where measurements were made. In order to obtain a more quantitative impression of the change in spectral flux from 1961 to 1962, the following numbers can be compared. The integral flux for helium nuclei with energies greater than 250 MeV/nucleon was 206 ± 13 particles/(m² sr.sec.) in 1961 and 221 ± 18 particles/(m² sr. sec.) in 1962. For the energy interval from 75 to 250 MeV/nucleon the flux was 12.7 ± 1.7 particles/(m² sr.sec.) in 1961 and 32.5 ± 4.1 particles/(m² sr.sec.) in 1962. Hence, between 1961 and 1962 there was a significant change in the helium particle flux below 250 MeV/nucleon and relatively little above.

Turning now to the protons it is known that there is a substantial secondary proton flux from interactions in the atmosphere above the detector. There is not space to describe the corrections in detail here, but the discussion of the method along with self-consistency arguments has been given in some detail in the earlier paper and will be treated further in the final publication of this work. Allowance has been made for the uncertainty in the interaction correction by drawing appropriately large errors in figure 2. This figure shows that there is a general tendency for the proton flux to increase during the period from 1961 to 1963, and that at least 1961 and 1962 the proton points lie above a reasonable extension of a smooth curve through the helium nuclei data multiplied by seven.

There is another set of data for which the analysis has just been completed which bears on this problem. Balasubraymanyan and McDonald [3] obtained a differential helium nuclei spectrum during the same period in 1963 as the proton data of this work. It agrees well with curve C above about 1.5 BV, but falls more quickly below this rigidity. These data are shown in figure 3, along with the results of this work in 1963 and

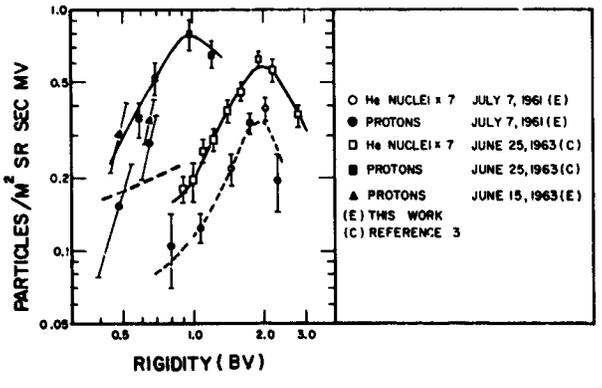


FIGURE 3.—Differential proton and helium fluxes compared to those of Balasubrahmanyan and McDonald.[3]

1961. The 1962 results lie in between those of 1961 and 1963, and were not included only to avoid unnecessarily complicating the figure. All these data are seen to be consistent with there being a difference between the proton differential

rigidity spectra and the helium one multiplied by seven below about 1.2 BV.

Hence, these combined results suggest strongly that there is a splitting of the proton differential spectrum and the helium nuclei one multiplied by seven below about 1.3 BV. This result does not contradict the earlier statements concerning the agreement in the rigidity spectra with those of the declining phase of cosmic rays because there is not sufficient helium data below 1.2 BV during this earlier period to justify a comparison in the 0.4 to 1.2 BV rigidity region. Also, although there is certainly the problem of not knowing the particle spectrum outside of the solar system, for the galactic spectra normally suggested, the splitting effect is at least qualitatively in agreement with the prediction of a modulation of the type proposed by Parker, since the protons have a higher velocity for a given rigidity than the helium nuclei and are therefore less suppressed.

REFERENCES

1. C. E. Fichtel, D. E. Guss, G. R. Stevenson, and C. J. Waddington, "Cosmic Ray Hydrogen and Helium Nuclei during a Quiet Time in 1961," to be published in the Physical Review.

2. For a summary of this work, see W. R. Webber, Progress in Elementary Particle and Cosmic Ray Physics Vol. VI, (1962). See particularly fig. 31a.

3. V. K. Balasubrahmanyan and F. B. McDonald, preceding paper, International Conference on Cosmic Rays, 1963.

MODULATION OF LOW-ENERGY GALACTIC COSMIC-RAY HYDROGEN AND HELIUM*

C. E. FICHTEL, D. E. GUSS, D. A. KNIFFEN, AND K. A. NEELAKANTAN†

A study of the energy spectrum of the cosmic-ray hydrogen and helium nuclei as a function of the period in the solar cycle provides one of the better means of gaining insight into the mechanism that modulates the galactic cosmic radiation. An examination of these nuclei in the low-energy region, where the modulation effects are greatest, has the particular merit of permitting the separation of velocity and rigidity effects, since for a given rigidity the helium nuclei have a markedly different velocity from that of the protons. Considerable information related to this problem had been obtained before 1960 for the most recent period of increasing solar activity [*Webber*, 1962].

In order to continue the study of the problem of the variation of the galactic cosmic-ray proton and helium nuclei spectrums and extend it to lower energies, a series of balloon flights, beginning in 1961, has been made from Fort Churchill, Canada. It is the aim of this letter to present the results of the nuclear emulsion studies of the hydrogen and helium nuclei obtained in 1962 and 1963 and compare them to those obtained in 1961 [*Fichtel et al.*, 1964].

The experimental data were obtained by carrying sets of nuclear emulsions aboard high-altitude skyhook balloons flown from Fort Churchill, Canada, where the earth's magnetic field allows all particles in the rigidity interval being studied to reach the top of the atmosphere. A second nuclear emulsion stack is released at the time the primary stack is rotated into the exposure position. The second stack provides the necessary information on the background particles collected during ascent and descent which must be subtracted from the particle flux measured in the primary stack, since nuclear emulsions are an integrating device.

There is not space here to describe in detail the corrections which must be applied to the data, but the discussion of the method along with self-consistency arguments has been given in some detail in an earlier paper [*Fichtel et al.*, 1964] and will be treated further in the final publication of this work. It is appropriate to mention, however, that there is a substantial secondary proton flux from interactions in the atmosphere above the detector, and allowance has been made for the uncertainty in the interaction correction in the calculation of the errors associated with the flux values that are quoted below.

In earlier work during the years 1955 through 1960, McDonald and Webber had shown that higher-energy data at higher-rigidity geomagnetic cutoffs suggested that the proton rigidity spectrum was the same as that for helium multiplied by seven [*Webber*, 1962]. Figure 1 shows that in the present data there is a general tendency for the proton points to lie above a reasonable extension of a smooth curve through the helium nuclei data multiplied by seven both in 1961 and 1962. In Figure 2 the 1963 proton data from the nuclear emulsions are plotted together with the proton and helium data of *Balasubrahmanyan and McDonald* [1964] at higher rigidity. Their data were obtained from counter experiments flown on June 24, 1963, when the high-latitude neutron monitor readings differed from those on June 15 and 16, 1963, when the emulsions were flown, by much less than 1 percent. In Figure 2 the splitting between the rigidity spectrums of the two components is quite marked. Hence, these results suggest strongly that there is

*Published in *Journal of Geophysical Research*, 69(15):3293-3295, August 1, 1964.

†Goddard Space Flight Center NAS-NASA Research Fellow, on leave from Tata Institute of Fundamental Research, Bombay, India.

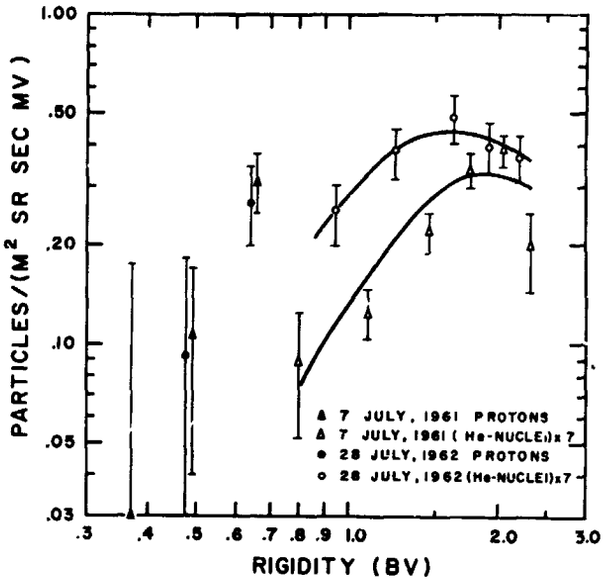


FIGURE 1.—Proton and He nuclei differential rigidity spectrums from 1961 and 1962 balloon flights at Fort Churchill.

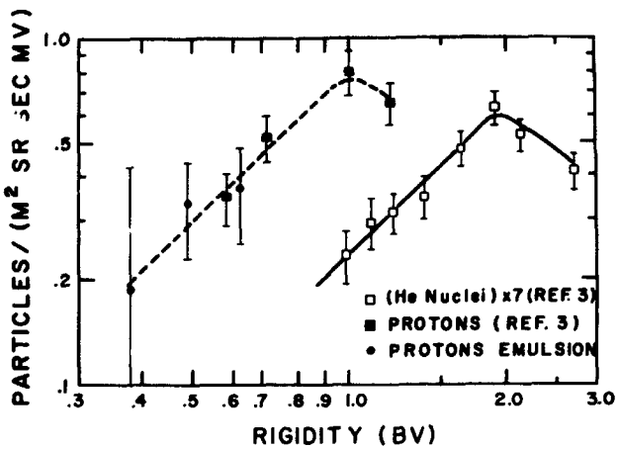


FIGURE 2.—Proton and He neclei differential rigidity spectrums from 1963 balloon flight from Fort Churchill.

a splitting below about 1.3 bv between the proton differential spectrum and that for helium nuclei multiplied by seven. This conclusion is not inconsistent with the earlier statements [*Fichtel et al.*, 1964] about the agreement in the rigidity spectrums of the proton and helium components during the declining phase of cosmic rays because there is not sufficient helium data below 1.3 bv during the earlier period to justify a comparison in the 0.4– to 1.3–bv rigidity region.

Although there is the problem of not knowing the particle spectrum outside the solar system, the splitting effect for the galactic spectrums normally suggested is at least qualitatively in agreement with the prediction of the modulation mechanisms currently proposed. Most of these models predict that protons will be less suppressed because they have a higher velocity for a given rigidity than helium nuclei. Therefore, unless the proton rigidity spectrum is very different from the helium spectrum in free space, the type of splitting in the spectrum which is observed would be predicted by these models. A more complete discussion of modulation models can be found elsewhere [*Webber*, 1962; *Fichtel et al.*, 1964].

REFERENCES

BALASUBRAHMANYAN, V. K., and F. B. MCDONALD, Solar modulation effects on the primary cosmic radiation near solar minimum, *J. Geophys. Res.*, **69**(15), August 1, 1964.

FICHTEL, C. E., D. E. GUSS, G. R. STEVENSON, and C. J. WADDINGTON, Cosmic ray hydrogen and helium nuclei during a quiet time in 1961, *Phys. Rev.*, **133**, B818, 1964.

WEBBER, W. R., *Prog. Elem. Particle Cosmic Ray Phys.*, **6**, 1962 (see particularly Figure 31*a*).

MAGNETIC EFFECTS OF THE QUIET TIME PROTON BELT*

R. A. HOFFMAN AND P. A. BRACKEN

An analysis of the magnetic effects of the quiet time proton belt has been performed with data obtained from a proton detector aboard Explorer XII. From the measured energy spectra, intensities, and pitch angle distributions of the protons with energies above 100 kev, the current density on a magnetic meridian plane was calculated using the theory of Akasofu and Chapman, modified to eliminate several simplifying assumptions. The algebraic sum of all the currents was 0.59 million amperes, and the magnetic moment of the current loop was $0.029 M_E$. From the electric current distribution the magnetic perturbations on the meridian plane were obtained. At the magnetic equator on the earth's surface the proton belt produced a 9 γ decrease in the field in comparison with the 38 γ decrease reported by Akasofu, Cain and Chapman based on some preliminary data from the same detector. The maximum perturbation appeared at $L=3.6$ on the magnetic equator, where the field attained a value of -23 γ.

I. INTRODUCTION

The various motions which a group of charged particles exhibit when trapped in the earth's magnetic field result in electric currents. The analysis of these movements yields the observation that the net currents are parallel to lines of magnetic latitude: in some regions of space they flow eastward, and in other regions westward. If one assumes longitudinal symmetry about the earth, these motions necessarily correspond to a toroidal ring current encircling the earth.

On the basis of geomagnetic evidence it has long been postulated that some type of ring current exists. The enhancement of the current at times was thought to be one cause of magnetic activity, but even during magnetically quiet times the current was never supposed to die away completely.

Thus, since the advent of the satellite era and the discovery of the earth's trapped radiation, there has been considerable expectation that the particles which constitute the ring current would be directly observed, and their species, energies and spatial distributions determined. In addition, magneticians have sought the location of the currents by attempting to measure their magnetic effects with satellite borne magnetometers. While the trapped radiation was discovered over five years ago, the endeavor to observe any significant ring current has been unsuccessful until recently.

In an initial survey of data received from a low energy proton detector aboard Explorer XII (launched August 16, 1961), *Davis and Williamson* (1963) noticed that the energy density of the 100 kev to 4.5 Mev protons amounted to an appreciable fraction of the energy density of the earth's magnetic field during a magnetically quiet period, and from 4 out to 8 earth radii the ratio of the two energy densities was almost constant. It was of obvious interest to determine the magnitude of the magnetic field produced by this population of protons. On the basis of some preliminary data furnished by Davis, *Akasofu, Cain and Chapman* (1962) calculated the magnetic disturbance for a model proton belt which resembled the measured proton distribution. The results predicted a decrease in the surface equatorial field of about 38 γ due to the quiet time belt. *Davis and Williamson* (1963) also observed a threefold increase in proton intensities at small pitch angles during a magnetic storm. Adopting the assumptions that (1) this increase was applicable to protons of all pitch angles, and (2) the average energy of the particles equalled the quiet

*Published as *Goddard Space Flight Center Document X-611-64-*. June 1964

time energy, *Akasofu and Cain* (1962a) calculated that an additional 80 γ decrease would be observed at the equator of the earth, sufficient to account for the -60 γ main phase of the aforementioned magnetic storm.

A considerably more thorough analysis of the magnetic effects of the quiet time proton belt has now been performed from the data obtained from the proton detector aboard Explorer XII, and the results are transmitted in this paper. In addition to the fact that these protons produce the greatest magnetic disturbance of any known population of trapped particles, it is also of interest to study the effects of the ring current on phenomena related to the earth's main magnetic field. The magnetic lines of force are distorted from the shapes calculated by the spherical harmonic analysis of the field at the earth's surface. There is a stretching out or strengthening of the field at distances beyond the ring current which affects the location of the magnetospheric boundary. Of interest to cosmic ray physicists is the influence of the Störmer cut-off rigidity for solar protons approaching the earth near the polar regions. Finally, the quiet time particle distributions are established in order to acquire a basis for determining storm time variations.

Thus, in this paper the proton data are also interpreted so as to determine the extent which the magnetic perturbations of the protons effect the earth's main field.

II. PROTON LIFETIMES

The proton ring current under discussion has been described as a "quiet time" current, there being arguments against consideration that its enhancement during a magnetic storm is the cause of the main phase. If one assumes that the lifetimes of these particles are limited solely by coulomb scattering and charge exchange, the decay time of an enhanced current during a magnetic storm would be an order of magnitude larger than the one day recovery time indicated by magnetograms at ground level (*Liemohn*, 1961). To obtain agreement between the observed decay time and calculated proton lifetimes, it is necessary to postulate that protons in the energy region 1 to 100 kev (below the threshold of the Explorer XII detector) constitute the storm time ring current (*Dessler, Hanson and Parker*, 1961). In light of the fact that at magnetically quiet periods the observed proton intensities remain stable to within a factor of two over a long period of time (*Davis, Hoffman and Williamson*, 1964), the electric current resulting from them is considered to exist at all times and is described as a quiet time ring current.

However, one must recognize the fact that an enhancement of the protons within the energy range of the detector might constitute the higher energy portion of the storm time particles. Since protons trapped at a given time do have finite lifetimes, there must be a source and probably an acceleration mechanism for replenishing the population. In view of the insufficiency of our knowledge concerning such processes, it is possible that the particular acceleration mechanism, and perhaps others, may influence the lifetimes of the particles. It has also been experimentally observed that during a magnetic storm some time variations do exist. For these reasons the storm time proton observations are being carefully analyzed and will be reported in a later paper.

III. ELECTRIC CURRENTS

Since the magnetic effects of the trapped protons are due to the electric currents they establish, it is first necessary to deduce the current distribution on a magnetic meridian plane from the experimental observations of the protons. For this calculation a slightly modified version of the theory developed by *Akasofu and Chapman* (1961) is used. Essentially the same notation has also been adopted and appears for reference in the Appendix.

In Alfvén's guiding center approximation describing the motion of charged particles in a strong magnetic field, the motion is analyzed into three parts: (a) the *gyration* about and in a plane perpendicular to a magnetic field line; (b) the *oscillation* essentially along a line of force between two mirror points; and (c) the east-west *drift* motion for positively charged particles on a magnetic shell. All of these motions may contribute to the flow of an electric current, i. To calculate the current produced in a gas permeated by a magnetic field, the gas pressure and the characteristics of the magnetic field must be known at every point in the gas.

We now limit the discussion to the steady state

case with a proton distribution symmetrical about the axis of the magnetic poles. While the experimental data will be analyzed in the B/B_e (or equatorial pitch angle), L coordinate system, the computations will assume a geomagnetic dipole field. If $\hat{k}$ is the eastward unit vector normal to a magnetic meridian plane at a particular point, the volume current densities arising from the three motions are obtained from the following formulae (*Parker*, 1957):

a) The current produced by the gyration:

$$\vec{i}_L=\hat{k}\left[\frac{c}{aBh_2}\left(\frac{\partial p_n}{\partial L}-\frac{1}{2}\frac{p_n}{p_m}\frac{\partial p_m}{\partial L}\right)+\frac{c}{BR_c}p_n\right] \tag{1}$$

where

$$h_2=\frac{\cos^3\varphi}{(1+3\sin^2\varphi)^{\frac{1}{2}}}=\frac{1}{\cos^3\varphi}\left(\frac{B}{Be}\right)^{-1} \tag{2}$$

and

$$\frac{1}{R_c}=\frac{3(1+\sin^2\varphi)}{aL\cos\varphi(1+3\sin^2\varphi)^{\frac{1}{2}}}. \tag{3}$$

b) The current due to the oscillations:

$$i_0=enw_s=0, \tag{4}$$

because as many particles are passing through a unit area perpendicular to B towards one mirror point as are moving towards the other mirror point, at least over a short period of time.

c) The current produced by the drift motion:

$$\vec{i}_D=\hat{k}\left(\frac{c}{2aBh_2}\frac{p_n}{p_m}\frac{\partial p_m}{\partial L}-\frac{c}{BR_c}p_s\right) \tag{5}$$

The total current is then

$$\vec{i}=-i\hat{k} \tag{6}$$

with

$$i=\frac{c}{BR_c}(p_s-p_n)-\frac{c}{aBh_2}\frac{\partial p_n}{\partial L} \tag{7}$$

The negative sign in equation 6 signifies that a positive term in i is a westward contribution to the current density, and a negative term is an eastward contribution.

Let us briefly consider just what details in the motions of the particles give rise to the three current terms of i. The first term involving p_s is due to the westward drift motion of protons which depends upon the centrifugal force mw_s^2/R_c associated with the motion along a curved field line.

The second term also arises from the curvature of the lines of force. The circular orbits due to the gyration are more crowded together on the concave side of the line than on the convex side. This is illustrated in Figure 1a, where between two

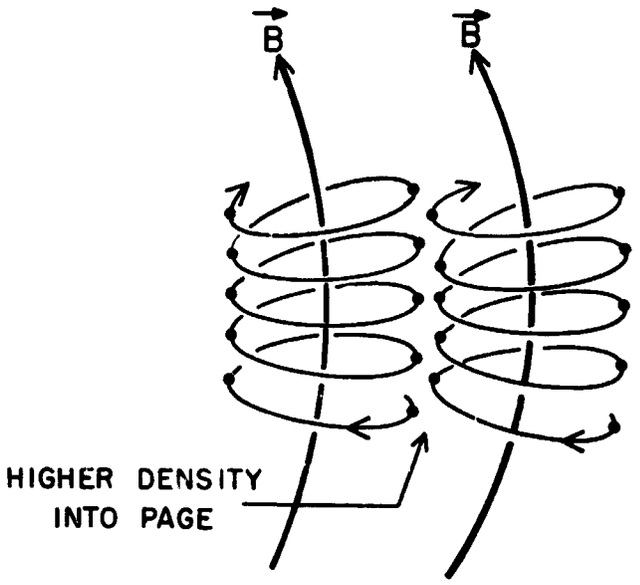


FIGURE 1a.—The source of the current due to the second term in equation 7. The eastward current between the two curved lines of force is caused by the crowding together of the circular orbits on the concave sides of the lines.

adjacent lines of force there is a higher density of particles per unit area moving into the page than out of it. For protons this results in a net eastward current at all points. Surprisingly, this term is larger than the term due to the drift motion if the pitch angle distributions are reasonably steep, as they are for the protons in the belt being discussed.

The final term is due to the fact that the radiation exists in a belt: the intensity increases with radial distance to a certain point, and then decreases. The effect in producing currents is illustrated in Figure 1b, in which the particle intensity is larger on the line of force at a greater distance from the earth, causing a more intense movement of charged particles in a direction into the page than out of it in the region between the lines of force where the two currents intersect. On the inner side of the belt the current

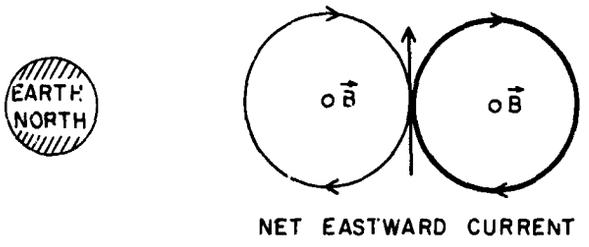


FIGURE 1b.—The source of the current arising from the last term in equation 7. If the particle intensity about the line of force further away from the earth exceeds that about the closer line, indicated by the relative weights of the gyration circles, a net eastward current results in the region between the two lines of force.

will be eastward, and on the outer side it will be westward.

The calculation of the current at all points on a meridian plane first requires the determination of the pressures p_s and p_n at each point. To simplify the equations we first consider protons of a single speed w.

The longitudinal pressure $p_s^* = Nmw_s^2$.

The lateral pressure $p_n^* = \frac{1}{2}Nmw_n^2$.

While it would be possible to calculate these pressures at each point in the magnetosphere from the experimental data, considerable simplification is obtained if they can be expressed in terms of a few parameters determined by the particle distributions. This is accomplished by first fitting the data to a special equatorial pitch angle distribution function with which the local intensities and pitch angle distributions can be obtained from the equatorial intensity and distribution by means of a single parameter depending only upon L. Then the current equation is integrated over the energy spectrum of the particles of interest at each L value; this integral turns out to be actually the total momentum flux at the equator. Hence, the local current at a given latitude depends only upon the two parameters representing the pitch angle distribution and equatorial momentum flux.

Let $F(L, s, \theta)$ denote the number density of protons at a point s on a magnetic shell L, whose speeds lie in dw at w, whose pitch angles lie in $d\theta$ at θ, and whose azimuth angle is arbitrary, since we assume symmetry in azimuth. Note that this definition of $F(L, s, \theta)$ is slightly different than that of *Akasofu and Chapman* (1961) in that it includes a differential angle in azimuth. Fortunately, we can also adopt Parker's special pitch angle distribution (*Parker*, 1957) for the functional form of the distribution in pitch angle, but with the conditions discussed in Section 4 "Proton Data" of this paper:

$$F(L,S,\vartheta, = C_\alpha\left(\frac{B_e}{B}\right)^{\frac{\alpha}{2}}\sin^\alpha\theta. \qquad (8)$$

This distribution has very useful and curious properties. It can be divided into two parts: the second part, $\sin^\alpha \vartheta$, is the local pitch angle distribution and has a form independent of B_e/B. Hence the ratio of p_n to p_s remains the same at every point on an L shell. Because the form function, $\sin^\alpha \theta$, is independent of B_e/B, the integral over all pitch angles, which would be evaluated to obtain the omnidirectional intensity at any value of B_e/B, is independent of this ratio of fields. Hence, the omnidirectional intensity varies down a line of force only as $(B_e/B)^{\frac{\alpha}{2}}$.

The constant C_α can be expressed in terms of the total density at the equator $N_e(L)$:

$$N_e(L) = C_\alpha \int_0^\pi \sin^\alpha\theta(2\pi\sin\theta)d\theta \qquad (9)$$

The integral in $N_e(L)$ will arise again in the theory and can be expressed in terms of gamma functions:

$$\int_0^\pi \sin^{\alpha+1}\theta d\theta = \sqrt{\pi}\frac{\Gamma\left(\frac{\alpha+2}{2}\right)}{\Gamma\left(\frac{\alpha+3}{2}\right)} \qquad (10)$$

Hence

$$C_\alpha = \frac{1}{2}\pi^{-3/2}\frac{\Gamma\left(\frac{\alpha+3}{2}\right)}{\Gamma\left(\frac{\alpha+2}{2}\right)}N_e(L) = A(\alpha)N_e(L) \qquad (11)$$

Therefore

$$F(L,s,\theta) = N_e(L)A(\alpha)\left(\frac{B_e}{B}\right)^{\frac{\alpha}{2}}\sin^\alpha\theta. \qquad (12)$$

Since a particle detector measures a flux rather than a density it would be well to express all equations in terms of fluxes.

$$F(L,s,\theta) = \frac{I_e(L)}{w}A(\alpha)\left(\frac{B_e}{B}\right)^{\frac{\alpha}{2}}\sin^\alpha\theta \qquad (13)$$

where $I_e(L)$ is the omnidirectional intensity at the magnetic equator as a function of L for protons of speed w. It must be noted that α is a function of L, and could also depend upon w.

Inserting this expression into the formulae for the gas pressures yields for protons with speed w:

$$p_s^* = m\int_0^{\pi} (w^2\cos^2\theta)\frac{I_e(L)}{w}A(\alpha)\left(\frac{B_e}{B}\right)^{\frac{\alpha}{2}} \sin^{\alpha}\theta(2\pi\sin\theta)d\theta \quad (14)$$

and

$$p_n^* = \frac{m}{2}\int_0^{\pi} (w^2\sin^2\theta)\frac{I_e(L)}{w}A(\alpha)\left(\frac{B_e}{B}\right)^{\frac{\alpha}{2}} \sin^{\alpha}\theta(2\pi\sin\theta)d\theta \quad (15)$$

Evaluating the integrals with the aid of equation 10, we have

$$p_n^* = \frac{1}{2}mwI_e(L)\Omega(\alpha)\left(\frac{B_e}{B}\right)^{\frac{\alpha}{2}} \quad (16)$$

where

$$\Omega(\alpha) = \frac{\alpha+2}{\alpha+3} \quad (17)$$

and

$$p_s^* = \frac{2}{\alpha+2}p_n^*, \quad (18)$$

which indicates again that the ratio of partial pressures is constant on an L shell.

The current can now be expressed in terms of $I_e(L)$ and α:

$$\begin{aligned} i^* &= \frac{c}{BR_c}\frac{\alpha}{\alpha+2}p_n^* - \frac{c}{aBh_2}\frac{\partial p_n^*}{\partial L} \\ &= -\frac{c}{BR_c}\frac{\alpha}{\alpha+3}\frac{1}{2}mwI_e(L)\left(\frac{B_e}{B}\right)^{\frac{\alpha}{2}} \\ &\quad -\frac{c}{aBh_2}\frac{1}{2}mw\left[\Omega(\alpha)\left(\frac{B_e}{B}\right)^{\frac{\alpha}{2}}\frac{\partial I_e(L)}{\partial L} + \left(\frac{B_e}{B}\right)^{\frac{\alpha}{2}}I_e(L)\right. \\ &\quad \left. +\frac{1}{2}I(L)\Omega(\alpha)\left(\frac{B_e}{B}\right)^{\frac{\alpha}{2}}\ln\left(\frac{B_e}{B}\right)\frac{\partial\alpha}{\partial L}\right] \end{aligned} \quad (19)$$

or

$$i^* = -\frac{mwc}{aH_0}\left[D^*(\alpha,\varphi)L^2I_e(L) + E^*(\alpha,\varphi)L^3\frac{\partial I_e}{\partial L}(L) + F^*(\alpha,\varphi)L^3I_e(L)\frac{\partial\alpha}{\partial L}(L)\right] \quad (20)$$

where H_0 is the field strength on the equator at the earth's surface, and

$$D^*(\alpha,\varphi) = \frac{3}{2}\left(\frac{B_e}{B}\right)^{\frac{\alpha}{2}+4}\frac{\alpha}{\alpha+3}\frac{(1+\sin^2\varphi)}{\cos^{19}\varphi},$$

$$E^*(\alpha,\varphi) = \frac{1}{2}\frac{\alpha+2}{\alpha+3}\left(\frac{B_e}{B}\right)^{\frac{\alpha}{2}}\cos^3\varphi,$$

$$F^*(\alpha,\varphi) - E^*(\alpha,\varphi)\left[\frac{1}{(\alpha+2)(\alpha+3)} - \frac{1}{2}\ln\left(\frac{B}{B_e}\right)\right] \quad (21)$$

Equation 20 corresponds to equation 52 of *Akasofu and Chapman* (1961) with the added term depending upon α (L).

It is now necessary to integrate the current over the energy range measured by the detector. Since the scintillator has a low energy cut-off near 100 kev, the current density distribution arising from protons with energies greater than 100 kev will be calculated. This low energy limit must be kept clearly in mind in what follows.

We next make another simplifying assumption: that α is independent of energy, or the energy spectrum is a function only of L, not of pitch angle. The validity of this assumption will be discussed in Section 4, "Proton Data." This allows D^*, E^*, and F^* to be independent of energy.

The only terms remaining in the current equation 20 which depend upon energy as the velocity w and the equatorial omnidirectional flux of this velocity, $I(L)$, which is actually the omnidirectional differential energy spectrum, $\frac{dN}{dE}$. Let

$$\Psi(L) = \int_{100}^{\infty} mw\frac{dN}{dE}dE = 2m\int_{100}^{\infty}\sqrt{E}\frac{dN}{dE}dE,$$

which is a momentum flux, or twice the energy density of the protons.

Then

$$i = -\frac{c}{aH_0}\left[D(\alpha,\varphi)L^2\Psi(L) + E(\alpha,\varphi)L^3\frac{d\Psi(L)}{dL} + F(\alpha,\varphi)L^3\Psi(L)\frac{d\alpha(L)}{dL}\right] \quad (23)$$

Hence, the current at any point in space, L, φ, need be expressed only in terms of the pitch angle parameter α and the momentum profile ψ at the equator.

IV. PROTON DATA

The data relating to the low energy protons from which the current distributions have been calculated were acquired from one orbit of the satellite, pass 10A (outbound) and 10B (inbound) on August 27 and 28, 1961. At this time the major axis of the orbit was lying in the meridian plane which included the sun (see Figure 2).

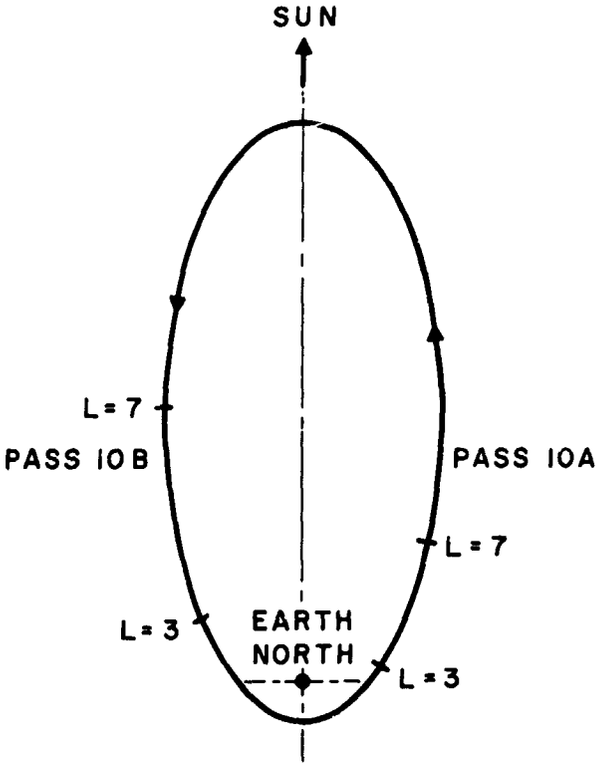


FIGURE 2.—The projection on the earth's equatorial plane of pass number 10 on August 27 and 28, 1961.

Since the heart of the proton belt ($L=3.5$) lies at a relatively low altitude in comparison with apogee ($L=13$) much of the data were collected near the twilight regions, although the outer edge of the belt was traversed on the sunlit side of the earth.

The planetary magnetic three-hour-range index K_p, plotted in Figure 3, indicated that no major disturbances were in progress. In fact the first appreciable magnetic disturbance during the lifetime of Explorer XII did not occur until pass 12.

This particular pass was chosen for thorough analysis because the orbit lay closer to the geomagnetic equator in the heart of the belt than any other orbit. Also the telemetry coverage was complete. When this study was initiated it was expected that a very typical proton distribution would be observed during the pass 10, since *Davis and Williamson* (1963) found no temporal variations larger than 30 percent during nineteen passes of Explorer VII distributed before and after this pass, but not including it. However, later comparisons with this original data indicate that the proton intensities were enhanced by as much as a factor of two at L values smaller than the maximum intensity at L of 3.5 during pass 10. The maximum was also somewhat sharper than observed on the average pass with a more rapid decrease in intensity with increasing L beyond the maximum. This means that the belt was lying at slightly smaller L values than the average belt.

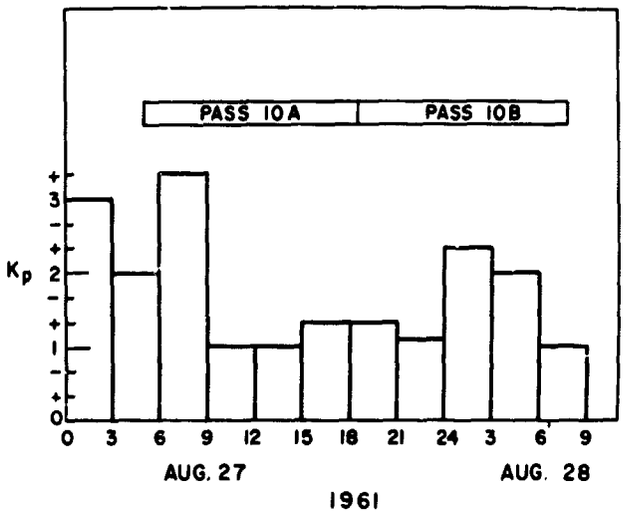


FIGURE 3.—The planetary magnetic three-hour-range index K_p for the period during which the data used in this study were acquired.

The details of the low energy proton detector have been described by *Davis and Williamson* (1963). Let it suffice here to produce a table containing the spectra employed in calculating the momentum integral (Table I), as well as some examples of pitch angle distributions. In general the energy spectrum was described as a doubly sloped exponential spectrum, illustrated in Figure 4 in the integral form which was actually measured by the detector. At L values below 2.2 the data points were best fit with power law

TABLE I.—*Proton Energy Spectra*

	30°–90° Equatorial Pitch Angles				0°–30° Equatorial Pitch Angles			
L	N_{OI}	E_0	E'	E_1	N_{OI}	E_0	E'	E_1
1.8	POWER LAW SPECTRA				POWER LAW SPECTRA			
2.0								
2.2	.....	.....	.....	662 [1]	4.69×10^{7}	625	1650	... [3,4]
2.5	2.04×10^{8}	635	1000	... [2]	1.46×10^{8}	515	1650	... [3]
2.8	2.17×10^{8}	456	1000	... [2]	3.07×10^{8}	390	900	725 [5]
3.0	4.20×10^{8}	303	900	550 [3]	4.15×10^{8}	255	900	320 [3]
3.25	5.07×10^{8}	222	1000	480	4.22×10^{8}	205	900	480
3.5	6.10×10^{8}	193	880	300	4.64×10^{8}	185	900	226
3.7	4.88×10^{8}	196	880	196	3.40×10^{8}	176	∞	.....
3.9	3.60×10^{8}	195	880	247	2.50×10^{8}	145	900	260
4.6	3.00×10^{8}	110	255	187	1.68×10^{8}	105	∞	....
5.0	2.80×10^{8}	73	255	139	1.73×10^{8}	61.5	255	90
5.5	1.62×10^{8}	68	350	[illegible]	1.14×10^{8}	52	∞	.....
6.0	2.05×10^{8}	50	375	[illegible]	2.20×10^{8}	36	∞	.....
7.05	1.35×10^{8}	33.5	.....	∞	6.13×10^{7}	33	∞	.....

Notes: All energies in kev.

[1] E_1 =662 kev from 470 to 1000 kev; for $E>1000$ kev, power law with γ for integral spectrum =4.2. From 140 to 470 kev, no protons.
[2] For $E>1000$ kev, power law with γ =4.2.
[3] For $E>1650$ kev, power law with γ =4.2.
[4] Below E =255 kev, no protons.
[5] For $E>2000$ kev, power law with γ =4.2.

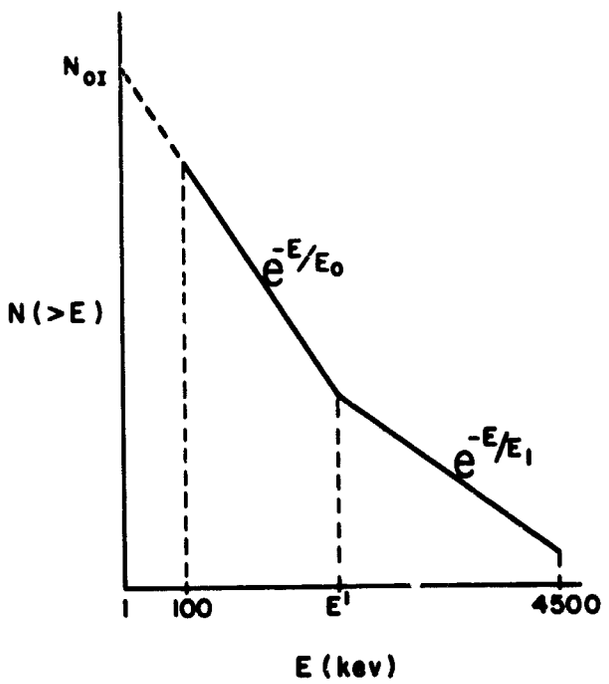


FIGURE 4.—Double sloped integral proton spectrum employed in the calculation of the momentum integral.

spectra, and up to L of 3 a steep power law spectrum was included for the higher energy protons, as suggested by the data of *Bame et al* (1963) and *Naugle and Kniffen* (1963). The spectra listed in Table I were differential before insertion into the momentum integral.

In the derivation of the formulae from which the current distributions may be computed, considerable simplification was attained by employing as the functional form of the pitch angle distribution, $\sin^{\alpha}\theta$. The measured angular distributions were found to fit this form reasonably well provided the entire range of pitch angles, 90° to 0°, was divided into two regions, 90° to 30°, and 30° to 0°, each with an independently determined value of α. Some sample fits of the data to this function appear in Figure 5.

Figure 5a illustrates the extrapolation to 90° of the distribution based on data points with pitch angles only between 30° and 56°. At low L values the orbit carried the satellite away from the magnetic equator making the measurement of protons with large pitch angles impossible. Figure 5b is an excellent example of the necessity of using the two independent values of α. An

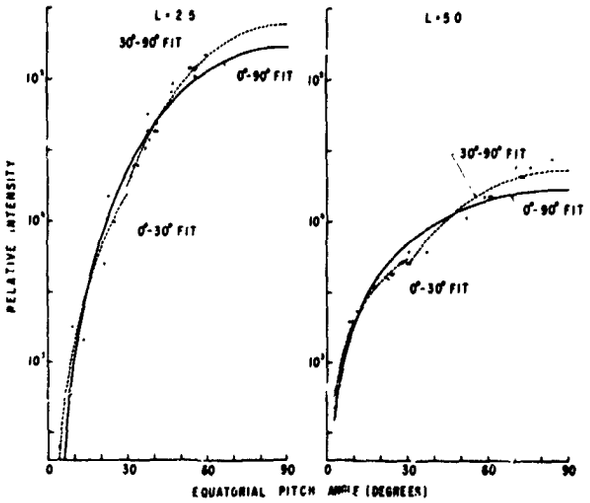


FIGURE 5.—Sample fits of the equatorial pitch angle distribution to the function sin α for the regions 90° to 30° and 30° to 0°.

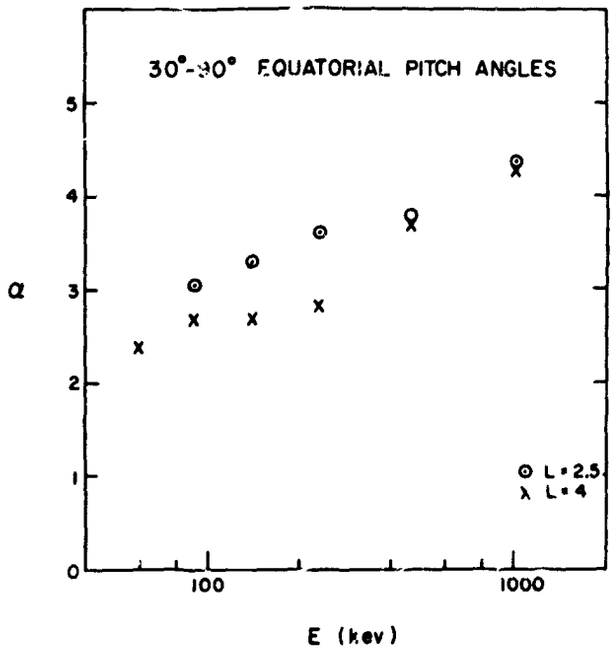


FIGURE 6.—The pitch angle parameter α as a function of energy for L values of 2 and 4.

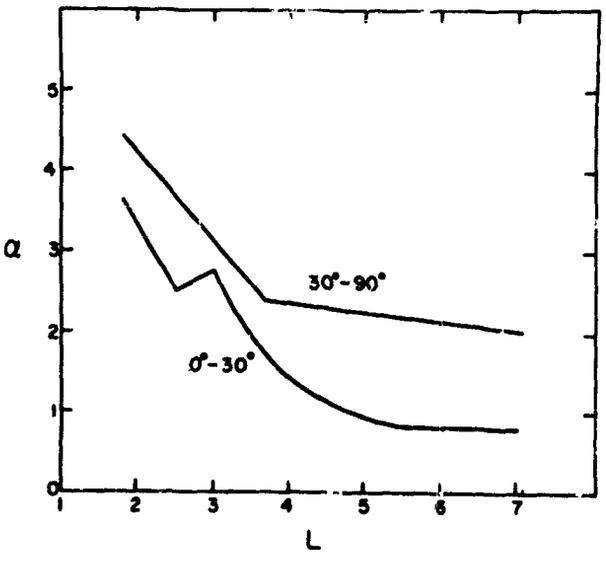


FIGURE 7.—The values of α as a function of L for the two pitch angle ranges derived for the particles having the average momentum on the given L shell.

attempt at a least squares fit of all data from 0° to 90° produced a curve (solid line) with a large hump at the middle pitch angles and a very low value at large pitch angles.

Another assumption which considerably simplified the derivation of the current distribution formulae involved the independence of α upon energy. In reality the value of α increases as a function of increasing energy, as illustrated by the data in Figure 6 for protons of equatorial pitch angles between 90° and 30°. The data in the figure were obtained from the output of the detector which measured the incident energy flux of protons rather than the particle flux. The energies at which the points are plotted are the effective low energy cut-offs for the integral energy spectrum measured at this value of L. By choosing a constant value of α for a given L, the functions D^*, E^*, and F^* (equation 21) became energy independent and could be removed from the integration over energy. Hence, the value of α (L) selected is approximately the value derived for the particles having the average momentum on the given L shell. This value of α as a function of L is plotted in Figure 7 for each of the two pitch angle ranges, 90° to 30°, and 30° to 0°.

The calculation of N_{OI}, employed in the determination of the momentum integral ψ (see Figure 4 for spectral shape of the ψ integral), required some special attention. This quantity, the integral omnidirectional intensity above zero energy, is equivalent to the directional intensities at 100 kev integrated over all pitch angles, which can be actually measured by the detector, multiplied by the factor exp $(100/E_0)$. However, the counting rates of the lowest energy channels of the detector aboard Explorer XII passed into

saturation about one earth radius on either side of the maximum intensity at 3.5 R_E, at least at large pitch angles. Hence, it was necessary to compute N_{OI} by either of two less direct methods.

If (1) a directional intensity above an energy E_c, $J(>E_c)$, is measured at a pitch angle θ_c, (2) the pitch angle distribution characterized by the parameter α is known, and (3) the energy spectrum (E_0) is measured (the latter two parameters could be determined by the current channels of the detector, which did not pass into saturation), then the value of N_{OI} can be calculated from the formulae

$$N_I(>E)=\frac{J(>E_c)}{G_0 \sin^\alpha \theta_c} 2\pi^{3/2} \frac{\Gamma\left(\frac{\alpha+2}{2}\right)}{\Gamma\left(\frac{\alpha+3}{2}\right)} \times \exp\left[(E_c-E)/E_0\right], \quad (24)$$

where G_0 is the telescope factor $=5.85\times10^{-3}$ cm^2−ster., the expression $2\pi^{3/2}$ times the gamma functions arises from integration over the pitch angles, $(\sin^\alpha \theta_c)^{-1}$ is actually an extrapolation of the directional intensity measured at θ_c to the intensity at 90° equatorial pitch angle, and $\exp[(E_c-E)/E_0]$ is an extrapolation in energy from the integral point at E_c to the energy E. The quantity N_I is then the integral omnidirection flux above the energy E, so N_{OI} would have $E=0$.

Therefore, when the counting rate channel at 100 kev was in saturation either of two extrapolations could be used to acquire N_{OI}: A higher energy channel (for example, the 880 kev step) would provide a non-saturated counting rate at a near 90° pitch angle, making $(\sin^\alpha \theta_c)^{-1}\simeq 1$ but requiring an extrapolation in energy $\exp[(E_c-E)/E_0]$, or a smaller pitch angle measurement from the 100 kev channel would provide a measurement of $J(>E_c)$ out of saturation, allowing $\exp[(E_c-E)/E_0]$ to be near one, but increasing the value of $(\sin^\alpha \theta_c)^{-1}$. Obviously the accuracy of either method depends upon how well the data can be characterized by the parameters α and E_0. The values of N_{OI} employed in the computation of ψ (L) were acquired by utilizing both methods. Table II shows a sample set of intensities calculated for an L value of 3.25 for particles with pitch angles 30° to 90°. The agreement between the various calculations of an N_{OI} at a given L value was within ±15 percent.

TABLE II

E_c(kev)	θ_c	α	E_0(kev)	N_{OI}
105	40°	2.5	222	5.01×10^8
470	30°	2.5	222	5.34×10^8
880	70°	2.5	222	4.87×10^8

From the two sets of values of α (L) necessary to describe the protons over the entire pitch angle range from 0° to 90°, a double set of N_{OI} is calculated, since this quantity is a function of α through equation 24. Hence, we obtain two independent profiles of ψ (L), one to be associated with protons having pitch angles 30° to 90°, the other with protons from 0° to 30°. These two profiles appear in Figure 8.

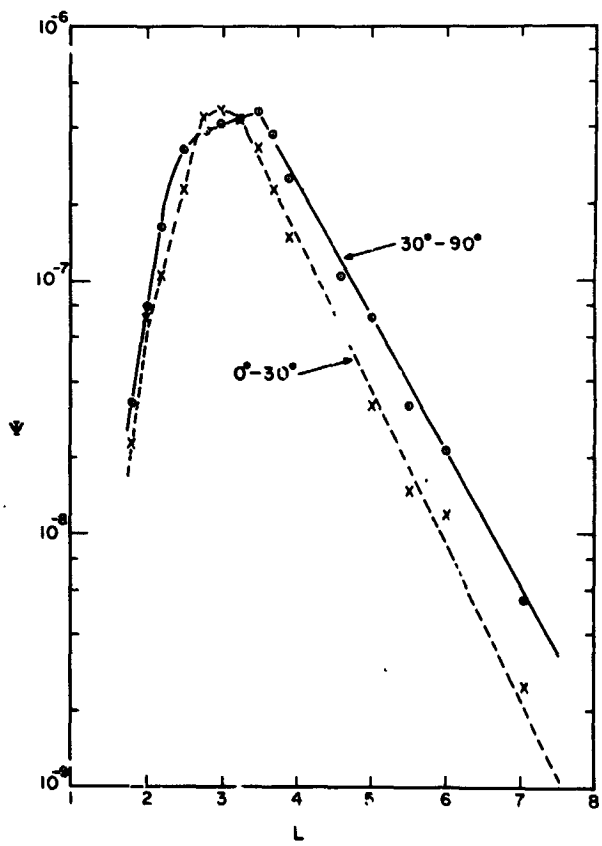


FIGURE 8.—The omnidirectional momentum flux ψ (L^3) for each of the two pitch angle ranges in units of gm/cm-sec^2 (or ergs/cm^3).

V. ELECTRIC CURRENT DISTRIBUTION

The data have now been compiled in a form sufficient for the calculation of the electric current distribution from equation 23. One must keep in mind that the current at a particular L, φ is actually a function of the local partial pressures p_s and p_n (equation 7), but we have replaced the problem of calculating them by utilizing the parameters α (L) and ψ (L). The set of these parameters which is applicable to the radiation at the point of calculation of i must be used. Since a proton with an equatorial pitch angle of 30° mirrors at a magnetic latitude of 32°, the calculation of i for latitudes less than 32° will employ the set of parameters describing protons with equatorial pitch angles from 30° to 90°, and the calculation of i for latitudes greater than 32° will utilize the parameters for protons of pitch angles 0° to 30°. There will be a discontinuity in i as a function of latitude at 32° because as this latitude is approached from the equatorial side, the parameters less accurately describe the actual radiation.

Two other discontinuities which arise in the distribution of i over a meridian plane are due to the approximation of α (L) and ψ (L) by the simple mathematical expressions listed in Table III. Except for ψ (L) in the region $L=2.25$ to 2.75 for protons with 30° to 90° pitch angles, and $L=2.75$ to 3.25 for 0° to 30° pitch angles, both α and ψ could be well approximated by either straight lines or exponentials. While the approximation curves are continuous, their derivatives contain discontinuities between the segments of the curves. Hence, since i depends upon the derivatives, it will contain discontinuities.

TABLE III

α	L Region	ψ	L Region
		30° to 90° E.P.A.	
$-1.07L+6.36$	1.5 to 3.7	$2.95\times10^{-11}\exp(L/0.256)$	1.5 to 2.4
$-0.12L+2.84$	3.7 to 10.0	$1.98\times10^{-7}\exp(L/4.12)$	2.7 to 3.5
.....	.....	$3.50\times10^{-5}\exp(-L/0.810)$	3.5 to 10.0
		0° to 30° E.P.A.	
$-1.62L+6.57$	1.5 to 2.5	$2.06\times10^{-12}\exp(5.18L)$	1.5 to 2.0
$0.52L+1.22$	2.5 to 3.0	$4.23\times10^{-10}\exp(2.52L)$	2.0 to 2.75
$27.5L^{-2.113}$	3.0 to 5.3	$5.91\times10^{-5}\exp(-1.476L)$	3.2 to 10.0
$-0.053L+1.16$	5.3 to 10.0	.....	.....

In the regions of Ψ not included by these formulae, tabulated values from the curves of Figure 8 were used.

The calculations of i in both pitch angles ranges have been performed for the region of L from 1.5 to 10 in intervals of 0.1 and latitude from the equator down an L shell to approximately 1000 km above the surface of the earth in 2° intervals by utilizing an IBM 7094 computer at Goddard Space Flight Center. Actually the calculations were terminated at a latitude within 2° of the latitude where the line of force attained a height of 1000 km. This convention was adopted so that a constant latitude increment could be used in the numerical integration involved in the magnetic field calculations. A contour plot of the current distribution on a magnetic meridian plane appears in Figure 9.

The contributions of each of the terms in equation 23 to the total current are displayed in Figure 10 for two different L shells, one on the inner side of the belt where the total current is eastward, the other on the outer side of the belt where the total current is westward. One notices from an inspection of Figures 9 and 10 that the westward portion

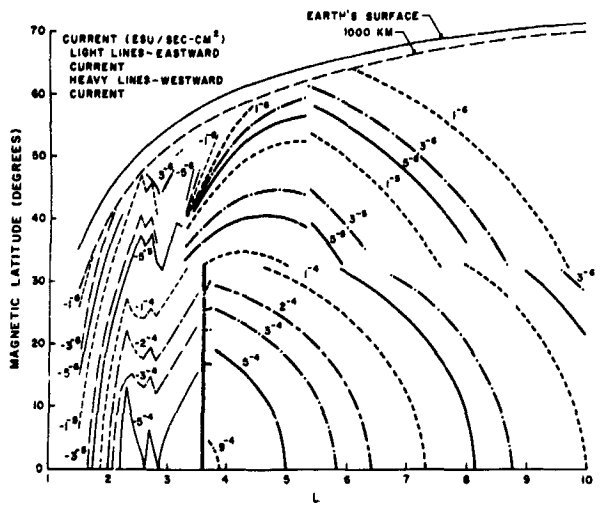


FIGURE 9.—A contour plot of the total electric current in a magnetic meridian plane derived for protons with energies above 100 kev. The currents, expressed as X^{ν} signify $X \cdot 10^{\nu}$. To convert currents from esu/cm²-sec to amps/km², multiply by three.

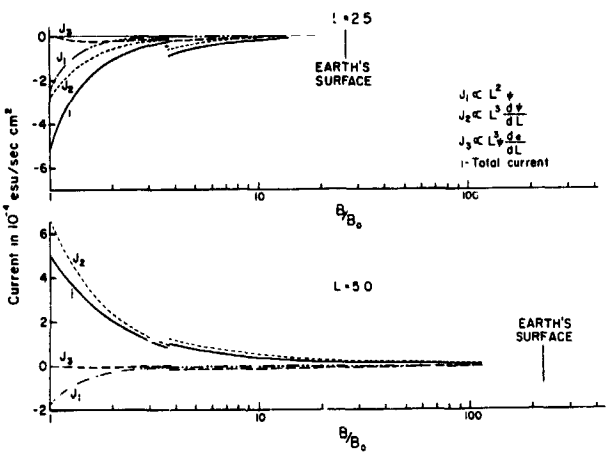


FIGURE 10.—The contributions of each of the terms in equation 23 to the total current for two different L shells.

of the ring current, producing a southward perturbing magnetic field on the earth's surface, arises from the second term J_2, which depends upon the gradiant of the momentum flux on the outer side of the belt.

Since the discontinuities in i appearing in these figures originate in the mathematical expressions used to fit the parameters appearing in the equation for i, a smoothed version of the current contour plot to illustrate the dominant features of the current distribution appears in Figure 11.

An integration has been performed to obtain the

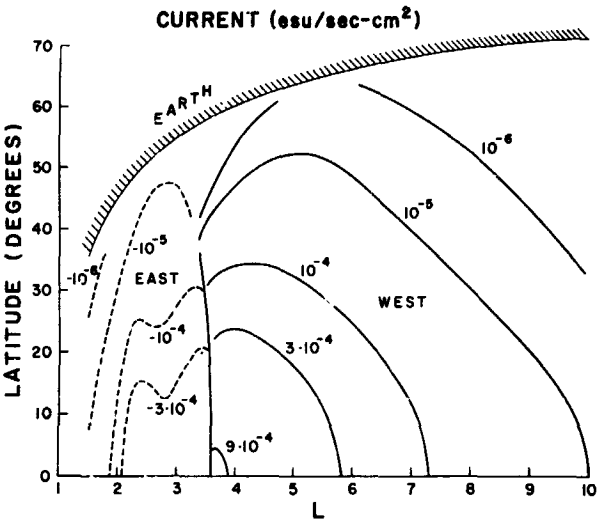


FIGURE 11.—A smoothed version of the electric current contour plot to illustrate the dominant features of the electric current distribution.

total eastward and westward currents, their algebraic sum, and the magnetic moment of the ring current. These values are tabulated in Table IV.

TABLE IV

	esu/sec	amps
Eastward Current	0.57×10^{15}	0.19×10^{6}
Westward Current	2.34×10^{15}	0.78×10^{6}
Total Current	1.77×10^{15}	0.59×10^{6}

Magnetic Moment $0.029M_E$ ($M_E=8.1\times10^{25}$ gauss-cm³)

VI. MAGNETIC FIELD OF THE RING CURRENT

The magnetic field due to the ring current may be denoted by its components H_p and H_z respectively perpendicular and parallel to the dipole axis. These components refer to a point (r, φ) specified by its polar coordinates (φ is latitude) or alternatively by a point (af, φ) in L space where $f=L\cos^2\varphi$.

The ring current region is then divided into elements dS_3 specified by (r', φ') in polar coordinates or by (af', φ') in L space, where $f'=L'\cos^2\varphi'$. Then (*Stratton*, 1941),

$$H_p=-\frac{2}{(acf\cdot\cos\varphi)}\int\int(f\sin\varphi-f'\sin\varphi')\Big[E(k^2)-K(k^2)+(2f\cdot f'\cos\varphi\cdot\cos\varphi')\frac{E(k^2)}{F_{-}^{2}}\Big](i/F)dS_3 \quad (25)$$

and

$$H_z = -\frac{2}{(ac)}\int\int\left[K(k^2)-E(k^2)+2f'\cos\varphi'(f'\cos\varphi' - f\cos\varphi)\frac{E(k^2)}{F_-^2}\right](i/F)dS_3 \quad (26)$$

Here $K(k^2)$ and $E(k^2)$ denote the complete elliptic integrals of the first and second kind respectively, and

$$k^2 = \frac{(4f\cdot f'\cos\varphi'\cos\varphi)}{F^2},$$

$$F_-^2 = f^2+f'^2-2f\cdot f'\cos(\varphi-\varphi'),$$

$$F^2 = f^2+f'^2+2f\cdot f'\cos(\varphi+\varphi').$$

An element of cross sectional area of the ring current region in the meridian, dS_3, is expressed as $dS_3 = a^2L'\cos^4\varphi' dL' d\varphi'$.

The foregoing formulae appear as in *Akasofu and Chapman* (1961) except for the correction of several misprints.

The components H_p and H_z were calculated separately at each point in the specified current region for particles in both pitch angle intervals. The results were then combined to give the total field components at each point in the current region for all ring current particles.

The integrations involved in calculating these field components were performed numerically on the IBM 7094 computer at Goddard, and the results appear in Figure 12. As with the current calculation, increments of 0.1 in L and 2° (0.0349 radian) in latitude were used.

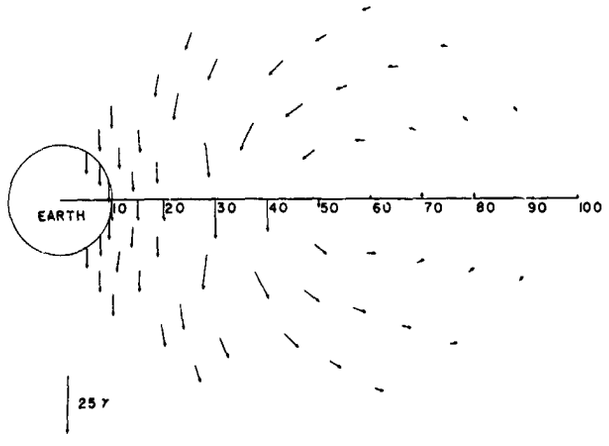


FIGURE 12.—Vector magnetic field on a meridian plane due to the proton ring current.

A discontinuity in the components H_p and H_z arises when $L=L'$ and $\varphi=\varphi'$, since in this case $F_-^2=0$, which appears in the denominator of the argument for each field component. The contribution to the components from the element of integration causing this discontinuity was taken to be zero each time such a discontinuity appeared.

The coordinate system used indicates a negative H_z is in a southerly direction and a negative H_p points toward the earth.

VII. DISCUSSION

(a) Vector Field of the Ring Current

At the magnetic equator on the surface of the earth the magnetic field due to the ring current attained a value of 9 γ in a southern direction, a value considerably smaller than the 38 γ obtained by *Akasofu, Cain and Chapman* (1962) from the preliminary data from the same detector. A reasonable error analysis on the field strength appears virtually impossible to perform. Purely an estimate would place a confidence of about 50 percent (±5 γ) on the field calculations.

The distribution of field vectors from the ring current which appears in Figure 12 indicates a field fairly uniform on the inside of the proton belt and parallel to the dipole axis as expected. In the region of the belt the field shows considerable curl and achieves a maximum value of 23 γ on the equator at an L of 3.6. Outside the belt the field becomes parallel to the earth's field, but has the extremely small value of 2.5 γ. Hence, the ring current field due to these protons above an energy of 100 kev would have little influence on the shape of the boundary of the magnetosphere.

(b) Second Order Calculation

While these protons with energies above 100 kev are showing some magnetic influence, the magnitude is not sufficient to perturb appreciably the magnetic field of the magnetosphere which arises from sources internal to the earth. In fact it appears that the distortion of the field even at the altitudes of the ring current due to the compression of the field by the solar wind is comparable to the disturbance from the ring current (*Cahill and Amazeen*, 1963). Hence the assumption of analyzing the data in the L, equatorial pitch angle coordinate system for the ring current

calculations will produce results as accurate and physically meaningful as would an attempt at a second order calculation (*Beard*, 1962; *Akasofu, Cain and Chapman*, 1961), in which the trapped protons are assumed to move in a magnetic field differing from the internal field by the field of the ring current itself.

(c) Cut-off Rigidity of Solar Protons

In order to explain the observations of low-rigidity solar protons near the earth at latitudes well below those predicted by normal Störmer theory, investigators have included the magnetic effects of a ring current as well as the termination of the geomagnetic field (*Kellogg and Winckler*, 1961; *Akasofu, Lin and Van Allen*, 1963; *Ray*, 1964). In order that a ring current contribute significantly to a reduction in the cut-off rigidity, its magnetic moment must be some appreciable fraction of the magnetic moment of the earth. However, the moment due to the quiet time protons with energies above 100 kev is calculated to be only 0.029 that of the earth, a value much too small to affect solar proton trajectories.

(d) Model Belt Parameters

On the basis of some very preliminary data from the proton detector aboard Explorer XII, the magnetic effects of a model belt that resembled the measured belt were calculated (*Akasofu, Cain and Chapman*, 1962). Since then the properties of a similar belt have been applied to various problems of a geophysical nature (see especially the later articles by Akasofu; *Vestine*, 1963). For this work the following parameters have been utilized to characterize the distribution of protons:

- r_0: distance where the number density attains its maximum value,
- g_1: Gaussian parameter for the inner side of the belt: $n=n_0 \exp(-g_1^2 z^2)$, where $z=(r-r_0)/a$,
- g_2: Gaussian parameter for the outer side of the belt,
- α: pitch angle parameter,
- n_0: maximum number density,
- E: average energy of all protons in the belt.

In table V are listed the values adopted in the literature and a set of suggested values which more accurately produce the results obtained in this more thorough study.

TABLE V

Parameter	Value in Current Literature	Revised Value
r_0	3.0 and 3.2	3.5
g_1	2.990	1.0
g_2	0.419	0.90
α	2.0	2.5
n_0	0.6/cc	0.46/cc
E	500 kev	300 kev

(e) Lower Energy Protons

One must keep clearly in mind in this discussion of the magnetic properties of the trapped protons that only those with energies above 100 kev, the low energy cutoff for the direct detection of proton particle fluxes by the detector, have been considered. However, in another mode of operation the detector could measure proton energy fluxes down to energies of some tens of kev, depending upon the steepness of the spectrum. Beginning at about an L of 5, large fluxes of protons were apparently present with energies extending down to at least 50 kev, and the ratio of the flux between 50 and 100 kev to that above 100 kev became increasingly large with range.

While a complete saturation of the field beyond five earth radii with very low energy protons would significantly contribute to a quiet time ring current, the maximum field produced at the earth's surface by such a particle distribution is insufficient to account for large magnetic storms.

For a rough estimation of this maximum field first assume that the particle energy density is limited by the energy density of the field on the magnetic equator of an L shell. Next assume that the particle intensity is isotropic ($\alpha=0$) and hence constant everywhere on an L shell to latitudes of 60°, which is essentially the surface of the earth for $L\geq 5$. Using equations 23 with $1/h_2$ add $1/R_c$ evaluated at the equator, their maximum values, and with

$$\Psi(L)=2\frac{H_0^2}{8\pi L^6}$$

the current is

$$i(L)=\frac{cH_0}{2\pi aL^4}\cos^3\varphi$$

This produces a field at the center of the earth, and approximately at the surface, of

$$H_z=-\frac{1\times10^4}{L^3}\gamma,$$

where L is the minimum distance of the belt from the center of the earth. For $L=5$, H_z is -76 γ, which is one, and probably two, factors of two too large if compared with a field derived from any realistic distribution of particles exactly calculated.

To account for a large storm of 300 γ (excluding inductive effects in the earth) the field would have to be saturated down to at least an L of 3.2, and probably more like 2.5 for a more reasonable particle distribution. The apparent necessity for such a high intensity, low altitude belt leads one to question whether the current equation 7 includes all the necessary terms to account for the ring current as it naturally occurs. One is tempted to suggest that electric fields may assist in the production of an increased westward movement of the protons already observed. It certainly would seem reasonable to attempt a measurement of electric fields in this region of space.

It is interesting to note that the lifetime of 50 kev protons at 5 earth radii is only on the order of a week (*Liemohn*, 1961). In order to sustain a quiet time belt of such particles it would be necessary to have a source which operates continually.

In passing it is perhaps well to mention that the trapped electrons observed to date produce negligible magnetic fields. Since the magnetic properties of a particle depend upon the square root of its mass, electron fluxes of the order of 10^{10} electrons/cm^2-sec would be required to produce magnetic effects equivalent to those of the protons observed. This flux is two orders of magnitude greater than any reported in the literature.

VIII. CONCLUSION

The protons with energies above 100 kev observed by the low energy proton detector flown by Davis and Williamson aboard Explorer XII are the first group of particles discovered in the trapped radiation region with intensities sufficient to produce significant magnetic effects. However, for applications to other geophysical problems, such as the distortion of field lines, the change in the Störmer cut-off rigidities, and the quiet time remains of a storm time ring current, the observed particle intensities are probably disappointingly small.

For future studies of the proton ring current two improvements in the data acquisition would be desirable. The orbit of the satellite should be more nearly equatorial over all L values, and the detectors should be capable of measuring protons down to nearly one kev in energy.

While this detector has provided information which has advanced our knowledge of the ring current problem, it appears that the complete understanding of the phenomenon will require some years of continued research with new particle detectors and possible complementary measurements of electric fields or a more complete understanding of other source mechanisms for the electric currents.

ACKNOWLEDGEMENTS

The authors appreciate the opportunity of working with the excellent data collected from the proton detector flown aboard Explorer XII by Mr. Leo R. Davis and Mr. James M. Williamson, and for very helpful discussions with them. One of us (R. A. H.) performed most of this study while the recipient of a NAS-NASA Resident Research Associateship, the support of which is gratefully acknowledged.

NOTATION

a	earth's radius, in cm.
c	speed of light, in cm/sec.
r	radial distance from the center of the earth to a point of interest, in cm.
L	McIlwain's coordinate, in earth radii.
B	magnetic field vector, in gauss.
p_m	magnetic pressure $=B^2/8\pi$.
m	mass of a trapped particle in grams.
e	magnitude of charge of a trapped particle, in esu.
w	velocity of a particle, in cm/sec.

w_s component of w along B.
w_n component of w normal to B.
E kinetic energy of a particle, in kev.
N number density of particles.
θ pitch angle of a trapped particle, i.e. the angle between w and B, or its complement if larger than 90°.
p_s pressure of the gas (trapped particles) along B.
p_n pressure of the gas normal to B.
φ geomagnetic latitude.
Note: the subscript "e" indicates that a quantity is being evaluated at the magnetic equator.

REFERENCES

AKASOFU, S. -I., and J. C. CAIN, A model storm-time proton belt, Paper presented at the 43rd annual meeting, A.G.U., Washington, April 25–28, 1962a.

AKASOFU, S. I. and J. C. CAIN, The magnetic field of the radiation belts, *J. Geophys. Research*, **67**, 4078–4080, 1962b.

AKASOFU, S. -I., J. C. CAIN, and S. CHAPMAN, The magnetic field of a model radiation belt, numerically calculated, *J. Geophys, Research*, **66**, 4013–4026, 1961.

AKASOFU, S. -I., J. C. CAIN, and S. CHAPMAN, The magnetic field of the quiet-time proton belt, *J. Geophys. Research*, **67**, 2645-2647, 1962.

AKASOFU, S. -I., and S. CHAPMAN, The ring current, geomagnetic disturbance, and the Van Allen radiation belts, *J. Geophys. Research*, **66**, 1321–1350, 1961.

AKASOFU, S. -I., W. C. Lin, and J. A. VAN ALLEN, The anomalous entry of low-rigidity solar cosmic rays into the geomagnetic field, *J. Geophys. Research*, **68**, 5327–5338, 1963.

BAME, S. J., J. P. CONNER, H. H. HILL, and F. E. HOLLY, Protons in the outer zone of the radiation belt, *J. Geophys. Research*, **68**, 55–64, 1963.

BEARD, D. B., Self-consistent calculation of the ring current, *J. Geophys. Research*, **67**, 3615–3616, 1962.

CAHILL, L. J., and P. G. AMAZEEN, The boundary of the geomagnetic field, *J. Geophys. Research*, **68**, 1835–1843, 1963.

DAVIS, L. R., R. A. HOFFMAN, and J. M. WILLIAMSON, Observations of protons trapped above 2 earth radii, Paper presented at the 45th annual meeting, A.G.U., Washington, April 21–24, 1964. To be published.

DAVIS, L. R. and J. M. WILLIAMSON, Low-energy trapped protons, *Space Research III, Proc. Third Intern. Space Sci. Symp., Washington*, 1962, ed. W. Priester, 365–375, North-Holland Pub. Co., Amsterdam, 1963.

DESSLER, A. J., W. B. HANSON, and E. N. PARKER, Formation of the geomagnetic storm main-phase ring current, *J. Geophys. Research*, **66**, 3631–3638, 1961.

KELLOGG, P. J., and J. R. WINCKLER, Cosmic ray evidence for a ring current, *J. Geophys. Research*, **66**, 3991–4001, 1961.

LIEMOHN, H., The lifetime of radiation belt protons with energies between 1 kev and 1 mev, *J. Geophys. Research*, **66**, 3593–3595, 1961.

NAKADA, M. P., J. E. MAZUR, and D. A. KNIFFEN, Variations of the proton energy spectrum with position in the inner Van Belt., *J. Geophys. Research*, **68**, 4065–4078, 1963.

PARKER, E. N., Newtonian development of the hydromagnetic properties of ionized gases of low density, *Phys. Rev.*, **107**, 924–933, 1957.

RAY, E. C., Cosmic-ray cut-offs at high latitudes, *J. Geophys. Research*, **69**, 1737–1741, 1964.

STRATTON, J. A., *Electromagnetic Theory*, McGraw-Hill Book Co., New York, 1941.

VESTINE, E. H., Low-level geomagnetic ring-current effects, *J. Geophys. Research*, **68**, 4897–4907, 1963.

N64-28...

AURORA*

BENGT HULTQVIST†

1. INTRODUCTORY DESCRIPTION

Descriptions of aurora of great antiquity are known. Some of the most famous are those by Aristotle (Meterologica, Book I, Chap. 5 (ca 340 B.C.)), Pliny (Historia Naturalis, Book II, Chaps. 27 and 33 (ca. 77 A.D.)), and Seneca (Quaestiones Naturales, Book I, Chaps. 14 and 15 (ca. 63 A.D.)). In central Europe the rare auroras were considered as warnings for coming calamities and the descriptions from the Middle Ages, especially, are characterized by superstitions and fears. In northern Europe where auroras are more common such superstitions did not influence the objective observation to the same extent. A remarkably good description can be found in the Norwegian chronicle The King's Mirror of about 1250. It is partly quoted by Störmer (1955) and he also quotes some good descriptions of auroral events of more recent time.

The name aurora borealis, or northern dawn, seems to have been first used by the French philosopher Gassendi in 1621, (Jacka and Paton, 1963) and is an appropriate description of the usual appearance of the northern lights when they are observed in medium latitudes. In analogy with aurora borealis the southern lights have been called aurora australis. "The northern dawn seen from middle latitudes is mostly the upper portion of a display which is overhead at higher latitudes. It is only when aurora moves equatorwards on the rare occasions of the great events that middle latitude observers can see those parts of the display which, being below the horizon, are so frequently concealed from them. Aurora then belies its title, since it is no longer an unspectacular glow on the horizon, like the dawn, and it is more aptly called "The Merry Dancers," the name by which it is known in the northern isles of the Britain" (after Jacka and Paton, 1963). In the last years, aurora has been made by man with the use of high-altitude nuclear explosions. The first artificial aurora was observed in 1958 (Cullington, 1958, Fowler and Waddington, 1958).

This review will concentrate on auroral characteristics which are influenced by the geomagnetic field and on results obtained in the last few years. Excellent extensive reviews of auroral phenomena have appeared, (Harang, 1951; Störmer, 1955; Chamberlain, 1961) some fairly recently, and the reader is referred to them for details on, e.g., the auroral spectrum, older morphological studies, correlations with other phenomena, etc. The emphasis here will be on the morphology, which is largely determined by the geomagnetic field. Rocket and satellite studies of primary auroral particles will also be treated in some detail, since they are not contained in the extensive treatises mentioned above. Finally, proposed theoretical models for various auroral phenomena will be briefly discussed in relation to the observations.

In this first subsection a short description of some auroral characteristics will be given.

1. Intensity and Energy Considerations

Aurora is one of the most intense luminous

TABLE 1.

Source	Flux (ergs cm^{-2} sec^{-1})
Sun	$1.4 \cdot 10^{6}$
Full moon	$3000 \cdot 10^{-3}$
Strong aurora	$1000 \cdot 10^{-3}$
Total starlight	$1.8 \cdot 10^{-3}$
Airglow (visible)	$16 \cdot 10^{-3}$
OH (infrared)	$19 \cdot 10^{-3}$
Lyman-α	$10 \cdot 10^{-3}$
Celestial sources (1230–1350Å)	$0.1 \cdot 10^{-3}$
Cosmic rays	$3.8 \cdot 10^{-3}$

*Published as *Goddard Space Flight Center Document X-611-64-97*, April 1964.

†Goddard Space Flight Center. NASA-National Academy of Sciences—National Research Council Senior Post-Doctoral Resident Research Associate on leave of absence from Kiruna Geophysical Observatory, Kiruna, Sweden.

phenomena in the sky, as is illustrated in Table 1 (after Friedman, 1960, except for the aurora value). The very strongest auroras are even more intense than the value given in Table 1. They are fully comparable to the full moon. The majority of auroras are, however, weaker and the whole scale down to the intensity of the airglow exists.

For estimation of brightness of aurora by visual observers, the following four classes of an International Brightness Coefficient (IBC) are used: I, the brightness of the aurora is equal to that of the Milky Way; II, brightness equal to that of thin moonlit cirrus clouds; III, brightness equal to that of moonlit cumulus clouds; and IV, the aurora provides a total illumination at the ground equivalent to full moonlight. These definitions are extremely crude. Seaton (1954) made a first effort to place the IBC scale on an absolute basis. His results have been slightly modified by Hunten (1955), who proposed the following definitions in terms of number of photons of λ 5577Å (generally the most important emission in aurora) per cm^2 (column) and sec.

IBC I	$10^9 = 1\ kR$
II	$10^{10} = 10\ kR$
III	$10^{11} = 100\ kR$
IV	$10^{12} = 1000\ kR$

10^6 photons/cm^2 (column) sec is called 1 Rayleigh (1R) in accordance with the proposal of Hunten et al. (1956). (For a discussion of the Rayleigh unit see also Chamberlain, 1961, Appendix II). The Rayleigh unit is not limited to a special wave length as is the IBC.

The energy influx required to produce visible aurora has been measured (McIlwain, 1960; O'Brien and Taylor, 1964) and has also been computed from optical emissions (Omholt, 1959; Chamberlain, 1961). The results show reasonable agreement. 0.2–1 percent of the input energy seems to be converted to visible light. The input energy required to produce an IBC I aurora is of the order of a few ergs/cm^2 sec (see e.g. Hultqvist, 1964c). Measurements by means of the Injun 3 satellite show that the average energy input rate required to sustain auroras around the world is of the order of 10^{18} ergs/sec (10^{11} watts) (O'Brien and Taylor, 1964). The average energy flux of the solar wind within a cylinder of radius 10 earth radii is of the order of 10^{20} ergs/sec. Thus in the average about 1 per cent of the solar wind energy has to be transformed into particle precipitation. However, during very strong auroras an appreciable fraction of the relevant solar-plasma energy flux must be converted into kinetic energy of precipitated particles, and the energy density of the solar beam may even be unsufficient on some occasions (Mogilevskiy, 1963). Large auroral events thus contribute much energy to the upper atmosphere. Most of this is probably radiated away (only some 15 percent transformed to heat according to Chamberlain, 1961). It seems quite possible that the temperature of the upper atmosphere in the auroral zones is primarily determined by the particle influx (Bates, 1960b; Chamberlain, 1961; Maeda, 1963).

Recent observations of the drag on Injun 3 indicate that the upper atmosphere temperature in the auroral zone is essentially the same as at the equator on geomagnetically disturbed days (Jacchia and Slowley, 1964). The heating accompanying geomagnetic perturbations in the auroral zones seems to be 4 or 5 times greater than the heating experienced during these perturbations in low latitudes.

2. Definition and Classification of Aurora

In connection with intensity the definition of aurora has to be considered. It has been fairly common to mean by aurora sporadic emissions of intensity at least a few times the airglow intensity but not to include emissions weaker than that. From a physical point of view this is not a good definition. Airglow and aurora, the production mechanisms of which are different, may well be thought of as being of similar brightness.

A definition based on the mechanism of excitation and emission would certainly be preferred. Along this line aurora is often defined as emissions in which the intensity ratio between the 3914 Å nitrogen band—which is produced by collision excitation—and the oxygen atom line at 5577 Å is of the order of unity. In airglow it is usually much less than unity. Unfortunately, however, the excitation mechanisms are complex and are not well-known, which makes it impractical to use them in definitions. One parameter seems to be of widely different importance for airglow and for

aurora: the geomagnetic field. Roach and Roach (1963) have defined aurora as emissions, the geographical location of which is determined by the geomagnetic field. The auroral phenomena then also include the stable red arcs, which are mostly subvisual and located in middle latitudes (see the review of Roach and Roach, 1963). We will in this review consider as aurora all emissions which are produced in one way or the other by particles of energy well above thermal and we will also include the above mentioned medium-latitude red arcs (m arcs), for which it is not known whether they are produced by particles (of very low energy) or not.

The aurora is an extremely complex phenomenon. The various types that occur can be grouped in the following six classes, for instance (due to Elvey, 1964)

1. Polar glow
2. Discrete polar cap auroras
3. High red aurora (Störmer's type III)
4. Intermediate red arcs (Störmer's type II)
5. M-arcs
6. Ordinary displays (Störmer's type I)

The polar glow aurora was discovered by Sandford (1961b). It is produced by the same solar protons in the energy range from some 10 Mev to several 100 Mev, which give rise to polar cap absorption. This type of aurora is thus fairly rare and has a pronounced maximum of occurrence frequency at solar activity maximum. The glow is distributed over the whole polar caps and is often subvisual.

The discrete polar cap auroras have several characteristics which are quite different from ordinary auroral-zone aurora (Störmer's type I). Their occurrence frequency is higher at low solar activity than at high and they also show a negative correlation with magnetic activity. The spectrum is, however, similar to that of ordinary auroral displays.

Class No. 3—high red aurora—is the most common type in low latitudes but it also occurs in the auroral zone sometimes.

The intermediate arcs are very rare. They often look grey due to faintness. The average lower height limit is 196 km (which is between the values for ordinary auroral displays and high red aurora; there-from their name). The red 6300 Å line is the most prominent spectroscopic feature. For more details see Störmer (1955).

The so-called ordinary auroral displays are, by far, the most common type of aurora and the rest of this review will deal mainly with them. Not only visual light is emitted by the aurora but emission also takes place in the ultraviolet (see for instance Chamberlain, 1961) and at radio-frequencies (cf e.g. Hartz, 1958, Murcray and Pope, 1960, Egan and Peterson 1961, and Martin and Helliwell, 1960).

In this review we will include among the auroral phenomena not only emissions of electromagnetic radiation but also the so-called radio aurora. The term radio-aurora refers to the effect on the propagation of radio waves, especially in the VHF and UHF frequency ranges, by the ionization in the atmosphere produced simultaneously with the excitation. Such ionization gievs rise to radio echoes of a special type.

3. The Development of an Auroral Display

The auroral displays show an enormous variety in their temporal development. One can, however, talk about typical sequences of events (see e.g. Heppner, 1954, Akasofu, 1963d). Equatorward of the auroral zones the display may begin with a glow along the poleward horizon. The glow then rises from the horizon and forms one or more quiet and mostly homogeneous bows of light (arc), oriented approximately east-west. An "arc" may remain relatively quiescent for hours, drifting slowly towards north or south. Suddenly it may brighten here and there and "rays" begin to appear across the arc. The arc is then likely to fold and so to lose its regular bow shape and to form an irregular "band." If the rays are very long, the band assumes the appearance of a great "drapery," waving like a curtain in the sky. Rapidly moving and fading irregular forms may appear and disappear all over the sky. As the display dies down, waves of light may surge upwards from the horizon in quick succession causing existing auroral forms to brighten as the waves pass over them (flaming aurora). A continuous glow (veil) may extend over a large part of the sky, serving sometimes as a background to the active forms. At other times, diffuse patches of light, closely resembling clouds, appear here and there in the sky. These

may come and go, often with regular pulsations, until twilight sets in. There is a tendency for reformation into arcs after the diffuse surface phase. The development pattern is closely correlated with the details of the coincident magnetic disturbance. The described sequence is typical equatorward of the auroral zone, as mentioned. On the poleward side, and especially over the central polar caps the development is quite different. This will be discussed in section 3. For some good descriptions of individual auroral events, see for instance Störmer (1955).

4. Auroral Forms

In the last few decades the auroral forms have mostly been classified in accordance with the recommendations of 1930 of the International Union of Geodesy and Geophysics. These recom-



FIGURE 1.1.—Auroral arc.

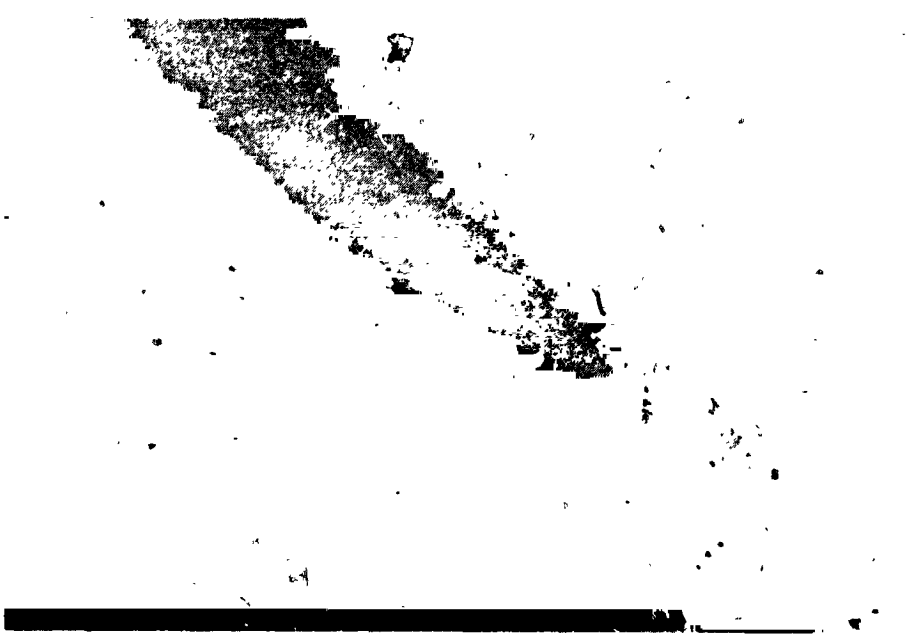


FIGURE 1.2.—A band.

mendations were worked out by Störmer (1930a, b) and are described in detail also in his book (Störmer, 1955).

The Committee for the "International Quiet Sun Year" (IQSY), 1964–1965, of the "Comité International de Géophysique" (CIG) has adopted a new scheme of nomenclature for describing auroras. It will be briefly reviewed below, following Jacka and Paton (1963).

An aurora is described in terms of 4 parameters (6 if brightness and color also are to be included).

1. Auroral forms are divided into three groups

1.1. Band-like forms: These are characterized by a continuous lower border.

(i) When the form appears as a simple, slightly curving arch (Fig. 1.1) or part of an arch, with a regular lower border it is called an arc (A).

(ii) When the lower border is irregular and contains kinks or folds (Fig. 1.2) the form is called a band (B).

1.2. Diffuse forms: These may take the form of a patch (*P*) or a veil (*V*).

(i) The patch is simply a blob of illumination with an ill-defined boundary; it often resembles an isolated cloud.

(ii) The veil refers to the uniform luminosity that may cover a large part of the sky. It may occur alone or as a background to other forms.

1.3 Rays (*R*): These are shafts of light aligned in the direction of the earth's magnetic field (cf. Fig. 1.3). There may be a single ray, a bundle of rays close together or many scattered rays (subscripts 1, 2, 3 may be added to *R* to indicate short, medium, and long rays respectively).

2. Structure is described by the following three adjectives

2.1 Homogeneous (*H*): This term implies an apparent lack of internal structure in the form. The brightness is uniform or it changes only gradually across the form. A homogeneous arc (*HA*), and horseshoe-shaped bands are shown in Fig. 1.4.

2.2. Striated (*S*): The form contains rather irregular striations or filaments aligned roughly parallel with the lower border (Fig. 1.5). Striation is usually recognizable only when the form is nearly overhead.

2.3. Rayed (*R*): Rayed forms are characterized by the appearance within them of rays which are aligned along the lines of force of the geomagnetic field (Fig. 1.6).

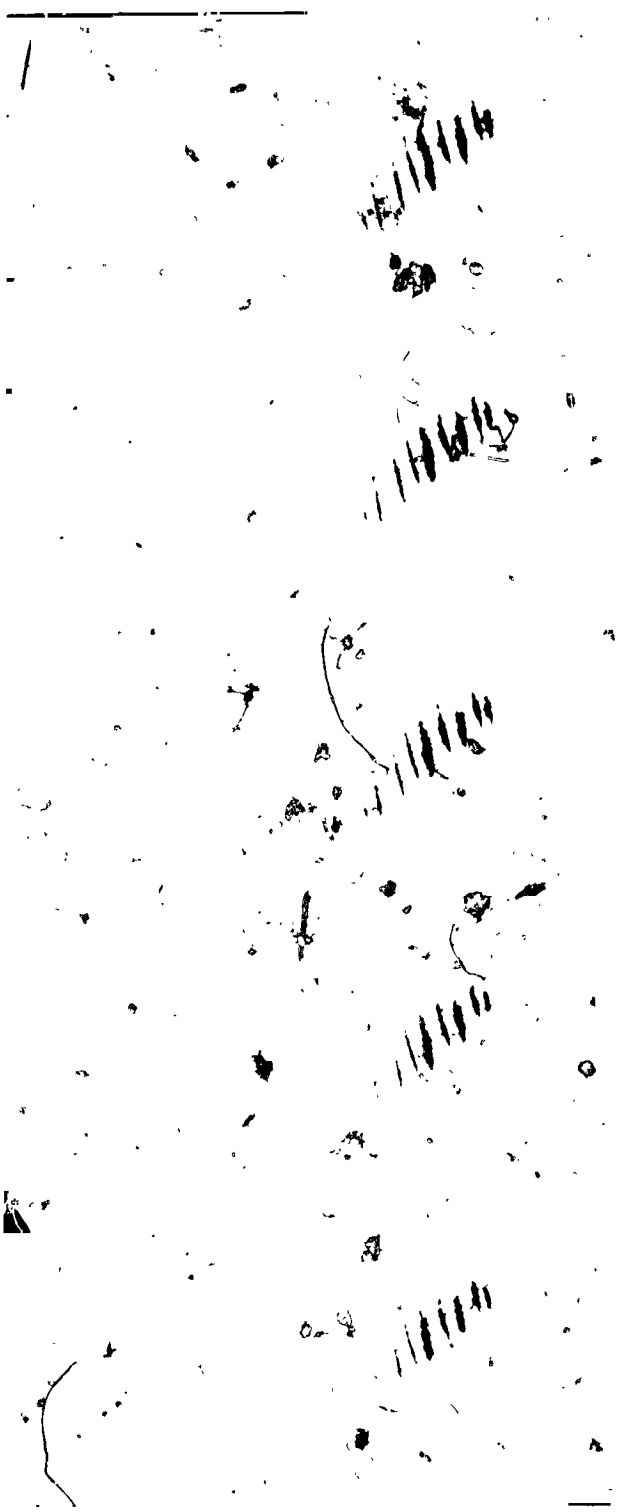


FIGURE 1.3.—Auroral rays photographed with an exposure time of 1/60 sec. 24 pictures per sec. (Courtesy T. N. Davis and G. T. Hicks).



FIGURE 1.4.—Homogeneous auroral arc and bands.



FIGURE 1.5.—Striated auroral arc.

3. *Qualifying parameters: One of three qualifying parameters may, when relevant, be added to the description*

3.1 Multiple (m): This term is used when there occur two or more associated and roughly parallel forms of the same kind. (A subscript numeral indicates the actual number of forms, e.g. m_2HB signifies two closely situated homogeneous bands).

3.2. Fragmentary (f): This term is used when only a portion of an arc or band is present.

3.3 Coronal (c): When any rayed form is viewed in the direction of the lines of magnetic force the rays appear to converge. Homogeneous forms

may also present a similar coronal appearance where they pass over the magnetic zenith.

4. *Condition describes the behavior of a form or of the whole display*

4.1 Quiet (q): The form undergoes only very slow changes in position or shape.

4.2. Active (a): is a form which moves or changes its shape rapidly. Such forms are frequently bright.

(i) a_1—refers to the movement of folds or irregularities along the boundary of a band.

(ii) a_2—refers to the rapid change of shape of the lower border of a form while the form itself may also move rapidly across the sky.

(iii) a_3—indicates the occurrence of a rapid movement of rays horizontally along a form.

(iv) a_4—refers to a display as a whole in which forms fade rather quickly, while new, similar or different forms appear in other parts of the sky. A photograph of such an event is seen in Fig. 1.7.

4.3. Pulsing (p): this describes a condition of fairly rapid and often rythmic fluctuation in brightness. The period of the fluctuation may range from a fraction of a second to the order of minutes.

(i) p_1—pulsating forms: The affected form or forms display a variation in brightness, the phase of which is uniform throughout the forms.



FIGURE 1.6.—Rayed bands.



FIGURE 1.7.—Photograph of an active aurora. The forms move and fade rapidly and new forms appear continuously in other parts of the sky.

(ii) p_2—flaming aurora. This kind of pulsing extends over a large area of the sky which appears lit by surges of light sweeping upwards over it.

(iii) p_3—flickering: A large part of the display undergoes very rapid and more or less irregular changes in brightness as if lit up by flickering flames. This occurs only rarely.

(iv) p_4—streaming: An irregular variation of brightness that progresses rapidly along the horizontal extent of homogenous forms.

Most auroras can thus be fairly well described by four symbols—or sometimes five—with subindices. For instance $a_3\,cfR_2B$ signifies a fragment (f) or a rayed band (R_2B) whose rays are of medium length (R_2) are moving rapidly in a horizontal direction along the form (a_3) and form a corona in the magnetic zenith. Symbols indicating the brightness and color may be added after the symbol for the form (for details see Jacka and Paton, 1963).

The auroral arcs and bands at the auroral zone are characterized by very long extension in east-west direction and a remarkable thinness in north-south direction. From the photographic recordings of aurora made during IGY continuous auroral arcs of lengths 4000–5000 km have been reported (Feldstein et al., 1962, Akasofu, 1963). The values mentioned were limited by the coverage of the net of recording cameras, so arcs may be still longer. Recent measurements of the thickness of 40 zenithal arcs have given values varying between 3.5 and 18.2 km with an average value of 9.1 km (Kim and Volkman, 1963). The thickness was found to increase with increasing magnetic activity. On the other hand, for an active auroral curtain the thickness has been measured to be as narrow as 150–350 m (Akasofu, 1961). No significant diurnal or seasonal variation was found. The vertical extension of the arcs and bands is generally a few ten km. It has been found to vary with solar activity, the extension being smaller and the atmospheric depth greater at the end of a solar cycle (Elvey, 1957). When the sunspot activity starts to increase again, the auroras often appear less deep in the atmosphere and have extension to great heights—the red auroras of type A.

The stable midlatitude red arcs, discovered by Barbier (1958), differ completely from ordinary auroral arcs. They cover some hundreds of kilometers in latitude and in height. Present evidence suggests that they may occur in complete rings around the world (cf. Roach and Roach, 1963).

The rays are characterized by great length and thinness. The length has been extensively studied by Störmer (1955). It may amount to several hundred km, but a more common value is between 100 and 150 km. Especially the sunlit rays can be very long and have been found to extend up to 1100 km sometimes. The diameter of the rays is of the order of 1 km (see e.g. the recent study of Akasofu, 1963b).

Wilcke (1777) first noticed that the auroral rays are aligned approximately along the earth's magnetic field lines. Vegard and Krogness (1920) found that the elevation of the "radiant point," i.e. the point towards which the rays in a corona appear to converge, is systematically one degree or so lower than the magnetic zenith. This difference is too great to be due to the curvature of the field lines. (The rays would have to be at about 1000 km to show a one degree displacement). Störmer (1938) found the auroral zenith to wander around a little in the sky. Abbott (1958) measured a rate of motion for the radiant point of at least 1° in 5 minutes of time. The radiant point was found on either side of the normal (undisturbed) magnetic zenith, but measurements of the disturbed field showed that the instantaneous magnetic zenith was 7 or 8° lower than the radiant point. No relationship between the motions of the magnetic zenith and the auroral zenith was found. Cole (1963) has studied in detail the geometry of the radiant point and found that it commonly differs by 0.5° from the "local auroral zenith," defined by him as the direction of a (hypothetical) ray, regarded as a segment of a straight line, passing through the observer. The difference may amount to as much as several degrees. Observations of the local auroral zenith have not yet been reported. Presuming the rays are oriented along the field lines such observations would give valuable information about the detailed direction of the geomagnetic field lines at auroral heights, which is difficult to obtain with comparable accuracy in other ways.

Detailed studies have shown that fine structures exist not only in auroral forms seen rayed to the eye or the camera but also in what is normally

called homogeneous forms (Nadubovich and Starkov, 1961). Dark filaments of a minimum width of 150 m and an average length of about 20 km have been found. They have been observed most often in the morning hours. The duration is several minutes.

There is a fine structure in aurora not only in space but often also in time. Rapid variations in auroral emission rate have recently been observed and studied by Campbell (1960a, b), Campbell and Rees (1961, Campbell and Leinbach (1961), Iyengar and Shepherd (1961), Omholt (1962), Skrynnikov (1962), Berger (1963), and Gustafsson (1964).

For a more detailed review of older investigations of the characteristics of various auroral forms, see e.g. Störmer (1955) and Chamberlain (1961).

5. Emission Spectrum

But for a very brief introductory survey, the auroral spectrum will not be dealt with in this review. For a detailed review see e.g. Bates (1960) and Chamberlain (1961).

The aurora seen in high latitudes is usually green to the eye. The colour is due to a forbidden transition ($'S \rightarrow 'D$) in the oxygen atom. The wavelength is 5577 Å, which is close to the maximum sensitivity of the eye. The transition probability is very low corresponding to a lifetime of the excited level of about 3/4 sec (Omholt and Harang, 1955).

In lower latitudes the aurora is often seen to be red (high red aurora in the nomenclature used). The spectral characteristic is then the 6300, 6364, 6391 Å triplett of O, which is produced in cascade ($^1D \rightarrow ^3P$) after the 5577 Å line. The 6300 line is also forbidden with a lifetime of the excited level of about 110 sec (Omholt, 1960). Due to the long lifetime, deactivation is important for this line and it is therefore mostly seen in auroras at high altitudes. The long lives of the metastable levels from which the λ 5577 Å and λ 6300 Å lines are emitted give rise to colour variations in active auroras as these lines are emitted for some time after the excitation has ceased.

There exist auroras even in high latitudes with anomalously increased 6300 Å emission. One famous example of this was the one on 11 Feb. 1958. Most other auroral spectra belong to one of three types, between which the boundaries are, however, not very sharp (Krasovskii, 1961). The first type is the low-latitude spectrum mentioned above, in which the ratio of the intensities of the 6300 Å and 5577 Å lines is high, and where these two emissions strongly dominate over e.g. the nitrogen molecule bands. In the next type, lines of neutral and ionized atoms dominate but in addition emissions of ionized and neutral nitrogen molecules are seen. This type of spectrum is observed in both high and low latitudes. Finally, the third type is characterized by numerous intense molecular bands. It is remarkable that in such spectra the green line of atomic oxygen and the first and second positive and negative band systems of N_2 are correlated in great detail. This third type of spectrum most frequently appears in high latitudes. The difference between the spectra are apparently accounted for by the depth of penetration of the exciting agent into the atmosphere, the depth increasing from the first to the third type.

Auroras located at great atmospheric depths usually have a red lower border (red aurora of type B). This is due mainly to the first positive bandsystem of the neutral nitrogen molecule. The reason for the domination of the N_2 emissions at great depths is probably that the $'S$ level of OI is deactivated by collisions to an appreciable amount.

The sunlit auroras have a spectrum different from that of the auroras located in the dark atmosphere. They have a bluish-white color which is produced by resonance emission of primarily the first negative band-system of the positively charged nitrogen molecule (N_2^+).

Of special interest are the hydrogen line emissions. They were first found by Vegard (1939). Analysis of their line contour showed that protons are entering the upper atmosphere with velocities up to 4000 km/sec (Meinel, 1951). They are seen mostly in homogeneous forms. More recently, weak hydrogen glow has been observed preceding the appearance of bright auroras (Romick and Elvey, 1958, and others). Auroras have been observed in which the $H\alpha$ line was so strong compared to other spectral emissions that this aurora might have been produced by proton impact alone (Omholt et al. 1962). In any case it appears certain that in those auroras the

protons supplied, relatively, at least one order of magnitude more energy than in other types of aurora, for which electrons are certainly the main energy contributors. The hydrogen glow is barely visible or even invisible to the eye. It occurs before the electron aurora and, when this has started, on the equatorside of it before midnight. After midnight no hydrogen emission is observed south of auroral forms but rather north of these. A distinct dark region is often seen between the proton aurora and the main forms (Stoffregen and Derblom, 1962). For more details see the reviews of Chamberlain (1961), Krasovskii (1961) and Galperin (1963).

In the midlatitude stable red arcs, mentioned earlier, no other spectroscopic feature than the 6300 Å line has been clearly identified. The location of these arcs is definitely controlled by the geomagnetic field. They are namely lined up along projections of circles from the geomagnetic equatorial plane to the earth's surface along the geomagnetic field lines. It is not clear how they can be produced by collison excitation without any other spectroscopic features being excited. According to Dalgarno (1964), the excitation may be a heating effect caused by low energy particles. On the other hand, if the excitation is chemical it is difficult to understand how the magnetic field control comes in. Megill and Carleton (1964) have proposed that local electric fields may be responsible for the excitation. A number of other proposals for excitation mechanisms are reviewed by Roach and Roach (1963).

6. Methods of Recording Visual Aurora for Morphological Studies

Almost all data concerning the morphology of aurora was taken by visual observers up to IGY. During IGY a large net of so called allsky-cameras, photographing the whole sky automatically every minute, was in operation. Various allsky-cameras have been constructed and used by Gartlein (1947), Lebedinskii (1955), Stoffregen (1955), Park (1957) and Davis and Elvey (see Elvey and Stoffregen, 1957, for a detailed description). A set of allsky-camera pictures are shown in Fig. 1.8. Most allsky films taken hitherto have been 16mm black and white film. The Soviet and Canadian cameras used, however, 35 mm film. In the last few years color film has begun to be used (Sandford and Heiser, 1959; Gadsden, 1960; Sandford, 1961). The committee for the "International Quiet Sun Year," 1964–65, of CIG has recommended the use of 35 mm film, which provides more details of the aurora. For detailed synoptic studies over large areas it is highly desirable that the net of cameras will be denser in the future than it was during IGY. Experience has shown that for most part of the polar regions the possibilities of obtaining large scale synoptic patterns from the IGY records are very limited. A most important contribution to the solution of the morphology problem may be given by very high altitude satellites, taking photographs showing most of the polar caps and auroral regions. Automatic devices for analyzing allsky-camera films have been constructed (Elvey and Belon, 1957; Nagata and Kaneda, 1961).

Davis, Deehr, and Leinbach (1960) have made an evaluation of allsky-camera observations by comparing them with simultaneous visual and photometric observations. They arrived at the following conclusions, among others.

(a) The US model allsky-camera operated during IGY (and probably most other types as well) was slightly less sensitive than an experienced visual observer in detection of very weak diffused forms and is about equal in detecting very weak but well defined forms.

(b) The allsky-camera is much superior to the visual observer in accurately locating and recording the shape of most auroral forms.

(c) The eye is more able to see the details of very bright or fast-moving forms. If the observer were to watch only a small part of the sky, his observations would, in general, be superior to those of the allsky-camera in that small region. However, if the entire sky is to be observed, the allsky-camera is superior in the sense of (b) above.

(d) A greater degree of continuity in auroral forms is often indicated in the compressed image of the sky on all-sky photographs than is apparent to the visual observer.

(e) During the times when visual observers report pulsating aurora, the over-all appearance of the display seems to be quite distinctive and often readily identified on the allsky film.

The exposure time of the ordinary allsky-cameras is generally around 20 sec. By means of image intensifier and TV camera it is possible

to take auroral pictures 24 times a second. Such films allow studies of two to three orders of magnitude more rapid time variations than has earlier been possible. An instrument for rapid filming of aurora was used successfully for the first time in the autumn of 1963 (Davis and Hicks, 1963).

It has recently become possible to observe visual aurora in daytime (Noxon, 1963).

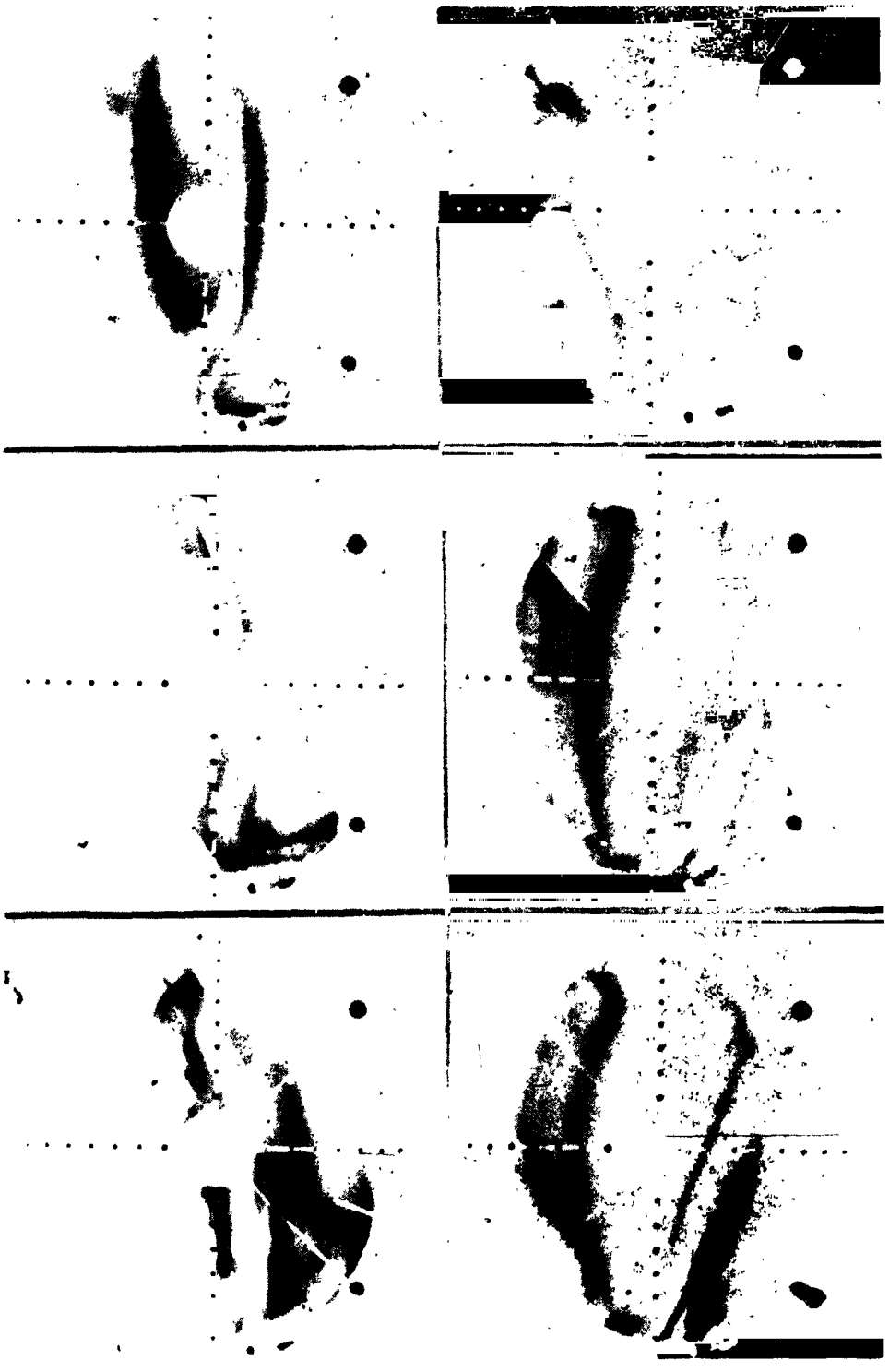


FIGURE 1.8.—Allsky camera pictures taken at Kiruna on 29 Jan. 1957. (Courtesy W. Stoffregen)

2. MORPHOLOGY OF VISUAL AURORA

1. Geographical Distribution of Auroral Activity

(a) Introduction

That the auroral activity does not increase continuously towards the poles seems to have been noted in print first by Muncke (1837). The dependence of auroral occurrence frequency on latitude was studied by Loomis (1860, 1868) in USA and by Fritz (1868, 1873, 1874, 1881) in Germany. Fritz (1881) made a detailed analysis of a large amount of data on aurora originating from very old times up to about 1878, from which he could delineate the first detailed map of the location of the northern auroral zone.

Fritz's auroral zone was defined as the locus of points where the probability of seeing an aurora in any part of the sky during any part of the night was maximal. It is thus a line, having no width. The curve connecting points with equal probability of seeing aurora in any part of sky at any time of night is called an isochasm. The auroral zone is thus the maximum isochasm. It has been recommended by Chapman (1953) and others to use the probability of seeing aurora overhead in evaluating geographical statistical distribution curves. Lines connecting points of equal probability of auroral occurrence in the zenith are called isoauroral lines or isoaurores.

It has become customary in recent years to think of the auroral zone as having a finite, but generally unspecified and variable, width. In this review, we will use the definition by which the width of the auroral zone is equaled to the distance between the points of half-maximum amplitude of a particular measure of auroral activity. The position of the peak of the auroral zone will be taken as that position at which there is maximum occurrence of overhead auroras, as determined by observations taken over a period of several months. With this definition we cannot talk about variation of the location of the auroral zone through one night or from one night to another, as is often done in the literature. One is then using the notation auroral zone for the area in which aurora exists at a given moment of time. Below, "auroral precipitation area" or "zone" or "momentaneous auroral occurrence zone" will be used for this unique-time distribution of visual auroral activity.

Both the width and the location of the center of the auroral zone is highly dependent on the type of activity-measure employed. The most commonly used measure is the occurrence frequecny of auroral activity, generally defined as the ratio of the number of half-hour (or full-hour) intervals in which aurora could be seen overhead to the total number of half-(or full-) hour intervals. The intensity of the aurora is not taken into account in this measure. Davis (1962 a, b) also employed auroral incidence, defined by the total number of auroral lower borders appearing each hour—as determined by counting the lower borders at 5-minute intervals—in a 1° latitude band passing through the station. It gives similar but somewhat sharper distributions.

The following terminology on auroral geography, suggested by Chapman, is widely used. The auroral regions (north and south) extend from geomagnetic latitudes (ϕ_m) 60° to the poles; the subauroral belts between 45° and 60° magnetic latitude; and the minauroral belt between 45° N and 45° S. The auroral regions include the aurorol zones and the auroral caps (the polar regions within the auroral zones). Alternatively, the regions on the equatorial side of the auroral zones are named cis-auroral regions, while those on the polar side which cover the polar cap, are called the transauroral regions (Jacka and Paton, 1963).

(b) Observed locations of the auroral zones and latitudinal distribution of visual auroral activity

The northern auroral zone delineation of Fritz (1881) and Vestine's (1944) modification of it are shown in Fig. 2.1 together with a zone determined by Feldstein (1960) on the basis of photographic recordings taken at 39 Arctic stations and visual observations made at three sites during the IGY, 1957–58.

While Fritz's auroral zone was based almost entirely on unsystematic auroral observations from several hundred years B.C. up to about 1878, Vestine had at his disposal more systematic observations, particularly those from the two polar years 1882–3 and 1932–3. Both delineations, however, are based on material obtained over very long periods of time with secularly changing geomagnetic field, varying solar activity, partly unknown meteorological conditions, etc. In contrast, Feldstein's curve is based on observations

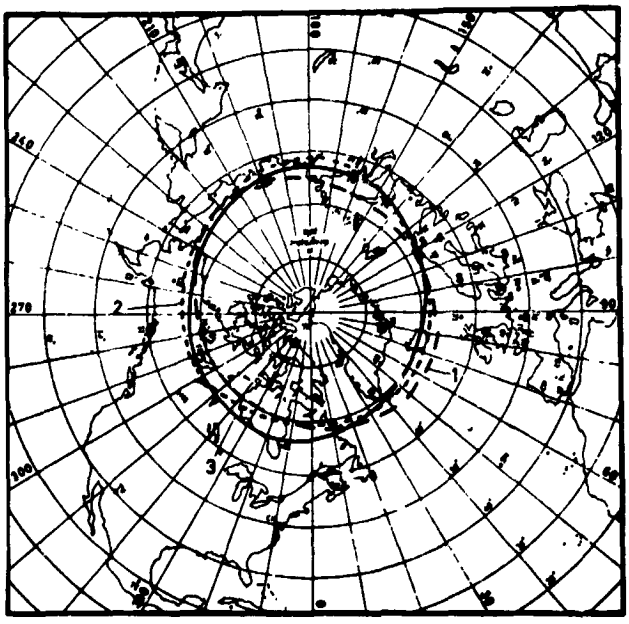


FIGURE 2.1.—Northern auroral zones determined from observations (on a map with geomagnetic co-ordinate system). Curve 1 is that of Fritz (1881), Curve 2 is Vestine's (1944) and Curve 3 Feldstein's (1960).

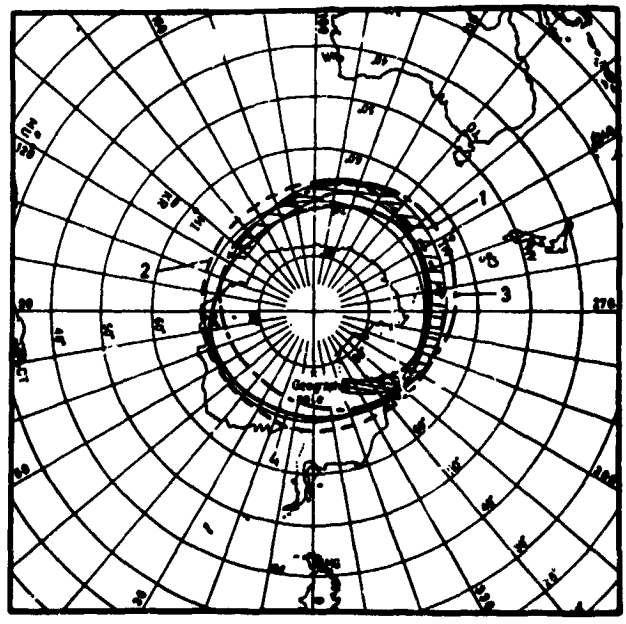


FIGURE 2.2.—Southern auroral zones determined from observations (on a map with geomagnetic co-ordinate system). Zone 1 is according to White and Geddes (1939), Zone 2 to Vestine and Snyder (1945), Zone 3 to Bond and Jacka (1960) and Zone 4 to Feldstein (1960).

made within a few years with the use of a fairly uniform technique at all stations. The auroral indices employed in his determination are more accurate than those previously used, as only the aurora in the zenith has been utilized.

The greatest difference between Vestine's and Feldstein's northern auroral zones is in the region of Hudson Bay, in North America, where Feldstein's curve reaches about 4° farther south than Vestine's.

The first attempt to investigate the auroral statistics in the southern hemisphere was made by Boller (1898). White and Geddes (1939) estimated the location of the zone on the basis of scattered visual observations of aurora australis, and Vestine and Snyder (1945) defined a zone based on geomagnetic observations in the Antarctic. These two zones are shown in Fig. 2.2 together with Bond's and Jacka's (1960) and Feldstein's (1960) zones. The IGY material used by Feldstein was taken at 20 stations on the Antarctic continent and at two stations on islands situated south of Australia. The technique employed in the evaluation was similar to that used in the Arctic.

A recent study of the detailed latitudinal variation of auroral occurrence frequency is that of

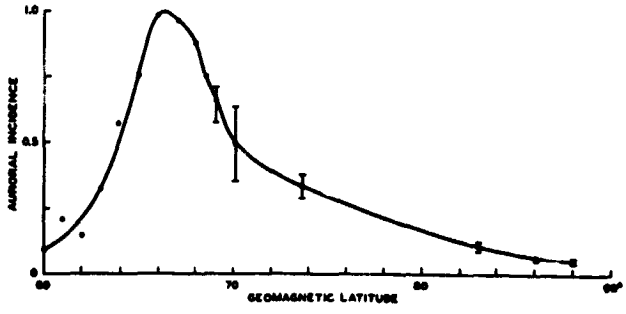


FIGURE 2.3.—(After Davis, 1962b) The incidence of auroral forms at geomagnetic latitudes 60–88°.

Davis (1962 b), shown in Fig. 2.3. It refers in time to the solar activity maximum in 1957–58. The "incidence" of auroral forms, as defined above, was deduced from allsky-camera records taken in Alaska (for the geomagnetic latitude, ϕ_m, less than 69°) and in northern Canada (for $\phi_m > 69°$). The indicated spread for the points representing incidence at high latitude is due to the uncertainty involved in connecting the Alaskan data with those of the high-latitude Canadian stations.

The curve has its maximum between 66° and 67° geomagnetic latitude in Alaska. The width of the zone, as defined above, is about 6° in latitude. The scale for the "incidence" in Fig. 2.3

is a relative one. The hourly occurrence frequency was found by Davis (1962 a) to be between 60 and 70 percent at about 66° geomagnetic latitude at solar activity maximum. Already at $\phi_m = 60°$ it was down to 10 percent, while on the inner side of the auroral zone, the decrease was slower (c.f. Fig. 2.3). Feldstein and Solomatina (1963) found that the frequency of auroral occurrence in the zenith during IGY was about 55 percent on magnetically quiet days and about 90 percent on disturbed days in the auroral zones. The index of activity used was the radio of the number of half-hour intervals with auroras in the zenith to the total number of half-hour intervals of observation, as recorded in the form of ascaplots. The last mentioned investigators also found that the probability of auroral occurrence in the boreal and austral auroral zones are very closely the same.

Nothing was known before IGY about the detailed relations between auroral phenomena in the two hemispheres, but in the last few years some results on this have appeared. They are, however, not consistant. While Jacka (1961) did not find any correlation between the occurrence of aurora at College and MacQuaire Island, which are nearly conjugate points (off by 600–700 k
and Fillius (1960) reported correlation between the general auroral activity at Ellsworth Antarctic station and the conjugate area in USA, but very small correlation (if at all significant) between the detailed pictures in the two areas, DeWitt (1962) observed good detailed correlation of form and motion and of variation of luminosity between auroral events observed at Campbell Island and MacQuaire Island in the southern hemisphere and at Farewell and Kozebue in the northern one. Breakup of the auroras was found to be simultaneous in the conjugate areas. The conjugacy of the observation points employed by DeWitt was certainly better than for any of the other two investigations quoted above. It seems therefore most probable that auroras usually occur as conjugate pairs at either end of geomagnetic field lines, as has long been suspected. More observational data would be of great value.

The occurrence of aurora in very low latitudes (subtropical and tropical) is extremely rare. Even in central Europe and southern USA the probability of seeing aurora is only a few thousandths of that at the center of the auroral zone. Instead, these medium and low latitude auroras are generally very spectacular and have often been noted and recorded also in ancient times. Recent summaries of older observations of auroras extending into the minauroral region have been made by Schove (1955), Matsushita (1956) and Chapman (1957 a, b). Several auroras occurring at very low latitudes have been described rather fully by Chapman (1953, 1957 c), Abbott (1951), and Abbott and Chapman (1959 a). A type of red auroral arcs occurring rather regularly in subauroral latitudes will be discussed later in this chapter.

(c) Comparison of observational auroral zones with various theoretical curves

There are good reasons for postulating circular symmetry in the geomagnetic equatorial plane outside the earth's surface of distance parameters representing statistical averages for corresponding variables of those processes, in and near the equatorial plane, which are associated with the occurrence of auroral phenomena in or near the two auroral zones. Only the dipole field affords a measurable contribution in the equatorial plane at a distance of several earth radii from the earth's surface, and the dipole field is circularly symmetric. Thus the configuration in the geomagnetic equatorial plane will not be influenced by the earth's rotation, provided minor effects of the different directions of the geomagnetic and the rotation axes are disregarded. The equatorial plane process will therefore be seen in the same average distance range at a given geomagnetic time from all geomagnetic meridians on the earth. It follows that—if the processes in the equatorial plane and in the auroral zones are connected by the real magnetic field lines, as seems physically plausible—the only longitudinal dependence of e.g. the averages of a great many auroral locations, observed at fixed points on the earth and equally distributed with regard to geomagnetic time, will be that introduced by the deviation of the earth's magnetic field from a dipole field. This deviation is appreciable at and near the earth's surface.

What has been said here should be true, even if the geomagnetic field is heavily distorted in its outer regions. If that is so, the distance at which the process in the equatorial plane take place may

be quite different from that of the intersection with the equatorial plane of that dipole line which intersects the earth's surface at the same place as the real field line connecting the auroral processes in the earth's atmosphere and in the equatorial plane.

According to this view the configuration in one hemisphere will correspond to the projection of the corresponding one in the other hemisphere along the geomagnetic field lines.

The only assumption made in the foregoing is that an electron or an ion moves from the equatorial plane to one of the auroral zones along a geomagnetic field line or on some surface defined by field lines. Such an assumption seems to be well founded. No other details of the actual physical processes are involved.

The projections of circles in the equatorial plane to the earth's surface along the real field lines, were calculated by Hultqvist (1958) for purposes of comparison with experimental isochasms. The projections are practically identical with the L curves of McIlwain (1961). Hultqvist used a perturbation method consisting of integration along the dipole line of the deflection from a dipole field line of the higher approximation line due to the effect of the second to the fifth terms in the spherical harmonic development of the earth's magnetic field. The calculations yielded deflection vectors for a large number of dipole lines. These gave the deflection for the point of intersection of the higher approximation field line—coinciding with the pertinent dipole line in the equatorial plane well outside the earth's surface—and the earth's surface from the corresponding point of intersection of the dipole line itself.

The calculations were carried out for 36 equidistant longitudes for each of 15 geomagnetic colatitudes, most of them less than 30 degrees. The results of Vestine's et al. harmonic analysis of the earth's magnetic field for epoch 1945, the most recent epoch for which data were then available, were used in the computation.

The projections of equatorial plane circles are ovals with the longest diameter approximately in the plane containing the 170° and 350° geomagnetic meridians. Two circle projections are shown in Fig. 2.4 (northern hemisphere) and Fig. 2.5 (southern hemisphere). They correspond to the two circles in the geomagnetic equatorial plane

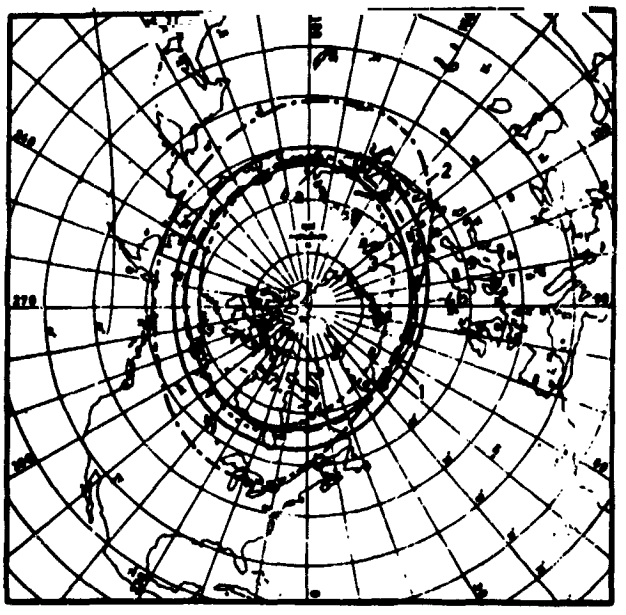


FIGURE 2.4.—Comparison of Feldstein's (1960) northern observational zone (Curve 1) with a number of "theoretical" auroral zones: 2 is according to Gartlein and Sprague (1960), 3 to Quenby and Webber (1959) and 4a and b are projections of two circles in the geomagnetic equatorial plane (corresponding to colatitudes 22 and 25 degrees respectively; Hultqvist, (1958).

having radii of 7.13 and 5.60 earth radii, for which the projections along the dipole lines are the geomagnetic colatitude circles 22° and 25° respectively.

Hultqvist pointed out that the experimental northern auroral zone of Vestine accorded well with the circle projections except over the North American continent, where the calculated curve extended about 4° farther south than the observational one. The IGY data as presented by Feldstein and illustrated in Fig. 2.4 show that the configurational agreement between the IGY northern auroral zone and the projections of circles is good. The maximal difference in latitude between Feldstein's zone and that projection which conforms best to it amounts to about one degree.

Since the network of observation stations in and near the northern auroral zone was quite dense during the IGY, it seems probable that the error in location of the experimentally determined auroral zone is only of the order of 1° in latitude. The geomagnetic data used in the calculations are for epoch 1945, which may account for minor inconsistencies with the observations of 1957–58. It is probable, however, that this error also is less than one degree of latitude. It is thus evident

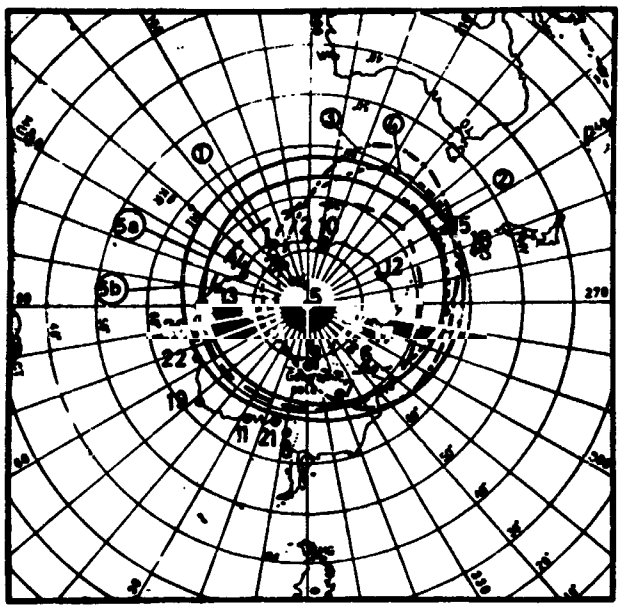


FIGURE 2.5.—Comparison of Feldstein's (1960) southern observational zone (Curve 1) and that of Bond and Jacka (1960) (Curve 2) with different "theoretical" auroral zones: 3 is according to Gartlein and Sprague (1960), 4 to Quenby and Webber (1959) and 5a and b are projections of two circles in the geomagnetic equatorial plane (corresponding to colatitudes 22 and 25 degrees, respectively; Hultqvist, 1958). The numbered points show the approximate locations of the stations at which the IGY observational material—the basis of Feldstein's auroral zone—was collected. The numbering is that of Feldstein (1960).

that the IGY northern auroral zone is, within the limits of probable error, in full configurational accord with the calculated projections of circles in the equatorial plane.

Feldstein's auroral zone in the southern hemisphere is shown in Fig. 2.5 together with the projections of the circles of radii 5.60 and 7.13 earth radii. The experimental curve differs quite significantly from the computed ones. While it follows the course of the latter in the geomagnetic longitude range of 320° E—90° E, it is much nearer to the Antarctic continent for the rest of the longitude circle where it is located over the sea. In fact, Vestine's and Snyder's (1945) auroral zone agrees better with the computed one than Feldstein's does.

A question of interest is whether the discrepancy between the IGY southern auroral zone and the projections of circles is a real one or attributable merely to inaccuracy of the observational data and/or of the computed curves. The errors of the geomagnetic data, used in the computations, are certainly greater for the Antarctic than for the Arctic. As mentioned earlier, the computations were based on Vestine's et al. harmonic analysis of the geomagnetic field for epoch 1945, when the southern hemisphere data were very sparse. Although this fact may partially account for the discrepancy, it seems likely that only a minor part of the 6 latitude degrees difference over the sea is parimarily attributable to errors in the harmonic analysis of the geomagnetic field, especially as a comparison with recent numerical computations by Dudziak et al. (1963) on the basis of the spherical harmonic coefficients for epoch 1960 by Jensen and Cain (1962) show only minor deviations.

All observational stations in the discrepant region were situated on or close to the coast of the Antarctic continent far inside the auroral zone. This means in effect that the auroras occurring in the auroral zone were visible from these stations at low elevation, a circumstance which implies considerable error in the observational results.

Indeed, since publication of Feldstein's paper, Bond and Jacka (1960) have communicated a report in which they note that the latitude of the auroral zone in the region of MacQuarie Island, south of Australia (point 15 on Fig. 2.5), is only one or two degrees more than that of the station. A small part of their auroral zone—the one near MacQuarie Island—is shown in Fig. 2.5 as curve No. 2. Taking this and the above mentioned considerations into account, it would appear that the auroral zone is fairly consistent with the computed curves, with which it shows agreement over the Antarctic continent. This seems so much the more plausible in view of the close accord between observed auroral zone and circle projections for the northern hemisphere. Indeed, it would be surprising if the two hemispheres showed entirely different degrees of agreement between the two types of curves.

Although, as mentioned above, there are good reasons for postulating circular symmetry in the geomagnetic equatorial plane for the processes producing aurora in the earth's atmosphere, it is certainly not the only possible approach. If e.g. the drift around the earth of the primary auroral particles is accordance with the adiabatic invariants is the determining process for the configuration

of auroral occurrence curves, it is perhaps reasonable to expect that the isoaurores should be identical with the lines of constant second adiabatic invariant. Vestine and Sibley (1960) computed such curves. They are very nearly identical with the projected equatorial-plane circles and it is therefore not possible to judge between the two approaches on that basis. Vestine and Sibley also calculated, using the spherical harmonic analysis of Finch and Leaton (1957), the mirror height of electrons for which the second adiabatic invariant, $J = \int_{\parallel} d\ell/v$, has the value 15.7. Their heights range from 57 km to 420 km in the northern hemisphere and from -15 km to 633 km in the southern hemisphere.

Figure 2.4 and 2.5 include, for purposes of comparison, curves showing two other types of more or less theoretical auroral zones: the zones suggested by Gartlein and Sprague (1960), which are almost identical with the isoclines 76°, and Quenby's and Webber's (1959) curves, which are based on calculations of vertical cut-off rigidities for cosmic ray particles in the earth's magnetic field, and take into account both the dipole and non-dipole parts of the internal field. Neither of these two curves conforms as closely to observed data as do the circle projections.

What conclusions can be drawn from the similarity between observed isoaurores and projections of circles in the earth's geomagnetic equatorial plane into the earth's atmosphere along the geomagnetic field lines? Not too many, in fact, since no details of the actual physical processes are involved. The geomagnetic field can be strongly distorted and we still expect to find the circular symmetry for the statistical data.

However, it can be concluded from the data presented above that local conditions close to the earth probably play a negligible role for the average auroral distribution over the earth's surface. The adiabatic drift of the primary particles around the earth seems not to be important for the shape of the insoaurores. If this process were important, one would expect the mirroring height to have a profound influence on the precipitation rate of particles and therefore on the frequence of occurrence of aurora. Since the mirror height varies very much with longitude for one $J = \text{const}$ shell, important variations in intensity and auroral occurrence frequency for various points along such a curve would be expected. This is not observed. That the drift motion is not important in this respect is also expected because of the fact that most primary auroral electrons need a time very much longer than the time scale of auroras in order to drift around the earth.

There are other statistical parameters than the isoaurores that may be expected to show circular symmetry in the equatorial plane. One is the average southern extent of the aurora a a definite degree of magnetic disturbance. Such curves have been evaluated by e.g. Gartlein et al. (1960) and Feldstein (1963). A comparison of

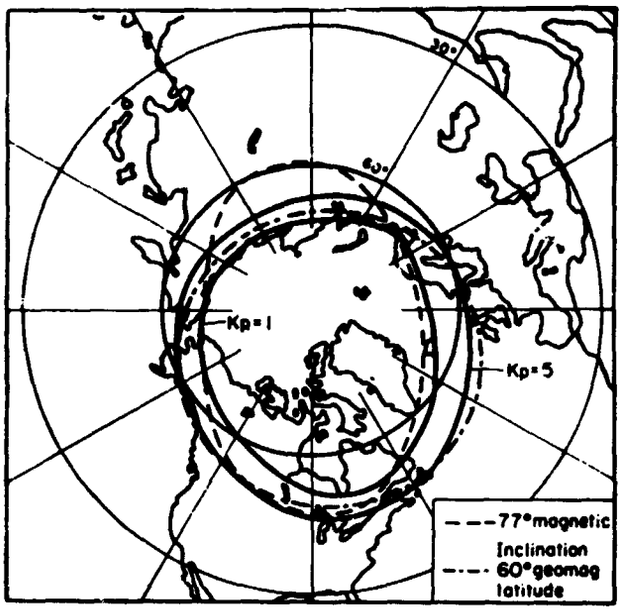


FIGURE 2.6.—Southern extent of aurora borealis for two different levels of geomagnetic activity (after Gartlein, Gartlein and Sprague, 1960).

the curves in Fig. 2.6 with the circle projections in Fig. 2.4 shows a good agreement. The southern extent increase with increasing K_p.

(d) Dependence of auroral zone location on the solar cycle

The pole distance of the auroral zone has been found to be dependent on the phase of the solar activity cycle. Sheret and Thomas (1961) made a detailed analysis of the auroral observations made at Halley Bay, Antarctica, in the years 1956–59. The position of the auroral zone, obtained from observations of quiet arcs, was the following in the four years.

TABLE 1.

Year	1956	1957	1958	1959
median south geomagn. lat	70.8	71.8	71.9	71.0

The auroral zone was thus closer to the pole at the solar-activity maximum in 1957 than before and after. This result is opposite to that of Feldstein (1962), who compared the location of the auroral zone in the Asian Arctic at solar activity maximum (1957–58) and at solar minimum (1954–55). He found the auroral zone to be 2.5—3° closer to the geomagnetic pole at the minimum of the solar cycle than at maximum. These variations in the location are discussed also in subsection 6 on motions.

(e) Auroral type effects

The distribution of auroral activity described above refers to all types and forms of aurora. Most of the basic data are obtained from black and white allsky-camera films of fairly long exposure times, which makes it impossible to distinguish between various spectroscopic characteristics as well as between quiet and rapidly varying auroral forms.

It is, however, of great theoretical interest to know whether there are significant differences in location between different types of aurora, especially between auroras caused by positive and negative primary particles. This is because relations—at least qualitative—between the locations of proton and electron precipitation areas is contained in several theoretical models. For instance, it seems difficult to think of electromagnetic acceleration processes if possitive and negative particles are precipitated together along the geomagnetic field lines.

In fact, it has been found that hydrogen line emitting aurora (see p. 13) usually is located on the equator side of the electron produced aurora before midnight (cf. the reviews of Omholt, 1963, and Galperin, 1963). (After midnight no hydrogen emission is observed south of auroral forms, but rather north of these; Stoffregen and Derblom, 1962). There is often a distinct dark region (up to 1° wide in latitude) between the "proton aurora" and the main forms. One would therefore expect the isoaurores and the auroral zone for "proton aurora" alone to be located one or a few degrees farther from the poles than the ordinary auroral zones (which certainly are the electron precipitation zones).

As mentioned in section one, there are several different types of auroral phenomena. The latitudinal distribution of the occurrence probability for high red arcs, for instance, certainly has its maximum in the subauroral regions. No detailed statistical studies of their geographical distribution exist as yet. Stoffregen (1962) has investigated the occurrence of different auroral forms in the geomagnetic latitude band 58–76° N (at the longitude of Scandinavia). His results, which he considered as a rough approximation, are shown in Fig. 2.7. As can be seen in the figure, the

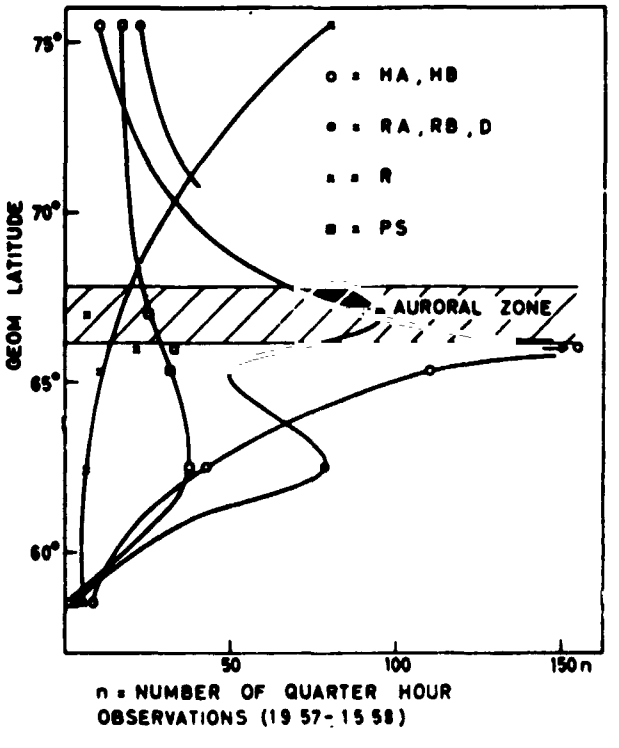


FIGURE 2.7.—(After Stoffregen, 1962) Approximate latitude distribution for different auroral types.

pulsating auroras occur most frequently some 5 degrees south of the auroral zone. This was also found by Heppner et al. (1952).

The low latitude aurora mentioned above is different from the stable subauroral red arcs described on p. 10. The red arcs populate sub-

auroral mid-latitude regions with shape very similar to the projections of circles along the field lines from the equatorial plane, described above (see Roach and Roach, 1963). The regions have their centers somewhere between the circle projections corresponding to equatorial plane circles of radii 2.5 and 3 earth radii (i.e. $L=2.5$ and 3 respectively). These red arcs seem not to exist in the auroral zones.

There certainly exist still other differences in the latitudinal distribution of the occurrence frequency for different types of aurora. The aurora inside the auroral zone has several characteristics that distinguish it from the aurora seen at and outside the auroral zone. In general, the high-latitude forms are relatively weak in intensity and more fleeting than those seen at the auroral zone. The forms consist, in part, of weak diffuse arcs and of bands that frequently show ray structure. Ray bundles are common, and these are usually arranged so as to appear as part of arcs, bands or draperies with variable intensity (cf. Davis, 1962 b).

(f) Inner auroral zones

Alfvén (1955) proposed on theoretical grounds that there should exist secondary auroral zones inside the ordinary ones, at geomagnetic colatitudes between 5 and 10 degrees. Since this proposal was made, a large number of attempts have been made to determine from observations if the inner auroral zones exist or not. At this time, it does not seem possible to draw firm conclusions. It is clear that the zone, it if exists, is much less pronounced than the main zones (which is not unexpected from Alfvén's theoretical point of view).

Observations in favor of and against the existence of the inner zones will be briefly reviewed.

Nikolskiy (1963) has discussed the Soviet data relevant to the existence of inner auroral zones. Most data is geomagnetic. What is generally found is a morning maximum in magnetic activity between 75 and 80° N geomagnetic latitude, most pronounced in the summer. The shape of the zone, found by Nikolskiy, is a spiral in a circular latitude-time coordinate system, unwinding in a clockwise direction. Meek (1957) found that the maxima of ionospheric disturbances are concentrated along a similar spiral.

On the other hand Lassen (1959), observing at Godhaven (about 80° geomagnetic latitude), could not find any magnetic evidence for an inner zone of high activity but instead he observed a type of zenithal aurora occurring around 6 h. local time, which seemed not to come up from south (i.e. from the main auroral zone) but to be produced sporadically at great elevation in the sky seen from Godhaven. The elevation did not change significantly in the course of the morning. Lassen found that the morning-type of aurora is uncorrelated not only with the local magnetic activity but with the planetary state of disturbance too. Analyzing data obtained in other parts of Greenland and in the northernmost Canada, Lassen concluded that morning aurora of the described kind is observed in the zenith only along a curve coinciding with the projection of the circle with radius 20 earth radii in the geomagnetic equatorial plane ($L=20$) on the earth's surface.

Witham et al. (1960), for instance, investigated the magnetic disturbance characteristics. They, and many other investigators of various parameters, could not reach any definite conclusion about the existence of an inner zone of high disturbance.

Davis (1962 b) analyzed allsky-camera films taken at Thule ($\phi_m=88°$), Alert ($\phi_m=86°$), Resolute Ray ($\phi_m=83°$), Baker Lake ($\phi_m=74°$) and Churchill ($\phi_m=69°$). The scaling was made by counting the number of auroral forms within zenith distance 80° on each fifteenth frame of the film from each station. The total number of forms was recorded. The result can be seen in Fig. 2.3. There is no indication from this diagram of an inner auroral zone. However, the lack of stations in the critical region between 74° and 82° prevents the making of a strong statement of the absence of a zone there. It appears that if there is an inner auroral zone, it must lie between 76° and 81° latitude over northern Canada, and that the enhancement of the number of auroral forms within this zone must not be great.

2. Diurnal Variation of Visual Auroral Activity

As does the geographical distribution of aurora, the diurnal variation of auroral activity varies considerably with the measure of the activity

used. In addition, it depends on the type of auroral display and perhaps on season and phase of the sunspot cycle. Very much still remains to be done before our knowledge of the diurnal variation of auroral activity is consistent and complete.

The most widely used measures of auroral activity are "occurrence frequency" and "incidence," defined in the previous subsection. Fig. 2.8 (after Davis, 1962 a) shows the occurrence

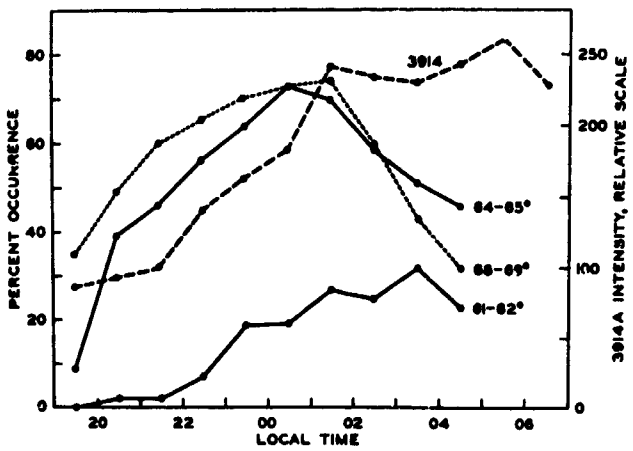


FIGURE 2.8.—(After Davis, 1962a) Percentage hourly occurrence in a latitude belt of width 1° centered over each of the stations: Barrow (68°–69°), College (64°–65°), and Farewell (61°–62°). The curve (3914) is drawn from data by Murcray (1959) and represents relative all-sky 3914Å intensity over College during 1955 and 1956.

frequency of visual aurora in a latitude belt of width 1° centered over each of the Alaskan stations Barrow (68–69°), College (64–65°), and Farewell (61–62°). Geomagnetic midnight occurs 1 to 2 hours after local midnight, the displacement from local midnight being largest for Barrow and smallest for Farewell. These curves demonstrate a shift of the auroral occurrence toward earlier local time with increasing latitude. In Fig. 2.8 there is also a curve marked 3914. It is drawn from data by Murcray (1959) and represents relative allsky 3914 Å intensity over College during 1955 and 1956. As can be seen in the figure, this curve has its point of gravity located several hours later than the others. This is due to an important difference between the auroral intensity distribution in premidnight and postmidnight hours. Before midnight the auroral forms stand out sharply above the low intensity of the background, but after midnight the intensity of the background is enhanced so that individual auroral forms are not easily distinguished.

Davis and DeWitt (1963) have evaluated the per cent of hours in which aurora was observed at Byrd, Antarctica, close to the southern auroral zone, for every hour of day and night in a two-months period of southern winter. Their result is shown in Fig. 2.9. The auroral occurrence

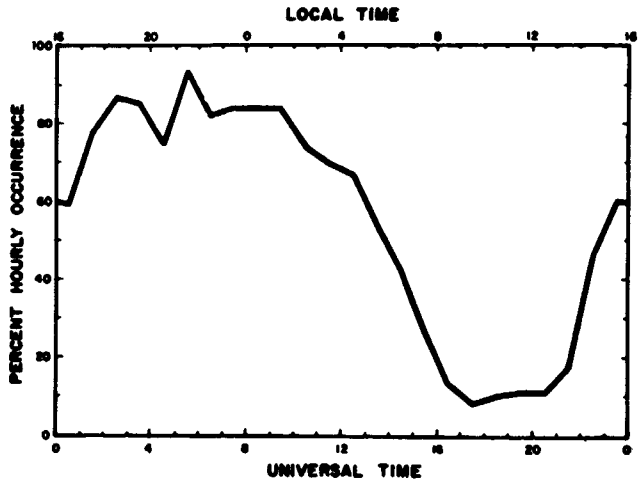


FIGURE 2.9.—(After Davis and DeWitt, 1963) Diurnal variations of hourly occurrence of visual aurora within 150 km of Byrd Station for all clear, dark hours in May and June 1960.

frequency has a maximum around geomagnetic midnight (at about 2200 local time) and a deep minimum around geomagnetic noon.

Reports on diurnal variation of aurora have also been given by Tromholt, 1882; Carlheim-Gyllenskiöld, 1886; Paulsen, 1893; Vegard, 1912; Chree, 1927; Sverdrup, 1927; Lee, 1930; Fuller, 1933, 1935; Currie and Edwards, 1934; Davies, 1935 a, 1950; Stagg, 1937; Stetson and Brooks, 1942; Jacka, 1953; Elvey et al., 1955; Murcray, 1959; Hale, 1959; Lassen, 1959, 1961; Malville, 1959; Chamberlain and Thorson, 1960; Feldstein, 1960; Feldstein and Solomatina, 1961; Loginov, et al., 1962 and others. A large number of these investigators have found diurnal variation curves similar to the occurrence curves in Fig. 2.8 and 2.9 with miximum fairly close to geomagnetic midnight.

There have been many suggestions of a less pronounced, secondary maximum in the early morning hours but it has not been possible to

find at some stations (Hulbert, 1931). Studies of ascaplots for IGY by Feldstein (1960) and Feldstein and Solomatina (1963) have resulted in the curves of Fig. 2.10 showing the geomagnetic time of maximum auroral occurrence frequency as function of the distance of the observation place from the observed auroral zone (curve 3 in Fig. 2.1), which distance is an approximately linear function of the geomagnetic latitude, corrected for the higher terms in the spherical harmonic expansion. There are two spiral-shaped curves in Fig. 2.10 unwinding in opposite directions. At the auroral zone, Feldstein found only one maximum (around geomagnetic midnight). Also, at geomagnetic latitudes around 78° only one maximum occurred (on the dayside) and at latitudes in between there were two maxima. The same result has been presented by Malville (1964). Similar functions of latitude have been found for the time of maxima of various other disturbance measures: e.g. for geomagnetic distrubances by Meek (1955), Burdo (1957) and Nikolskiy (1957), for sporadic-E ionization by Hagg et al. (1959), for redio aurora by Forsyth et al. (1960) and Egeland et al. (1962) and for radio wave absorption by Thomas and Piggot (1960).

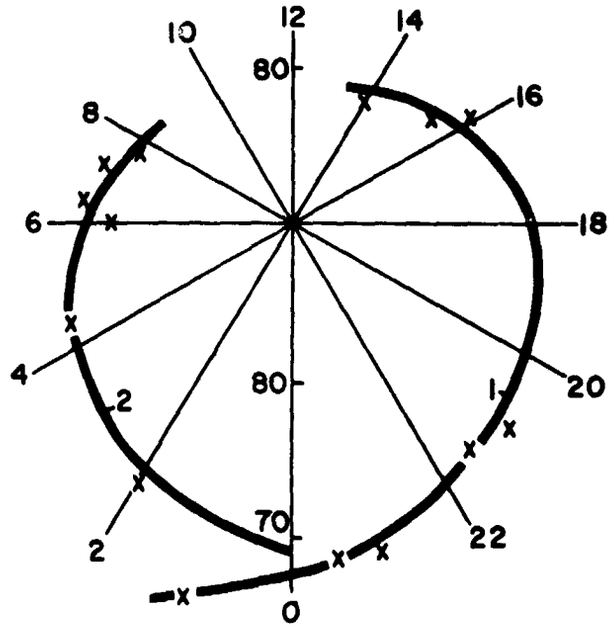


FIGURE 2.10.—(After Feldstein and Solomatina, 1963) Geomagnetic times of maximum frequency of auroral occurrence in the zenith as function of geomagnetic latitude corrected for the non-dipole part of the geomagnetic field.

On the basis of their observational results, Feldstein and Solomatina (1961) and Khorosheva (1961, 1962) (see also Feldstein, 1964) have proposed that the precipitation curve for aurora at a given time has the shape of the two combined spirals in Fig. 2.10. An instantenous auroral zone of a similar shape has been proposed by Akasofu, (1963 d). Support for this shape are among other observations those of Khorosheva (1961), who mapped 17 cases of momentary auroral distribution over the eastern hemisphere and found parts of oval-shaped curves similar to that in Fig. 2.10. However, the simple regularity of Fig. 2.10 was not found by Davis (1962 b) and the results of the case studies of some strong auroral events made by Akasofu (1960, 1962, 1963) and Akasofu and Chapman (1962) gave instantaneous configurations of the aurora, which were very closely similar to the projections of equatorial-plane circles on the earth's surface along the geomagnetic field lines (see previous subsection). The extensive auroral studies by Morozumi (1963) at the South Pole (geomagnetic latitude 78° S) did not show one, but two peaks in the diurnal variation curve for auroral occurrence. Instead for a maximum at magnetic noon there was a minimum.

A theoretical motivation for an oval auroral zone, being at higher latitude at noon than at midnight, is obtained when one considers the drift motion of particles around the earth in a magnetosphere, which is compressed by the solar wind on the dayside. The mirror points are at lower latitudes in the night than in the day. On this theoretical basis oval instantaneous auroral zones have been proposed by Rees and Reid (1959) and by Liemohn (1960). Detailed calculations by Malville (1960) show, however, that the expected effect is much less than the observed (less than 2° as compared with 8° observed). O'Brien (1963) therefore, concludes that if the diurnal variations are related at all, it is only in the sense that they are effects of a common and as yet unknown cause. More observational data are evidently needed before a conclusion about the existence or nonexistence of the oval instantaneous auroral zone can be reached.

That the main maximum in the diurnal distribution of auroral activity occurs close to geomagnetic noon at geomagnetic latitudes of about 78°, as Feldstein found, need not necessarily mean

that the corresponding processes in, or close to, the equatorial plane take place on the dayside, namely if the topology of the geomagnetic field is the one proposed by Johnson (1960) and others. A fairly evident hypothesis is that the neutral points or lines of that model are responsible for the daytime maximum in auroral occurrence.

Two spiral arcs of a kind described above are incorporated in the qualitative model for magnetic disturbances and associated phenomena proposed by Axford and Hines (1961).

That the diurnal variation differs for different auroral forms has been found by e.g. Fuller (1935) at College, Alaska, which is close to the auroral zone (cf. Fig. 2.11). Strong and brilliant forms have some predominance during the evening hours, and the faint and quiet forms in the morning. The hydrogen glow generally precedes the appearance of bright auroras, as first found by Romick and Elvey (1958) near the auroral zone, and the diurnal variation curve for this type of aurora therefore has its maximum well before midnight. Pulsating aurora occurs most often in the morning hours (Heppner, 1954).

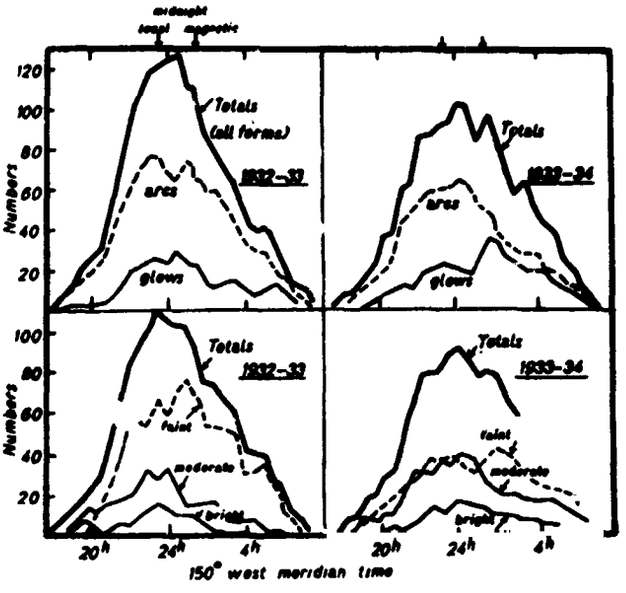


FIGURE 2.11.—(After Fuller, 1953) Diurnal variation of different auroral forms and of auroras of different intensitities at College, Alaska.

3. Other Time Variations of Visual Auroral Activity

(a) 27 day recurrence tendency

Auroras tend to repeat with a period of about 27 days, which is the synodic rotation period of the sun. As the moon considerably influences the possibilities of observing aurora, the apparent repetition tendency of auroral activity is enforced by the synodic lunar period (29.5 days), which is fairly close to the solar one. When the analysis is corrected for the lunar effect, a small but definite indication of a recurrence maximum around 26 and 27 days remains, which evidently is due to the rotation of the sun (Meinel et al., 1954). By the second and third solar rotations, no recurrence tendency remains.

(b) Seasonal variation

In the subauroral regions the annual curve of auroral occurence frequency has two maxima at the equinoxes. Fig. 2.12 shows the monthly aurora occurrence frequency distribution found by Meinel et al., (1954) at Yerkes Observatory.

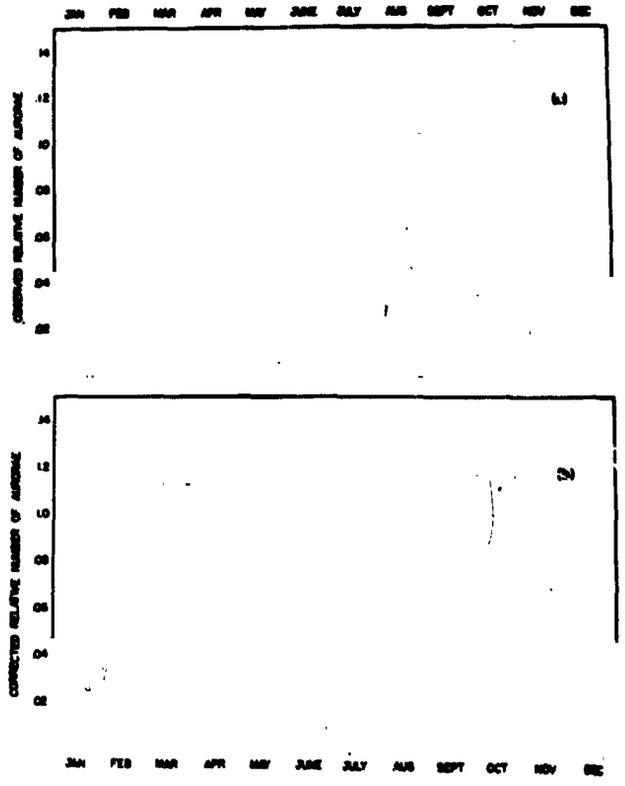


FIGURE 2.12.—(After Meinel et al., 1954) (a) Fraction of all auroras observed in each calendar month at Yerkes Observatory; (b) monthly frequency distribution corrected for cloudiness and number of dark hours in each month.

The existence of spring and fall maxima were known already to Mairan (1733). It was proposed by Cortie (1912) that the similar maxima in geomagnetic disturbances are due to the earth being at highest heliographic latitudes in March and September. If this is true, the angular diameters of the solar plasma beams do usually not exceed about 5° as seen from the sun. This is certainly not so for the very largest solar events, however.

The dependence of the seasonal variation on latitude is obscured by the change with latitude of the ratio of the lengths of day and night. If this change is ignored, as it usually is, then the summer minimum becomes progressively deeper with increasing latitude, while the winter minimum becomes progressively shallower, disappears and is replaced by a maximum, which is then the only one in the distribution (cf. e.g. Chree, 1911). Gartlein and Moore (1953) found that the ratio of the mean of the numbers of overhead auroras in North America occurring in June and December to that for auroras in March and September is a steadily increasing function of geomagnetic latitude. Between about 58 and 60° N the level of activity appears to be constant throughout the year. Feldstein and Shevnina (1961) investigated the seasonal variation in auroral occurrence frequency in two hours centered on local midnight—thereby eliminating the influence of the diurnal variation—for six stations between 63 and 83° geomagnetic latitude in the northern hemisphere and for five stations between 66° and 90° geomagnetic latitude in the southern one. They found the same seasonal variation as described above at and close to the auroral zones, but no perceptible seasonal variation in auroral frequency on the central auroral caps.

(c) Sunspot cycle variation

Fig. 2.13, after Tromholt (1902), shows the correlation between the number of days with aurora per year, observed in Norway between 1761 and 1877 and the sunspot number. The correlation is good and the maxima seem to coincide. Meinel et al., (1954) found a similar correlation when analyzing auroral observations made at Yerkes Observatory between 1897 and 1951. The maximum auroral activity occurred, however, during those 5 eleven year sunspot cycles some 2 years after the sunspot maximum. Egedal (1937) too, found a time delay of the maximum occurrence frequency of aurora in relation to the solar activity in Denmark during the 3 solar cycles from 1897 to 1929, but the delay was only 1 year.

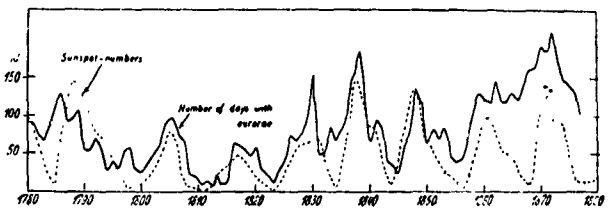


FIGURE 2.13.—(After Tromholt, 1902) Variation of the number of auroral days, N, in each year, observed in Norway, and the sunspot number between 1761 and 1877.

4. Orientation of Auroral Arcs

Hitherto various characteristics or auroral "activity," i.e. of statistical parameters describing the probability of occurrence of aurora, have been described. Now we will go somewhat more into detail. In this section the orientation of auroral arcs, which can be readily measured from photographic recordings, will be discussed.

(a) Average orientation

There exist fairly much data on the average orientation of auroral arcs. Even the very early studies of aurora showed that quiet auroral arcs tend to be oriented in a preferred direction, approximately perpendicular to the horizontal component of the geomagnetic field. A number of these early investigations were discussed by Fritz (1881). Vegard and Krogness (1920) referred the directions to the geomagnetic meridian and the Norwegian school made a large number of accurate direction determinations in northern Scandinavia.

In the majority of the pre-IGY investigations average values of all observed quiet arcs and bands were estimated. The fact that the observations were often made only during a certain period of the night—and in different periods for different investigations—was not considered.

A fairly extensive collection of such data is contained in Fig. 2.14 (after Hultqvist, 1962 b). Also included are a number of computed projections of circles in the equatorial plane on the earth's surface along the geomagnetic field lines (Hultqvist, 1958).

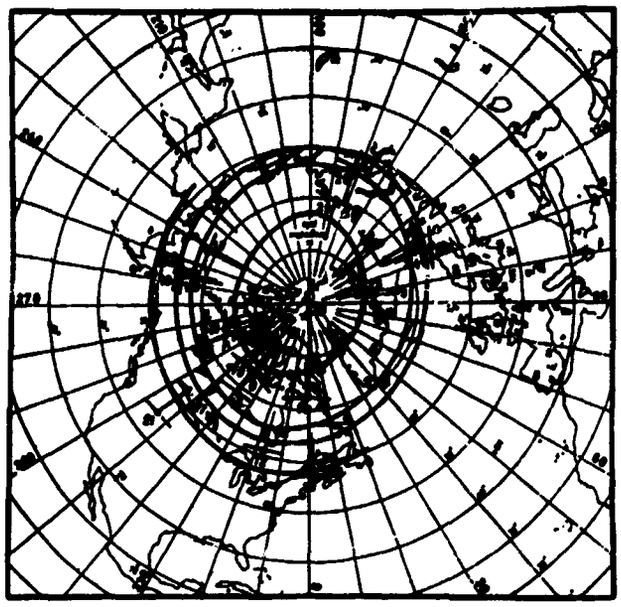


FIGURE 2.14.—Directions of quiet auroral arcs in the northern hemisphere. 1. Haldde and Bossekop; 2. Cap Thordsen; 3. Jan Mayen; 4. Nain; 5. Kingua Fjord; 6. Fort Rae; 7. Sagastyr (the data for points 1-7 are after Vegard and Krogness, 1920); 8. Chesterfield; 9. Coppermine; 10. Cape Hope's Advance; 11. Saskatoon; 12. Coral Rapids; 13. Aroostook; 14. Gjöahavn (the data for points 8-14 are taken from Currie and Jones, 1941; the dotted lines at Chersterfield and Coppermine refer to all types of aurora); 15. Baker Lake; 16. Churchill; 17. Barrow; 18. Fort Yukon; 19. College; 20. Farewell; 21. Choteau (the data for points 15-21 are after Davis, 1961; the values are averages for the hour around geomagnetic midnight); 22. Godhavn (the data are after Lassen, 1959, and refer to the early morning hours); 23. the region of Micardbu (approximate data after Störmer, 1944); 24. Tromsö (the data are after Harang, 1945, and refer to geomagnetic midnight); 25. Wiese; 26. Dixon; 27. Cape Schmidt (data for points 25-27 are after Feldstein, 1960, and refer to local midnight); 28. Thule (data of Harang, taken from Störmer, 1955); 29. Kiruna (refers to geomagnetic midnight). The map has a geomagnetic coordinate system.

Despite the above-mentioned reservations concerning the observational material, it is evident from Fig. 2.14 that, in general, the alignment of the arcs conforms to the direction of the relevant circle projection quite closely, and that the agreement with the circle projections is better than with the geomagnetic latitude circles. The observed arc azimuth is in most cases, greater than 90° in those quadrants where the azimuth of the circle projections exceeds 90°, and less than 90° in quadrants where this is under 90°. Thus, it has been verified experimentally that on the eastern and western coasts of Greenland the observed average azimuths deviate in opposite directions from 90°, as do the azimuths of the circle projections.

Starkov and Feldstein (1960) and Evans and Thomas (1959) investigated the alignment of arcs at two places where the isoclines and the geomagnetic latitude circles differ greatly (Dixon near the northern auroral zone and Halley Bay near the southern one). They found that the preffered direction of alignment is far more consistent with geomagnetic latitude circles, and therefore with the higher approximation circle projections, than with the isoclines.

The available observational results for the southern hemisphere are relatively exiguous. A collection of IGY and subsequent data is contained in Fig. 2.15. The arrows indicate results of

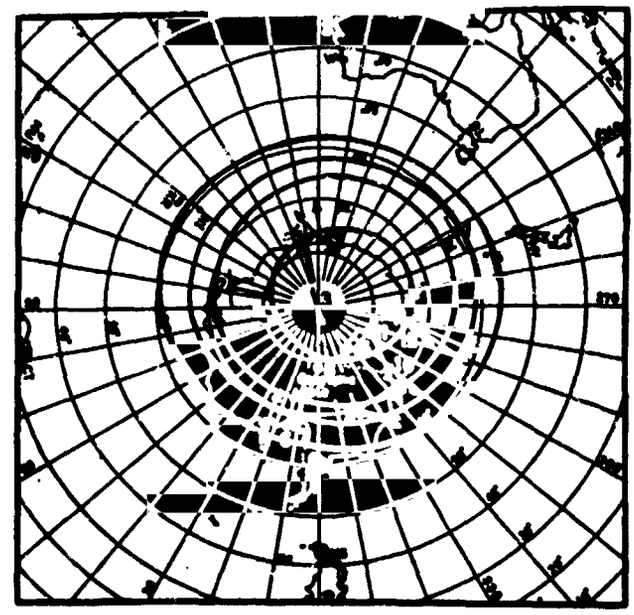


FIGURE 2.15.—Directions of quiet auroral arcs and sighting directions to the highest points of quiet arcs and bands (arrows) in the southern hemisphere. 1. Macquarie Island. 2. Mawson (data taken from Bond and Jacka, 1960); 3. Vastok; 4. Oasis (The data, after Feldstein, 1960 refer to local midnight); 5. Halley Bay; 6. Shackleton (data from Evans and Thomas, 1959); 7. Ellsworth; 8. South Pole; 9. Byrd; 10. Little America; 11. Scott Base (the long arrows at points 11 and 12 are after Hatherton and Thomas, 1959); 12. Hallett (short arrow); 13. Wilkes (the data for points 7-13 are from Gartlein et al., 1960, the two arrows at points 8 and 13 show the average directions for the two years 1957 and 1958, the 1957 arrow being the one nearest the meridian at both points); the map has a geomagnetic coordinate system.

TABLE 2

No.	Station	Geomagn. Coord. °N	Geomagn. Coord. °E	Observed azimuth (degr.)	Approx. measure of deviation	Azimuth of circle projection ("theor. azimuth")	Diff. obs. azimuth and "theor. azimuth"
1	Baker Lake	73.8	314.8	94	Diff. max-min ~60°	93	1
2	Churchill	68.8	322.5	99	Diff. max-min ~60°	95	4
3	Barrow	68.6	241.2	92	Diff. max-min ~60°	84	8
4	Fort Yukon	66.6	256.8	90	Diff. max-min ~80°	87	3
5	College	64.6	256.5	86	Diff. max-min ~60°	87	−1
6	Farewell	61.4	253.2	83	Diff. max-min ~60°	86	−3
7	Choteau	55.5	306.0	88	Diff. max-min ~70°	95	−7
8	Schmidt	62.6	226.6	80	Diff. max-min ~30°	80	0
9	Dixon	63	162	98	Stand. deviat. $< \pm 10°$ for magn. quiet periods	98	0
10	Vise	69	176	103	$\sim \pm 15°$	100	3
11	Pyramid	---	---	111	$\sim \pm 15°$	113	−2
12	SP-7 1957–58	---	---	96	---	96	0
13	SP-7 1958–59	---	---	0	---	0	0
14	Tromsö	67	117	98	Diff. max-min ~27°	106	−8
15	Kiruna	65.3	115.6	106	Stand. deviat. ±5°	105	1

observations of directions, as sighted to the auroral point of highest elevation. Two arrows for any single station represent the results of observations during each of the two years 1957 and 1958. The differences in direction of the arrows for the two years demonstrate the uncertainty of the results.

As will be seen from Fig. 2.15, the agreement between observed directions with those of the projected equator plane circles is fairly good for all places except Byrd (No. 9), and the agreement is certainly better than for the geomagnetic latitude circles.

Because of the existence of a diurnal variation in the orientation of the auroral arcs, increased accord between the directions of the arcs and of the circle projections is to be expected if values for a short period (e.g. 1 hr.) of the night are compared. The results of such a comparison for local midnight at twelve stations, most of them in or near the northern auroral zone, are presented in Table 2 and Fig. 2.16. A comparison with Fig. 2.14 clearly shows that the concordance of the arc directions with the circle projections is better for the selected material of Fig. 2.16 than for the unselected values of Fig. 2.14.

Table 2 contains azimuths of quiet auroral arcs at local midnight in the northern hemisphere (measured from northern geomagnetic meridian towards the east). The observed azimuths for stations 1–7 are after Davis (1961), those for stations 8–13 are after Feldstein (1960), and that for Tromsö is after Harang (1945).

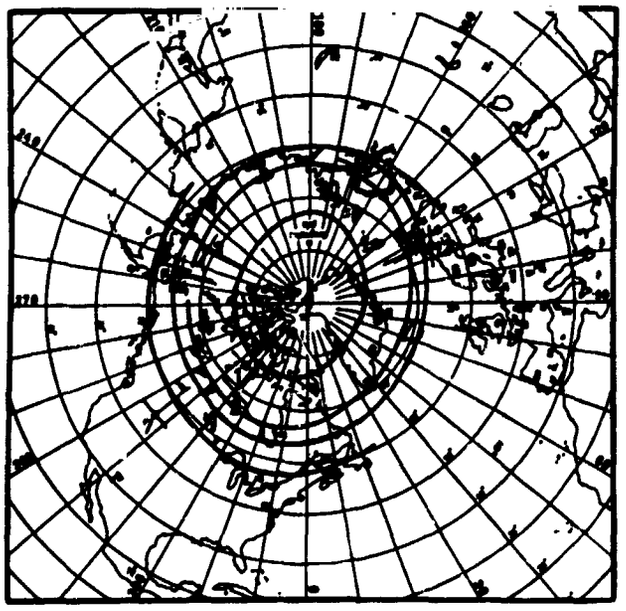


FIGURE 2.16.—Directions of quiet auroral arcs and bands at local midnight in the northern hemisphere. 1. Baker Lake; 2. Churchill; 3. Barrow; 4. Fort Yukon; 5. College; 6. Farewell; 7. Choteau (data for points 1-7 taken from Davis, 1961); 8. Cape Schmidt; 9. Dixon; 10. Wiese (data taken from Feldstein, 1960); 11. Tromsö (after Harang, 1945); 12. Kiruna. The map has a geomagnetic coordinate system.

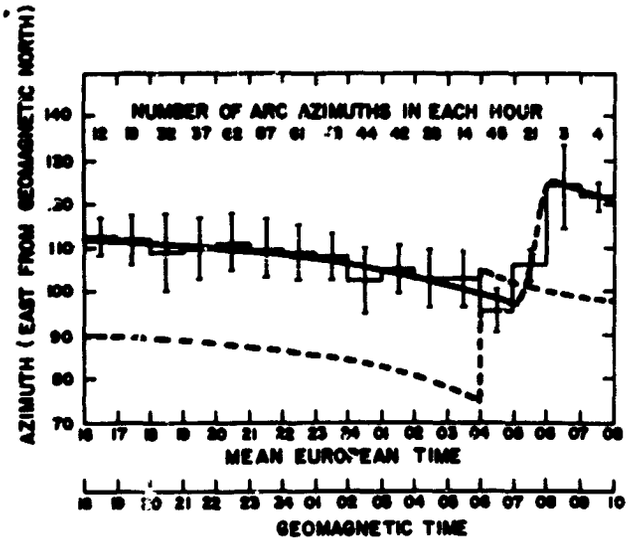


FIGURE 2.17.—Observed directions of quiet auroral arcs at Kiruna as function of MET and of approximate geomagnetic time. For each hour the average direction and the standard deviation are given. The lower, dashed curve is Alfvén's theoretical diurnal variation curve (see Alfvén, 1950).

(b) Diurnal variation in the orientation in and near the auroral zones

Although it was evident from the observations of Bravais at Bossekop in northern Norway as early as 1938–40 that the directions of the arcs and bands show a diurnal variation, very little attention was paid to this phenomenon until the nineteen-thirties, and even now only few studies exist.

From allsky-camera recordings taken at Kiruna ($\phi_m = 65.3°$ N) in the period December 1958 to end of 1962, the directions of 559 quiet auroral arcs were estimated. Only arcs reaching from horizon to horizon or, in the case of the most distant ones, covering an azimuth angle of view of about 100° were used in the analysis. The error in the measured direction of an arc probably does not exceed 4°.

The result of the alignment investigations is shown in Fig. 2.17.

For each hour of Mean European Time (MET) the average direction and the standard deviation were computed, for which procedure the direction value of each arc was weighted with the arc's duration. The durations vary over a wide range—from less than a minute to about an hour.

The total average azimuth for the whole material is 109°. A typical standard deviation for a 1 hr. period is 5°.

At the top of Fig. 2.17 is shown the number of arcs used for each hour. The number is maximal around geomagnetic midnight. In the morning hours the aurora is usually unsuitable for direction determination; as a rule, it is diffuse and situated far to the north. That is the reason why the observational material is smallest for that part of the night.

The azimuth charge in the morning is so rapid as to resemble a discontinuity of the type predicted by the Alfvén theory (see e.g. Alfvén, 1950).

In fact, the "discontinuity" can probably be observed only in longitudes where geomagnetic time runs ahead of local time, since daylight will preclude auroral observations in or close to the auroral zones until 8 or 10 o'clock geomagnetic time in other longitudes.

A number of similarly obtained diurnal variation curves for auroral arc orientation, based on observations in or fairly near the northern auroral zone (up to a geomagnetic latitude of 80°), are shown in Fig. 2.18.

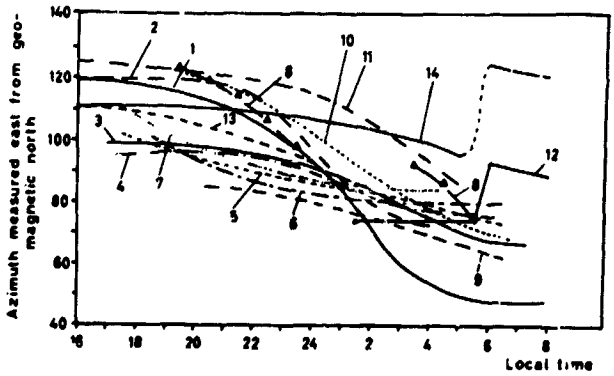


FIGURE 2.18.—Observed (smoothed) diurnal variation curves for the orientation of quiet auroral arcs and bands in and near the northern auroral zone. 1. Baker Lake (geom. lat. 74°); 2. Churchill (geom. lat. 69°); 3. Barrow (geom. lat. 69°); 4. Fort Yukon (geom. lat. 67°); 5. College (geom. lat. 65°); 6. Farewell (geom. lat. 61°); 7. Choteau (geom. lat. 56°) (curves 1-7 are after Davis, 1962b); 8. Dixon (geom. lat. 63°); (after Starkov and Feldstein, 1960); 12. Godhavn (geom. lat. 80°) (after Lassen, 1959); 13. Tromsö (geom. lat. 67°) (after Harang, 1945); 14. Kiruna (geom. lat. 65°).

The wide dispersion of the location and form of the different diurnal variation curves is striking. The amplitude of the diurnal variation also varies greatly. Here a regularity emerges, the amplitude increasing rapidly with latitude. Thus, the auroral arc direction varies at Baker Lake (geomagnetic latitude 74°) over a range of 70°, while at Choteau (geomagnetic latitude 56°) the range is only 20° (and at Kiruna, 65° N, 15°, if the final rapid increase in azimuth is ignored).

A "discontinuity" in the morning hours has been observed only at Godhaven and at Kiruna. As mentioned above, observation is likely to be possible only in the eastern geomagnetic hemisphere, where the geomagnetic time is running ahead of the local time. The difference between geomagnetic and local times is greatest in the region of northern Scandinavia (about 2 hours), where it is possible to observe aurora up to 1000 hrs. geomagnetic time. The reason why such "discontinuity" has not been observed at the Siberian stations could be that the geomagnetic times differs less from local time at those sites than it does at Godhavn and Kiruna.

For the southern hemisphere there is still a paucity of published observational data concerning the diurnal variation of arcs in and near the auroral zone. One such diurnal variation curve, of Evans and Thomas (1959) for Halley Bay, is shown in Fig. 2.19. From this curve it can be seen that at the southern auroral zone the azimuth —measured east from geomagnetic north—increases with time instead of diminishing as in the northern hemisphere. The range of variation (about 15°) is similar to that for the stations of lowest latitude in Fig. 2.18. This lends support to the assumption that the northern and southern momentaneous auroral occurrence curves are geomagnetically conjugated.

Comparison of the curve for Halley Bay in Fig. 2.19 (No. 1) with that for Oasis (No. 2 in Fig. 2.19) of geomagnetic latitude 78° shows that the range of variation increases with latitude in the southern hemisphere also.

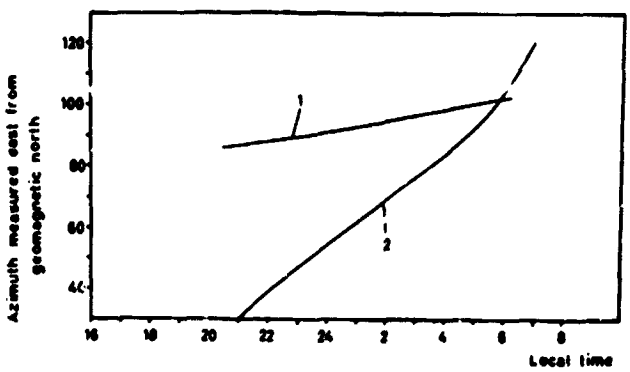


FIGURE 2.19.—Observed (smoothed) diurnal variation curves for the orientation of quite auroral arcs in the southern hemisphere. 1 Halley Bay (geom. lat. 66°S; outside the auroral zone) (after Evans and Thomas, 1959); 2. Oasis (geom. lat. 78°S; far inside the auroral zone) (after Feldstein, 1960).

(c) Discussion of the diurnal variation near the auroral zones

In order to ascertain the degree to which the differences between the curves in Fig. 2.18 could be attributed to referral of the direction of the geomagnetic meridian, i.e. to overlooking of the effect of the non-dipole part of the geomagnetic field, Hultqvist (1962 b) made a correction for this factor in respect of several stations near the northern auroral zone. The direction, instead of being referred to the geomagnetic meridian, was referred to the perpendicular to the projection of circles in the geomagnetic equatorial plane on to the earth's surface along higher approximation geomagnetic

field lines. The azimuths of these circles at a number of stations are given in Table 2. By changing the time scale to the geomagnetic one, the curves of Fig. 2.20 were obtained.

A comparison of Figs. 2.18 and 2.20 shows that the agreement between the North American stations on the one hand and Cape Schmidt and Kiruna on the other, was much improved by this modification.

Fig. 2.20 also includes the theoretical curve of Alfvén (1939, 1940, 1950). A comparison shows that the time derivative sign is the same, but that all the experimental azimuths are higher than the theoretical ones before midnight and that the amplitude of the diurnal variation for most stations is greater than theoretically expected.

The "discontinuity" observed at Kiruna has the same amplitude as that on the theoretical curve but is displaced about 1½ hour in the positive time direction.

From the observed diurnal variation curves of Fig. 2.20, a third type of auroral precipitation curve can be evaluated by integration according to formula (1)

$$\theta = \theta_0 \exp\left(-\int_{\lambda_0}^{\lambda} \cot a \, d\lambda\right) \qquad (1)$$

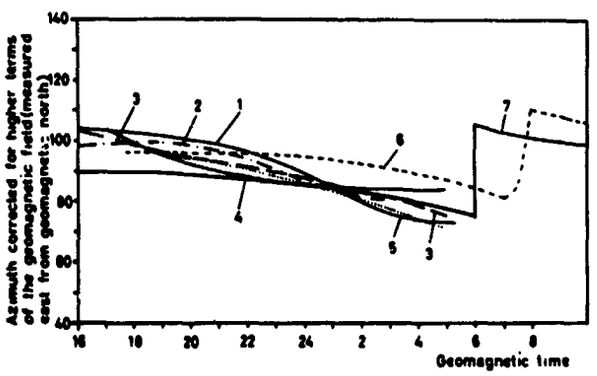


FIGURE 2.20.—Observed diurnal variation curves (smoothed) for the direction of quiet arcs, corrected for the non-dipole part of the geomagnetic field, given as a function of geomagnetic time for a number of stations situated in or near the northern auroral zone. 1. Barrow (geom. lat. 69°); 2. Fort Yukon (geom. lat. 67°); 3. College (geom. lat. 65°); 4. Farewell (geom. lat. 61°); (curves 1-4 are based on data taken from Davis, 1961); 5. Cape Schmidt (geom. lat. 63°) (based on a curve presented by Feldstein, 1960); 6. Kiruna (geom. lat. 65°); 7. Alfvén's theoretical diurnal variation curve (cf. e.g. Alfvén, 1950).

where θ is the geomagnetic colatitude,
λ is the geomagnetic time angle,
a is the azimuth of the auroral arcs measured from geomagnetic north towards east,
λ_0 is the geomagnetic time angle at the start of the observational curve, and
θ_0 is an integration constant that determines the polar distance of the curve.

This curve is Alfvén's theoretical I curve. It is defined by the orientation of the auroral arcs around the auroral zone, considered to be fixed in relation to the sun.

From the Kiruna diurnal variation curve of Fig. 2.20 the I curve of Fig. 2.21 has been derived. The whole curve is situated outside the circle with which it coincides at 1800 hours, in contrast to Alfvén's theoretical curve, which is inside the circle. The difference is due to the fact that the observation curve in Fig. 2.20 is situated almost completely above azimuth 90°. The difference between the experimental and theoretical curves

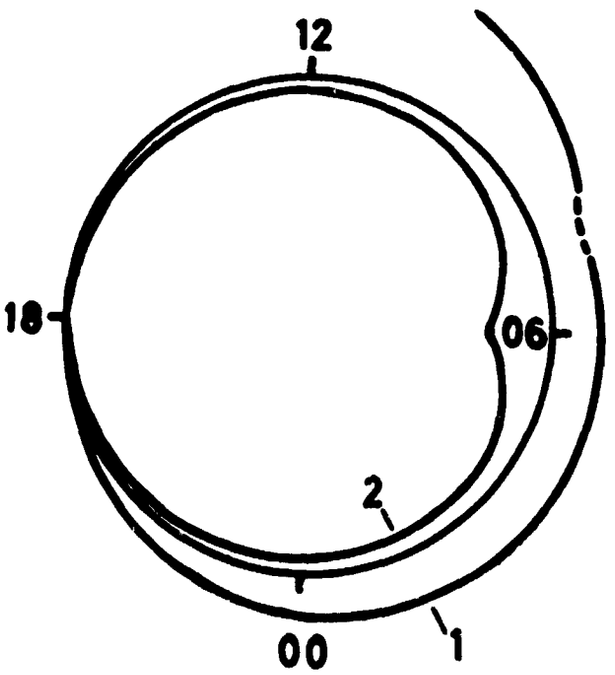


FIGURE 2.21.—Curve 1 is the auroral zone corresponding to Alfvén's I curve, evaluated from the Kiruna curve of Fig .17 for diurnal variation in the orientation of quiet auroral arcs. It is corrected for the non-dipole part of the geomagnetic field. Curve 2 is Alfvén's theoretical I curve (see Alfvén, 1950).

varies between 7 and 9°, which means that the theoretical curve is situated outside the standard deviation "region" of the observational material.

The very high amplitudes of the diurnal variations found for high latitude stations are not contained in Alfvén's model, according to which the greatest deviation of the arc azimuth from 90° should be at 0600 hours geomagnetic time, where the I curve has a singular point and the azimuth changes from 105° to 75° in the auroral zone.

Even for L = 70 earth radii, corresponding to a geomagnetic auroral latitude of above 80°, the azimuth is only 76° at 0600 hours.

Khorosheva (1961, 1962) investigated recordings of aurora made simultaneously at widely different longitudes in the arctic of the eastern hemisphere and found that the auroras form a single, physically related band which changes synchronously in brightness, width and latitude over its entire extension. On the basis of these observations, Khorosheva and Feldstein proposed that the momentaneous auroral precipitation zone forms an oval-shaped curve around the geomagnetic axis pole, at the normal distance of the auroral zone (about 23°) on the night side but only 10–12° from the pole on the day side, i.e. the zone is similar in shape to the combination of two spiral arcs shown in Fig. 2.10. The proposal (Feldstein, 1964) also implies that the auroral arcs are lined up along the momentaneous precipitation zone of the described shape. This would explain the existence of two maxima in the diurnal variation curve for auroral occurrence at latitudes inside the ordinary auroral zones. It would, however, mean that the aximuth of the auroral arcs, measured from the perpendicular to the auroral zone (curve 3 in Fig. 1), would be smaller than 90° after midnight, with the most rapid change in direction around midnight. That is usually not observed.

(d) Diurnal variation on the central auroral caps

On the central auroral caps (inside about 80° geomagnetic latitude) the diurnal variation of the orientation of quiet auroral arcs has been reported to be entirely different from the one described above. Observed (smoothed) diurnal variation curves are shown in Fig. 2.22 for both the northern and southern central caps. Curve 2 in Fig. 2.19 also belongs to this category.

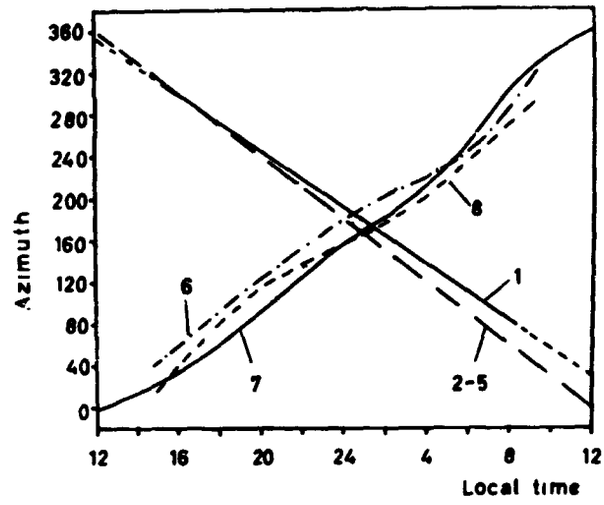


FIGURE 2.22.—Observed diurnal variation curves (smoothed) for the direction of quiet arcs at stations deep inside the auroral zones. The azimuth is measured towards the east from geographic north for curves 1-5 and from geomagnetic north for curves 6-8. 1. Dumon d'Urville (geom. lat. 76°S) (after Weill, 1958); 2. Wilkes (geom. lat. 77°S); 3. Dumont d'Urville; 4. Scott Base (geom. lat. 79°S); Hallet (geom. lat. 74°S) (the curves for points 2-5 are after Denholm and Bond, 1961); 6. Thule (geom. lat. 88°N); 7. Alert (geom. lat. 86°N); 8. Resolute Bay (geom. lat. 83°N) (the curves for points 6-8 are after Davis, 1961).

If the azimuth in Fig. 2.22 is taken from geographic north, the diurnal variation curves imply that the arcs are continuously pointing to the sun while the earth rotates below the auroral pattern. The first to report such a 360° diurnal variation of the auroral arc direction seems to have been Mawson (1914), but it was forgotten and rediscovered by Weill (1958). In the last few years it has also been reported by Davis (1960, 1961, 1962 b), Feldstein (1960), and Denholm and Bond (1961). Consistent data showing this 360 degrees rotation of the arcs during the 24 hours of day and night have thus been presented by several investigators and there seems to be no doubt that the effect is real. There are, however, a number of inconsistencies between observational results found by different observers (see Hultqvist, 1962 b) and there is, therefore a need for more observations.

It has been suggested by Cole (1959, 1960) and Davis (1960, 1961, 1962 b) that the geometric pattern of aurora over the polar caps is the same as the pattern of the ionospheric current system which is equivalent to the local-time-dependent

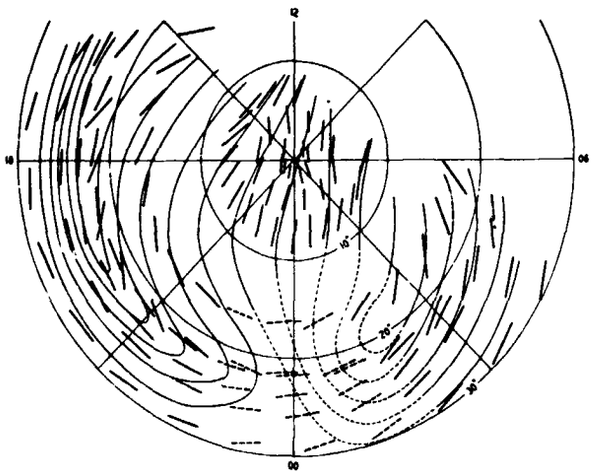


FIGURE 2.23.—(After Davis, 1962b) The alignment of auroral forms in a polar coordinate system with geomagnetic colatitude and approximate geomagnetic time as polar and azimuthal coordinates. The dashed lines represent the discontinuous post break-up aurora.

component of the geomagnetic disturbance—the DS current system (see Fig. 2.23). These views were incorporated as important constituents of the qualitative model of geomagnetic storms and associated phenomena by Axford and Hines (1961). According to this model the DS currents and the auroral arcs should have the same direction over the auroral caps.

The angle between auroral arcs and the direction of the horizontal currents, corresponding to the observed magnetic disturbance, has been evaluated by Sobouti (1961), Davis and Kimball (1962), Pudovkin and Yevlashin (1962), Feldstein (1964) and others for auroral and transauroral latitudes. Near the auroral zone, all investigators found the directions being identical within measuring accuracy. While Feldstein's results in the near polar regions were that the directions mentioned form an angle of, in the average, about 50°, Sobouti (1961) did not see any difference at Resolute Bay (geomagnetic latitude 83.2° N). A disagreement between directions of auroral arcs and the ionospheric current lines has also been found in the Antarctic (see Cole, 1963). A difference between the two directions introduces difficulties for theoretical models which require that the geographical pattern of auroral arcs and of the DS current system should be identical (like the model of Axford and Hines, 1961). It is, however, to be remembered that the horizontal geomagnetic disturbance component shows the direction of the ionospheric current only for simple current patterns.

(e) Other variations in arc orientation

Jensen and Currie (1963) found a pronounced seasonal variation in the auroral arc orientation in the years 1949–51 at Saskatoon (lat. 52.1° N) and Chesterfield (lat. 63.3° N) Canada. The peak to peak variation in azimuth over a year amounted to about 30°, with a maximum azimuth of 115° in February at Saskatoon and minima in October-November and April-May. The seasonal-variation investigation of Jensen and Currie (1953) seem to be the only one made.

Those authors also studied the dependence of the arc direction on the activity of the auroral display. No significant difference between quiet and active displays could be found. When grouping their observational material for auroral arc orientation with regard to magnetic disturbance level, Jensen and Currie obtained for Saskatoon an average azimuth of $93° \pm 1°$ (standard deviation) for the highest degree of activity, $100° \pm 1°$ for median disturbance and $100° \pm 2°$ for the lowest degree of disturbance. More investigations of the kinds undertaken by Jensen and Currie, made for different geomagnetic longitudes, would be of great value.

No theoretical models of aurora contain any seasonal variations in the auroral activity at all. The difference in direction between the earth's rotation axis and the dipole axis has been neglected so far.

5. The Height and Vertical Extent of Visual Aurora

The height of auroras was studied extensively especially in the first two to three decades of this century by the Norwegian school of auroral physics, with Störmer as the originator. Detailed reports of the methods used and the results obtained can be found in the books by Harang (1951) and Störmer (1955). Observations in Canada and Alaska have been reported by McLennan et al. (1931), Currie (1934, 1955), Fuller and Bramhall, (1937), and McEwen and Montalbetti (1958). Southern hemisphere observations have been published by Geddes (1939). Only a very short review will be given here.

The height can be measured only if the auroral form contains some well defined features which can be identified on photographs taken from two places at a distance from each other comparable to the height of the aurora (paralactic photography).

In most auroras the lower border is well defined and the height given in most reports is the height of the lower limit of the aurora, as seen on photographic pictures. Fig. 2.24 shows the height distribution of lower limits for different auroral forms (after Vegard and Krogness, 1920). The differences between the various types represented in the figure are fairly small.

The dependence of the heights of the lower borders of auroral arcs on their intensity is shown in Fig. 2.25 (after Harang, 1951). As can be seen from Fig. 2.25, there is a difference of 20 km

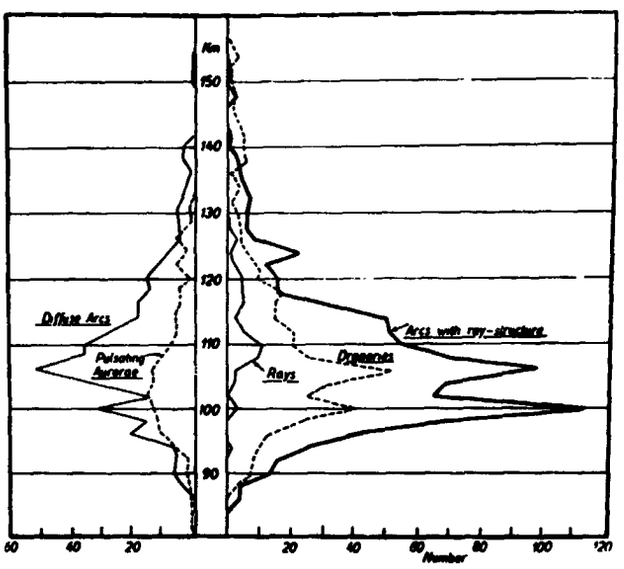


FIGURE 2.24.—(After Vegard and Krogness, 1920) Distribution of lower limits for different auroral forms.

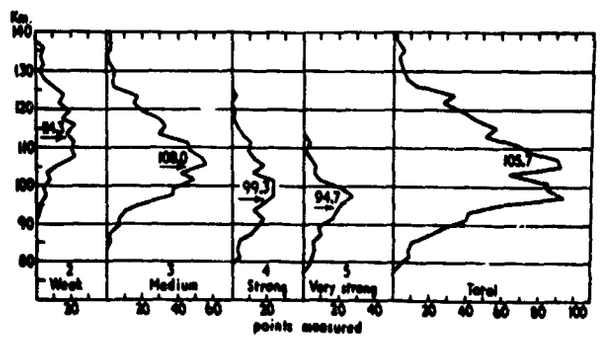


FIGURE 2.25.—(After Harang, 1944) Lowering of the heights of arcs with increasing intensities, as observed from Tromsö 1929-1938.

in height between weak and very strong arcs. Harang found a similar but less strong dependence on intensity also for other forms.

The results described above were obtained in or close to the northern auroral zone. The height of subauroral zone aurora has been extensively studied by Störmer (cf. Störmer, 1955) over southern Scandinavia. Compared with the auroral heights determined in the vicinity of the auroral zones the height diagram for southern Norway shows a far greater extension towards increasing heights. The lower limits are, however, the same in both cases. According to Barbier (1963) the majority of auroras seen over central Europe have lower limits situated above 250 km altitude.

In addition to the type effects shown in Fig. 2.24, there are a number of others. Störmer (see Störmer 1955) determined the height of the lower limit of sunlit auroras and found a height distribution completely different from those in Fig. 2.24. The distribution was very extended in height, with its maximum of probability at about 300 km.

Aurora of type B, i.e. aurora with the lowest part red due to emissions from the first positive band system of N_2 and first negative system of O_2^+, usually occurs at altitudes for the lower border of between 80 and 100 km. The lowest height of aurora reported was found by Harang and Bauer (1932) for a type B display. Numerous measurements were made on an intense arc and they obtained some points extending down as low as 95 km with an average height of 70 km.

On the other hand, red aurora of type A (with the 6300 Å O line dominating) is observed mostly at heights greater than those shown in Fig. 2.24. The medium latitude aurora mentioned above, which has a lower limit greater than 250 km, is mostly of this kind.

Störmer has reported a type of homogeneous arc, lying in the dark atmosphere near 200 km, which may be of a different kind from the aurora observed at 100 km level (Barbier, 1963; see also p. 6).

Finally, the quiet mid-latitude red arcs have been found to occur at heights between 390 and 560 km, but most often in the interval 400-450 km (see the review by Roach and Roach, 1963).

The extent along the field lines of an auroral form is quite different for different forms. Fig. 2.26 (after Vegard; cf. Harang, 1951, p. 35) shows

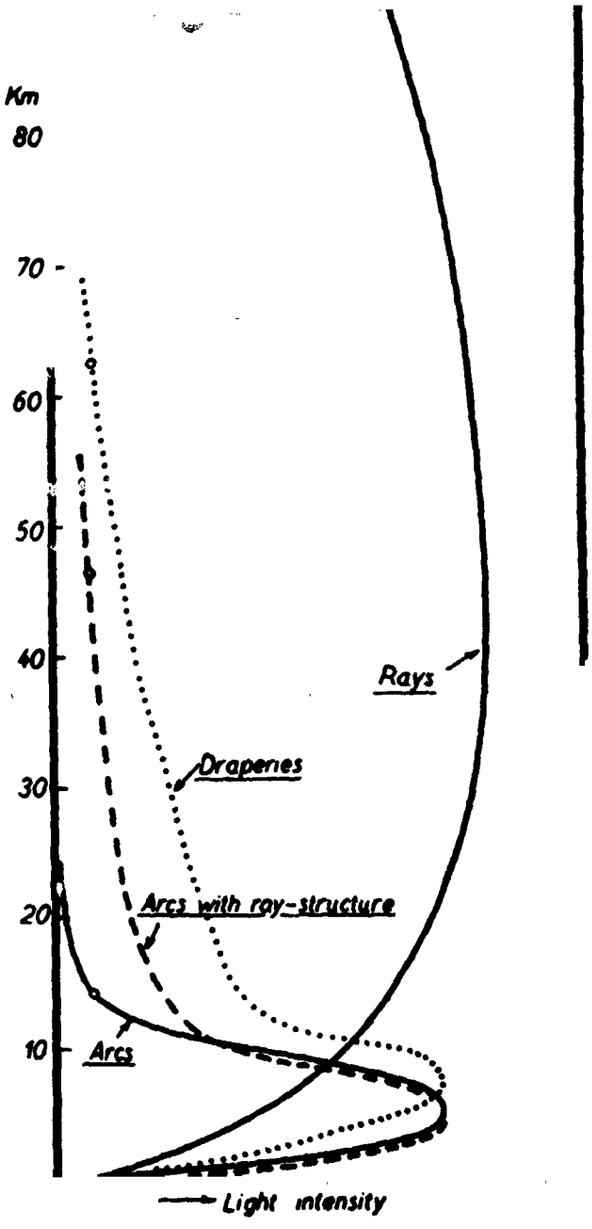


FIGURE 2.26.—(After Harange, 1951) Distribution of light intensity along different auroral forms.

the mean variation of light intensity in vertical direction for different auroral forms. An ° indicates the upper limit of photographic impression. Some average numerical parameters have been collected in Table 3 (after Vegard and Krogness, 1920).

Here l_1, l_2 and l_3 are the distances from the lower limit to the height of maximum luminosity, the upper limit of strong luminosity, and the upper limit of faint luminosity, respectively. It is evident from Fig. 2.26 and Table 3 that the rays differ entirely from the other auroral forms in respect of luminosity distribution with height.

There is a solar-cycle variation in the height of the lower boundary of aurora as well as in vertical extent (Elvey, 1957). Both height and vertical extension are lower in years with solar-activity-minimum than at high solar activity.

TABLE 3

Type	l_1	l_2	l_3
arcs	6.4 km	14.0 km	14.0 km
arcs with ray-structure	5.8 km	13.6 km	46.7 km
draperies	9.1 km	16.3 km	63.4 km
rays			137.0 km

6. Motions

There are many different types of motions associated with the auroral activity and individual auroral forms. Some of the motions for which observational data are available will be discussed here. From the point of view of the theoretical models of aurora, we are interested in how the auroral precipitation zones change location with time under various conditions as well as in how individual complete forms, or irregularities in the forms, move. These types of motion will be discussed in this subsection, where we will deal with the statistical properties. In subsection 7 examples of motion in connection with changes of momentaneous auroral geographical distributions will be discussed.

(a) Changes in the geographical pattern of auroral activity

It has been mentioned already before in this review that the location of the auroral zone shifts towards lower latitudes when the solar activity increases (although inconsistency exists in the available data). The first one who deduced this motion from observational data seems to have been Tromholt (1882). Davies (1950) evaluated a change of about 3 degrees of latitude between solar maximum and solar minimum. Tromholt

(1882) and Davies (1950) also arrived at a seasonal variation in the polar distance of the auroral zones. This is large around the equinoxes and small at solstizes. This is supported also by the recent study by Feldstein and Shevnina (1961), but the opposite result was found by Sheret and Thomas (1961).

With the statistical definition of the auroral zone adopted earlier in this section, seasonal variation in its location are the shortest that can exist. We will, however, now consider briefly the average behavior of the momentaneous auroral precipitation zone as a function of geomagnetic disturbance level (this momentaneous region of auroral activity is often called the auroral zone in the literature). There exist two types of data of relevance in this respect: those which concern the equator-ward boundary of the region where auroras occur, and those referring to the region of maximum occurrence frequency for the aurora at a given moment of time.

Gartlein et al. (1960) found the southern boundary of aurora in the northern hemisphere to be at almost 10° lower latitude for $K_p=5$ than for $K_p=1$, the $K_p=5$ boundary passing through Stockholm, southernmost Alaska and the Great Lakes in North America (see Fig. 6). Akasofu and Chapman (1963) found the southern limit of latitude of northern quiet auroral arcs to be approximately a linear function of Dst(H), i.e. of the ring current field, for 16 magnetic storms studied. The latitude variation with the ring current amounted to more than 10°, the southern limit reaching 49° geomagnetic latitude for the strongest storms investigated.

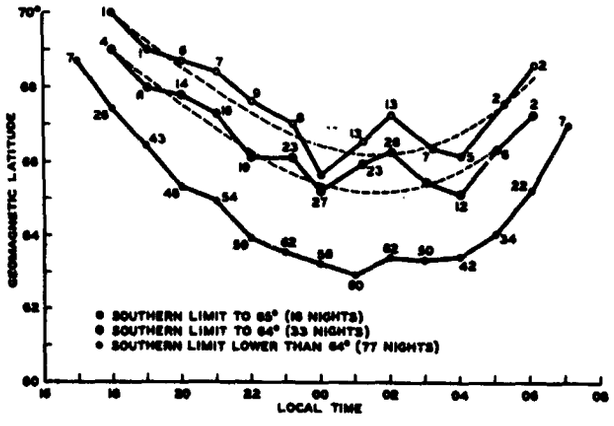


FIGURE 2.27.—(After Davis, 1962a) Diurnal variation of average southern extent for each of three groups of displays extending to geomagnetic latitudes 65°, 64°, and beyond 64°. The number beside each point gives the number of displays contributing to that particular mean position.

The southern border of auroral activity in Alaska, taken to be the lowest latitude position at which defined auroral forms appear, was found by Davis and Kimball (1960) to move regularly in the course of the night in the manner shown in Fig. 2.27. The extreme southern extent of the smoothed curves occurs, as can be seen in Fig. 2.27, around geomagnetic midnight (which is 1-2 hours after local midnight in Alaska). Davis (1962 b) states that the decrease in the number of discrete forms in low latitudes in the morning must be due to the actual disappearance of the forms in place—either because of a decrease of auroral luminosity or to a tendency for the forms to diffuse —and not to withdrawing to the north.

Jacka (1953) studied the location of the region of maximum auroral occurrence around MacQuarie Island in the southern hemisphere. He found that the southern auroral precipitation zone moves towards the equator with increasing K_p as $\Delta\phi = 2.95 - 0.605\,K_p$, where $\Delta\phi$ is the latitude departure of the homogeneous arcs from MacQuarie Island. From Jacka's relation, it can be seen that the latitude of the auroral precipitation zone can vary with as much as 5–6°. This is the same amount as found by Feldstein and Solomatina (1961).

A question of interest is the following: is the detailed shape of the distribution of auroral activity within the auroral precipitation zone changed when its center or southern boundary moves in latitude. Studies of this have been reported in the last few years. Some investigations of individual events will be reviewed in subsection 7, below. Here the average behavior will first be discussed.

Feldstein and Solomatina (1961) reported that at the same time as the auroral precipitation zone withdraws from the pole, there is an expansion of the zone, also towards higher latitudes, but they did not present quantitative results. Davis and Kimball (1960) and Davis (1962 b) give some details. Fig. 2.28, after Davis (1962 b) shows several measures of auroral activity vs. geomagnetic latitude for weak and strong displays. For curve A the maximum K index at College during

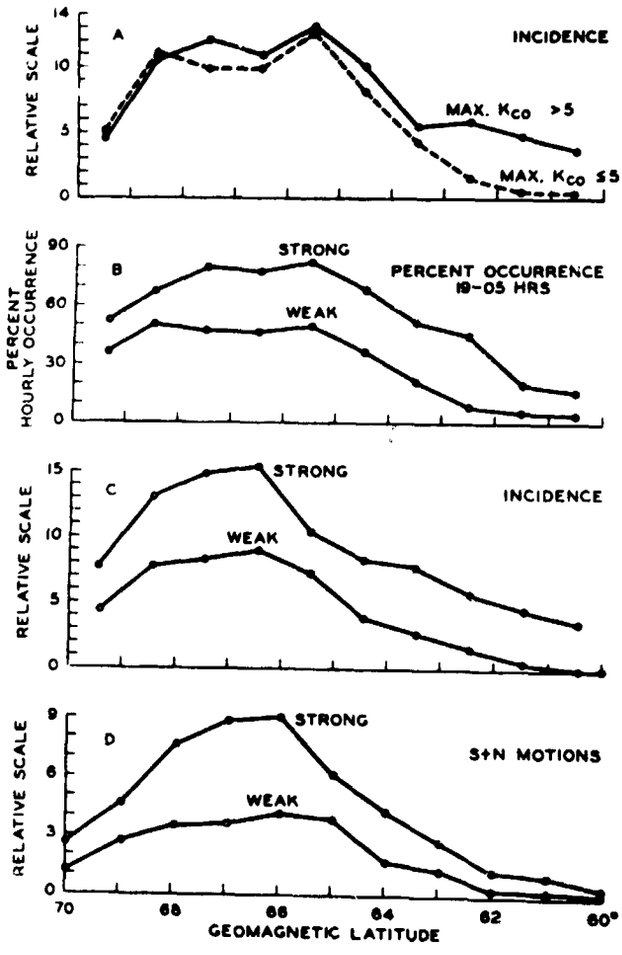


FIGURE 2.28.—(After Davis, 1962b) Several measures of auroral activity versus latitude for weak and strong displays observed in Alaska in the season 1957-58. (A) Curves of total incidence for 25 displays during which the maximum K-index for College did not exceed 5 (dashed line) and for 25 displays for which it exceeded 5 (solid line). (B) Average percentage hourly occurrence of overhead auroras (local time interval 19-05[h]) during 24 strong and weak displays. (C) Total incidence of forms during the 24 strong and 24 weak displays used to draw curves B. (D) Total south and north moving forms observed throughout 31 strong and 30 weak displays.

the display has been the criterion for classing. Davis points out that the small difference between strong and weak events at high latitudes, shown in Fig. 2.28, does not mean that there is no connection between magnetic and auroral activity in those latitudes, but only that the highest K index achieved during a display at an auroral-zone station is a poor indicator of the auroral activity at the auroral zone. In Fig. 2.28 B, C, and D the division of the displays into strong and weak was made on the basis of the total number of forms observed in Alaska between geomagnetic latitudes 60° and 70° during each display. In Fig. 2.28 D the total number of auroral forms moving north or south through the zenith of the station has been taken as measure of auroral activity.

Fig. 2.28 clearly indicates that the position of the maximum activity is at essentially the same geomagnetic latitude (66° to 67°) for both small and large displays, but at the lower latitude positions the relative increase of activity for the large displays is considerably greater than at higher latitudes.

(b) Motion of individual auroral forms

A drift of distinct luminous auroral features must be due either to a drift or change of the source of the impact particles or to a change in the electric and magnetic fields experienced by the particles on their way to the atmosphere. The studies of the drift motion of auroral forms is therefore, of course, of considerable interest.

The available observational data in general give the motion in N–S and E–W direction. Considering the deviation from the E–W direction of both geomagnetic isolatitude lines and the average orientation of the auroral forms, it would be preferable in the future to have the motions measured parallel and perpendicularly to the observed auroral zone or to the individual auroral arcs. In this review the motion will, however, have to be divided into N–S and E–W direction.

With regard to the motions of individual auroral forms there are two cardinal observations, which are valid irrespective of direction. First, all parts of rays, which can be several hundred km long, appear to move at the same velocity, aligned with the geomagnetic field in the first approximation, and secondly, auroral forms and inhomogeneities have not been reported to overtake one another.

(i) *Meridional Motion*

The motion of individual auroral forms varies with latitude and local time.

Davis and Kimball (1960) found, when scaling allsky-camera films from Alaska to determine the meridional motions, that the direction of first motions observed each evening at the various latitude positions followed definite trends. On many nights, there existed a geomagnetic latitude

between 60° and 70°, north of which the first motions were to the north, and south of which the first motions were to the south. This latitude they referred to as the "latitude position of origin." According to Davis (1962 b), the auroral forms do not necessarily form at this latitude. Rather, they seem to develop within a few degrees of the "latitude position of origin" and have a direction of first motion determined by whether the forms develop on the poleward or equatorward side of the "latitude position of origin." Davis and Kimball found the "latitude position of origin" to be between 60° and 70° on 86 of the 180 nights when aurora was observed over Alaska during the 1957–1958 season. On 58 other nights the aurora first moved south across the whole latitude band 60°–70°. Approximately 10 nights were found on which either the first motions were without pattern or there were more than one latitude of origin.

The regularity mentioned was for the first direction of motion seen in the evening. Later on in the night both northward and southward moving forms are observed. In fact, Davis and Kimball (1960) and Davis (1962 b) found that during IGY the number of auroral forms moving southward within the geomagnetic latitude band 60° to 70° over Alaska was greater than the number of north-moving forms at any time in the night. Bhattacharyya (1961) observed near Ottawa, well to the south of the auroral zone, that the number of equatorward moving forms outnumbered those in the reverse direction only up to 0200 hours (75th meridian time). Fillius, Gartlein and Sprague found a similar result at the southern auroral zone (according to Cole, 1963). The data presented by Davis (1962 b) differs from those of Bhattacharyya and Fillius et al, in that Davis gave the number of forms within a 10° latitude range, whereas the others presented observations from one place.

Older observational results of local time dependence of meridional motion in subauroral latitudes have been reviewed by Harang (1951) and Störmer (1955.) They can be roughly summarized in the following way: the motion is on the average equatorwards before midnight and polewards after. The apparent discrepancy between the results of Davis and Kimball (1960) and most older results may be due to the difficulty to distinguish, without the use of allsky films, between motion of the equatorward boundary of auroral activity and actual meridional movement of individual forms.

Kim and Currie (1958) analyzed the N–S motion south of the northern auroral zone with regard to velocity. They found the velocity to be greatest (about 160 m/sec) at the latitude of the auroral zone. The number of occasions of equatorward movement exceeded those polewards in every 50 m/sec speed range up to 670 m/sec except in the range 0–50 m/sec.

(ii) *Longitudinal Motion*

At the auroral zones and in subauroral latitudes both east and west longitudinal motions of auroral forms occur during the night, westerly predominating early in the night and easterly in the morning. (Meinel and Schulte, 1953; Meek, 1954; Bless et al., 1955; Malville, 1959; Evans, 1959, 1960; Kim and Currie, 1960; Bond, 1960; Stoffregen, 1961; Davis and DeWitt, 1963). Most investigators are agreed that east–west motions are in general much faster—by at least a factor of 3—than north-south motions, amounting to several hundred m/sec for average disturbance levels. Kim and Currie (1960) observed a general increase of speed with proximity to the auroral zone and also with increasing geomagnetic disturbance level.

Davis (1962 b) and several before him found that during the simplest type of auroral display, one accompanied by positive disturbance of the horizontal magnetic field in the evening and negative disturbance in the morning, the auroral motion is westward during the positive disturbance and eastward during the negative. The reversal from westward to eastward auroral motion is abrupt to within 30 minutes and occurs at the time of, or some minutes after, the change of sign of the horizontal magnetic disturbance.

Mawson (1914) and Stoffregen (1961 b) observed inside the southern and northern auroral zones, respectively, that the longitudinal motion in auroras was, in general, from the daylight to the dark side of the earth, i.e. opposite the general trend outside the zone.

According to Davis (1961, 1962 b), the moderate evening aurora at the auroral zone is typified by the existence of a partial or complete array of west-opening loop structures along which clockwise (as seen from above) motions of irregularities occur. These motions are superimposed on the

general westward drift of the auroral forms. Following the change from positive to negative horizontal disturbance, broken forms are prevalent, but now counter clockwise motions occasionally can be seen. Auroral displays accompanied by magnetic disturbance of variable sign have correspondingly complicated auroral motion.

The motion pattern observed by Davis agrees with Mawson's and Stoffregen's observations inside the auroral zones and it also fits well to the theoretical model of Axford and Hines (1961).

Recently, Omholt (1962) has reported east–west speeds of rays in a very active auroral curtain as high as 30 km/sec. These observations were made with a new technique. The data reported are only preliminary and more measurements of this kind will certainly add much to our knowledge about the auroral mechanisms. Still higher velocities of irregularities along auroral forms (up to 124 km/sec) have been observed by Davis and Hicks (1963) with the use of image amplifier technique, permitting 30 exposures/sec or more.

7. Dynamic Morphology of Individual Auroral Displays

A few synoptic studies of auroral displays have been published since IGY (but much fewer than would be desirable). This type of investigation is certainly one of the most important that can be made at the present stage of our knowledge about aurora. Those published have revealed several previously unknown characteristics of the auroral morphology. A brief review will be given here.

Khorosheva (1963) found that auroras observed simultaneously at different longitudes in the eastern hemisphere form a single, physically related band, which changes synchronously in brightness, width, and latitude over its entire extension.

Davis (1962 b) and Davis and Kimball (1962) analyzed synoptic auroral maps for Alaska and evaluated the results reviewed earlier concerning auroral motion and auroral configurations. Cood evidence for the configuration pattern of aurora being similar to that of the DS current system was found.

Synoptic maps covering still larger areas than those of Davis (1962 b), namely eastern Siberia, Alaska, and western and central Canada; have been prepared and analyzed by Akasofu (1960, 1962, 1963) and Akasofa and Chapman (1962). The disturbance periods studied are the very great ones. (Sept. 22–23, 1957, Feb. 11, 1958, Feb. 13, 1958; March 24, 1958).

A short report of the time history for the Feb. 13, 1958 aurora, given by Akasofu (1963), will be included here, because it illustrates better than has been done before the development of a large aurora on a world-wide scale. Fig. 2.29, a-e, shows snapshots of the situation at five different moments of time.

Three arcs extended across Alaska at 0830 UT but only one to the north of eastern Siberia (Fig. 2.29 a). Five minutes after the break-up, the auroras over Alaska were completely disrupted. A few arcs rapidly moved northward and one of them went to the north of Barrow (Fig. 2.29 b). To the west of Alaska a remarkable loop formation can be seen in Fig. 2.29 b. It seemed to be produced at the time of the breakup over Bering Strait. There is a great difference in the latitude range of the aurora between Alaska and Canada in Fig. 2.29 b. This is because the zone with aurora did not expand simultaneously over the longitude range shown. It started over Alaska. There was, however, a strong increase in brightness also over Canada when the breakup began in Alaska.

At 0900 UT the expansion had spread to central Canada (Fig. 2.29 c). A new breakup started at 0907 UT in Alaska and the broken bands became much brighter also in central Canada. At 1030 UT the auroral forms in Alaska and the Siberian Arctic Sea were quiet again and there were several faint arcs. The aurora was, however, still irregular in Canada and did not recover to this quiet phase, although it became fainter. The auroral precipitation zone was still expanding to the north there.

Some of the observations made by Akasofu in analyzing synoptic sequences similar to the one in Fig. 2.29 are summarized here.

(1) The auroral precipitation zone starts to move away from the pole when the ring current starts to build up. It may reach latitudes 10° (or more) less than that of the auroral zone. The width of the precipitation zone in quiet periods (between polar magnetic (DP) substorms) is not increased (5° wide or less).

(2) The breakup occurs when a DP substorm starts. It is characterized by a sudden increase

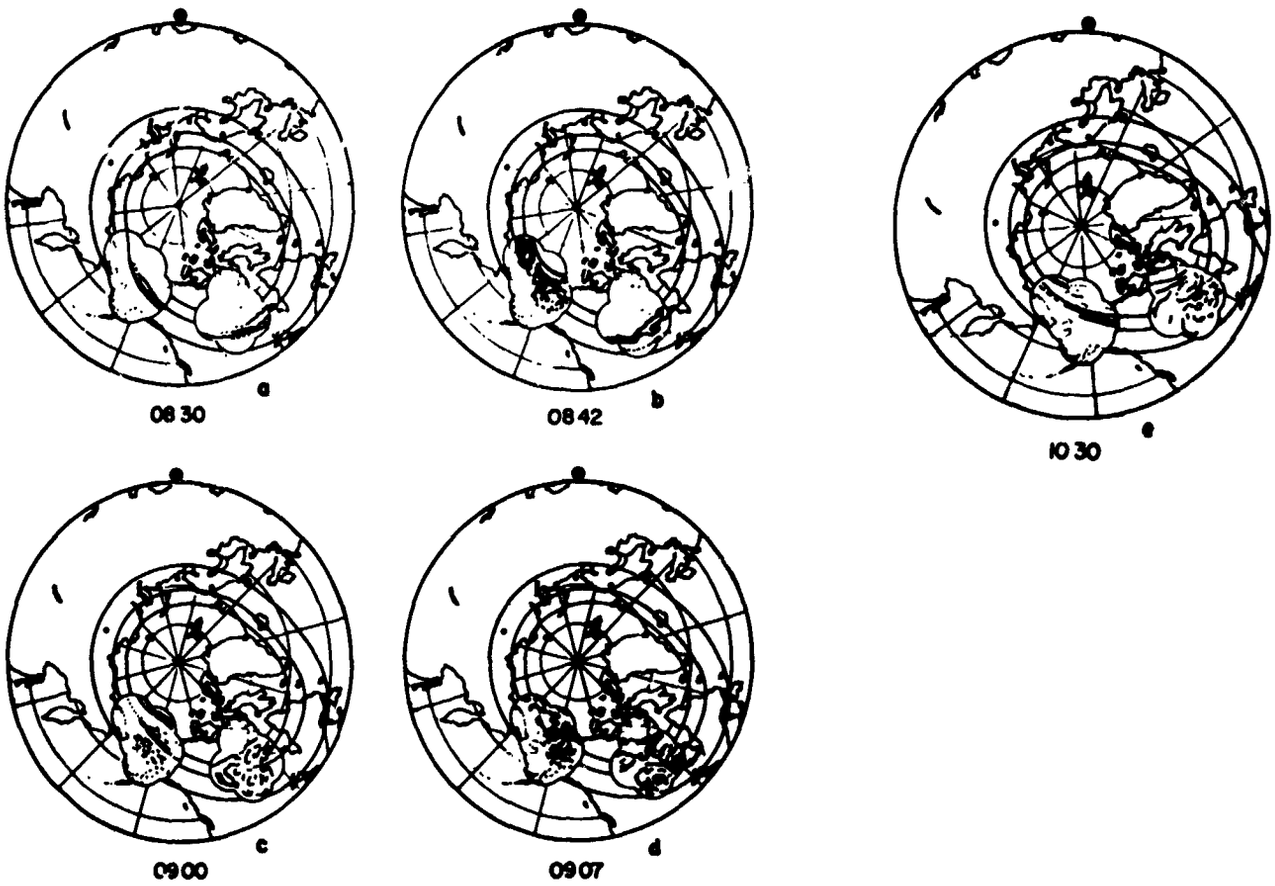


FIGURE 2.29.—(After Akasofu, 1963) The distribution of the auroras on February 13, 1958. The field of view of each allsky camera station is indicated by a circle of radius 500 km. The dark dot on the top of each map shows the direction of the sun. L-curves 4, 6 and 8 are also shown.

of the brightness and rapid and complicated folding or wavy motions of all arcs and bands over a large area on the morning side, together with a rapid expansion polewards of the precipitation zone. The northward expansion often starts around midnight or in the early morning and propagates into the morning side. It may amount to more than 10°.

(3) Breakup with poleward expansion has been observed in the auroral zone already 4 minutes after the sudden commencement.

(4) At the breakup, westward traveling folds are often formed. They propagate with velocities of 300–600 m/sec and have characteristic lengths of a few hundred km. Akasofu thinks it is possible the folds are identical with those observed and studied by Davis (1962 b).

(5) Some of the auroral forms in the auroral caps move up there from auroral zone latitudes during and after the breakups.

(6) Complete breakup of the quiet auroral forms are not so common on the auroral caps as in the auroral zones or on their equatorward side.

(7) An arc extending at least 5000 km in E–W direction was seen on March 24, 1958.

It is evident that a large number of the characteristics found by Akasofu do not fit into any theoretical model, in most cases because the existing models are not detailed enough. Some of the results presented above seem to disagree with some proposed models. The disrupted auroral forms shown in Fig. 2.29 are not similar to the DS-current pattern to be expected on the basis of Axford's and Hines' (1961) model for magnetic storms and aurora, just to mention one of the most recent ones. Also, the observation of large scale activation of aurora, with breakup and northward expansion of the auroral precipitation zone already 4 minutes after a sudden commencement, seems not fit their model.

Akasofu (1963 d) has worked out a model of simultaneous auroral activity over the entire polar region in terms of the auroral substorm, which corresponds to the polar geomagnetic substorm (DP).

3. CORRELATION WITH SOLAR AND GEOPHYSICAL PHENOMENA

In this section observed statistical relations between visual aurora and solar and geophysical phenomena will be briefly reviewed.

1. Solar Disturbances

Except for the investigations of the variation of the occurrence frequency of aurora with the solar cycle, which have been mentioned in Section 2 (see page 32), detailed investigations of the correlation between visual aurora and solar disturbances seem not to have been made. A large number of studies of relations between solar events and geomagnetic disturbances exist in the literature and are reviewed in Chapter 00 of this book. Since the correlation between occurrence of aurora and planetary magnetic activity is good (cf. below), most of the relevant conclusions reached for magnetic storms are also valid for aurora. The 27-day recurrence tendency seems to be less pronounced for aurora than for magnetic disturbances (Chamberlain, 1961).

2. Geomagnetic Disturbances

That worldwide magnetic disturbances occur at the times when aurora can be observed was discovered in 1741 by Celsius and Hiorter in Uppsala, Sweden (cf. Chapman and Bartels, 1940).

A detailed investigation of the correlation between the occurrence of aurora and the planetary magnetic activity has been made by Bartels and Chapman (1958). It has been known for long to aurora investigators that the start of a magnetic storm in the day almost invariably leads to the occurrence of aurora in auroral zone latitudes in the following night. This is true in subauroral latitudes only when the storm is a strong one—in which cases very little aurora may be seen in the auroral zone (see Section 2).

Over the central polar caps (geomagnetic latitude above 80°) the visual aurora is negatively correlated with the local as well as with the planetary K indices (Davis, 1963). In the region between geomagnetic latitudes 75 and 80° the relationship between auroral occurrence and geomagnetic activity is more complicated and of a transitional nature. Feldstein (1962 b) found that auroras are observed in the zenith at Murchison Bay (geomagnetic latitude 75°N) most frequently when there is a weak magnetic disturbance ($K=2$–3). When auroras appearing to the south were also taken into account the auroral occurrence frequency increased with the K-index up to high values of it. Morozumi (1963) found positive correlation between auroral occurrence and magnetic disturbance even at 78° geomagnetic latitude (at the South Pole).

There is not only a correlation between the general occurrence of aurora, but a fairly detailed relationship exists between the development of the visual aurora, as seen in auroral zone latitudes, and the development of the associated geomagnetic disturbance. These relations have been studied by Meek (1953, 1954b), Heppner (1954a, b, 1955), Zaborshchikov and Fediakina (1957), Fan (1958), Bless et al. (1959) and Akasofu (1963d).

Heppner (1954b) summarizes his observations at College (geomagn. lat. 64.5°N) in the following way. The magnetic disturbance and simultaneous auroral activity on a large majority of, and perhaps all, nights may be represented by means of two patterns, which apply individually or in combination. This is a consequence of recognizing that magnetic disturbances are made up of individual bay disturbances and that coincident with each bay there is a distinct sequence of auroral activity. The two patterns differ in respect to what happens when the disturbance in the horizontal component of the geomagnetic field changes from being positive to being negative. In the first pattern the aurora undergoes a distinct change in form and in the second it disappears at this time or recedes very far northward. The first pattern is more frequent than the second.

In the early part of the night the aurora is generally far north as seen from College and has quiet forms (arcs and glow). The luminosity advances slowly equatorwards. A positive magnetic bay is generally observed in this period. Close to the moment when a negative bay sets in, the aurora becomes active, moves southward and breaks up (first pattern) or disappears (second

pattern). In the first pattern of Heppner the recovery of the horizontal component takes place during the appearance of diffuse and pulsating aurora. In the majority of cases there is a short period following the negative bay in which the horizontal magnetic disturbance component is positive. A homogeneous arc is frequently formed again during this period.

A close correlation between rapid variations in the auroral luminosity and geomagnetic micropulsations has been found by Campbell and Nebel (1959), Campbell (1960 a,b,c), Campbell and Rees (1961) and Berger (1963) (see also Chapter).

3. Earth Currents

The occurrence during the day of induced voltages in long telegraph wires was used by Störmer in his extensive program for paralactic height measurements on aurora with good success for many years as a warning that aurora would occur in the next night. As the electric conductivity of the ground is generally not known with adequate precision it is difficult to relate the details of the induced electrical field and the resulting earth currents to the aurora in individual cases. The correlation of occurrence of the two phenomena is, however, good (cf. e.g. Hessler and Wescott, 1959). Freier (1961) has reported evidence for the accumulation of negative charges at the earth's surface during auroras, corresponding to an influx of negative charges into the upper atmosphere.

4. Radio Aurora

Reflections of radio waves of frequencies well above the critical frequencies of the ionosphere from ionization associated with aurora—radio aurora—were first investigated in 1938 by Harang and Stoffregen (1938, 1940). That radio wave propagation via auroral ionization is possible was, however, discovered earlier in the 1930's by amateur radio operators. After the second world war and particularly during the last ten years radio aurora has been studied by a number of researchers in North America, Britain, Norway, Sweden, Germany, Russia, New Zealand and the Antarctic. For a review of radio aurora with detailed references, see for instance Egeland (1962), Huitqvist and Egeland (1964) and Leadabrand (1964).

The diurnal variation of radio aurora as observed in the auroral zone has an afternoon maximum as well as a night maximum (Fig. 3.1). The seasonal variation in the auroral zone has maxima at the equinoxes. The long term occurrence frequency variation has its maximum about 2 years after the solar activity, as is the case with visual aurora.

Radio aurora is observed in the same general regions as optical aurora. With high power radars auroral type echoes can be recorded very often even in latitudes far from the auroral zones.

The reflection process has been found to take place in a fairly narrow height interval with mean value between 90 and 110 km, i.e. in the same height range as the lower border of visual aurora, by a large number of investigators. Also the motions of radio aurora correlate quite well with those of visible aurora, described in Section 2 (see p. 44). Fig. 3.2 shows an example of simultaneous recording of visual and radio aurora. As can be seen, there is a significant but not perfect correlation. The correlation is, however, good only at low elevation angles in a sector centered somewhere around geographic or geomagnetic north, in the northern hemisphere, and south in the southern hemisphere (cf. Fig. 3.3). Even extremely strong aurora overhead does not give any echoes at all.

The mentioned limitations of the regions in which auroral echoes are found is due to the fact that the reflection mechanism is aspect sensitive: reflections occur predominantly from a direction perpendicular to the magnetic field lines at the reflection point. This can be understood if the reflection mechanism is one of weak scattering from small irregularities strongly elongated along the lines of force, as well as if it is a critical reflection from larger magnetic-field-aligned irregularities. Recently it has been suggested that the irregularities may be ion sound waves (Buneman, 1963). The fine structure of the received echoes shows that a large number of reflecting centers are active at the same time.

As radio aurora has been observed for frequencies up to 2850 Mc/s (Groth et al., 1964) it seems most probable that weak scattering from small irregularities in the electron density distribution is the most important process at higher frequencies, whereas critical reflection—i.e. reflection

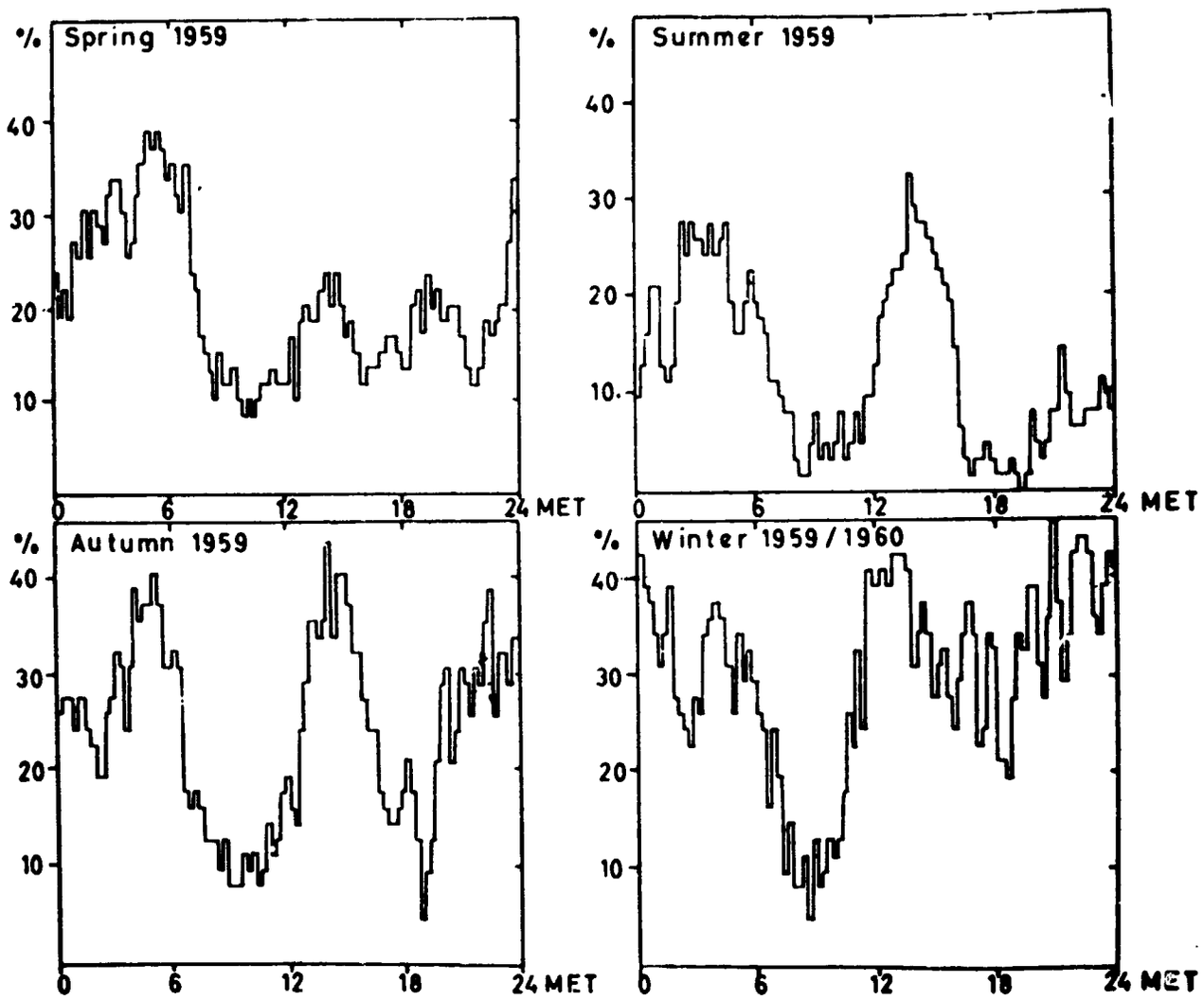


FIGURE 3.1.—Diurnal variation of the occurrence frequency for radio aurora at about 90 Mc/s, observed at Kiruna Sweden (geomagnetic latitude 65.3°). After Egeland (1962a).

from an electron density so large that the refractive index is zero—may be of importance in the lowermost part of the VHF band. How the irregularities are produced is far from clear. See the above mentioned review papers for more details.

5. Bremsstrahlung X-rays at Balloon Altitudes

In the auroral zone the probability of occurrence of x-rays has been found to be fairly well correlated to large scale magnetic activity. But the detailed correlation between aurora and precipitation of the energetic electrons, which generate bremsstrahlung, is in general quite poor there (cf. Anderson, 1960, 1962; Anderson and Enemark, 1960). It does, however, happen that auroral luminosity and x-ray flux follow each other extremely well (see Fig. 3.4, after Anderson and DeWitt, 1963). This seems to be fairly exceptional in auroral zone latitudes.

In subauroral latitudes strong x-ray events occur almost only during strong magnetic storms, and the correlation between the occurrence of visual aurora and bremsstrahlung at balloon heights is quite good (see the review by Winckler, 1962, for instance).

The balloon observations thus show that the high-energy electrons responsible for the bremsstrahlung production in general do not constitute the high energy tail of the electron spectrum producing the visible aurora in auroral zone latitudes, whereas they usually do in subauroral latitudes. The spectrum of the primary electrons is discussed in some more detail in Section 4.

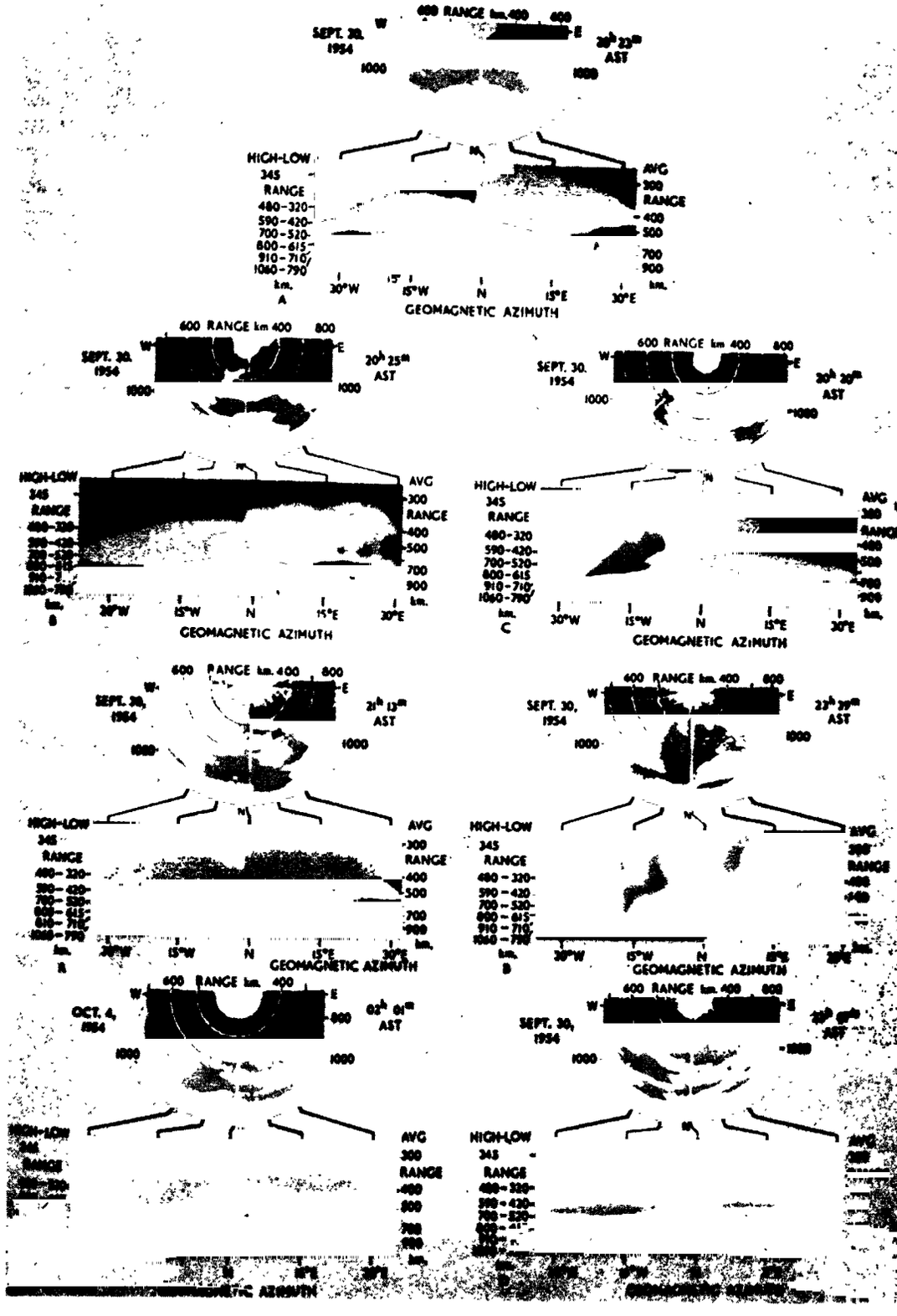


FIGURE 3.2.—Allsky-camera photographs of visible aurora compared with simultaneous 106 Mc/s radar echoes obtained at College, Alaska. After Bowles (1954).

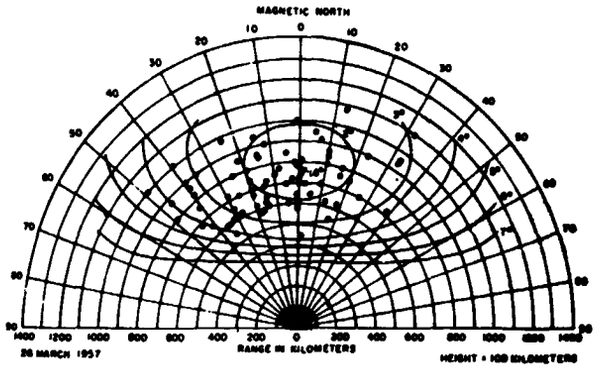


FIGURE 3.3.—Mass plot of 398/Mc/s echoes as a function of range and azimuth, College, Alaska, March 26, 1957. The points represent observations. The numbers attached to the various curves give the off-perpendicular angle with regard to the geomagnetic field lines at the reflection point. After Leadabrand, et al. (1959).

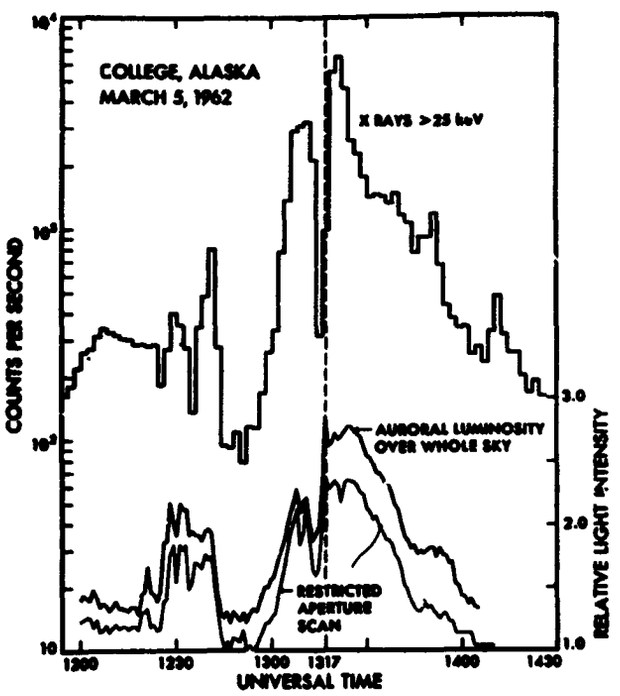


FIGURE 3.4.—Comparison of auroral luminosity obtained from allsky-camera pictures with x-ray flux at balloon altitude. After Anderson and DeWitt (1963).

6. Auroral Absorption

During aurora and magnetic storms the ionospheric absorption of radio waves recorded by means of riometers most often shows an irregular time structure. Although the period of absorption may last several hours, individual peaks often have durations of only a few minutes. Rates of increase and decrease may be as high as 3–2 db/minute, and the magnitude of the absorption can be 8–10 db at about 30 Mc/s for short periods. The character of the record differs definitely from that seen during SIDs or polar cap absorption events (PCA) (cf. Chapter). This type of absorption is usually called auroral absorption. Thus, the auroral absorption (AA) is defined as all absorption observed when there is no SID or PCA. It is then not surprising that Ansari (1963) has been able to show that AA consists of more than one type of absorption phenomenon, different with regard to the energy spectrum of the ionizing particles. One type, primarily occurring in the pre-breakup phase of aurora, is well correlated with the visual aurora (see also Gustafsson, 1964). It is limited approximately to the luminous forms and its intensity is generally fairly low ($\lesssim$1 db at about 30 Mc/s). During the break-up of the visual aurora the AA is intensified and is still mostly limited to the luminous forms.

After break-up, and practically only after magnetic midnight at College, strong fairly slowly varying absorption often occurs. It is sometimes associated with faint and diffuse aurora all over the sky, but there is no detailed correlation between light intensity and AA.

Even if there, thus, is a good statistical correlation between auroral absorption and visible aurora, bright coronas may appear without being associated with any measurable amount of absorption (see e.g. Kavadas, 1962). Ansari (1964) reported cases in which the absorption even decreased when an auroral form brightened up.

The auroral absorption has been found to have its maximum average value in a zone located a few latitude degrees further from the pole than the visual aurora (Gorbushina, 1962; Hartz, 1963; Holt, 1963; Basler, 1963).

The height where the auroral absorption takes place is a parameter of great importance for the understanding of the electron reactions in the lowermost ionosphere (see Hultqvist 1963b, 1964). The observation that the effect of sunrise and sunset on AA is much lower than expected on the basis of existing models (see e.g.Hultqvist,1963a) has been interpreted as indicating that all AA takes place above 90 km (Brown and Barcus, 1963). Such an interpretation is, however, in

contradiction with a large body of other experimental data (see Hultqvist, 1963b). Recent extensive multi-frequency riometer measurements (Little et al., 1963) which provide height information for the AA in the interval 35-75 km, have shown that in some 40 percent of 112 studied cases the height distribution of the absorption per km had its maximum at or below 72 km. In some 10 percent of the cases the peak was even at or below 60 km.

The multi-frequency riometer technique cannot tell about possible existence of appreciable AA in the *E*-layer in the presence of AA also below the *D*-layer, for instance. Analysis of recent satellite measurements of electron precipitation has shown that most of the auroral absorption in general is located between 60 and 90 km (Hultqvist 1964a, b) in good agreement with the riometer measurements mentioned. The satellite data show that AA is due mainly to electrons of energy above 40 kev, whereas the visual aurora is caused by electrons of energies between a few kev and a few ten kev (see the review by Hultqvist 1964c). This is illustrated by Figures 3.5 and 3.6.

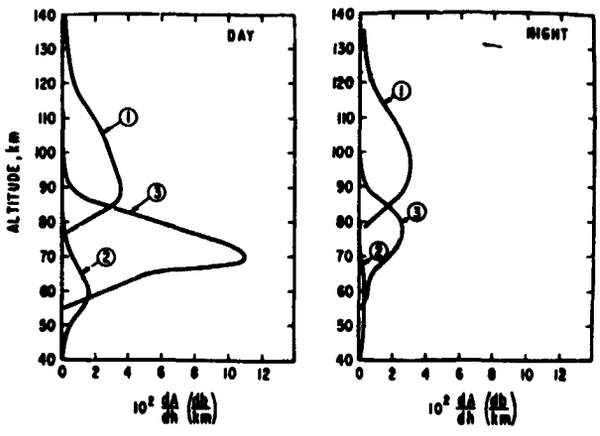


FIGURE 3.5. —Curves 1 show the height distribution of the absorption produced by the differential energy spectrum $N(E)=5.10^9\ e^{-E/5}$ electrons (cm² sec ster kev)$^{-1}$. The total absorption under the curve amounts to 1.04 db in the day and to 0.89 db at night. Curves 2 give the absorption due to the bremsstrahlung of the mentioned electron spectrum. Total absorption in the day is 0.27 db and in the night 0.061 db. Curves 3, finally, represent the absorption distribution produced by the differential energy spectrum $N(E)=7.10^4e^{-E/41}$ electrons (cm² sec kev)$^{-1}$ coming in along the field lines. Total daytime absorption is 1.9 db and the nighttime one is 0.52 db. After Hultqvist (1964a).

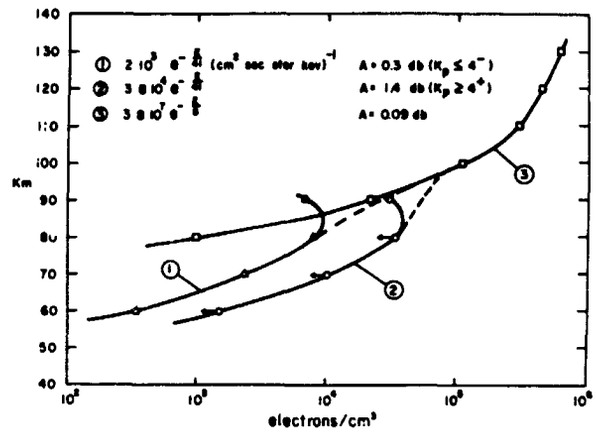


FIGURE 3.6.—Equilibrium electron density profiles due to three energy spectra of precipitated electrons. The spectra as well as the total absorption, A, at 27.6 Mc/s are given in the figure. Curve No. 2 gives an upper limit for the equilibrium electron density distribution produced by the spectrum $3.8\cdot10^4e^{-E/30}$ electrons (cm² sec ster kev)$^{-1}$ which is the equivalent spectrum for the average fluxes when $K_p>4$ as found by McDiarmid et al. (1963) a lower limit for this spectrum is curve No. 1. After Hultqvist (1964b).

Fig. 3.5 shows an estimation of the height distribution of the absorption produced by the primary electrons of a strong visible aurora of International Brightness Coefficient (IBC) between III and IV (curves nos. 1; spectrum of the electrons being that observed by McIlwain, 1960). The absorption due to the bremsstrahlung of those primary electrons (curves 2) and the absorption due to high energy electrons observed in the auroral zone with satellites by Mann et al. (1963) (curves 3). In Fig. 6 the equilibrium electron density profiles produced by the average precipitation fluxes of electrons in the auroral zone, as observed by means of the satellite Alouette (McDiarmid et al., 1963) and Injun 3 (O'Brien, 1964) are shown. For electrons of energy above 40 kev data are available for low ($K_p<4$) and high ($K_p>4$) magnetic activity. Curve 1 gives the average distribution for low magnetic activity, while curve 2 gives only an upper limit for the electron density profile when $K_p>4$. Curve 3 finally is due to the average electron flux in the energy range 1–40 kev.

The total absorption values computed from the satellite measured electron fluxes (0.3 db, <1.4 db, and 0.09 db for curves 1, 2, and 3, respectively) for the two ranges of magnetic activity are in quite

good accordance with the average absorption during the five most disturbed days and five most quiet days in each month, as observed over several years at College by Basler (1963). For the disturbed days he found a daily average of about 1 db in the summer and between 1 and 2 db in the winter and at equinoxes. This should be compared to the value $\sim$ 1 db of Fig. 3.6. For the quiet days the daily average observed by Basler was 0.3 db, which is identical to the value obtained from the satellite data.

The bremsstrahlung x-rays are most probably always negligible as compared to the primary electrons for producing AA (Ansari 1963, Hultqvist 1963, 1964, Brown, 1964).

The fluxes of the high-energy electrons causing AA and the low energy electrons producing aurora vary independently of each other. This explains the variable degree of correlation between aurora and AA. One can expect to observe aurora without AA and AA without aurora and all combinations in between. Ansari's (1963) results indicate that the electrons of energy above 40 kev are precipitated mainly after magnetic midnight. In the same way it is also understandable why auroral absorption has its maximum in the morning (Fig. 3.7), while aurora is most frequent in the middle of the night. The good correlation that has been found between AA and bremsstrahlung x-rays (see Brown, 1961, and Pfotzer et al., 1962, for instance) fits also well into the picture.

From the data presented above it seems reasonable to conclude that the absence of a significant influence of sunlight on the auroral absorption cannot be explained by the absorbing ionization being located, as a rule, above 90 km altitude.

For a discussion of other aspects of auroral absorption see e.g. the reviews of Ansari (1963), Holt (1963), and Hultqvist (1963c) and Chapter 0 of this book.

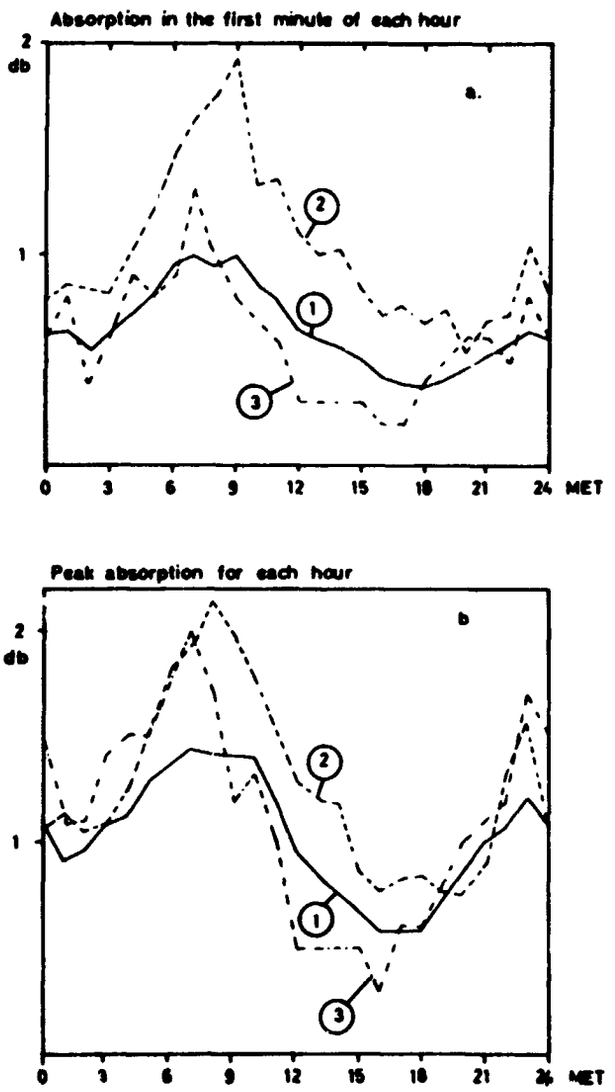


FIGURE 3.7.—Diurnal variation of the average intensity of 27.6 Mc/s cosmic noise absorption for the equinoxes at Kiruna. Months in which PCA's occurred are excluded. Curve 1 represents the average for the following months: April 1959, September 1959, October 1959, March 1960, October 1960, March 1961, April 1961, September 1961, and October 1961. Curve 2 is for September 1961 alone and curve 3 for September 1959. After Hultqvist (1963a).

7. Other Ionospheric Disturbances Associated with Aurora

The first phase of an auroral situation in the ionosphere is usually characterized by increasing height of the $F2$-layer. In about 100–110 km height an auroral sporadic E-layer is formed, which is quite similar to the daytime E-layer but has greater vertical extension. The critical frequency of this layer may rise up to 7–8 Mc/s and has usually well pronounced ordinary and extraordinary components. Multiple reflections are quite common, thus little absorption below this layer is often observed. In the second phase more absorption takes place, the minimum frequency of the E-layer increases and less multiples are recorded. Auroral forms with ray structure is often seen at this time. The third phase is characterized by high absorption of radio-waves,

often resulting in black-out and strong aurora with ray-structure (after Stoffregen, 1958).

During the second and third phases low-frequency reflections are often observed at abnormally low altitudes of down to 60 km (Stoffregen, 1958; Stoffregen et al., 1960; Gregory, 1961; Pedersen, 1962; and others.)

Only few detailed studies of the correlation between visual aurora and sporadic *E* in the auroral zones have been reported (Heppner et al., 1952; Knecht, 1956; Hunsucker and Owren, 1962). The most recent one is that of Hunsucker and Owren. They compared auroral type sporadic-*E* recorded with ionosondes in Alaska during the IGY with simultaneous allsky camera observations and other data for times of auroral activity. They found a high correlation between zenithal aurora and the critical frequency of the sporadic-*E* layer (*fEs*). From detailed studies of simultaneous visual aurora and ionospheric sounding data Hunsucker and Owren drew the following conclusions:

The motion of an auroral arc or band from a low elevation angle to a position near the zenith is accompanied by an increase in the value of *fEs*.

fEs is more than two times as great when there is aurora in the zenith as when there is not.

The highest value of *fEs* is observed when an auroral arc or band is present at the astronomical zenith (not the magnetic zenith).

A general survey of existing knowledge concerning sporadic-*E* phenomena of all kinds can be found in the paper by Thomas and Smith (1959). For other ionospheric disturbance effects associated with aurora, see Chapter 0.

8. Scintillation of Radio Waves from Radio Stars and Satellites

Although no statistical studies of the correlation between auroral occurrence and scintillation of radio-waves from point sources outside the ionosphere have been made, it can be said that a good correlation certainly exists. A correlation exists namely between scintillation and geomagnetic activity (cf. e.g. Little et al. 1962, and Liszka, 1963), as well as between occurrence of visual aurora and geomagnetic activity (see above) and a few detailed studies of situations in which the radio waves have propagated through an observed visual aurora have shown that the scintillation is higher when the waves pass through the aurora than when they do not. This has been found both for ordinary aurora (Little and Maxwell, 1952a; Benson 1960; Morcroft and Forsyth) and for the subauroral-latitude red arcs (Roach, 1963).

9. Radio-Noise Emission

There is some evidence for VHF radio noise being emitted by the auroral primary electrons. The situation is still somewhat unclear. If the reported emissions are of auroral origin, a strong solar cycle variation in their occurrence frequency seems to exist. Observations have been reported only for solar activity maximum. Leadabrand (1964) has given a review of published observational results. Synchrotron radiation from the auroral primaries seems to be a possible mechanism of generation (Hower, 1963).

Whereas, thus, the association of aurora with significant VHF noise emission still must be considered as somewhat doubtful or at least as of rare occurrence, observational results showing a close relationship between visual aurora and a type of very-low-frequency noise emissions, the so-called auroral hiss, have recently appeared. Hiss differs from the other types of VLF emissions in that discrete emissions are not distinguishable in the emission spectra. A class of hiss known as "auroral hiss" has been shown to be closely associated with the occurrence of aurora on the basis of ground observations (Martin and Helliwell, 1960; Jörgensen and Ungstrup, 1962; Morozumi, 1962) although VLF hiss emission is sometimes undetectable on the ground even in the presence of an intense and active aurora. Morozumi (1963) has found that VLF hiss usually occurs in the prebreakup phase of aurora and is well correlated with homogeneous auroral arcs and bands. He observed, however, also that intense auroral hiss sometimes occurs unaccompanied by auroras. Fig. 3.8 shows simultaneous records of VLF noise, visual emission, and precipitated electrons obtained on board Injun 3 (Gurnett and O'Brien, 1964). The agreement between the time functions of the various variables is quite good. The satellite was between 250 and 350 km while the records in Fig. 3.8 were taken. The photometer was looking along the field lines, and the measured electrons were on their way down into the atmosphere. The antenna on the satellite received the

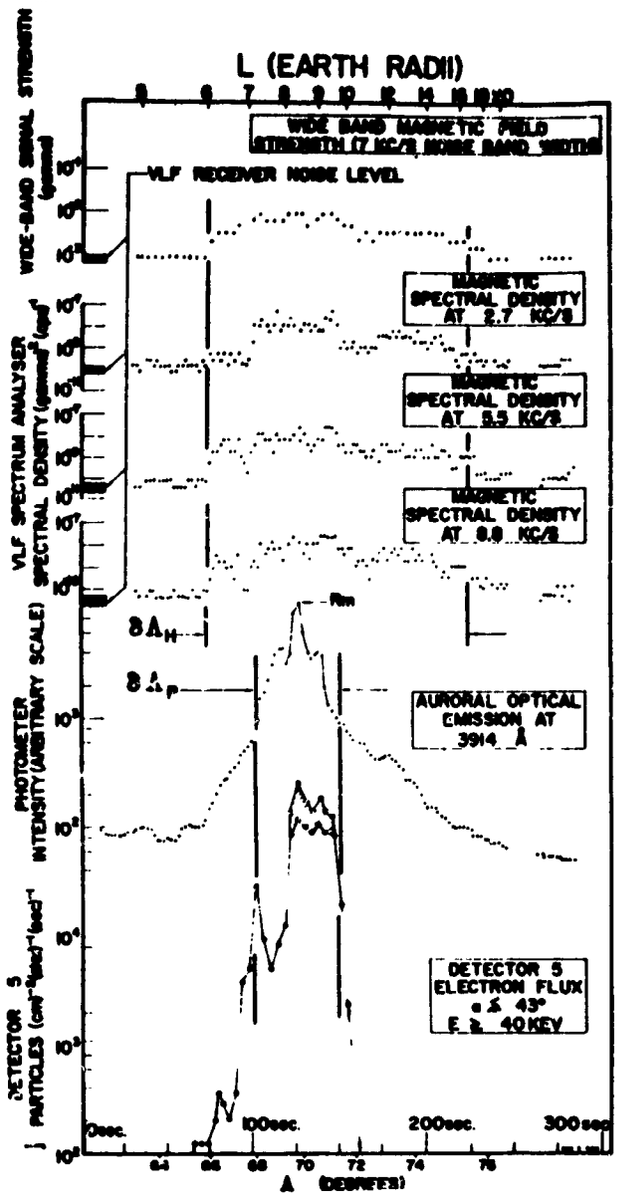


FIGURE 3.8.—A simultaneous observation of VLF electromagnetic emission, auroral luminosity and precipitated electrons on board Injun 3 on March 3, 1963, 0720-0725 UT. After Gurnett and O'Brien (1963).

VLF magnetic component perpendicular to the geomagnetic lines of force. The emission spectrum of the auroral hiss record in Fig. 3.8 is essentially a flat noise spectrum starting below 2.7 kc/s and extending above 8.8 kc/s. The Injun 3 measurements suggest that, at least sometimes, the same precipitated electrons cause both visual aurora and VLF hiss emission.

10. Auroral "Sound"

A large number of reports of aural observations of sound in connection with aurora exist. A review has recently been presented by Campbell and Young (1963). According to these reports, some of which have been given by experienced auroral researchers, a sound is heard in a few per cent of the active auroras, mainly in the years of maximum in the eleven-year solar activity cycle. The sounds are said to be at the upper frequency limit of the observers' auditory response.

Chapman (1931, 1932) has pointed out that the mean free path in the height interval where aurora occurs is too large compared with the wave length of sound waves in the audible frequency range, for transmission down to the earth's surface to be possible. The aurora has to reach down below 65 km for the sound waves to be able to propagate to the earth's surface. But if this occurred, the variations in luminous and acoustic intensity could not be simultaneous, as has been reported, because of the low sound propagation velocity (unless the aurora extended to the ground level). It seems probable that the reported acoustic effects are due to the electric fields that are induced in any suspended wire connected to the earth by the geomagnetic variations associated with strong auroras, as proposed by Harang (1951), Störmer (1955) and others. During very strong auroras these voltages may be of sufficient magnitude to produce discharges.

For infrasonic waves in the period range 10–110 seconds the ratio of wavelength to mean free path is large, and propagation to the ground from the auroral altitudes is physically possible. Such waves have also been observed (Chrzanowski et al., 1961, 1962). Maeda and Watanabe (1964) have pointed out that auroral activity is expected to be associated with the appearance of infrasonic waves due to periodic heating of the ionosphere by the primary auroral particles. Such correlation was found by Campbell and Young (1963) in Alaska, where they observed that every night of bright visual aurora, or ionospheric evidence of aurora, was associated with 10–110-sec-period pressure oscillations of about 1 to 10 dynes/cm^2 amplitude. Nights of no auroral activity showed no infrasonic effects.

4. DIRECT OBSERVATIONS OF ENERGETIC PARTICLES ASSOCIATED WITH AURORA

1. Introduction

The first definite observational evidence for fast charged particles entering the earth's upper atmosphere along the geomagnetic field lines was obtained from a spectrogram taken by Meinel (1950a, b, 1951). He pointed his instrument toward the magnetic zenith during a strong aurora and obtained an $H\alpha$ profile that was shifted—as well as broadened—with its peak corresponding to a doppler velocity of roughly 500 km/sec, but the short wavelength tail of the profile extended to more than 2000 km/sec. In the direction of the magnetic horizon the profile was not shifted but only broadened to about 500 km/sec on either side of the center. The broadening on the red side arises probably in part from scattering of light from other parts of the sky and in part from motions of protons away from the observer. The latter may be due to many small-angle scatterings of protons that initially were moving nearly perpendicularly to the magnetic field.

Proton fluxes as high as 10^8 $(cm^2\ sec)^{-1}$ with velocities below 500 km/sec have been deduced from spectroscopic measurements (Malville, 1960). That proton velocities much above 2000 km/sec have not been observed, does not mean that higher velocities do not exist in the incoming proton beam. The observed limit is due to the effect that much more rapid protons cannot take up electrons to form neutral hydrogen atoms which then can emit the hydrogen lines on its continued way.

In this section will be given a brief review of some recent direct measurements of particles related to aurora which have been made by means of balloon-, rocket-, and satellite-borne instruments.

2. Observed Fluxes and Energy Spectra of Electrons Precipitated Into the Upper Atmosphere

In the summer of 1953 a group from the State University of Iowa when carrying out a program for study of the low-energy end of the primary cosmic-ray spectrum recorded an anomalous soft radiation by means of rockoons launched from a ship west of Greenland (Meredith et al., 1955). The radiation was identified as x-rays of energy 10–100 kev. Some two dozen rockoon launchings at various latitudes showed that the latitude dependence of the radiation intensity was very similar to that for the occurence of visual aurora (Van Allen, 1957).

X-rays directly coincident with auroras were first observed at balloon levels in 1957 (Winckler et al., 1958). A few rockets were launched through visible auroral forms in 1958 (Davis et al., 1960; and McIlwain, 1960). The first proof that electrons are precipitated into the atmosphere from altitudes of 1000 km or more was obtained from Sputnik 3 in a period of about one minute on May 15, 1958, by Krasovskii et al. (1962). A number of reports of measurements on electrons precipitated into the atmosphere have been collected in Table 1. Nos. 1–3 are rocket measurements, no. 4 refers to the rockoon results of the Iowa group, nos. 5–9 are measurements obtained by means of balloons (only x-ray maesurements for which the existence, or nonexistence, and characteristics of visual aurora in the period of measurement have been reported are included in Table 1), nos. 10–21, finally are satellite results.

The early rocket measurements (Davis et al., 1060; McIlwain, 1960) as well as the satellite results (e.g., O'Brien, 1962a, 1964) have verified the conclusions of Omholt (1957, 1959) based on spectroscopic observations of aurora that the protons play a very minor role in the excitation of most auroras.

The first somewhat detailed satellite studies of particles that were definitely precipitated into the atmosphere were made by means of Injun 1 (O'Brien, 1962a, b). O'Brien and others working with satellite data define precipitated particles as such which would mirror at or below 100 km altitude (if the atmosphere had not been present). The satellite results have also shown that electrons on the average carry much more of the energy to the lower ionosphere than do protons (O'Brien, 1962b).

(a) Latitudinal Distribution and Other Spatial Characteristics

The recent simultaneous measurements on Injun 3 (O'Brien and Taylor, 1964) of flux of precipitated electrons and of auroral light emission below the satellite have shown that above auroras there are electron fluxes several orders of

TABLE 1.—*Precipitated Electrons*

No.	Reference	Flux (above the atmosphere)		Spectrum or energy range	Time of obs.	Place of observat. geomagn. lat.	Visible aurora		Magn. disturb.	Radio-wave absorpt.	Method of measur.	Remarks
		Number of particles	Energy				Intens.	Location with regard to part. measurement				
1a	Davis et al. (1960)	$N(E \geq 10$ kev$) = \geq 6 \cdot 10^2$ (cm^2 sec ster)$^{-1}$	0.5 –2.5 erg (cm^2 sec ster)$^{-1}$ 8 kev $< E <$ 100	$N(>E) \propto E^{-1}$ for 5 kev $< E <$ 50 kev	Jan. 26, 1958 0421 UT	Churchill, A = 68.8°N	Fading rayed structure and const. diffuse surface of IBC ~1; after break-up.	Passed through rayed structure during ascent, and through diffuse surface on both ways. Peak altit. 178 km.			Rocket scintill	Protons present. El. flux isotropic within ±15% between magn. zenith angles 9° and 81°
1b	Davis et al. (1960)		< 0.01 erg (cm^2 sec ster)$^{-1}$	8 kev $< E <$ 100 kev	Mar. 16, 1958 0454 UT	Churchill, A = 68.8°N		No penetrat. Weak arc moved to the north.			Rocket scintill.	Protons present.
1c	Davis et al. (1960)		< 0.06 erg (cm^2 sec ster)$^{-1}$	8 kev $< E <$ 100 kev	Mar. 22, 1958 0641 UT	Churchill, A = 68.8°N		No penetrat. Fading arc moved to south.			Rocket scintill.	Protons present.
1d	Davis et al. (1960)		< 0.02 erg (cm^2 sec ster)$^{-1}$	8 kev $< E <$ 100 kev	Nov. 16, 1958 0658 UT	Churchill, A = 68.8°N		No penetrat. Bright arc just to the north.			Rocket scintill.	Protons present.
1e	Davis et al. (1960)	$< 10^9$ (cm^2 sec ster)$^{-1}$ 30 ev $< E <$ 1000 ev			During 1a, b, c, d	Churchill, A = 68.8°N						
2a	McIlwain (1960)	$N(>3$ kev$) = 2.1 \cdot 10^8$ (cm^2 sec ster)$^{-1}$*	1.7 erg (cm^2 sec ster)$^{-1}$ $E > 3$ kev	$N(>E) = 3.9 \cdot 10^{8}$* exp ($-E/5$ kev) electrons (cm^2 sec ster)$^{-1}$	Feb. 22, 1958 0537 UT	Churchill, A = 68.8°N	Faint glow	Passed through the aurora. Peak altitude 120 km.			Rocket scintill.	Protons present.
2b	McIlwain (1960)	$N(E = 6$ kev$) = 5.10^{10}$ (cm^2 sec ster)$^{-1}$	2000 erg cm^{-2} sec^{-1}	Approx. mono-energetic $E = 6$ kev	Feb. 25, 1958 0551 UT	Churchill, A = 68.8°N	Bright arc; after break-up	Passed through the aurora. Peak altitude 120 km.	$\Delta X > 1000\gamma$		Rocket scintill.	No protons present.
3	McDiarmid et al. (1962)	$\leq 2.10^5$cm^{-2}sec^{-1} for $E > 30$ kev		$N(>E) \propto \exp$ ($-E/22$ kev) $E > 30$ kev	Oct. 28, 1960 1223 UT (about sunrise)	Churchill, A = 68.8°N			$\Delta Z > 1000$ before and after	~3 db (30 Mc/s)	Rocket GM-tube.	Protons present. El. flux very irregular. Peak alt. 150 km.
4	Van Allen (1957)	$10^3 - 10^4$ cm^{-2} sec^{-1}	0.01 –1 erg/cm^2 sec^1	10 –100 kev	Summers of 1953–1955	A = 54 –88°					Rockoons	Lat. dep. as for visual aurora.

TABLE 1.—*Precipitated Electrons—Continued*

No.	Reference	Flux (above the atmosphere)		Spectrum or energy range	Time of obs.	Place of observat. geomagn. lat.	Visible aurora		Magn. disturb.	Radio-wave absorpt.	Method of measur.	Remarks
		Number of particles	Energy				Intens.	Location with regard to part. measurement				
5	Anderson & Enemark (1960)	$\leq 2.10^8$ cm^{-2} sec^{-1} for E > 25 kev 10^8 cm^{-2} sec^{-1} for E > 25 kev	0.1 erg cm^{-2} sec^{-1} for E > 25 kev	$N(>E) \propto e^{-E/8}$ for E > 25 kev	Aug. 18, 1959 0745 UT	Churchill, A = 68.8°N	Strong active auroral forms overhead		$\Delta Z = -250\gamma$	≤ 1.5 db at 30 Mc/s	Balloon x-rays	X-rays present 40% of time. Poor correl. with visible aurora. Time average over days.
6	Winckler (1960)	8.10^8 cm^{-2} sec^{-1} for E > 40 kev			Sep. 23, 1957 1050 UT	Minneapolis A = 55°N	Very strong ray structure at 30° elevation		Weak dist. (during recovery phase of storm)		Balloon x-rays	
7	Bhavsor (1962)	$\leq 10^9$ cm^{-2} sec^{-1} for E > 22 kev		$N(>E) = 0.66.10^{15}$. E^{-4} cm^{-2} sec^{-1} for E > 22 kev	May 12, 1959	Minneapolis A = 55°N	Strong overhead aurora				Balloon x-rays	
8	Anderson (1962)	$\leq 10^7$ cm^{-2} sec^{-1} for E > 25 kev			Mar. 27-28, 1961 1100 UT	Fairbanks A = 64.5°N	No apprec. visual aurora				Balloon x-rays	No correl. visual aurora x-rays.
9	Anderson & DeWitt (1963)	5.10^7 cm^{-2} sec^{-1} for E > 25 kev	3 erg cm^{-2} sec^{-1}	$N(>E) = E^{-4}$	Mar. 5, 1962 1320 UT	Fairbanks A = 64.5°N	Strong glow over entire sky; aver. int. ~60 kR			~6 db (riometer)	Balloon x-rays	Very good correl. visual aurora x-rays
10	Krasovskii et al. (1962)		10^{-3} to > 20 erg/cm^2 sec ster	Equivalent energy most often 14 kev	1958	Over a very wide latitude range					Sputnik 3	Neither energy flux nor spectrum depend significantly on pitch angle.
11	O'Brien (1962a)	Aver. 10^8 (cm^2 sec ster)$^{-1}$ for E > 40 kev	$\geq 0.6.10^{-3}$ erg (cm^2 sec ster)$^{-1}$ for E > 40 kev		June 30–July 2, 1961	L = 2 –10			$K_p = 0 - 4$ No correl. with magn. act.		Injun 1, 12 passes during period of rapid spin	Occurs more on day side than on night side; uncorrel. with lat. and long. of observ.
12	O'Brien & Laughlin (1962)	6.10^7 (cm^2 sec ster)$^{-1}$ for E > 40 kev [10^{10} (cm^2 sec ster)$^{-1}$ for E > 10 kev]	400 erg (cm^2 sec ster)$^{-1}$ for E > 1 kev	$N(E) \propto E^{-4}$ for E > 10 kev	Sep. 25, 1961 2237:10 UT LT 1835	Lat. 59.0° Long. 303° L 8.82 Alt. 1011 km			Magn. disturbance		Injun I	Measured part. were not precip. but there is evidence that electron flux was isotropic.
		3.10^9 (cm^2 sec ster)$^{-1}$ for E > 40 kev [6.10^{10} (cm^2 sec ster)$^{-1}$ for E > 0]		$N(E) \propto \exp(-E/6)$ Very rapid variat. in spectrum	Sep. 25, 1961 2236:25 UT LT 1835	Lat. 57.3 Log 300° L 8.04 Act. 1011 km						

TABLE 1.—*Precipitated Electrons—Continued*

No.	Reference	Flux (above the atmosphere)		Spectrum or energy range	Time of obs.	Place of observat. geomagn. lat.	Visible aurora		Magn. disturb.	Radio-wave absorpt.	Method of measur.	Remarks
		Number of particles	Energy				Intens.	Location with regard to part. measurement				
13	O'Brien (1962b)	4.10^3 (cm² sec ster)$^{-1}$ 8.10^3 (cm² sec ster)$^{-1}$ 4.10^4 (cm² sec ster)$^{-1}$ 10^5 (cm² sec ster)$^{-1}$ 7.10^4 (cm² sec ster)$^{-1}$ 6.10^4 (cm² sec ster)$^{-1}$		E > 40 kev	June 29–Sep. 6, 1961	$\cos^{-1}(L^{-1/3})$ = 45–50° 50–55° 55–60° 60–65° 65–70° 70–75°					Injun I	Aver. intens. at 1000 km altitude. Only order of magnitude significant.
14	O'Brien et al. (1962)	3.10^3 (cm² sec ster)$^{-1}$ for E > 40 kev	~1 erg (cm² sec ster)$^{-1}$ for E > 40 kev	$N(E) \propto E^{-3}$	July 21, 1961 ~1600 UT	L ~ 5.3					Injun I	
		[$\sim 10^3$ (cm² sec ster)$^{-1}$ for E > 5 kev]	~9 erg (cm² sec stre)$^{-1}$ 1 kev < E < 40 kev	$N(E) \propto E^{-3}$							Injun I	
15	Mann et al. (1963)	$4 \cdot 10^3$ (cm² sec ster)$^{-1}$ for 80 < E < 106 kev		$N(E) \propto \exp(-E/5$ kev)	1500 UT Sep. 1— 0130 UT Sep. 2, 1961	1/4 to 1/3 of a great circle; in high latitudes			Correl. with magn. disturb.		Discoverer 29 magnetic el. spectromet.	Polor orbit with altitude varying between 160 and 610 km. Directional measurements along radius vector from earths center.
		$2 \cdot 10^3$ (cm² sec ster)$^{-1}$ for E > 80 kev		Two groups of spectra: $N(E) \propto \exp(-E/25 \pm 5)$ and $N(E) \propto \exp(-E/42 \pm 3)$	Aug. 30–Sep. 3, 1961	In the auroral zones			Correl. with magn. disturb.		Discoverer 29 magnetic el. spectromet.	
16	Stilwell (1963)	$\sim 10^3$ (cm² sec ster)$^{-1}$ for E > 10 kev	~2 erg (cm² sec ster)$^{-1}$ for E > 10 kev	$N(E) \propto \exp(-E/4)$ E > 10 kev Mean energy < 10 kev	Mar. 3, 1963 0721 UT LT–2220	L = 8 – 10	~20kR of γ3914Å	Along the field line the satell. is on			Injun 3 Rev. 994	Angular distrib. isotropic

TABLE 1.—*Precipitated E[illegible]s—Continued*

No.	Reference	Flux above the atmosphere)		Spectrum or energy range	Time of obs.	Place of observat. geomagn. lat.	Visible aurora		Magn. disturb.	Radio-wave absorpt.	Method of measur.	Remarks
		Number of particles	Energy				Intens.	Location with regard to part. measurement				
17	McDiarmid et al. (1963)	$3 \cdot 10^8$ (cm² sec ster)$^{-1}$ for E > 40 kev	Aver. value correspond to exponential equival. mean spectrum at A = 65°	Equiv. mean spectrum at A = 65° $N(E) = 2.10^8 \cdot e^{-E/kT}$ el (cm² sec ster kev)$^{-1}$ or	Oct. 1962–Jan. 1963	Invariant lat. A = 65°			$K_p \leq 4^-$		Alouette	Average values at peak of latitude distribution.
		$1.9 \cdot 10^8$ (cm² sec ster)$^{-1}$ for E > 250 kev	0.03 erg (cm² sec)$^{-1}$	$N(E) = 8 \cdot 10^8 E^{-2.8}$ el (cm² sec ster)$^{-1}$ for 40 < E < 250 kev)	Oct. 1962–Jan. 1963	Invariant lat. A = 60°			$K_p \leq 4^-$		Alouette	Average values at peak of latitude distribution.
		$3 \cdot 10^8$ (cm² sec ster)$^{-1}$ for E > 40 kev	Aver. value correspond to exponential equival. mean spectrum at A = 65°	Equiv. mean spectrum at A = 65° $N(>E) = 4 \cdot 10^4 e^{-E/38}$ el (cm² sec ster kev)$^{-1}$ or	Oct. 1962–Jan. 1963	Invariant lat. A = 65°			$K_p > 4^+$		Alouette	Average values at peak of latitude distribution.
		$2.6 \cdot 10^8$ (cm² sec ster)$^{-1}$ for E > 250 kev	0.3 erg (cm² sec)$^{-1}$	$N(>E) = 4.4 \cdot 10^{11} \cdot E^{-3.0}$ el (cm² sec ster)$^{-1}$ (for 40 < E < 250 kev)	Oct. 1962–Jan. 1963	Invariant lat. A = 60°			$K_p \geq 4^+$		Alouette	Average values at peak of latitude distribution.
18	O'Brien (1964) O'Brien and Taylor (1964)	Average: $4 \cdot 10^8$ cm^{-2} sec^{-1} for E > 40 kev	Average: 4 erg cm^{-2} sec^{-1} for E > 1 kev	Equiv. mean spectrum at A = 65° $N(E) = 7.8 \cdot 10^7 e^{-E/5.7}$ el (cm² sec kev)$^{-1}$ or $N(>E) = 1.4 \cdot 10^8 E^{-2.3}$ el. cm^{-2} sec^{-1} (for 1 < E < 40 kev)	First months of 1963	Invariant lat. A ~ 65°	$2\binom{+4}{-1.5}$ kR	On the same field line where the port. were measured			Injun 3	Average values at peak of latitude distribution. 100–1000 times less at subauroral lat. Variations of a factor of 10^3 in auroral zone.
19a	Sharp et al. (1964a)	Max. 10^8 el/cm² sec ster above 2 kev		$N(E) \propto e^{-E/a}$ in burst	March 1, 1962 10:40 UT	Over Alaska		Homogeneous arc			Tota energy scint. detect on board a short-liv d polar orbiting satellite.	Very little contribution from protons to energy flux
19b	Sharp et al. (1964b)	Max $8 \cdot 10^8$ el/cm² sec ster above 2 kev	Max. ~100 ergs/cm² sec ster above 2 kev	$N(E) \propto e^{-E/E_0}$ with $E_0 = 4 - 9$ kev	March 2, 1962 0918 UT	Over Alaska		Auroral arc				
19c	Sharp et al. (1964b)		Max. ~100 ergs/cm² sec ster above 2	$N(E) \propto e^{-E/E_0}$ with $E_0 = 3 - 5$ kev	March 2, 1962 1048 UT	Over Alaska					"	Very little contribution from protons to energy flux

TABLE 1.—*Precipitated Electrons—Continued*

No.	Reference	Flux (above the atmosphere)		Spectrum or energy range	Time of obs.	Place of observat. geomagn. lat.	Visible aurora		Magn. disturb.	Radio-wave absorpt.	Method of measur.	Remarks
		Number of particles	Energy				Intens.	Location with regard to part. measurement				
20a	Sharp et al. (1964c)		0.60 erg/cm² sec ster above 0.18 kev 0.035 erg/cm² sec ster above 10 kev 0.0087 erg/cm² sec ster above 31 kev	$N(E) \propto e^{-E/E_0}$ with E_0=2.2 kev between 0.18 and 10 kev	May 1963 LT~05	61.3°N 124.9°W geographic coord.					Total energy scint. detector and electrostatic retard. detector with postacceleration on board a short-lived polar orbiting satellite.	"Typical Orbits"
20b	Sharp et al. (1964c)		0.83 erg/cm² sec above 0.18 kev 0.065 erg/cm² sec ster above 10 kev	E_0=2.4 kev between 0.18 and 10 kev	May 1963 LT~05	65.3°N 141.3°W geographic coord.					"	"Typical Orbits"
20c	Sharp et al. (1964c)		0.21 erg cm² sec ster above 0.18 kev 0.084 erg/cm² sec ster above 10 kev	E_0=4.9 kev between 0.18 and 10 kev	May 1963 LT~05	68.7°N 156.5°W geographic coord.					"	"Typical Orbits"
20d	Sharp et al. (1964c)		0.34 erg/cm² sec ster above 0.18 kev 0.096 erg/cm² sec ster above 10 kev	E_0=4.0 kev between 0.18 and 10 kev	May 1963 LT~05	62.3°N 131.4°W geographic coord.					"	"Typical Orbits"
21a	Sharp et al. (1964c)		4.4 ergs/cm² sec ster above 0.080 kev 4.4 ergs/cm² sec ster above 1.5 kev 0.25 erg/cm² sec ster above 21 kev	E_0=4.6 kev between 1.5 and 21 kev Cutoff above 1.5 kev	2 Nov 1963 ~20 UT	60.5°S 76.4°E geographic coord.					"	Sample of a "representative orbit."

TABLE 1.—*Precipitated Electrons—Continued*

No.	Reference	Flux (above the atmosphere)		Spectrum or energy range	Time of obs.	Place of observat. geomagn. lat.	Visible aurora		Magn. disturb.	Radio-wave absorpt.	Method of measur.	Remarks
		Number of particles	Energy				Intens.	Location with regard to part. measurement				
21b	Sharp et al. (1964c)		1.1 erg/cm² sec ster above 0.080 kev 1.1 erg/cm² sec ster above 1.5 kev 0.04 erg/cm² sec ster above 21 kev	E_0 =4.1 kev between 1.5 and 21 kev Cutoff above 1.5 kev	2 Nov. 1963 ~20 UT	54.5°S 76.1°E geographic coord.					Total energy scint. detector and electrostatic retard detector with postacceleration on board a short-lived polar orbiting satellite.	Sample of a "representative orbit."
21c	Sharp et al. (1964c)		1.2 erg/cm² sec ster above 0.080 kev 1.2 erg/cm² sec ster above 1.5 kev 0.86 erg/cm² sec ster above 21 kev	E_0 =20 kev between 1.5 and 21 kev Cutoff above 1.5 kev	2 Nov. 1963 ~20 UT	71.5°N 68.6° E geographic coord.					"	Sample of a "representative orbit."
21d	Sharp et al. (1964c)		0.63 erg/cm² sec ster above 0.080 kev 0.45 erg/cm² sec ster above 1.5 kev 0.44 erg/cm² sec ster above 21 kev	E_0 =95 kev between 1.5 and 21 kev	2 Nov. 1963 ~20 UT	62.1°N 114 7°W geographic coord.					"	Sample of a "representative orbit."

*This value is obtained from the energy flux value in the way described by McIlwain (196C). It is, however, a factor of about 2π smaller than the value given by McIlwain himself.

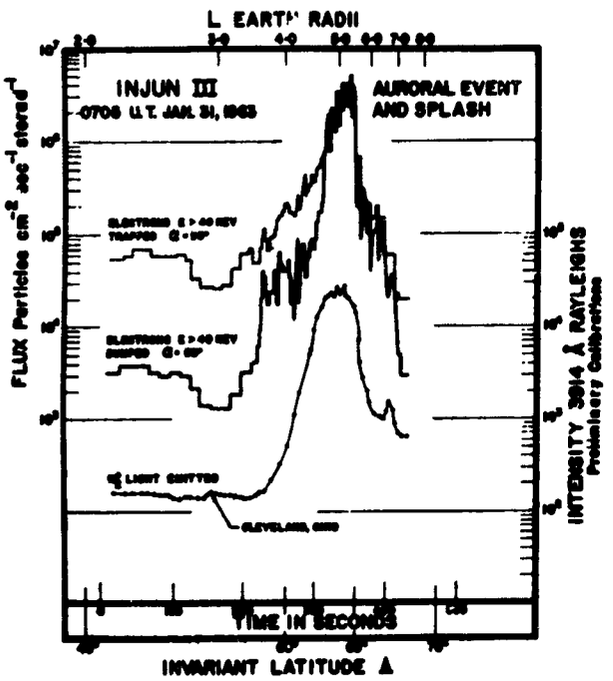


FIGURE 4.1.—Data from a northbound pass of Injun 3 over North America, which shows simultaneous detection of an aurora, of the precipitated electrons (pitch angle ~50°) partially responsible for causing it, and of trapped electrons. Note the approach to isotropy of the particle flux over the aurora. No attempt has been made to subtract the low-latitude contamination of the photometer signal, part of which was the detection of Cleveland, Ohio, and its surrounds. (After O'Brien and Taylor, 1964.)

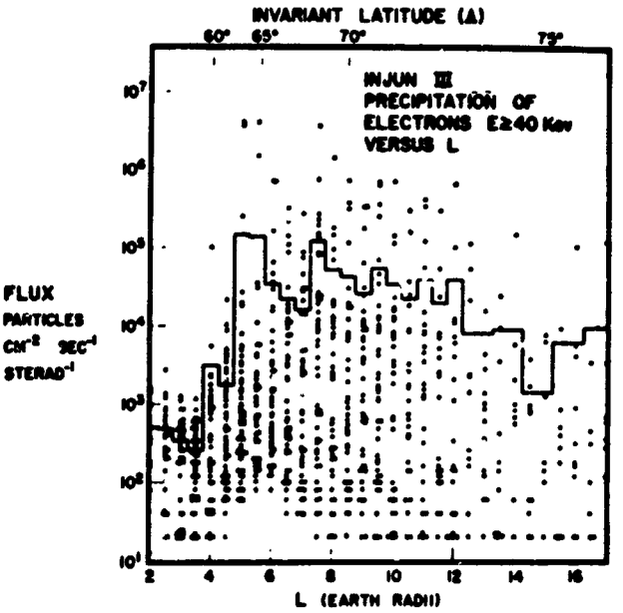


FIGURE 4.2.—Samples of precipitated fluxes over North America in January 1963. Each point is an 8-sec average of thirty-two measurements made at half-integral values of L. The solid line gives the average flux. (After O'Brien, 1964.)

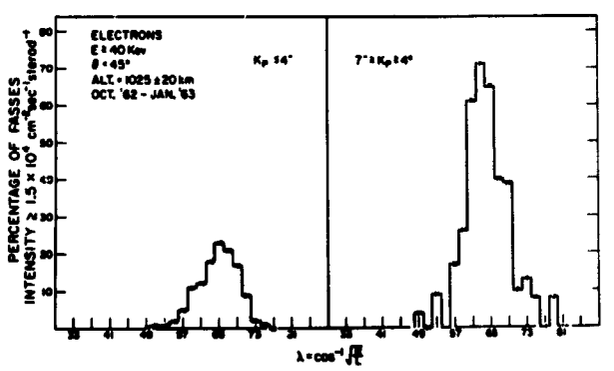


FIGURE 4.3.—Percentage of passes in which the intensity of precipitated electrons with energies greater than 40 kev is greater than 1.5 x 10^4 cm^{-2} sec^{-1} plotted against invariant latitude. (After McDiarmid et al., 1963.)

magnitude larger than outside the aurora (see Fig. 4.1).

Fairly extensive statistical data on the spatial characteristics of precipitated electrons have recently been published by O'Brien (1962a, b, 1964) and McDiarmid et al. (1963). Fig. 4.2 shows the scatter diagram of the Injun 3 observations of precipitated electrons of energy above 40 kev for the range of the invariant latitude Λ (defined by $\Lambda = \cos^{-1}(1/L)^{1/2}$ and at the most a few degrees different from geomagnetic latitude in the latitude range of interest here) between 45 and 76 degrees. As can be seen from the figure, the precipitated flux of electrons of $E \gtrsim 40$ kev has a maximum of about 10^5 (cm^2 sec ster)$^{-1}$ for Λ between 60 and 70 degrees. The flux is 2–3 orders of magnitude lower in subauroral latitudes. Fig. 4.2 also shows that some 10 degrees inside the auroral zone the precipitation rate is down by 1–2 powers of ten. Similar results have been found by McDiarmid et al. (1963) for electrons of energies above 40 kev and above 250 kev. Fig. 4.3 shows their data presented in a manner different from that of O'Brien (1964).

The histograms give, for two different ranges of magnetic activity, the percentage of passes in which the intensity of precipitated electrons with $E \gtrsim 40$ kev was above $1.5 \cdot 10^4$ (cm^2 sec ster)$^{-1}$ (corresponding to 0.2 db absorption at about 30 Mc/s). The flux of $1.5 \cdot 10^4$ (cm^2 sec ster)$^{-1}$ was one half of the maximum average intensity for

$K_p<4$ and 1/20 for $K_p>4$. Fig. 3 shows that during quiet geomagnetic conditions the precipitated intensity was greater than half the average value during some 25 percent of the passes through $\Lambda=65°$, while for moderately disturbed conditions —there were no strong storms represented in the data—the intensity exceeded 1/20 of the average value in about 72 percent of the passes over that latitude.

Simultaneous observations on two satellites indicate that at least sometimes the area of precipitation has a fairly limited longitudinal extent (O'Brien and Laughlin, 1962). On the other hand the precipitated flux seems on occasions to be uniform over as much as 80 degrees of longitude, i. e., over more than 4000 km (O'Brien, 1964).

The flux of precipitated electrons has its maximum close to the poleward boundary of the region of trapped electrons, as observed on the same satellite (McDiarmid et al., 1963, O'Brien, 1964).

Electron precipitation occurs simultaneously in magnetically conjugate areas at least sometimes, according to balloon observations of x-rays (Brown et al., 1963). This is in accordance with observations of aurora (cf. e.g. DeWitt, 1962).

A 100 fold change in the flux over two km has often been observed by Injun 3, which is in agreement with results of rocket measurements in aurora by Davis et al. (1960). They found the electron flux being concentrated in the visible auroral forms, whereas protons were found over a much larger volume than that occupied by the aurora.

The statistical latitude profile of the intensity of dumped electrons with $E\gtrsim 40$ kev has been found by O'Brien and Taylor (1964) to have its maximum at a few degrees lower latitude than the frequency of occurrence of visual aurora as observed on the same satellite. Sometimes the visual emission has been found to extend to higher latitudes than the precipitation of electrons with energy greater than 40 kev, indicating electrons of $E<40$ kev are being precipitated up to higher latitudes than the higher energy ones. The lower latitude limit of the aurora coincides with the lower latitude limit of appreciable precipitation (O'Brien, 1964; O'Brien and Taylor, 1964).

Precipitation was found all the time in the auroral zone by Injun 3 (O'Brien, 1964). It can be seen in Fig. 4.2 that at $L=6$ there was never

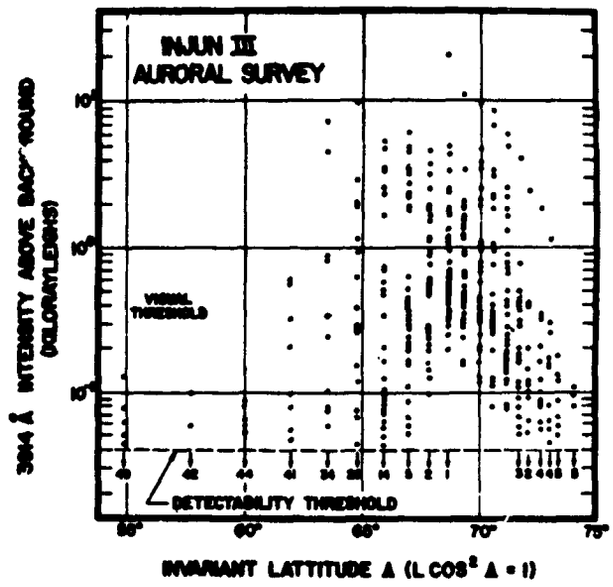


FIGURE 4.4.—Intensity of 3914 Å auroral light averaged over 4 sec at half-integral values of L in about fifty passes of Injun 3 early in 1963. (After O'Brien and Taylor, 1964.)

observed any flux lower than about 80 electrons $(cm^2\ sec\ ster)^{-1}$.

The corresponding continuous photon emission in the auroral zone is evident from Fig. 4.4. The minimum emission rate is, however, below visual threshold even in the auroral zone.

Figs. 4.2 and 4.4 also give a good impression about the enormous variability of precipitation and photon emission at all latitudes, but especially near the auroral zone. At $L=6$ there are values of about $6\cdot 10^6$ electrons $(cm^2\ sec\ ster)^{-1}$ shown in the figure. This is some 10^5 times the minimum value observed there

The latitudinal extent of the precipitation can be very restricted (e.g., over $L\sim 1$ earth radius) or very extensive (over $L>20$ earth radii). This is illustrated in Fig. 4.5.

(b) Time Characteristics of Precipitation

It is not possible to differ between time and space variations in satellite measurements. Balloon observations show, however, that time variations as rapid as of periods of a tenth of a second can occur in the electron flux (Winckler et al., 1962).

No systematic variation in the intensity of electron precipitation over periods as long as a day could be found by O'Brien (1964) from the

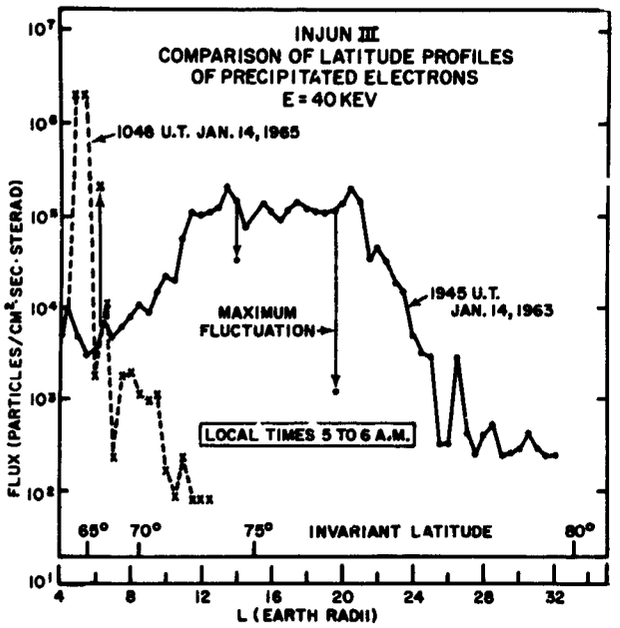


FIGURE 4.5.—Comparison of the latitude, or L profile, of precipitation for two successive passes at about the same local time. Arrows illustrate range of fluctuation of intensities at given locations. (After O'Brien, 1964.)

Injuns 1 and 3 measurements, in contrast with the case for trapped electrons. Sharp et al. (1964) observed a higher nighttime than daytime precipitation flux of electrons in the energy range 0.08–24 kev—as well as of protons—during the 5 days lifetime of an oriented polar-orbiting satellite. They also found the energy distribution of the electrons to be harder on the dayside than on the nightside of the earth (Johnson et al., 1964). With the enormous variability of the precipitation phenomenon a large amount of data seems, however, necessary for deducting statistically significant diurnal variation curves.

(c) Pitch Angle Distribution

Figure 4.1 shows that over the aurora, where the precipitation is intense the directional flux of precipitated electrons becomes equal to the flux of trapped electrons, which also is increased over the aurora. Figure 4.1 illustrates a general rule that has been found by O'Brien (1962a, b, 1964) namely that for electrons of $E \geq 40$ kev the pitch angle distribution approaches isotropy in the regions of intense precipitation. An isotropic pitch angle distribution in regions of strong precipitation has also been observed by Krasovsk" et al. (1962) at an energy of about 10 kev. No cases have been found in which the directional flux of precipitated electrons has been higher than the corresponding value of trapped electrons. The mentioned observations were made fairly close to the earth's atmosphere.

The tendency to isotropy seems to indicate that the acceleration of the electrons—if it is directed along the field lines—takes place far away from the atmosphere. This, as well as the mentioned observations of precipitation of electrons from above 1000 km in auroras, suggest that the role of the ionosphere in the production of the energetic electrons is not important. O'Brien (1964) found a flux upwards along the field lines, which was some 10 percent of the precipitated electron flux. He interpreted these observations as backscattering of electrons from the atmosphere.

(d) K_p Dependence of Precipitation

The K_p dependence of the precipitation of electrons of energy above 40 kev is illustrated in Fig. 4.3. The average intensity was ten times higher near the auroral zone when K_p was above 4 than when it was below 4. The data material for the higher K_p range in Fig. 4.3 does not contain data from any strong magnetic storm, so still higher values may be expected.

O'Brien (1964) found that the flux of precipitated electrons above 40 kev increased on the average by a factor of 5 for every step of K_p. A close correlation between precipitation intensity was also observed for the 0.8–24 kev energy range by Sharp et al. (1964). O'Brien's results, obtained on Injun 3 in a low orbit, are shown in Fig. 4.6

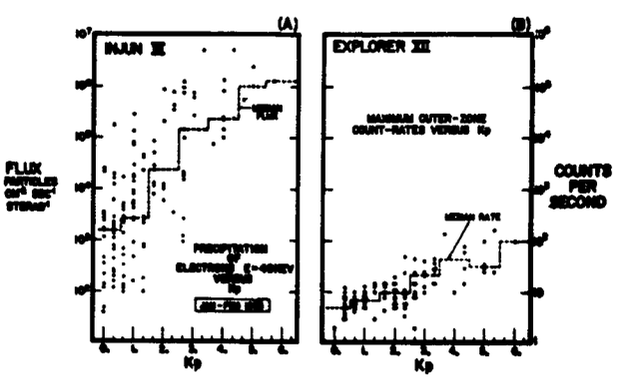


FIGURE 4.6.—Illustration that the flux of precipitated electrons (in A) varies more with K_p than does the omnidirectional flux (mainly of trapped electrons) in the equatorial plane (in B). Each point shows the maximum respective flux encountered on an outer-zone pass. (After O'Brien, 1964.)

together with the dependence of the omnidirectional flux above 40 kev as observed in the eqautorial plane far from the earth by Explorer 12 (Freeman, 1963). As can be seen, the K_p dependence is very much larger close to the earth (at one end of the field line) than in the equatorial plane (at the middle of the field line). If one assumes that both increases are due to a common acceleration mechanism it follows that it acts preferentially parallel to the geomagnetic field lines (O'Brien, 1964).

The change of omnidirectional flux above 40 kev with K_p, shown in Fig. 4.6b, is opposite to what has mostly been observed for electrons of $E \gtrsim 2$ Mev (Arnoldy et al., 1960; Hoffman et al., 1962).

The dependence of the poleward boundary of precipitation on magnetic activity has been studied by Maehlum and O'Brien (1963). They used data for trapped electrons, but since O'Brien (1964) has shown that the precipitation has its maximum close to the poleward boundary of the region of trapped electrons, their results may be interpreted in terms of possible extension of precipitated electrons.

Maehlum and O'Brien (1963) found that during magnetic storms there was a very sharp boundary of the region where trapped electrons of energy above 40 kev could be observed at about 1000 km altitude. For this boundary, measured in L, they used the symbol L_n. It can be seen as a function of the K_p index during one geomagnetic storm in Fig. 4.7. When K_p reached its maximum value of 9, L_n had its minimum value of 4. Maehlum and O'Brien (1963) also found that the poleward edge of strong radio wave absorption followed the L_n quite closely during the storm.

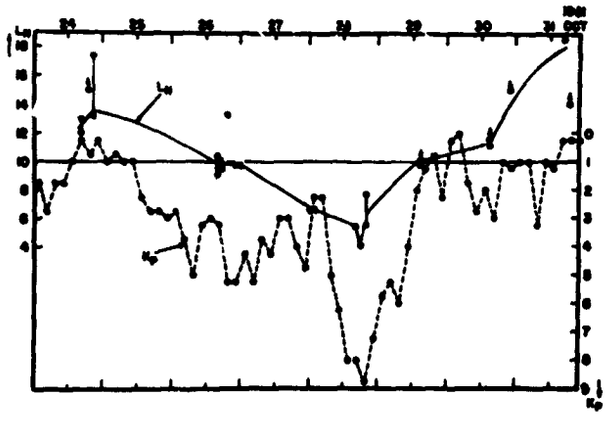


FIGURE 4.7.—Variation of the northern boundary of trapping with K_p during a geomagnetic storm, as observed by Injun 1 at a height of about 1000 km. (After Maehlum and O'Brien, 1963.)

The effect of K_p on L_n is similar to the equatorward movement of the region of visual aurora during strong magnetic storms, which has been studied in the last few years in detail for some storms by Akasofu (1962, 1963a,b) Akasofu and Chapman (1962) and Davis and Kimball (1962).

(e) The Spectrum of Precipitated Electrons

Up to now only rough measurements of the energy spectrum of the precipitated electrons have been reported. In most cases the proposed spectra have been obtained from two instruments with different energy characteristics. They thus are to be considered as rough equivalent spectra for measurements made in a defined way. In addition, it has been found that the equivalent spectrum is highly variable both in space and time (cf. O'Brien et al., 1962). Nonetheless the available spectral data are of great interest at the present stage of knowledge in the field, and they even make it possible to draw some interesting conclusions about ionospheric effects of the precipitated electrons (see Section 3).

Most spectra that have been reported hitherto can be divided in two categories if they are expressed in an exponential form. On one hand, the e-folding value, b, in the spectrum of precipitated electrons, written in the form $\alpha \exp(-E/b)$, has been found often to be in the range 2–8 kev (McIlwain, 1960; Stilwell, 1963; Sharp et al. 1964 a,b,c.) This is a very steep spectrum, but even steeper ones have been observed. McIlwain interpreted his rocket measurements in a strong aurora as indicating a monoenergetic flux of electrons with an energy of about 6 kev. Krassovskii et al. (1962) has observed steep spectra for precipitated electrons with a most common equivalent energy value of 14 kev.

The majority of the existing measurements have been made with Geiger tubes, for which the minimum detectable electron energy is about 40 kev. These measurements have mostly given e-folding values much higher than those mentioned above namely between 20 and 45 kev. (Davis et al. 1960; McDiarmid et al. 1960; O'Brien et al. 1962; Mann et al. 1963; McDiarmid et al. 1963; O'Brien and Taylor, 1964). Mann et al. (1963) found

from measurements during only some 20 orbits that the values within this range were grouped in two classes, 25 ± 5 kev and 42 ± 3 kev.

The balloon measurements of x-rays at 30–35 km altitude also, in general, give spectra of a fairly flat type, corresponding to e-folding energies in the range 20–45 kev mentioned (see e.g., Anderson and Enemark, 1960), although power law spectra often are found to fit the observations somewhat better than exponential spectra.

Recently the first direct measurements of precipitated electrons below 1 kev energy have been reported (Sharp et al. 1963, 1964c; Evans et al. 1964). They showed that there is generally not very much energy flux below 1 kev. This is also evident from observations of the luminosity distribution with height in aurora. (O'Brien and Taylor, 1964). It therefore seems probable that the spectrum, in general, does not increase very fast below 1 kev, which means that it is not of the power law type at these low energies.

Sharp et al. (1964c) reported a significantly harder spectrum on the dayside than on the nightside of the earth. This observation agrees with multifrequency riometer measurements of auroral absorption by Lerfald et al. (1964) according to which the absorbing ionization is located lower down in the atmosphere in the day than in the night.

Of special interest are the statistical data on electron precipitation, published recently. As mentioned earlier the precipitation is most intense near the auroral zone. O'Brien (1964) found an average flux of 4.10^5 electrons/cm^2 sec in the precipitation cone at an invariant latitude of 65°. Injun 1 had a CdS detector, measuring electrons of energy above about 1 kev (O'Brien, 1962b). O'Brien and Taylor (1964) stated that the flux of 4.10^5 electrons/cm^2 sec of energy greater than 40 kev was associated with an energy flux of 4 erg/cm^2 sec for electrons of $E \gtrsim 1$ kev. These data should be considered as accurate to a factor of about three. Using these two values an equivalent electron spectrum in the range 1–40 kev can be duduced. It is found to be $n(E) = 7.8 \cdot 10^7 e^{-E/5.7}$ electrons (cm^2 sec kev)$^{-1}$

For the energy values 40 and 250 kev similar average fluxes have been obtained from the Alouette measurements reported by McDiarmid et al. (1963). While the flux above 40 kev had its maximum at an invariant latitude of 65 degrees, the flux above 250 kev was maximal at about 60 degrees. McDiarmid et al. (1963) did not present any details for the latitudinal variation of this latter integral flux. By applying a rough correction factor of 0.1 for the latitude variation of the average flux above 250 kev from 60 to 65 degrees invariant latitude the following approximate equivalent spectra for the energy range 40–250 kev can be derived for two ranges of magnetic activity in the auroral zone, if isotropy is assumed for the upper hemisphere:

$$K_p < 4 : n(E) = 1.2 \cdot 10^4 e^{-E/41}$$

electrons (cm^2 sec kev)$^{-1}$,

$$K_p > 4 : n(E) = 2.4 \cdot 10^5 e^{-E/30}$$

electrons (cm^2 sec kev)$^{-1}$

(cf. Hultqvist 1964b). The applied latitude correction is probably too large and the resulting spectra are thus too steep rather than too flat. As mentioned, the spectra of precipitated electrons are very variable. There is also an important latitude variation in the steepness which increases with latitude (O'Brien et al. 1962; McDiarmid et al. 1963; O'Brien, 1964). On the other hand there seems to be no significant dependence of the spectrum on pitch angle or on the intensity of precipitation (O'Brien, 1964). The averaging on which the equivalent spectra given above are based has been made over a large number of measured fluxes. With a variation of a factor of 10^5 the use of averages may be questioned. It is also questionable whether average integral fluxes are the best values to use when the interest is in the average electron density produced by the particle influx. The average electron density is then defined by the averaging process, and this method of averaging is not identical with the averaging made in measuring the ionospheric absorption, for instance. It can, however, be shown that the average electron density produced by the precipitated electrons can be expressed in the average integral fluxes measured on satellites (Hultqvist, 1964). Since only two experimental integral flux values are available one has to fit a two-parameter energy relation for the average flux to these experimental data, as has been done above.

3. Observed Fluxes and Energy Spectra of Protons Precipitated Into the Atmosphere

The number of observations of protons on their way down into the atmosphere is much less than that for electrons. Some rocket results observed in passing through aurora or close to distinct auroral forms, or from launchings made under quite disturbed upper atmosphere conditions with no auroral information available, are collected in Table 2.

While electrons were observed only within the auroral form in the case of no. 1a, protons were found over much larger volumes. In the flights 1b, c, and d, when no auroral form was penetrated and no electrons were recorded, protons (and/or heavier ions) were found in appreciable amounts. The proton precipitation thus seems to have that wide spread and diffuse character that the hydrogen emissions generally show. This may possibly be due to protons taking up an electron also well outside the atmosphere. As a neutral atom it is not locked to the magnetic field lines and the particle flux may be spread out before reaching the atmosphere (Shklovskii, 1958). In all cases of Tables 1 and 2 in which both electrons and protons were measured, the energy flux of the electrons was about two orders of magnitude, or more, greater than the proton energy flux.

The satellites Injun 1 and Injun 3 did not measure any auroral event in which the electrons were not the dominant constituent in radiation of a given penetrability (O'Brien and Taylor, 1964). The possibility that there were as many protons/cm² sec with an energy about 40 kev, say, as there were electrons/cm² sec with the same energy could, however, not be excluded on the basis of the Injun measurements. Such protons would be stopped above the normal auroral heights and one would expect hydrogen emissions well above the ordinary aurora. They could, however, be quite weak, especially as the protons flux generally is spread out over a much larger surface than the electron flux, as mentioned above. It seems difficult to judge on these things on the basis of ground observations, as the hydrogen emissions are mostly diffuse and therefore unsuitable for height determination with standard ground techniques. Rocket measurements of the height distribution of the hydrogen emissions would be valuable.

4. Quantitative Relations Between Electron Precipitation and Photon Emission in Aurora

There are two direct measurements reported of the quantitative relations between precipitation and photon emission which will be discussed here, namely those of McIlwain (1960) and O'Brien and Taylor (1964). It is of interest to compare their results with what is expected on theoretical grounds.

McIlwain did not use any filter, but measured the photon flux integrated over the transmission curve of the photomultiplier. If we assume that ⅕ of the light was λ3914 Å, the measured photon flux corresponds to an emission rate of about 3 kilorayleigh at this wavelength. The fraction ⅕ is fairly arbitrary, but it seems reasonable (cf. e.g. Dalgarno, 1964) and is probably in error by less than a factor of 2 and ½, respectively, for an ordinary auroral-zone aurora. McIlwain's (1960) measured electron flux corresponds to an energy flux of 20 erg/cm² sec if the spectrum obtained was extrapolated to $E=0$. Thus, the resulting electron energy flux required per unit of λ3914 Å emission rate is 7 ergs cm^{-2} sec^{-1} per kR.

O'Brien and Taylor (1964) reported an average λ3914 Å intensity of $2\binom{+4}{-1.5}kR$ at the maximum of the latitude distribution i.e. in the auroral zone. As mentioned before the electron measurements gave a corresponding average of the electron flux above 40 kev of $4\cdot10^5$cm^{-2} sec^{-1} and an average energy flux for $E\gtrsim1$ kev of about 4 ergs cm^{-2} sec^{-1}, which values they consider as accurate to a factor of about three. In order to get a value directly comparable with that of McIlwain one would have to extrapolate the energy spectrum from 1 kev down to $E=0$ and evaluate the total energy flux. Using the equivalent exponential spectrum, we find 5 ergs cm^{-2} sec^{-1} for $E>0$. Thus the Injun 3 results give 2.5 ergs cm^{-2} sec^{-1} per kR. Considering the uncertainties in measurements and the method of evaluation, this is a good agreement.

What energy flux per kR does one expect on the basis of existing knowledge about the emission processes? Omholt (1957, 1959) Chamberlain (1961), Rees (1963) and Dalgarno (1964), among others, have discussed this. It has been shown by Stewart (1956) that the ratio between the

TABLE 2.—*Precipitated Protons (and other ions)*

No.	Reference	Flux (above the atmosphere)		Spectrum or energy range	Time of observ.	Place of observ. Geomagn. lat.	Visible aurora intensity and location with regard to part. measurement	Magn. disturb.	Radio-wave absorp.	Method of observ.	Remarks
		Number of particles	Energy								
1a	Davis et al. (1960)	$1.2 \cdot 10^5$ (cm² sec ster)$^{-1}$ $100 < E < 800$ kev	$3 \cdot 10^{-3}$ erg (cm² sec ster)$^{-1}$	$N(>E) = E^{-2.2}$	Jan. 26, 1958 0421 UT	Churchill, $\Lambda = 68.8°$	Passed through fading rayed structure during ascent and through diffuse surface of IBC 1 on both ways.			Rocket Scintill.	Definitely protons.
1b	Davis et al. (1960)	$4 \cdot 10^5$ (cm² sec ster)$^{-1}$ $100 < E < 800$ kev	$1.5 \cdot 10^{-3}$ erg (cm² sec ster)$^{-1}$	$N(>E) \propto E^{-1}$	Mar. 16, 1958 0454 UT	Churchill, $\Lambda = 68.8°$	No penetrat. of aurora. Weak arc moved to the north.			Rocket Scintill.	
1c	Davis et al. (1960)	$\sim 3 \cdot 10^5$ (cm² sec)$^{-1}$ ster)$^{-1}$ $100 < E < 800$ kev	$\sim 1.2 \cdot 10^{-3}$ erg (cm² sec ster)$^{-1}$	$N(>E) \propto E^{-1.7}$	Mar. 22, 1958 0641 UT	Churchill, $\Lambda = 68.8°$	No penetrat. of aurora. Fading arc moved to the south.			Rocket Scintill.	
1d	Davis et al. (1960)	$8.5 \cdot 10^5$ (cm² sec ster)$^{-1}$ $100 < E < 800$ kev	$2 \cdot 10^{-3}$ erg (cm² sec ster)$^{-1}$	$N(>E) \propto E^{-2.2}$	Nov. 16, 1958 0658 UT	Churchill, $\Lambda = 68.8°$	No penetrat. Bright arc just to the north.			Rocket Scintill.	
2a	McIlwain (1960)	$2 \cdot 10^5$ (cm² sec ster)$^{-1}$ for $E > 80$ kev		$N(>E) = 2.5 \cdot 10^5 \cdot \exp(-E/30)$ (cm² sec ster)$^{-1}$ for $80 < E < 200$ kev	Feb. 22, 1958 0537 UT	Churchill, $\Lambda = 68.8°$	Passed through a faint glow			Rocket Scintill.	Most probably protons.
2b	McIlwain (1960)	$< 4 \cdot 10^5$ (cm² sec ster)$^{-1}$ for $E > 100$ kev			Feb. 22, 1958 0551 UT	Churchill, $\Lambda = 68.8°$	Passed through bright arc (after break up of aurora)				
3	McDiarmid et al. (1962)	$3 \cdot 10^5$ cm^{-2} sec^{-1} for $E > 500$ kev			Oct. 28, 1960 1223 UT	Churchill, $\Lambda = 68.8°$		$\Delta Z > 1000\gamma$	~3 db (30 Mc/s)	Rocket GM tube	

excitation cross section for the λ3914 Å band of N^{2+} and the total ionization cross section is constant at least up to 200 ev energy and has the value 0.02. Assuming that this value is true over the whole energy range of interest, one finds that 50 electrons-ions pairs are produced for each 3914 Å photon. Since the mean energy expended by fast electrons in nitrogen per electron ion pair is 35 ev (at least for energies down to a few hundred ev; it is assumed the figure is correct down to zero energy) we find that $50 \times 35 = 1750$ ev is dissipated per 3914 Å photon. Since each photon has an energy of 3.2 ev, the efficiency with which energy is converted into 3914 Å radiation is $1.80 \cdot 10^{-3}$. When the initial energies of the fast electrons fall below perhaps 100 ev, these efficiencies must decrease sharply (Dalgarno and Griffing, 1958; Dalgarno, 1964). 1*kR* of 3914 Å photons corresponds to an energy influx of 2.8 erg/cm² sec, if the efficiency figure $1.8 \cdot 10^{-3}$ is employed. Dalgarno (1964) has used the value $1 \cdot 10^{-3}$. With this conversion efficiency we find the energy flux requirement to be 5.1 erg cm^{-2} sec^{-1} per *kR*. The agreement between these values and the experimental ones of McIlwain (1960) and O'Brien and Taylor (1964) is much better than expected, when the uncertainties in the analysis are taken into account.

The total efficiency of converting particle energy into photons of any energy is expected to be about 1 percent on theoretical basis (cf. Chamberlain, 1961). McIlwain found from his rocket measurements a value of only 0.2 percent, while preliminary Injun 3 results point towards 1 percent (O'Brien and Taylor, 1964).

5. Relation of Precipitated Electrons to the Outer Radiation Belt

The idea about the relations between the electrons that are precipitated into the atmosphere in auroras and elsewhere and the trapped radiation in the radiation belts, which was prevailing one or two years ago, was roughly that the precipitated particles were dumped from the large storage of energetic trapped particles in the magnetosphere through the influence of disturbing effects caused by the solar plasma, as, for example the magnetic disturbances. The number of trapped particles was thought to be sufficiently large to allow the precipitation rates observed to occur with only weak and perhaps independent processes of injection and acceleration required to replenish the particles in the radiation belt, and the occurrence of which is necessary to be assumed in any case for understanding the existence of the trapped radiation. This picture has been completely changed recently, through the investigations by O'Brien (1962b, 1964) of the measuring results of Injun 1 and Injun 3 and also through a number of balloon studies of x-ray fluxes in the lower atmosphere (Winckler et al., 1962; Anderson, 1964).

O'Brien (1962b) estimated the lifetime of the trapped electrons in the outer radiation belt assuming that the source was stopped but the loss mechanisms operated at the same rate as observed by Injun 1. He found that the outer zone beyond $L \sim 2$ would drain empty of electrons in a few hours. Similar average lifetimes have also been evaluated from x-ray measurements (cf. e.g. Winckler et al. 1962 and Anderson, 1964). Sometimes precipitation rates several orders of magnitude above the average (more than 1000 ergs/cm² sec) have been observed (Krasovskii et al. 1961; O'Brien and Laughlin, 1962, Winckler et al., 1962). Thus Winckler et al. (1962) recorded an electron burst reaching 10^{11} electrons $(cm^2\ sec)^{-1}$, a flux which would have used up the total energy of trapped particles in the field tube in half a second, but yet it persisted for about 100 seconds.

O'Brien (1964) has also shown that the flux of trapped radiation increases when precipitation takes place instead of diminishing, which would be expected if the trapped particles were simply dumped into the atmosphere. An acceleration mechanism seems to influence precipitated and trapped particles simultaneously.

Finally, O'Brien (1964) has demonstrated that the precipitation is highly energy dependent. There was no significant (<10 percent) precipitation or change in the flux of electrons of energy greater than 1.5 Mev observed in the middle of a strong burst of electrons of energy above 40 kev. This demonstrates that the precipitation could not be due simply to a lowering of the mirror point through a decrease in the geomagnetic field, since such a mechanism would be active over the whole spectrum.

Taken together, this new evidence clearly demonstrates that the precipitated electrons which produce aurora and ionospheric ionization are not

produced by simple dumping of trapped electrons into the atmosphere. An acceleration machanism must be involved. This mechanism must be one that can act with full strength very quickly (in a fraction of a second) and it should not change the flux of Mev electrons more than 10 percent when the simultaneous changes for trapped electrons of energy above 40 kev is a hundredfold, and of the precipitated electron flux above 40 kev is more than three orders of magnitude.

5. THEORETICAL MODELS

1. Introduction

In the last few years many new observational facts have been added to the picture of the aurora. Important new information has been obtained primarily from satellites, but also through detailed synoptic studies of individual auroral storms on the basis of allsky camera recordings taken during IGY.

Some important examples of such new information, some of which have been mentioned earlier in this review, are the following. Electrons are precipitated into the atmosphere over a very wide latitude range, with the maximum intensity in the auroral zone where some precipitation always takes place. The flux of trapped electrons of energy above 40 kev increases simultaneously with the precipitated flux. The increase is such that isotropy over the upper atmosphere is approached. The precipitated flux has never been found to be greater than the trapped flux.

The precipitation mechanism has been found to be energy dependent. During strong precipitation of electrons of energy greater than 40 kev the flux of Mev electrons was not significantly changed. During strong storms the poleward boundary, L_N, of the precipitation zone is very distinct and may be located at L-values as low as 4. During the quiet phase no aurora is seen on the poleward side of a certain L-value, probably identical with the L_N boundary. Following the break up, the aurora may extend rapidly polewards from the pre-break-up position as far as 10° of latitude or even more. Then it returns slowly toward the low L boundary again. It has also been definitely established that the energy content of the particles precipitated into the atmosphere in a period of the order of a minute may be orders of magnitude greater than the total energy content of the particles trapped in the corresponding field tube. Both virtually monoenergetic electrons and exponential and power law spectra have been observed in aurora. Evidence has accumulated that the local properties of the geomagnetic field in the upper atmosphere is not very important for the aurora. This indicates that drift motion around the earth of the particles which at some stage are precipitated into the atmosphere is not an important process. It has been proven that discharge acceleration within the ionosphere is not necessary for auroral production (O'Brien, 1962b).

Special chapters in this book are devoted to the description of some theoretical models of the relations between solar disturbances and geophysical effects, thereamong aurora. The reader is referred to them for details about the models.

The development of our experimental knowledge about the aurora and of our understanding of the physical processes in it is very rapid at present. It is, therefore, not a suitable stage for summing up and comparing experimental data with theoretical models. In this section only a very short summary of some main lines of proposed theories will be given in connection with a comparison of the models with some observational data which seem to be significant today.

2. Some Specifications of the Theoretical Problem

There is a general agreement that auroras, as well as magnetic storms, are caused by plasma streams which are emitted from disturbance centers on the sun. The problem of the interaction between the solar plasma and the geomagnetic field is one of great complexity, which has not as yet been solved in quantitative detail in a rigorous way. Simplified models have been used and the broad picture obtained differs very much depending on the theoretical approach.

Observations by Mariner II (Snyder and Neugebauer, 1963) and other space vehicles have shown that the interplanetary space usually has a number density, N, of charged particles of the order 10 cm^{-3}, the equivalent temperature, T, obtained from the velocity dispersion of the plasma, for instance, is typically a few times 10^5 degrees Kelvin, there exists a magnetic field, B, of

the order of a few gamma, mostly directed about 45° from the sun-earth line in the equatorial plane (see e.g. Ness, 1964). This means that the mean free path for Coulomb collision is a few astronomical units (see e.g. Parker, 1962, Kellogg, 1962, and Alfvén and Fälthammar 1963). In the plasma beams from disturbance centers on the sun somewhat higher values (but less than one order of magnitude higher) of N, B and T have been observed. A plasma of the mentioned characteristics is a "low density plasma" when the system sun-earth is considered (cf. e.g. Alfvén and Fälthammar, 1963). For such a plasma the ordinary fluid theory cannot be applied, as the statistically defined parameters have no meaning. Especially, the electric conductivity is not defined and the hypothesis that the magnetic field lines are "frozen" into the plasma may possibly not be applicable. Electric fields may exist along the geomagnetic field lines (Persson, 1963, Alfvén and Fälthammar, 1963). No macroscopic theory for the interaction of a low density plasma with the geomagnetic field seems to have been developed.

The plasma contained in the magnetosphere is probably also a low-density plasma, at least during disturbed conditions according to Alfvén and Fälthammar (1963) and Alfvén et al. (1964).

Although the theoretical basis for fluid-dynamic models of the interaction between the solar plasma and the geomagnetic field, which have been proposed, thus must be considered as weak at the present time, some of the characteristics of these models agree well with observations. They all require that the geomagnetic field shall be enclosed in a finite region around the earth, the so-called magnetosphere. A discontinuity in the magnetic field has been observed by Explorers XII, XIV and XVIII (Cahill and Amazeen, 1963; Freeman et al., 1963; Ness, 1964) at distances of about 10 earth radii on the side facing the sun, where it is expected to be according to the fluid-theory models.

The discussion below will be limited to a comparison of some observations with results of two fundamentally different types of theoretical approach, namely the fluid-dynamic (macroscopic) one, leading to a closed magnetosphere, and a microscopic model in which the plasma penetrates into the geomagnetic field to some extent.

3. Closed Magnetosphere Models

Chapman and Ferraro (1930, 1931, 1932, 1933) first discussed a model containing a confinement of the geomagnetic field lines in a finite volume around the earth. They considered a neutral unmagnetized plasma beam approaching the earth with a velocity of about 1000 km/sec and applied the boundary condition that the magnetic pressure is equal to the dynamic pressure of the plasma. Their model did not contain any details about the generation of aurora. The basic concept treated by Chapman and Ferraro has in recent years been further developed. Accurate calculations of the shape of the boundary of the magnetosphere have been made, the most recent and accurate one being that of Mead and Beard (1964). Extensive qualitative models of the geophysical effects caused by the solar plasma have been worked out on the basis of the Chapman-Ferraro model. The most comprehensive one is that of Axford and Hines (1961). In this model the acceleration of the particles is due to the plasma in the interface between magnetosphere and solar beam being convected into the interior of the magnetosphere, during which process it is adiabatically compressed. The theory is described in detail in Chapter

The closed magnetosphere models are characterized by a distinct boundary between the magnetosphere and the interplanetary medium. No model contains any details about how particles can enter the magnetosphere. For stable flow no particles can pass the boundary, except possibly at the neutral points. As the energy of the particles precipitated into the atmosphere with all probability must originate in the solar beam, one has made hypothesis about how the energy transport takes place over the boundary. Magnetohydrodynamic waves have been proposed. Axford and Hines assumed that some type of viscous interaction takes place at the boundary of the magnetosphere. Other types of instabilities have also been invoked (cf. e.g. Gold, 1962). It seems to be common to all of the proposed ways of energy transport over the magnetospheric boundary that only a minor fraction of the energy content in the solar wind is transferred to the magnetospheric plasma.

The existing fluid theories all are based on the assumption that the magnetic field lines are

equipotential lines and are "frozen" into the plasma. Therefore, acceleration of particles inside the magnetosphere must in these models be due to other effects than electrostatic fields along the lines of force. Of known acceleration processes the Fermi and betatron mechanisms and possibly electromagnetic radiation in the VLF band may be of importance in the magnetosphere (see Kaufmann, 1963, for a review). All these mechanisms require a large number of acceleration steps, and therefore fairly long time, and they produce a wide energy spectrum due to statistical fluctuations in the process. It has also been proposed that electric fields perpendicular to the magnetic field lines may cause acceleration of electrons and ions along neutral lines, i.e. lines where the geomagnetic field is zero (see Kaufmann, 1963). The existence of neutral lines has been proposed and their properties discussed by Dungey (1958, 1961, 1963) and Akasofu and Chapman (1961).

Since the discovery of the radiation belts and up till recently it was fairly generally believed that the trapped radiation was the source of those particles which produce the aurora, as was mentioned in the previous section. The density of trapped radiation was then thought to be three or four orders of magnitude greater than now (cf. e.g. O'Brien, 1963b). It was considered possible that some of the slow acceleration mechanisms mentioned above might be responsible for the acceleration of the particles from their energy in the solar wind to that observed inside the magnetosphere, although no quantitative theories existed neither for the particle transport into the magnetosphere nor for the acceleration of the particles in the outer radiation belt. An important unsolved problem was and is still how the electrons, which are most important in aurora, are energized, since most of the mentioned acceleration mechanisms give most energy in the ions.

The precipitation of the primary aurora particles into the atmosphere was thought to be produced by more or less transient instabilities of various kinds. A number of such were proposed by Akasofu and Chapman (1961), Chamberlain (1961, 1963), Kern (1962) and others. Other studies of possible auroral effects associated with adiabatically invariant motions of trapped charged particles, some of which are scattered into the atmosphere, were those of Chamberlain et al. (1960), Kern and Vestine (1961) and others.

The recent observations mentioned in the introduction to this section have completely changed the picture in regard to the importance of the trapped radiation as source for the aurora.

The fact that precipitated energies have been found sometimes to be several orders of magnitude greater than the total particle energy in the corresponding magnetic field tube (see previous section) shows that dumping without acceleration cannot be the important process for precipitation of electrons into the atmosphere, at least on some occasions. The observations of O'Brien (1964) that the flux of trapped particles always increases when precipitation takes place and is never less than the precipitated flux demonstrates clearly that the trapped radiation is not the source of the precipitated ones but is rather produced by the same acceleration mechanism that accelerates the precipitated electrons.

The observation of energy dependence of the precipitation mechanism (O'Brien, 1964) demonstrates that the precipitation is not due to a lowering of the mirror point by some geomagnetic disturbance, since such a lowering would affect the whole energy spectrum, contrary to what has been observed.

The acceleration mechanism responsible for the production of trapped as well as precipitated particles according to this mentioned picture is unknown. An electrostatic voltage of the order of 10 kilovolt along the magnetic field lines was mentioned as a possible source of acceleration by O'Brien (1964). That electrostatic field acceleration is important in aurora is supported by the observation by McIlwain (1960) of monoenergetic electrons of 6 kev energy in strong aurora. In the closed magnetosphere models that have been presented electrostatic fields along magnetic field lines cannot exist. The mechanism must therefore be an unknown one.

In the described "splash-catcher" model of O'Brien (1962b, 1964) the precipitated electrons are "fresh," i.e. they have existed with the energy they have at precipitation only for a very short time. This means that drift motion around the earth of the source of auroral primaries is not important, which is in agreement with the earlier

mentioned independence of the isoauroral curves on the mirror height.

The recent observations described above have thus modified the early closed magnetospheric models considerably. Some of the observations are even difficult to accommodate in such models. This is true for the very high energy fluxes of precipitated electrons which have been observed. While, in the average, only of the order of one percent of the solar wind energy has to be converted into kinetic energy of precipitated electrons (O'Brien and Taylor, 1964) an appreciable fraction of the solar wind energy has to be given off in electron precipitation during strong magnetic storms. This is difficult to associate with a closed magnetosphere, where energy is transported over the boundary by second order effects.

Another important difficulty with models of the closed magnetosphere type which have been proposed hitherto, is that the solar plasma is supposed to be unmagnetized, whereas observations by a number of space vehicles have shown that the interplanetary plasmas are magnetized. Alfvén (1964) has pointed out that the difference between the two cases with unmagnetized and magnetized solar plasma is more than a question of mathematical method. As the magnetic flux passing through a certain mass of gas must be fairly constant during the motion, a plasma with an initial magnetization may be brought into a stronger field if it is compressed. Thus a plasma with any value of the initial magnetization may be brought into a strong magnetic field. If the initial magnetization is zero this is not the case. The assumption $B=0$ leads therefore to conclusions which are fundamentally different from the assumption $B\neq 0$ even if B is very near zero.

The entire surface of the closed magnetosphere maps into one neutral point in each hemisphere. These are not in the auroral zones and there are no obvious reasons for the precipitation of particles taking place preferably in the auroral zones. Some complicated mechanism, like the one devised by Axford and Hines (1960), is needed to get the aurora in the proper regions of the earth.

The topology of the magnetospheric field lines is very complicated in some regions of the closed models. Aurora shows a fairly high degree of order on the earths surface with arcs often being aligned along L-curves (see e.g., Akasofu, 1963; Akasofu and Chapman, 1962 and Hultqvist, (1962). This corresponds to high degree of order also in the equatorial plane with the processes taking place at about constant distance from the earth, if the field is axisymmetric. If such an auroral arc is projected to the equatorial plane along the field lines in the closed magnetospheric model, the equatorial cross section of which is shown in Fig. 5.1, curves are found which seem to have no relation to simple physical processes. They differ significantly from particle drift surfaces, for instance (Hones, 1964).

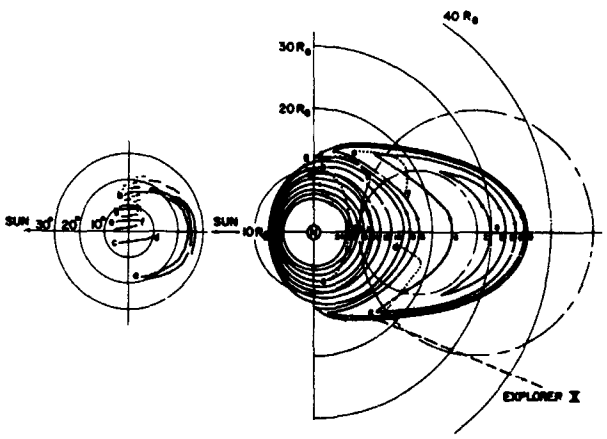


FIGURE 5.1.—Projections of auroral arcs along the geomagnetic field lines to the equatorial plane in a closed magnetosphere model. The letters a-h in the two figures indicate corresponding points in the earth's atmosphere and in the geomagnetic equatorial plane (Courtesy, Dr. E. W. Hones, Jr.).

4. Model with the Solar Plasma Penetrating into the Geomagnetic Field

Alfvén (1939, 1940, 1950, 1955; Alfvén and Fälthammar, 1963; Alfvén et al., 1964) has pointed out that a low-density magnetized plasma can penetrate into the geomagnetic field through the action of the electric field that is associated with a moving magnetic field. The depth of penetration is determined by the electric and magnetic fields and the temperature in the incoming plasma and not by the dynamic pressure as in hydromagnetic models. The penetration is stopped at a forbidden region, from which electrons and protons may be accelerated down into the atmosphere by an electric field. The acceleration field is thought to be due to charge separation,

which takes place during the invasion of the geomagnetic field. The particles from the forbidden zone moving along the field lines reach the auroral latitudes, if the properties of the plasma are in agreement with space vehicle observations. The particles may be accelerated to energies of tens or hundreds of kev.

In the first model of Alfvén it was assumed that no space charges are collected around the earth, but that immediate discharging takes place along the magnetic lines of force as soon as space charges start to build up. Karlsson (1963) has investigated the case of maximal space charge accumulation. He assumed that no discharging at all takes place along the field lines. In that way he found the general picture derived by Alfvén is preserved when space charges are taken into account. The forbidden zone is, however, smaller and more nearly circular than in the case studied by Alfvén. The general picture is thus not influenced very much by the amount of space charge accumulation.

There seems to be no basic difficulties in incorporating the above mentioned large fluxes of precipitated particles, observed by Winckler et al. (1962) and others, in the Alfvén model. This mechanism may well be active in the presence of simultaneous dumping of trapped particles due to disturbances of the geomagnetic field. For the mechanism to meet the requirements that O'Brien (1962b) has derived from the Injun 1 data, it is necessary that a rapid diffusion of particles scattered into trapped orbits takes place from the forbidden zone into the inner part of the radiation belt. If the quantitative requirements can be met is not known to this author.

Together with other large-scale charge-separation mechanisms, Alfvén's has the difficulty that electrons and protons are expected to be precipitated in two distinctly separated areas over the earth. Spectral observations from the ground as well as satellite observations of particle precipitation (Evans et al., 1964) show that electrons and protons often are precipitated simultaneously in the same auroral form.

Electrostatic acceleration of electrons and protons into the atmosphere is an inherent feature of Alfvén's model. The observations of O'Brien (1962b, 1964) thus fit well into his model in this respect. On the other hand it does not contain the observed discontinuity in the geomagnetic field at about 10 earth radii from the center of the earth, which have been observed by means of Explorers XII, XIV, and XVIII. Alfvén's model was made with the magnetic field of the solar plasma parallel to the earth's magnetic field in the equatorial plane. This direction was originally chosen in order to have the model fit observed diurnal variations in auroral occurrence. Dungey (1961, 1963) has discussed the case with the magnetic field of the solar plasma being antiparallel to the geomagnetic field in the equatorial plane. In that case neutral points or lines occur. The plasma does not penetrate further than to the neutral line and is precipitated from there along the field lines into the atmosphere. Recent laboratory experiments with magnetized plasma beams moving towards and interacting with a dipole magnetic field (Alfvén et al., 1963) have verified that there is a fundamental difference between the cases with parallel and antiparallel magnetic fields in the equator plane.

The observed discontinuity in the magnetic field may thus be due to Dungey's mechanism. Magnetic field measurements on space vehicles seem, however, to indicate that the interplanetary magnetic field lines mostly form only a small angle with the geomagnetic equatorial plane and the component parallel or antiparallel to the geomagnetic field lines is generally small and variable in magnitude and sign (Ness, 1964). In spite of this the discontinuity in the field seems always to have been observed by Explorer XVIII, at least on the side of the earth facing the sun. This does not seem to fit Dungey's model.

It is probable that within the near future space vehicles will provide experimental information that will make it possible to determine which of the two above-mentioned types of approach is relevant to nature—if any.

The theories discussed concern the large scale configuration of the auroral precipitation pattern. For the breakup of aurora and the extremely complicated phenomena that then occur on a global scale (see p. 49) no comprehensive theoretical models exist. The theories for this certainly will have to be sought in the realm of plasma instabilities, a field where the theoretical understanding of many basic phenomena still is poor.

ACKNOWLEDGEMENT

I am grateful to many colleagues at Goddard Space Flight Center for having read the manuscript, or part of it, and proposed improvements.

REFERENCES

ABBOTT, W. N., The aurora of August 19, 1950, photographed in Greece, J. Atm. Terr. Phys., 1, 343–344, 1951.

ABBOTT, W. N., Displacement of the radiant point during the auroral disturbance of September 22, 1957, Can. J. Phys., 36, 643–648, 1958.

ABBOTT, W. N. and S. CHAPMAN, On the aurora of August 19, 1950, photographed from Greece, J. Atm. Terr. Phys., 14, 111–122, 1959.

AKASOFU, S. I., J. Atm. Terr. Phys., 19, 10, 1960.

AKASOFU, S. I., J. Atm. Terr. Phys. 21, 287, 1961.

AKASOFU, S. I., Large-scale auroral motions and polar magnetic disturbances–II. The changing distribution of the aurora during large magnetic storms, J. Atm. Terr. Phys., 24, 723–727, 1962.

AKASOFU, S. I., The dynamical morphology of the aurora polaris, J. Geophys. Res., 68, 1667–1673, 1963.

AKASOFU, S. I., The auroral rays, J. Atm. Terr. Phys., 25, 163–165, 1963 b.

AKASOFU, S. I., Deformation of magnetic shells during magnetic storms, State University of Iowa, Report SUI, 63–16, 1963 c.

AKASOFU, S. I., The development of the auroral substorm. Paper presented at the American Geophysical Union, Third Western National Meeting, Boulder, Dec. 1963, 1963 d.

AKASOFU, S. I. and S. CHAPMAN, A neutral line discharge theory of the Aurora Polaris, Phil. Trans. Roy. Soc. London, A, 253, 359–406, 1961.

AKASOFU, S. I. and S. CHAPMAN, Large-scale auroral motions and polar magnetic disturbances–III. The aurora and magnetic storm of February 11, 1958, J. Atm. Terr. Phys. 24, 785–796, 1962.

AKASOFU, S. I. and S. CHAPMAN, The lower limit of latitude (US sector) of northern quiet auroral arcs, and its relation to Dst (H), J. Atm. Terr. Phys., 25, 9–12, 1963.

ALFVÉN, H., Theory of magnetic storms, I. Kungl. Sv. Vet.–Akademiens Handl. (3), 18, No. 3, 1939.

ALFVÉN, H., Theory of magnetic storms. II, III. Kungl. Sv. Vet.–Akademiens Handl. (3), 18, No. 9, 1940.

ALFVÉN, H , "Cosmical Electrodynamics.", The Clarendon Press, Oxford, 1950.

ALFVÉN, H., On the electric field theory of magnetic storms and aurora, Tellus 7, 50–64, 1955.

ALFVÉN, H., Personal communication, 1961.

ALFVÉN, H., Hydromagnetics of the magnetosphere, Space Science Reviews (in print), 1964.

ALFVÉN, H. and C. A. FÄLTHAMMAR, Cosmical electrodynamics (2nd. ed.) The Clarendon Press, Oxford, 1963.

ALFVÉN, H., L. DANIELSON, C. A. FÄLTHAMMAR, and L. LINDBERG, On the penetration of interplanetary plasma into the magnetosphere, Paper presented at the Plasma Space Science Symposium at the Catholic University, Washington, D.C., 1963.

ALFVÉN, H., L. DANIELSON, C. A. FÄLTHAMMAR, L. and LINDBERG, On the penetration of interplanetary plasma into the magnetosphere (in print), 1964.

ANDERSON, K. A., Balloon observations of x-rays in the auroral zone I, J. Geophys. Res. 65, 551–564, 1960.

ANDERSON, K. A., Relation of balloon x-rays to visible auroras in the auroral zone, J. Phys. Soc. Japan, 17, Suppl. A–I, 237–241, 1962.

ANDERSON, K. A., Balloon measurements pertaining to the dynamics of the outer radiation belt above the auroral zone. Paper presented at the Lockheed Symposium on Auroras, Palo Alto, January, 1964.

ANDERSON, K. A. and D. C. ENEMARK, Balloon observations of X-rays in the auroral zone II, J. Geoph. Res., 65, 3521–3538, 1960.

ANDERSON, K. A. and R. DEWITT, Space-time association of auroral glow and X-rays at balloon altitude, J. Geoph. Res., 68, 2669–2675, 1963.

ANSARI, Z. A., The spatial and temporal variations in high latitude cosmic noise absorption and their relation to luminous aurora, University of Alaska Sci. Rep., No. 4, NSF Grant No. G14133, 1963.

ARNOLDY, R. L., R. A. HOFFMAN, and J. R. WINCKLER, Observations of the Van Allen Radiation Regions during August and September, 1959, Part I, J. Geoph. Res., 65, 1361–1376, 1960.

ARNOLDY, R. L., R. A. HOFFMAN, J. R. WINCKLER, and S. I. AKASOFU, Observations of the Van Allen radiation regions during August and September, 1959, J. Geoph. Res. 67, 3673–3686, 1962.

AXFORD, W. T and C. O. HINES, A unifying theory of high-latitude geophysical phenomena and geomagnetic storms, Can. J. Phys., 39, 1433–1464, 1961.

BARBIER, D., Ann. Geophys., 14, 334, 1958.

BARBIER, D., Introduction à l'étude de la luminesence atmosphérique et de l'aurore polaire. In Geophysics, the Earth's environment (Eds. DeWitt, Hieblot and Lebeau) Gordon and Breach, New York pp. 303–372, 1963.

BARTELS, J. and S. CHAPMAN, Geomagnetic activity data, in connection with the aurora in lower latitudes. Akad. der Wiss-Göttingen, Beiträge zum Internationalen Geophysikalischen Jahr, Heft 1, 1958.

BASLER, R. P., Radio wave absorption in the auroral ionosphere, J. Geophys. Res., 68, 4665–4681, 1963.

BATES, D. R., The auroral spectrum and its interpretation, in "Physics of the upper atmosphere" (Ed. Ratcliffe) pp. 297–353, Academic Press, Inc., New York, 1960.

BENSON, R. F., Effect of line-of-sight aurora on radio star scintillations, J. Geophys. Res., 65, 1981–1985, 1960.

BERGER, S., Giant pulsations in the magnetic field and pulsating aurora, Planet Space Sci., 11, 867–868, 1963.

BHATTACHARYYA, B. K., Can. J. Phys., 39, 350, 1961.

BHAVSAR, P. D., Scintillation-counter observations of auroral X-rays during the geomagnetic storm of May 12, 1959, J. Geoph. Res., 66, 679–690, 1961.

BLESS, R. C., C. W. GARTLEIN and D. S. KIMBALL, East-west motions in the aurora, Astrophys, J., 122, 205–206, 1955.

BLESS, R. C., C. W. GARLTEIN and D. S. KIMBALL, and G. SPRAGUE, Auroras, magnetic bays, and protons, J. Geoph. Res., 64, 949–953, 1959.

BOLLER, W., Das Südlicht, Beitr. angew. Geophys., 3, 56–130, 1898.

BOND, F. R., Austral. J. Phys., 13, 477, 1960.

BOND, F. R., and F. JACKS, Distribution of auroras in the southern hemisphere, Austral. J. Phys., 13, 611–612, 1960.

BOWLES, K. L., Doppler shifted radio echoes from aurora, J. Geoph. Res., 59, 553–555, 1954.

BROWN, R. R., Balloon observations of auroral-zone X-rays, J. Geoph. Res., 66, 1379–1388, 1961.

BROWN, R. R., Features of the auroral electron energy spectrum inferred from observations of ionospheric absorption. Arkiv för Geofysik (in print), 1964.

BUNEMAN, O., Excitation of field aligned sound waves by electron streams, Phys. Rev. Letters, 10, 285–287, 1963.

BURDO, O. A., Canadian Defence Research Board Translation by E. R. Hope T321 R (1959), 1955.

CAHILL, L. J., and P. G. AMAZEEN, The boundary of the geomagnetic field, J. Geophys. Res., 68, 1835–1843, 1963.

CAMPBELL, W. H., Magnetic micropulsations and the pulsating aurora, J. Geoph. Res., 65, 784, 1960 a.

CAMPBELL, W. H., Magnetic micropulsations, pulsating aurora, and ionospheric absorption, J. Geoph. Res., 65, 1833–1834, 1960 b.

CAMPBELL, W. H., Natural electromagnetic energy below the ELF range, J. Res. Nat. Bur. Standards, 64D, 409–411, 1960 c.

CAMPBELL, W. H., and B. NEBEL, Micropulsation measurements in Alaska, Nature, 184, 628, 1959.

CAMPBELL, W. H., and H. LEINBACH, Ionospheric absorption at times of auroral and magnetic pulsations, J. Geoph. Res., 66, 25–34, 1961.

CAMPBELL, W. H., and J. M. YOUNG, Auroral-zone observations of infrasonic pressure waves related to ionospheric disturbances and geomagnetic activity, J. Geophys. Res., 68, 5909–5916, 1963.

CAMPBELL, W. H., and M. H. REES, A study of auroral coruscations, J. Geoph. Res., 66, 41–55, 1961.

CARLHEIM-GYLLENSKIÖLD, Expl. int. polaires 1882–83. Exp. Suédois. Aurore boreale, Stockholm, 1886.

CHAMBERLAIN, J. W., and H. M. THORSON, The nightly variation of aurora at a subauroral station, J. Geophys. Res., 65, 133–136, 1960.

CHAMBERLAIN, J. W., Theory of auroral bombardment, Astrophys. J. 134, 401–424, 1961.

CHAMBERLAIN, J. W., Plasma instability as a mechanism for auroral bombardment, J. Geoph. Res., 68, 5667–5674, 1963.

CHAMBERLAIN, J. W., J. KERN, and E. H. VESTINE, Some consequences of local acceleration of auroral primaries, J. Geoph. Res. 65, 2535–2537, 1960.

CHAPMAN, S., The audibility and lowermost altitude of the aurora polaris, Nature 127, 341–342, 1931.

CHAPMAN, S., Low altitude aurora, Nature 130, 764–765, 1932.

CHAPMAN, S., Polar and tropical aurorae and the isoauroral diagram, Proc. Indian Acad. Sci. 37, 175–188, 1953.

CHAPMAN, S., The aurora in middle and low latitudes, Nature 179, 7–11, 1957 a.

CHAPMAN, S., Aurora in middle and low latitudes, Ann. IGY, Pt. II, 25–40, 1957 b.

CHAPMAN, S., Auroral observations in India and Pakistan, Bull. Natl. Inst. Sci. India, No. 9, 180–192, 1957 c.

CHAPMAN, S., and V. C. A. FERRARO, A new theory of magnetic storms, Nature, 126, 129–130, 1930.

CHAPMAN, S., and V. C. A. FERRARO, A new theory of magnetic storms, Terr. Magn., 36, 77–97, 171–186, 37, 147–156, 421–429, 38, 79–96, 1931, 1932, 1933.

CHAPMAN, S., and J. BARTELS, "Geomagnetism," Clarendon Press, Oxford, 1940.

CHREE, C., Aurora Polaris, in "Encyclopedia Britannica," 11th. edn., Vol. 2, p. 927, New York, 1911.

CHREE, C., Proc. Phys. Soc., 39, 389, 1927.

CHRZANOWSKI, P., G. GREEN, K. T. LEMMON, and J. M. YOUNG, Traveling pressure waves associated with geomagnetic activity, J. Geoph. Res. 66, 3727–3733, 1961.

CHRZANOWSKI, P., et al., Infrasonic pressure waves associated with magnetic storms, J. Phys. Soc. Japan, 17, Suppl. A-2, 9–13, 1962.

COLE, K. D., Proc. Antarctic Sympos., Buenos Aires, Nov. 1959.

COLE, K. D., Austral. J. Phys. 13, 484, 1960.

COLE, K. D., The directions of auroral rays, Austral. J. Phys. 16, 32–39, 1963.

COLE, K. D., Motions of the aurora and radio-aurora and their relationships to ionospheric currents, Planet. Space Sci., 10, 129–164, 1963.

CORTIE, A. L., Sunspots and terresterial magnetic phenomena, 1898–1911, The cause of the annual variation in magnetic disturbances, Monthly Nat. Roy. Astr. Soc. 73, 52–60 and 136, 1912.

CULLINGTON, A. L., A man-made or artificial aurora, Nature 182–1365, 1958.

CURRIE, B. W., Summary of some auroral height measurements and observations at Chesterfield, Terr. Magn. 39, 293–297, 1934.

CURRIE, B. W., Auroral heights over west-central Canada, Can. J. Phys. 33, 773–779, 1955.

CURRIE, B. W., and H. W. EDWARDS, Summary of some auroral height-measurements and observations at Chesterfield, Canada, Terr. Magn. 39, 293–297, 1934.

CURRIE, B. W., and C. K. JONES, Directional and diurnal characteristics of auroras at some places in Canada, Terr. Magn. 46, 269–278, 1941.

DALGARNO, A., Ann. de Géophys. (in print), 1964.

DALGARNO, A., and W. G. GRIFFING, Proc. Roy. Soc. A248, 415, 1958.

DAVIES, F. T., The diurnal variation in magnetic and auroral activity at three high-latitude stations, Terr. Magn. 40, 173–182, 1935.

DAVIES, F. T., Visual auroral observations in Canada, 1943–47, Bull. Intern. Assoc. Terr. Magn. Elec., Intern. Un. Geod. Geoph. No. 13, 255–273, 1950.

DAVIS, L. R., O. E. BERG, and L. H. MEREDITH, Direct measurements of particle fluxes in and near auroras, Space Research (Ed. H. K. Kallmann Bijl), pp. 721–735, North-Holland Publ. Co., Amsterdam, 1960.

DAVIS, T. N., An investigation of the morphology of the auroral displays of 1957–58, Geophysical Institute, College, Sci. Rep. No. I, NFS Grant No. G14782, 1961.

DAVIS, T. N., The morphology of the auroral displays of 1957–1958, I, Statistical analyses of Alaska data, J. Geoph. Res. 67, 59–74, 1962 a.

DAVIS, T. N., The morphology of auroral displays of 1957–1958, 2. Detail analyses of Alaska data and analyses of high-latitude data, J. Geoph. Res. 67, 75–110, 1962 b.

DAVIS, T. N., Negative correlation between polar-cap visual aurora and magnetic activity, J. Geophys. Res. 68, 4447–4453, 1963.

DAVIS, T. N., and D. S. KIMBALL, Incidence of auroras and their north-south motions in the northern auroral zone, Geophys. Insti. College, Alaska, Sci. Rept. No. 4, NSF Grant No. Y/22.6/327, 1960.

DAVIS, T. N., and D. S. KIMBALL, The auroral display of February 13–14. 1958, Univ. of Alaska, Sci. Rep. No. 2, NSF Grant No. 14782, 1962.

DAVIS, T. N. and R. N. DEWITT, Twenty-four hour observations of aurora at the southern auroral zone, J. Geophys. Res. 68, 6237–6241, 1963.

DAVIS, T. N., and G. T. HICKS, Cinema photography of variations in auroral forms with an image Orthicon Television System. Paper presented at AGU Third Western National Meeting, Boulder, 26–28 Dec. 1963.

DENHOLM, J. V., and F. R. BOND, Orientation of polar auroras, Austral. J. Phys. 14, 193–195, 1961.

DEWITT, R. N., The occurrence of aurora in geomagnetically conjugate areas, J. Geophys. Res. 67, 1347–1352, 1962.

DUDZIAK, W. F., D. D. KLEINECKE, and T. J. KOSTIGEN, Graphic displays of geomagnetic geometry, RM 63TMP–2, DASA 1372, General Electric Co., Santa Barbara, Calif., 1963.

DUNGEY, J. W., "Cosmic electrodynamics," Cambridge University Press, Oxford, 1958.

DUNGEY, J. W., Phys. Rev. Letters, 6, 47–48, 1961.

DUNGEY, J. W., Interactions of solar plasma with the geomagnetic field, Planet. Space Sci. 10, 233–237, 1963.

EGAN, R. D., and A. M. PETERSON, Auroral noise at HF, J. Geophys. Res. 65, 3830–3832, 1960.

EGEDAL, J., Observations of aurorae from Danish light-vessels during the years 1897–1937, Publ. Det Danske Meteorol. Inst., Tillaeg, Copenhagen, 1937.

EGELAND, A., Studies of auroral reflections in the VHF band. I. Experimental investigations, with special regard to time variations, fading rate, azimuthal distributions and polarization characteristics, Arkiv för Geofysik 4, 103–169, 1962 a.

EGELAND, A., Studies of auroral reflections in the VHF band. II. Comparison of experimental results with theoretical models, Arkiv för Geofysik 4, 171–209, 1962 b.

EGELAND, A., J. ORTNER, and B. HULTQVIST, A study of the statistics of VHF oblique auroral reflections, Ann. Géoph. 18, 23–44, 1962.

ELVEY, C. T., Aurora borealis, in "Advances in electronics and electron physics," Vol. IX, pp. 1–42, Academic Press, Inc., New York, 1957.

ELVEY, C. T., Auroral morphology. Paper presented at the Lockheed Symposium on auroras at Palo Alto, January, 1964.

ELVEY, C. T., H. LEINBACH, J. HESSLER, and J. NOXON, Preliminary studies of the distribution of auroras in Alaska, Trans. Am. Geoph. Un. 36, 390–394. 1955.

ELVEY, C. T., and A. BELON, Description of the allsky-camera, its method of operation, an instrument (Ascagraph) for measuring the film. Sci. Rept. No. 1, IGY Project No. 1.1, NSF Grant No. Y/1.1/44, Geophys. Inst., College, Alaska, 43 pp., 1957.

ELVEY, C. T., and W. STOFFREGEN, Auroral photography by allsky-camera, Ann. IGY 5, Pt. II, 117–151, 1957.

EVANS, S., J. Atm. Terr. Phys. 16, 191, 1959.

EVANS, S., Proc. Roy. Soc. A 256, 234, 1960.

EVANS, S., and G. M. THOMAS, The southern auroral zone in geomagnetic longitude sector 20° E, J. Geophys. Res. 64, 1381–1388, 1959.

EVANS, J. E., R. G. JOHNSON, R. D. SHARP, and J. B. REAGAN, Precipitated proton fluxes at 300 km altitude over the auroral zones. Paper presented at the American Geophysical Union, Spring meeting in Washington, D.C., April 1964.

FAN, C. Y., Time variation of the intensity of auroral hydrogen emission and the magnetic disturbance, Astroph. J. 128, 420–427, 1958.

FELDSTEIN, Y. I., Geographical distribution of aurora and azimuth of auroral arcs. Investigation of the aurora, The Academy of Sciences of the USSR, Moscow, No. 4, 61–78, 1960.

FELDSTEIN, Y. I., Changes in the auroral zone location in connection with solar activity cycle, Geomagnetism and Aeronomie (English transl.), 2, 476–478, 1962.

FELDSTEIN, Y. I., Auroras and magnetic activity in the neighborhood of the pole, Geomagnet. and Aeronomie (English Transl.) 2, 706–708, 1962 b.

FELDSTEIN, Y. I., Auroral morphology. I. The location of the auroral zone, Tellus (in print), 1964 a.

FELDSTEIN, Y. I., Auroral morphology. II. Aurora and geomagnetic disturbances, Tellus (in print), 1964 b.

FELDSTEIN, Y. I., and N. F. SHEVNINA, Seasonal variations in auroral frequency, Geomagn. and Aeromonie (US transl.) 1, 812–814, 1961.

FELDSTEIN, Y. I., and E. K. SOLOMATINA, Auroras in the southern hemisphere, Geomagn. and Aeron. (US transl.) 1, 475–479, 1961.

FELDSTEIN, Y. I., O. V. KHOROSHEVA, and A.T. LEBEDINSKY, Investigations of auroral phanetary distributions, J. Phys. Soc. Japan 17, Suppl. A–1, 249–254, 1962.

FELDSTEIN, Y. I., and N. F. SHEVNINA, The position of the auroral zone in the southern hemisphere, Geomagn. Aeronomie (US transl.) 2, 240–242, 1962.

FILLIUS, R. W., IGY General Report No. 12, 8, 1960.

FINCH, H. F., and B. R. Leaton, The earth's main magnetic field—epoch 1955.0, Monthly Notices Roy. Astron. Soc., Geophys. Suppl., 1, 314–317, 1957.

FORSYTH, P. A., F. GREEN, and W. MAH, Can. J. Phys. 38, 770, 1960.

FOWLER, P. H., and C. J. WADDINGTON, An artificial aurora, Nature 182, 1728, 1958.

FREEMAN, J. W., State Univ. of Iowa Publ. SUI-63-20, 1963.

FREEMAN, J. W., J. A. VAN ALLEN, and L. J. CAHILL, Explorer 12 observations of the magnetospheric boundary and the associated solar plasma on September 13, 1961, J. Geoph. Res. 68, 2121-2130, 1963.

FREIER, G. D., Auroral effects on the earth's electric field, J. Geophys. Res. 66, 2695-2702, 1961.

FRIEDMAN, H., The sun's ionizing radiations, in "Physics of the upper atmosphere" (Ed Ratcliffe) pp. 133-218, Academic Press, Inc., New York, 1960.

FRITZ, H., Die Gewitter and Hydrometeore in ihrem Verhalten gegenüber den Polarlichten, Vierteljahrschrift der Naturforschenden Gesellschaft in Zürich 13, 337-373, 1868.

FRITZ, H., Verzeichnis beobachteter Polarlichten, Akad. Wissenschaffen, Wien, 1873.

FRITZ, H., Die geographische Verbreitung des Polarlichtes, Petermanns Geographische Mitteilungen 20, 347-358, 1874.

FRITZ, H., Das Polarlicht, Brockhaus, Leipzig, 1881.

FULLER, V. R., Auroral observations at the Alaska Agricultural College and School of Mines for the years 1931-1932, Terr. Magn. 38, 207-238, 1933.

FULLER, V. R., A report of work on the aurora borealis for the years 1932-1934. Terr. Magn. 40, 269-275, 1935.

FULLER, V. R. and E. H. BRAMHALL, Auroral research at the University of Alaska, 1930-1934, Misc. Publ. Univ. Alaska, Vol. III, 1937.

GADSDEN, M., Colour photography of the aurora. J. Atm. Terr. Phys. 17, 347-49, 1960.

GALPERIN, Y. I., Proton bombardment in aurora. Planet. Space Sci. 10, 187-193, 1963.

GARTLEIN, C. W., Unlocking secrets of the northern lights. Nat. Geograph. Mag. Nov., pp. 673-704, 1947

GARTLEIN, C. W., Ann. Géophys. 15, 31, 1959.

GARTLEIN, C. W., and R. K. MOORE, J. Geoph. Res. 56, 85, 1953.

GARTLEIN, C. W., A. E. GARTLEIN, and G. SPRAGUE, The aurora and the local magnetic field. IGY, WDC, A, U.S. National Academy of Sciences, IGY General Report No. 12, 57-67, 1960.

GARTLIEN, C. W., B. NACK, and G. SPRAGUE, IGY, WDC, A, U.S. Nat. Acad. Sci., IGY General Rept. No. 12, 38, 1960.

GARTLEIN, C. W., and G. SPRAGUE, Auroral occurrence, IGY, WDC, A, U.S. National Academy of Sciences, IGY General Report No. 12, 68-73, 1960.

GEDDES, M., The photographic determination of the height and position of aurorae observed in New Zealand during 1937. New Zealand J. Sci. Tech. 20, No. 6B, 1939.

GOLD, T., Magnetic Storms, Space Sci. Rev. 1, 100-114, 1962.

GORBUSHINA, G. N., The geographical distribution of anomalous absorption in the northern hemisphere. Geomagnetism and Aeronomi (English translation), 2, 225-231, 1962.

GREGORY, J. B., Radio wave reflections from the mesosphere, J. Geophys. Res. 66, 429-445, 1961.

GROTH, L. H., L. J. ANDERSON, C. C. EASTERBROOK, and L. R. BURDETTE, S-Band auroral radar returns, J. Geophys. Res. 69, 194-196, 1964.

GURNETT, D. A., and B. J. O'BRIEN, High-latitude geophysical studies with satellite Injun 3 5. Very-low-frequency electromagnetic radiation, J. Geophys. Res. 69, 65-89, 1964.

GUSTAFSSON, G., Ionization in the D-region during auroral break-up events., Planet. Space Sci., (in print), 1964.

GUSTAFSSON, G., and J. ORTNER, Kiruna Geophysical Observatory, Sci. Rep. No. 2, Contract No. AF61(052)-288, 1962.

HAGG, E. L., D. MULDREW, and E. WARREN, J. Atm. Terr. Phys. 14, 345, 1959.

HALE, D. P., Characteristics of the visual aurora at Byrd Station, Antarctica, during 1957, J. Atm. Terr. Phys. 17, 65-70, 1959.

HARANG, L., A study of auroral arcs and draperies, Geofys. Publikasj. 13, No. 14, 1-15, 1944.

HARANG, L., A study of auroral arcs and draperies, Terr. Magn. 50, 296-306, 1945.

HARANG, L., The aurorae, John Wiley and Sons, Inc., New York, 1951.

HARANG, L., and W. BAUER, über einen Nordlichtbogen in weniger als 80 km Höhe über der Erde, Gerl. Beite. Geophys. 37, 109-115 1932.

HARANG, L., and W. STOFFREGEN, Scattered reflections of radiowaves from a height of more than 1000 km. Nature, 142, 832-833, 1938.

HARANG, L., and W. STOFFREGEN, Echoversuche auf Ultrakurzwellen. Hochfreq. und Elektroak. 55, 105-108, 1940.

HARTZ, T. R., Auroral radiation at 500 Mc. Can. J. Phys. 36, 677-682, 1958.

HATHERTON, T., and G. M. THOMAS, J. Geogh. Res. 64, 1381, 1959.

HEPPNER, J. P., Time sequences and spatial relations in auroral activity during magnetic bays at College, Alaska, J. Geoph. Res. 59, 329-338, 1954 a.

HEPPNER, J. P., A study of relationships between the aurora borealis and the geomagnetic disturbances caused by electric currents in the ionosphere. Thesis, California Institute of Technology, Pasadena, 1954 b.

HEPPNER, J. P., Note on the occurrence of world-wide S.S.C.'s during the onset of negative bays at College, Alaska, J. Geoph. Res. 60, 29-32, 1955.

HEPPNER, J. P., E. C. BYRNE, and A. E. BELON, The association of absorption and E_s ionization with aurora at high latitudes. J. Geoph. Res. 57, 121-134, 1952.

HERLOFSON, N., Particle diffusion in the radiation belts. J. Phys. Soc. Japan 17, Suppl. A2, 151-153, 1962.

HESSLER, V. P., and E. M. WESCOTT, Correlation between earth-current and geomagnetic disturbance. Nature 184, 627, 1959.

HOFFMAN, R. A., R. L. ARNOLDY, and J. R. WINCKLER, J. Geophys. Res. 67, 4543, 1962.

HOLT, O., Some experimental studies of the ionospheric D region at high latitudes. Norwegian Defence Res. Establ. Report No. 46, 1963.

HOLT, O., and A. OMHOLT, J. Atm.Terr. Phys. 24, 467, 1962.

HONES, E. W., JR., Personal Communication, 1964.

HOWER, G. L., Synchrotron emission as the source of uhf auroral noise. Stanford Electronics Laboratories, Stanford University, No. SEL-62-149-(SUP), Technical Rept., (1403-1), Sci. Rept. (1), 1963.

HULBERT, E. O., On the diurnal variation of the aurora polaris. Terr. Magn. 36, 23-40, 1931.

HULTQVIST, B., The geomagnetic field lines in higher approximation. Arkiv för Geofysik 3, 63-77, 1958.

HULTQVIST, B., On the orientation of auroral arcs. J. Atm. Terr. Phys. 24, 17-30, 1962.

HULTQVIST, B., Circular symmetry in the geomagnetic equatorial plane. Planet. Space Sci., 8, 142-150, 1962 b.

HULTQVIST, B., Sunrise and sunset effects on cosmic noise absorption associated with aurora and magnetic storms. Planet. Space Sci. 11, 371-383, 1963 a.

HULTQVIST, B., On the height distribution of the ratio of negative ion and electron densities in the lowest ionosphere. J. Atm. Terr. Phys., 25, 225-240, 1963 b.

HULTQVIST, B., Studies of ionospheric absorption of radio waves by the cosmic noise method. In "Radioastronomical and Satellite Studies of the Atmosphere (Ed. J. Aarons), North-Holland Publishing Co., Amsterdam, 1963 c.

HULTQVIST, B., On the height of auroral absorption. Planet. Space Sci., (in print), 1964 a.

HULTQVIST, B., On the height of auroral absorption II. Planet. Space Sci., (in print), 1964 b.

HULTQVIST, B., Aurora and the lower ionosphere in relation to rocket and satellite measurements of precipitated particles. Paper presented at the COSPAR Fifth International Space Science Symposium in Florence, May 1964. (1964 c).

HULTQVIST, B., and A. EGELAND, Radio aurora. Space Science Review, (in print), 1964.

HUNTEN, D. M., Some photometric observations of auroral spectra. J. Atm. Terr. Phys. 7, 141-151, 1955.

HUNTEN, D. M., F. E. ROACH, and J. W. CHAMBERLAIN, A photometric unit for the airglow and aurora. J. Atm. Terr. Phys. 8, 345-346. 1956.

IYENGAR, R. S., and G. G. SHEPHERD, Observations of auroral luminosity fluctuations. Can. J. Phys. 39, 1911, 1961.

JACCHIA, L. G., and J. SLOWLEY, Atmospheric heating in the auroral zones: A preliminary analysis of the atmospheric drag of the Injun 3 satellite. J. Geophys. Res. 69, 905-910. 1964.

JACKA, F., Austral. J. Phys. 6.2, 1953.

JACKA, F., The southern auroral zone as defined by the position of homogeneous arcs. Austral. J. Phys. 6, 219-228, 1953.

JACKA, F., Ann. IGY 11, 145, 1961.

JACKA, F., and J. PATON, IQSY Instruction Manual No. 3 Aurora, CIG-IQSY Committee, London, 1963.

JENSEN, R. E., and B. W. CURRIE, Orientations of auroral displays in west-central Canada. J. Geophys. Res. 58, 201-208, 1953.

JENSEN, D. C., and J. C. CAIN, Unpublished, presented at April 1962 meeting of the American Geophysical Union, Washington, D.C., 1962.

JOHNSON, F. S., The gross character of the geomagnetic field in the solar wind. J. Geoph. Res. 65, 3049-3051, 1960.

JOHNSON, R. G., J. E. EVANS, R. D. SHARP, and J. B. REAGAN, Paper presented at the American Geophysical Union Meeting in Washington, D.C., April 1964. (1964).

JÖRGENSEN, T. S., and E. UNGSTRUP, Ionosphere Laboratory, Royal Techn. Univ. Denmark, Copenhagen, Sci. Rept., 1962.

KAISER, T. R., Radio investigations of aurorae and related phenomena. In "The Airglow and the Aurora" (E. B. Armstrong and A. Dalgarno, eds.), pp. 156-173, Pergamon Press, London, 1956.

KARLSSON, E., Streaming of a plasma through a magnetic dipole field. Phys. Fluids, 6, 708-722, 1963.

KAUFMAN, R., Experimental tests for the acceleration of trapped particles. J. Geoph. Res. 68, 371-386, 1963.

KAVADAS, A., J. Atm. Terr. Phys. 23, 170. 1962.

KELLOGG, P. J., Flow of plasma around the earth. J. Geophys. Res., 67, 3805-3811, 1962.

KERN, J. W., A charge separation mechanism for the production of polar auroras and electrojets. J. Geoph. Res. 67, 2649-2665, 1962.

KERN, J. W., and E. H. VESTINE, Theory of auroral morphology. J. Geoph. Res. 66, 713-723, 1961.

KHOROSHEVA, O. V., The space and time-distribution of auroras and their relationship with high-latitude geomagnetic disturbances. Geomagn. Aeronomi (U. S. translation), 1, 615-621, 1961.

KHOROSHEVA, O. V., The diurnal drift of the closed auroral ring. Geomagn. and Aeronomie (English translation) 2, 696-705, 1962.

KIM, J. S., and B. W. CURRIE, Horizontal movements of aurora. Can. J. Phys. 36, 160-170, 1958.

KIM, J. A., and B. W. CURRIE, Can. J. Phys. 38, 1366, 1960.

KIM, J. S., and R. A. VOLKMAN, Thickness of zenithal auroral arcs over Fort Churchill, Canada. J. Geophys. Res. 68, 3187-3190. 1963.

KNECHT, R. W., Relationships between aurora and sporadic E-echoes at Barrow, Alaska. J. Geoph. Res. 61, 59-69, 1956.

KRASOVSKII, V. I., Some results of investigations of aurora and night airglow during the IGY and IGC. Planet. Space Sci. 8, 125-141, 1961.

KRASOVSKII, V. I., YU. M. KUSHNIR, G. A. BORDOVSKII, G. F. ZAKHAROV, YE. and M. SVETLITSKII, Detection of corpuscular fluxes by means of the Third Soviet earth satellite. Ann. IGY, Vol. XII, Part II, 722-724. (Published in "Artificial Earth Satellites" Issue No. 2, Moscow, 1958.)

KRASOVSKII, V. I., I. S. SHKLOVSKII, YU. I. GALPERIN, E. M. SVETLITSKII, YU. M. KUSHNIR, and G. A. BORDOVSKII, The detection of electrons with energies of approximately 10 kev in the upper atmosphere. Planet. Space Sci. 9, 27-40. 1962.

LASSEN, K., Existence of an inner auroral zone. Nature, 184, 1375-1377, 1959.

LASSEN, K., Day-time aurorae observed at Godhavn 1954-56, Det Danske Meteor. Inst., Meddel. No. 15, 1961.

LEADABRAND, R. L., Electromagnetic measurements of auroras. Paper presented at the Lockheed Symposium on Auroras, Palo Alto, January 1964.

LEADABRAND, R. L., L. DOLPHIN, and A. M. PETERSON, Preliminary results of the 400-Mc radar investigations of auroral echoes at College, Alaska. IRE Trans. Antennas and Prop., AP-7, 127–136, 1959.

LEBEDINSKII, A. I., On certain applications of wide-angle mirror cameras (translated title), Doklady Akad. Nauk. SSSR, 102, 473–475, 1955.

LEE, A. W., Auroral observations at Lerwick Observatory, 1924–29. Meteorology Office Professional Notes, No. 56, London, 1930.

LERFALD, G. M., C. G. LITTLE, and R. PARTHASARATHY, D-region electron density profiles during auroras. J. Geophys. Res. (in print), 1964.

LIEMOHN, H., The outer radiation belt and aurorae. Nature, 188, 394–395, 1960.

LISZKA, L., Satellite cintillation observed in the auroral zone. Arkiv for Geophysik. 4, 211–225, 1963.

LITTLE, C. G., and A. MAXWELL, Scintillation of radio stars during aurora and magnetic storms. J. Atm. Terr. Phys. 2, 356–360. 1952.

LITTLE, C. G., and H. LEINBACH, Some measurements of high-latitude ionospheric absorption using extraterrestrial radio waves. Proc. Inst. Radio Engrs. 46, 335–348, 1958.

LITTLE, C. G., G. C. REID, E. STILTNER, and R. P. MERITT, An experimental investigation of the scintillation of radio stars observed at frequencies of 223 and 456 Mc/s from a location close to the auroral zone. J. Geoph. Res. 67, 1763–1884, 1962.

LITTLE, C. G., G. M. LERFALD, and R. PARTHASARATHY, Personal communication, 1963.

LOGINOV, G. A., M. I. PUDOVKIN, and R. A. SKRYNNIKOV, Diurnal variation of the intensity of aurorae and S_D variation. Geomagnetism and Aeronomy (English translation) 2, 709–713, 1962.

LOOMIS, E., The great auroral exhibition of August 28 to September 4, 1859 and the geographical distribution of auroras and thunder storms. Am. J. Sci. and Arts. 30, 79–100, 1860.

LOOMIS, E., Sciences physiques et naturelles. Arch. Sci. Phys. et Nat. 31, 273–285, 1868.

MAEDA, K., Auroral dissociation of molecular oxygen in the polar mesosphere. J. Geophys. Res. 68, 185–197, 1963.

MAEDA, K., and T. WATANABE, Pulsating auroral and infrasonic waves in the polar atmosphere. J. Atm. Sci. 21, 15–29, 1964.

MAEHLUM, B., and B. J. O'Brien, Study of energetic electrons and their relationship to auroral absorption of radio waves. J. Geoph. Res. 68, 997–1010, 1963.

MAIRAN, J. J. D. DE, Traité Physique et Historique d l'Aurore Boréale. Imprimerie Royale, Paris (2nd ed. 1754), 1733.

MALVILLE, J. M., Antarctic auroral observations. Ellsworth Station, 1957. J. Geophys. Res. 64, 1389–1393, 1959.

MALVILLE, J. M., Narrow hydrogen emissions in the aurora. Planet. Space Sci. 2, 130–132, 1960.

MALVILLE, J. M., The effect of the initial phase of a magnetic storm upon the outer Van Allen belt. J. Geophys. Res. 65, 3008, 1960.

MALVILLE, J. M., Diurnal variation of high-latitude auroras. J. Geophys. Res. 69, 1285–1292, 1964.

MANN, L. G., S. D. BLOOM, and H. I. WEST, Jr., The electron spectrum from 90 to 1200 kev as observed on Discoverer satellites 29 and 31. Space Research III (Ed., W. Priester), pp. 447–462, North-Holland Publ. Co., Amsterdam, 1963.

MARTIN, L. H., and R. A. Helliwell, Association between aurorae and very-low frequency hiss observed at Byrd Station, Antarctica. Nature, 187, 751, 1960.

MATSUSHITA, S., Ancient aurorae in Japan. J. Geophys. Res. 61, 297–302, 1953.

MAWSON, D., Scientific Report Australasian Antarctic Expedition 1911–14, Vol. 11, Part 1: Records of the Aurora Polaris, 1914.

McDIARMID, I. B., D. C. ROSE, and E. BUDZINSKI, Direct measurement of charged particles associated with auroral zone radio absorption. Can. J. Phys. 39, 1888–1900, 1961.

McDIARMID, I. B., J. R. BURROWS, E. E. BUDZINSKI, and M. D. WILSON, Some average properties of the outer radiation zone at 1000 km. Can. J. Phys. 41, 2064–2079, 1963.

McEWEN, D. J., and R. MONTALBETTI, Parallactic measurements on aurorae over Churchill, Canada. Can. J. Phys. 36, 1593–1600, 1958.

McILWAIN, C. E., Direct measurement of particles producing visible auroras. J. Geoph. Res. 65, 2727–2747, 1960.

McILWAIN, C. E., Coordinates for mapping the distribution of magnetically trapped particles. J. Geophys. Res. 66, 3681–3691, 1961.

McLENNAN, J. C., H. S. WYNNE-EDWARDS, and H. J. C. IRETON, Height of the polar aurora in Canada. Can. J. Res. 5, 285–296, 1931.

MEAD, G. D., and D. B. BEARD, The shape of the geomagnetic-field solar-wind boundary. Goddard Space Flight Center, Rep. X-640-63-239, 1963.

MEEK, J. H., Correlation of magnetic auroral and ionospheric variations at Saskatoon. J. Geoph. Res. 58, 445–456, 1953.

MEEK, J. H., Correlation of magnetic auroral and ionospheric variations at Saskatoon. Part 2. J. Geoph. Res. 59, 87–92, 1954 b.

MEEK, H. H., The location and shape of the auroral zone. J. Am. Terr. Phys. 6, 313–321, 1955.

MEEK, J., J. Am. Terr. Phys., Suppl. 120, 1957.

MEINEL, A. B., On the entry into Earth's atmosphere of 57 kev protons during auroral activity. Phys. Rev. 80, 1096–1097, 1950 a.

MEINEL, A. B., Evidence for the entry into the upper atmosphere of high-speed protons during auroral activity. Science 112, 590, 1950 b.

MEINEL, A. B., Doppler-shifted auroral hydrogen emission. Astrophys. J. 113, 50–54, 1951.

MEINEL, A. B. and D. H. SCHULTE, A note on auroral motion. Astrophys. J. 117, 454–455, 1953.

MEINEL, A. B., B. J. NEGAARD, and J. W. CHAMBERLAIN, A statistical analysis of low-latitude aurorae. J. Geoph. Res. 59, 407–413, 1954.

MEREDITH, L. H., M. B. GOTTLIEB, and J. A. VAN ALLEN, Direct detection of soft radiation above 50 kms in the auroral zone. Phys. Rev. 97, 201–205, 1955.

MOGILEVSKIY, On the nature of the action of solar corpuscular flux upon the Earth's magnetosphere (translated title) Geomagnetism i Aeronomiya 3, No. 6, 1002, 1963.

MOORCROFT, D. R., and Forsyth, P. A., On the relation between radio star scintillations and auroral and magnetic activity. J. Geoph. Res. 68, 117–124, 1963.

MOROZUMI, H. M., A study of the aurora australis in connection with and association between VLFE hiss and auroral arcs and bands observed at the South Geographic Pole. Thesis, State University of Iowa, 62–14, 1962.

MOROZUMI, H. M., Semidiurnal auroral peak and VLF emissions observed at the South Pole, 1960. IG Bulletin No. 73, July 1963, pp. 16–23. (Paper presented at the Second Western National Meeting of the American Geophysical Union at Stanford Univ., Dec. 1962).

MUNCKE, G. W., Nordlicht. Physikalisches Wörterbuch 7, 113–268, 1837.

MURCRAY, W. E., Some properties of the luminous aurora as measured by a photoelectric photometer. J. Geophys. Res. 64, 955–959, 1959.

MURCRAY, W. B., and J H. Pope, Radiation from protons of auroral energy in the vicinity of the earth. J. Geoph. Res. 65, 3569–3574, 1960.

NADUBOVICH, Y. A. and G. V. STARKOV, The fibrous structure of faint auroral homogeneous arcs. Geomagn. Aeron. (U. S. transl.) II, 57–61, 1961.

NAGATA, T. and E. KANEDA, A photoelectric analyzer of all-sky camera auroral photographs. J. Geophys. Res. 66, 2259–2261, 1961.

NESS, N. F., Personal communication, 1964.

NIKOLSKIY, A. P., The world-wide distribution of magneto-ionospheric disturbance and aurora (translated title). Doklady Akad. Nauk. SSSR, 115, 84–87.

NIKOLSKIY, A. P., Experimental proof of the existence of a second zone of magnetic disturbance in the east Arctic. Geomagnetism and Aeronomy (U. S. transl.) 1, 833–837, 1963.

NOXON, J. F., Observation of daytime aurora. J. Am. Terr. Phys. 25, 637–645, 1963.

O'BRIEN, B. J., Direct observations of dumping of electrons at 1000-km altitude and high latitudes. J. Geoph. Res. 67, 1227–1233, 1962 a.

O'BRIEN, B. J., Lifetimes of outer-zone electrons and their precipitation into the atmosphere. J. Geoph. Res. 67, 3687–3706, 1962 b.

O'BRIEN, B. J., A large diurnal variation of the geomagnetically trapped radiation. J. Geoph. Res. 68, 989–995, 1963.

O'BRIEN, B. J., Review of studies of trapped radiation with satellite-borne apparatus. Space Science Reviews 1, 415–484, 1963.

O'BRIEN, B. J., High-latitude geophysical studies with satellite Injun 3.3. Precipitation of electrons into the atmosphere. J. Geophys. Res. 69, 13–43, 1964.

O'BRIEN, B. J. and C. D. LAUGHLIN, An extremely intense electron flux at 1000-km altitude in the auroral zone. J. Geoph. Res. 67, 2667-2672, 1962.

O'BRIEN, B. J., C. D. LAUGHLIN, J. A. VAN ALLEN, and L. A. FRANK, Measurements of the intensity and spectrum of electrons at 100-km altitude and high latitudes. J. Geoph. Res. 67, 1209–1225, 1962.

O'BRIEN, B. J., and H. TAYLOR, High-latitude geophysical studies with satellite Injun 3.4. Auroras and their excitation. J. Geophys. Res. 69, 45–63, 1964.

OMHOLT, A., Astrophys. J. 126, 461, 1957.

OMHOLT, A., Geoph. Publikasjoner 20 (11), 1, 1959.

OMHOLT, A., The time delay of the red (OI) lines in the aurora. Plan. Space Sci. 2, 246–48, 1960.

OMHOLT, A., Velocities of very active auroral rays. Planet. Space Sci. 9, 285–286, 1962.

OMHOLT, A., Observations and experiments pertinent to auroral theories. Planet. Space Sci., 10, 247–262, 1963.

OMHOLT, A. and L. HARANG, Measurement of the mean lifetime of the metastable ^1S-state of the oxygen atom in the upper atmosphere during auroral displays. J. Atm. Terr. Phys. 7, 247–253, 1955. J. Geophys. Res. 61, 161, 1956.

OMHOLT, A., W. Stoffregen, W., and H. DERBLOM, Hydrogen lines in auroral glow. J. Atm. Terr. Phys. 24, 203–209, 1962.

PARK, F. R., An all-sky camera for auroral research. Bull. Radio and Electr. Eng. Div., Nat. Res. Council, Ottawa, 7, 1–3, 1957.

PARKER, E. N., Kinetic properties of interplanetary matter. Planet. Space Sci. 9, 461–475, 1962.

PAULSEN, A., Obs. int. polaires 1882–83. Exp. Danois, Copenhague, 1893.

PEDERSON, A., Time, height, and latitude distribution of D layers in the subauroral zone and their relation to geomagnetic activity and aurorae. J. Geophys. Res. 67, 2685–2694, 1962.

PERSSON, H., Electric field along a magnetic line of force in a low-density plasma. Phys. of Fluids 6, 1756–1757, 1963.

PFOTZER, A., A. EHMERT, H. ERBE, E. KEPPLER, B. HULTQVIST, and J. ORTNER, a contribution to the morphology of X-ray bursts in the auroral zone. J. Geophys. Res. 67, 575–585.

PUDOVKIN, M. T. and L. S. YEVLASHIN, Spatial relationship between auroras and electric currents in the ionosphere. Geomagnetism and Aeronomy, (English translation), 2, 557–560, 1962.

QUENBY, J. J., and W. R. WEBBER, Cosmic ray cut-off rigidities and the earth's magnetic field. Phil. Mag. (8), 4, 90–113, 1959.

REES, M. H., Auroral ionization and excitation by incident energetic electrons. Planet. Space Sci. 11, 1209–1218, 1963.

REES, M. H., and G. C. REID, Nature 184, 539, 1959.

ROACH, J. R., Effects of radio wave propagation through mid-latitude 6300Å auroral arcs. J. Res. Nat. Bur. Standards, 67D, 263–271, 1963.

ROACH, F. E., and J. R. ROACH, 6300Å auroral arcs in mid-latitudes. Planet. Space Sci. 11, 523–545.

ROMICK, C. J., and C. T. ELVEY, J. Atm. Terr. Phys. 12, 283, 1958.

ROSSER, W. G. V., B. J. O'BRIEN, J. A. VAN ALLEN, L. A. FRANK, and C. D. LAUGHLIN, Electrons in the earth's outer radiation zone. J. Geoph. Res. 67, 4533-4542, 1962.

SANDFORD, B. P., Aurora and airglow from colour film observations. J. Atm. Terr. Phys. 21, 177.

SANDFORD, B. P., Nature 190, 245. Arctic Inst. of North America, Res. Paper No. 18, August 1961, 1961b.

SANDFORD, B. P., and P. HEISER, Colour photography of the aurora. Nature 184, 541, 1959.

SCHOVE, D. J., The Sunspot cycle, 649, B.C. to A.D. 2000. J. Geophys. Res. 60, 127–146.

SEATON, M. J., Excitation processes in the aurora and airglow. I. Absolute intensities, relative ultraviolet intensities, and electron densities in high latitude aurorae. J. Atm. Terr. Phys. 4, 285–294, 1954.

SHARP, R. D., J. E. EVANS, R. G. JOHNSON, and J. B. REAGAN, Paper presented at American Geophys. Union Third Western Nat. Meeting, Boulder, December 1963.

SHARP, R. D., J. E. Evans, R. G. Johnson, and J. B. REAGAN, Paper presented at COSPAR Fifth Space Sci. Symp. Florence, May 1964 a.

SHARP, R. D., J. E. EVANS, W. L. IMHOF, R. G. JOHNSON, J. B. REAGAN, and R. V. SMITH, Satellite measurements of low-energy electrons in the northern auroral zone, submitted for publication to J. Geophys. Res., 1964 b.

SHARP, R. D., et al., Personal communication, 1964 c.

SHERET, M. A., and G. M. THOMAS, Auroral observations at Halley Bay, Antarctica, during 1959. Nature 189, 826, 1961.

SHKLOVSKII, I. S., Ann. Geophys. 14, 414, 1958.

SKRYNNIKOV, Short-period variations of the intensity of auroral luminosity. Geomagnetism and Aeronomy (Engl. transl.) 2, 894–896, 1962.

SNYDER, C. W., and M. NEUGEBAUER, Interplanetary solar-wind measurements by Mariner II. Paper presented at the fourth COSPAR Space Science Symposium, Warsaw, June 1963.

SOBOUTI, Y., The relationship between unique geomagnetic and auroral events. J. Geophys. Res. 66, 725–737, 1961.

STAGG, J. M., British Polar Year Expedition, Fort Rae, N. W. Canada, 1932-33, Vol. 1, p. 127, Royal Soc., London, 1937.

STARKOV, G. V., and Y. I. FELDSTEIN, Investigations of the Aurora, No. 4, p. 56. Academy of Science of the USSR, Moscow, 1960.

STETSON, H. T., and C. F. BROOKS, Terr. Magn. 47, 21, 1942.

STEWART, D. T., Proc. Phys. Soc. A69, 437, 1956.

STILWELL, D. E., Observations of intense, low energy electron fluxes in the outer zone during January and March, 1963. Thesis, State University of Iowa, SUI-63-28, 1963.

STOFFREGEN, W., All-sky camera auroral research during the Third Geophysical Year 1957–1958. Tellus 7, 509–517, 1955.

STOFFREGEN, W., Radio reflections on low frequencies from 75–90 km height during intense aurora activity. J. Atm. Terr. Phys. 13, 167–169, 1958.

STOFFREGEN, W., Results quoted by Harang, L. and J. TRÖIM, Studies of auroral echoes. Planet. Space. Sci. 5, 33–45, 1961.

STOFFREGEN, W., The east-west drift of auroral forms determined from all-sky camera films. J. Atm. Terr. Phys. 21, 257–260, 1961b.

STOFFREGEN, W., H. DERBLOM, and A. OMHOLT, Some characteristics of the D-region ionization during auroral activity. J. Geophys. Res., 65, 1699–1704, 1960.

STOFFREGEN, W. and H. DERBLOM, Auroral hydrogen emission related to charge separation in the magnetosphere. Planet. Space Sci. 9, 711–716, 1962.

STÖRMER, C., "Photograph c Atlas of Auroral Forms and Scheme for Visual Observations of Aurorae", 1st. ed., A. W. Bröggers Boktrykkeri, Oslo. Publication sponsored by the International Union of Geodesy and Geophysics.

STÖRMER, C., Photography of polar aurora. In "Supplement to the Photographic Atlas of Auroral Forms" pp. 4–11, A. W. Bröggers Boktrykkeri, Oslo. Publication sponsored by the International Union of Geodesy and Geophysics.

STÖRMER, C., Photographische Höhenmessungen und Spectra des grossen Nordlichtes vom 25. bis 26. January 1938. Naturwissenschaffen 26, 633–638, 1938.

STÖRMER, C., Results of the photogrammetric measurements of the aurora borealis during the Norwegian-French polar expedition to North-East Greenland, 1938–1939. Geophys. Publ. XII, No. 13, 1944.

STÖRMER, C., The Polar Aurora, Oxford at the Clarendon Press, 1955.

SVERDRUP, H. V., Auroral results, Maud-expedition, 1918–1925. Rep. Dept. Terr. Magn., Carnegie Inst. Wash. 6, No. 175, 1927.

THOMAS, J. A. and E. K. Smith, A survey of the present knowledge of sporadic-E ionization. J. Atm. Terr. Phys. 13, 295–314,1959.

THOMAS. L. and W. R. PIGGOTT, Some aspects of the incidence of polar blackout during IGY. In "Some ionospheric results obtained during the IGY" (Ed. Beynon), Elsevier Pub. Co., pp. 62–71.

TROMHOLT, S., Sur les periodes des aurores boréales. Danish Meteor Inst. Yearbook for 188?, 1882.

TROMHOLT, S., Katalog der in Norwegen bis Juni 1878 beobachteten Nordlichter. Kristiania (Oslo), 1902.

VAN ALLEN, J. A., Direct detection of auroral radiation with rocket equipment. Proc. Natl. Acad. Sci. 43, 57–62, 1957.

VEGARD, L., On the properties of the rays producing aurorae borealis, Phil. mag. 23, 211–237, 1912.

VEGARD, L., Hydrogen showers in the auroral region. Nature 144, 1089–1090, 1939.

VEGARD, L. and KROGNESS, O., The position in space of the aurora polaris from observations made at the Haldde Observatory 1913–191? Geophys. Publik. 1, 1–172.

VESTINE, E. H., The geographic incidence of aurora and magnetic disturbance, northern hemisphere. Terr. Magn. 49, 77–102, 1944.

VESTINE, E. H., J. Geophys. Res. 65, 260–262, 1960.

VESTINE, E. H., and E. J. SNYDER, The geographic incidence of aurora and magnetic disturbance, southern hemisphere. Terr. Magn. 50, 105–124, 1945.

VESTINE, E. H., and W. L. SIBLEY, The geomagnetic field in space, ring currents and auroral isochasms. J. Geophys. Res. 65, 1967–1979, 1960.

WEILL, G. W., Aspects de l'Aurore observeé à la base Dumont d'Urville on Terre Adélie. Compt. Rendus, 246, 2925–2927, 1958.

WHITE, F. W. G., and M. Geddes, The Antarctic zone of maximum auroral frequency. Terr. Magn. 144, 367–377, 1939.

WILCKE, J. C., Von den jährlichen und täglichen Bèwegungen der Magnetnadel in Stockholm. Svensk. Vet. Akad. Handl. pp. 273–300, 1777.

WINCKLER, J. R., Balloon study of high-altitude radiations during the IGY. J. Geoph. Res. 65, 1331–1359, 1960.

WINCKLER, J. R., Atmosphere phenomena, energetic electrons, and the geomagnetic field. J. Res. Nat. Bur. Standards 66D, 127–143, 1962.

WINCKLER, J. R., L. PETERSON, R. ARNOLDY, and R. HOFFMAN, X-rays from visible aurorae at Minneapolis. Phys. Rev. 110, 1221–1231, 1958.

WINCKLER, J. R., P. D. BHAVSAR, and K. A. ANDERSON, A study of the precipitation of energetic electrons from the geomagnetic field during magnetic storms. J. Geoph. Res. 67, 3717–3736, 1962.

WITHAM, K., E. LOOMER, and E. NIBLETT, J. Geophys. Res. 65, (1960)

ZABORSCHIKOV, F. VA., and N. T. FEDIAKINA, Interrelations of aurora, radiowave propagation, magnetic and ionospheric disturbance (translated title). Preblemij Arktiki 2, 149–159, 1957.

N65-33705

AURORA AND THE LOWER IONOSPHERE IN RELATION TO SATELLITE OBSERVATIONS OF PARTICLE PRECIPITATION*

BENGT HULTQVIST†

The direct observations of electron precipitation by means of rockets and satellites is reviewed. Upon comparison of the observed energy fluxes with those expected on the basis of optical measurements of aurora, the agreement is found to be good. It is demonstrated that the satellite observations make understandable the observed variable degree of correlation between visual aurora and auroral absorption of radio waves. The precipitated electrons contribute significantly to the night time ionosphere not only in the auroral zone but also over the polar caps and in sub-auroral latitudes. It does not seem impossible that the observed precipitation of electrons is the main source of nighttime ionization in the lower ionosphere. The airglow is briefly discussed in relation to observed particle precipitation. Finally, the recent demonstrations of the insufficiency of the Van Allen belt as a source for the precipitated electrons is briefly reviewed.

INTRODUCTION

The first rocket investigations of particles precipitated into the atmosphere were made by the Iowa group in the early 1950's. In IGY and thereafter a few more direct rocket measurements of high energy particles in auroral altitudes have been reported. Although of extreme value as exploratory studies, these few rocket measurements suffer from the weakness of being very limited in space and time. Some data about the rate of precipitation of electrons having energies greater than about 25 kev have been obtained from balloon observations of X-rays, but it is only in the last two years that systematic satellite measurements of precipitated particles, primarily by the two Injun satellites (O'Brien, 1962, a, b, 1964) and Alouette (McDiarmid, et al., 1963) have provided statistical data, on the basis of which some rough estimates of the average influence on the ionosphere can be made. In addition to precipitated electrons Injun 3 also measured the emission rate below the satellite of 3914Å and 5577Å photons along the magnetic field line upon which the satellite was located. The measurements have recently given very valuable information about detailed relations between electron precipitation and photon emission rate (O'Brien and Taylor, 1964).

One primary purpose of this review is to summarize some of the knowledge about relations between particle influx into the atmosphere and the resulting aurora and ionospheric effects. It should be said, however, that most reports on satellite observations published hitherto concern particles of fairly high energies (greater than 40 kev for electrons). For particles in the lowest energy range (i.e., from a fraction of 1 kev for electrons and some tens of kev for protons) the available experimental results are still scanty. What is going to be said about the influence of observed precipitation on the ionosphere above 100 km will therefore be most preliminary.

OBSERVATIONS OF PRECIPITATED ELECTRONS

The early rocket measurements (e.g., Davis et al., 1960; McIlwain, 1960) as well as the satellite results (e.g., O'Brien, 1962a, 1964) have verified the conclusions of Omholt (1957, 1959), based on spectroscopic observations of aurora, that the protons play a very minor role in the excitation of most aurorae and in producing ionization in the

*Published as Goddard Space Flight Center Document X-611-64-107, May 1964.

†Goddard Space Flight Center, NASA—National Academy of Sciences—National Research Council Senior Post-Doctoral Resident Research Associate on leave of absence from Kiruna Geophysical Observatory, Kiruna, Sweden.

lower ionosphere. Therefore only electrons will be discussed in this review.

The first somewhat detailed satellite studies of particles that were definitely precipitated into the atmosphere were made by means of Injun 1 (O'Brien, 1962a, b). O'Brien and others working with satellite data define precipitated particles as those particles which would mirror at or below 100 km altitude (if the atmosphere had not been present).

(a) Latitudinal Distribution and Other Spatial Characteristics

The recent simultaneous measurements on Injun 3 (O'Brien and Taylor, 1964) of flux of precipitated electrons and of auroral light emission below the satellite have shown th... above aurorae there are electron fluxes several orders of magnitude larger than outside the aurora. (Fig. 1).

Fairly extensive statistical data on the spatial characteristics of precipitated electrons have recently been published by O'Brien (1962a, b, 1964) and McDiarmid et al. (1963). Figure 2 shows the scatter diagram of the Injun 3 observations of precipitated electrons of energy above 40 kev for the range of the invariant latitude Λ (defined by $\Lambda = \cos^{-1}(1/\sqrt{L})$ and at the most a few degrees different from geomagnetic latitude in the latitude range of interest here) between 45 and 76 degrees. As can be seen from the figure, the precipitated flux of electrons of $E \gtrsim 40$ kev has a broad maximum of about 10^5 electrons/cm² sec ster for Λ between 60 and 70 degrees. The flux is 2–3 orders of magnitude lower in subauroral latitudes. Figure 2 also shows that some 10 degrees inside the auroral zone the precipitation rate is down by 1–2 powers of ten.

Similar results have been found by McDiarmid et al. (1963) for electrons of energies above 40 kev and above 250 kev. Figure 3 shows their data presented in a manner different from that of O'Brien (1964). The histograms give, for two different ranges of magnetic activity, the percentage of passes in which the intensity of precipitated electrons with $E \lesssim 40$ kev was above $1.5 \cdot 10^4$ electrons/cm² sec ster (corresponding to 0.2 db absorption at about 30 Mc/s). The flux of $1.5 \cdot 10^4$/cm² sec ster was one-half of the maximum

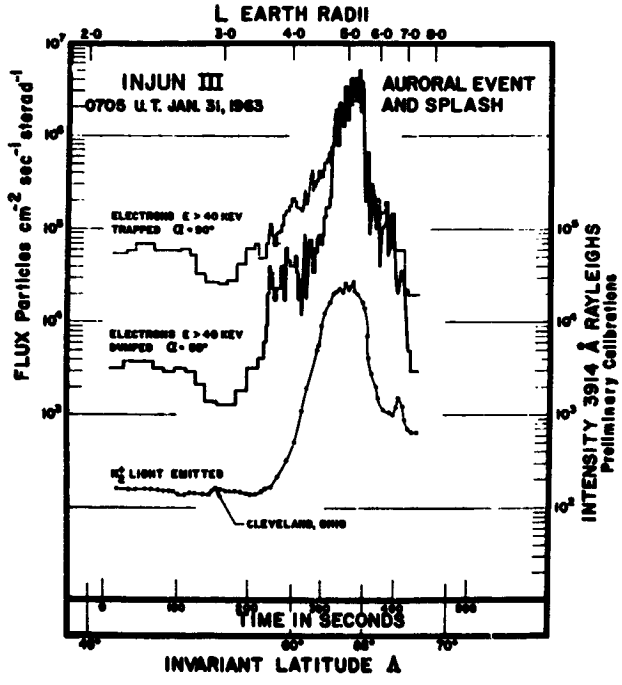


FIGURE 1.—Data from a northbound pass of Injun 3 over North America, which shows simultaneous detection of an aurora, of the precipitated electrons (with pitch angle 50°) partially responsible for causing it, and of trapped electrons. Note the approach to isotropy of the particle flux over the aurora. No attempt has been made to subtract the low-altitude contamination of the photometer signal, part of which was the detection of Cleveland, Ohio, and its surroundings. (After O'Brien and Taylor, 1964.)

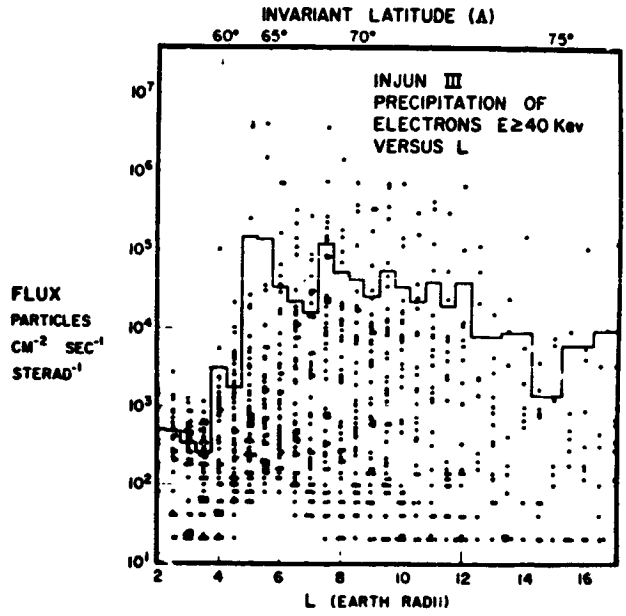


FIGURE 2.—Samples of precipitated fluxes over North America in January 1963. Each point is an 8-sec average of thirty two measurements made at half-integral values of L. The solid line gives the average flux. (After O'Brien, 1964.)

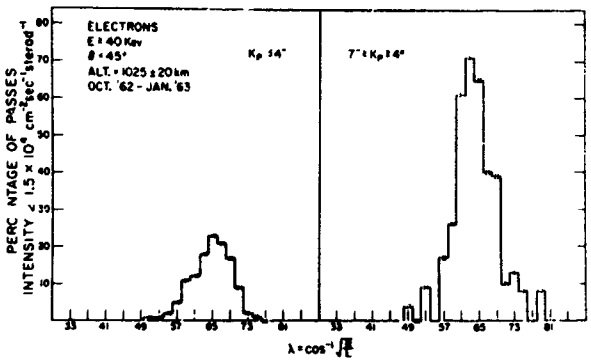


FIGURE 3.—Percentage of passes in which the intensity of precipitated electrons with energies greater than 40 kev was greater than 1.5×10^4cm^{-2} sec^{-1} sterad^{-1} corresponding to 0.2 db of 30 Mc/s absorption, plotted against invariant latitude. (After McDiarmid et al., 1963.)

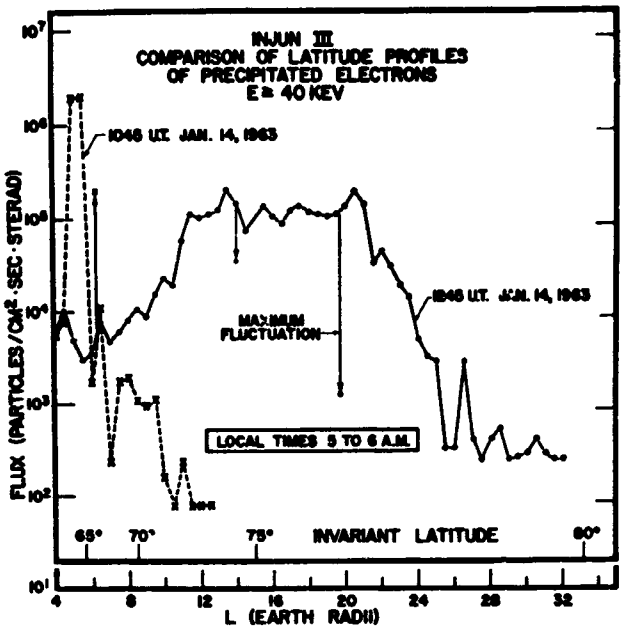


FIGURE 4.—Comparison of the latitude, or L profile, of precipitation for two successive passes at about the same local time. Arrows illustrate range of fluctuation of intensities at given locations. (After O'Brien, 1964.)

average intensity for $K_p<4$ and 1/20 for $K_p>4$. Figure 3 shows that during quiet geomagnetic conditions the precipitated intensity was g[illegible] than half the average value during some 2[illegible] percent of the passes through $\Lambda=65°$, while for moderately disturbed conditions—there were no strong storms represented in the data—the intensity exceeded 1/20 of the average value in about 72 percent of the passes over that latitude.

Simultaneous observations on two satellites indicate that at least sometimes the area of precipitation has a fairly limited longitudinal extent (O'Brien and Laughlin, 1962). On the other hand the precipitated flux seems on occasions to be uniform over as much as 80 degrees of longitude, i.e., over more than 4000 km (O'Brien, 1964).

The latitudinal extent of the precipitation can be very restricted (e.g., over a range of $L\backsim1$ earth radius) or very extensive (over an L interval >20 earth radii). This is illustrated in Figure 4.

The flux of precipitated electrons has its maximum close to the poleward boundary of the region of trapped electrons, as observed on the same satellite (McDiarmid et al., 1963; O'Brien, 1964).

Electron precipitation occurs simultaneously in magnetically conjugate areas at least sometimes, according to balloon observations of X-rays (Brown et al., 1963). This is in accordance with observations of aurora (cf. e.g. DeWitt, 1962).

A 100 fold change in the flux over two km has often been observed by Injun 3 which is in agreement with results of rocket measurements in aurora by Davis et al. (1960). They found the electron flux being concentrated in the visible auroral forms whereas protons were found over a much larger volume than that occupied by the aurora.

The statistical latitude profile of the intensity of dumped electrons with $E\gtrsim40$ kev has been found by O'Brien and Taylor (1964) to have its maximum at a few degrees lower latitude than the frequency of occurrence of visual aurora as observed on the same satellite. Sometimes the visual emission has been found to extend to higher latitudes than the precipitation of electrons with energy greater than 40 kev, indicating electrons of $E<40$ kev are being precipitated up to higher latitudes than the higher energy ones. The lower latitude limit of the aurora coincides with the lower latitude limit of appreciable precipitation (O'Brien, 1964; O'Brien and Taylor, 1964).

Precipitation was found all the time in the auroral zone by Injun 3 (O'Brien, 1964). It can be seen in Figure 2 that at $L=6$ there was never observed any flux lower than about 80 electrons (cm^2 sec ster)$^{-1}$.

The corresponding continuous photon emission in the auroral zone is evident from Figure 5. The

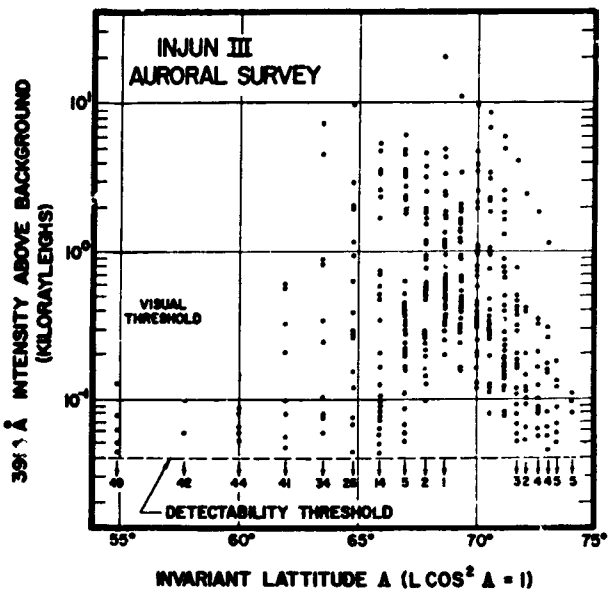


FIGURE 5.—Intensity of 3914Å auroral light average over 4 sec at half-integral values of L in about fifty passes of Injun 3 early in 1963. (After O'Brien and Taylor, 1964.)

minimum emission rate is, however, below visual threshold even in the auroral zone.

Figures 2 and 5 also give a good impression about the enormous variability of precipitation and photon emission at all latitudes, but especially near the auroral zone. At $L=6$ there are values of about $6 \cdot 10^6$ electrons/cm^2 sec ster shown in the figure. This is some 10^5 times the minimum value observed there.

(b) Time Characteristics of Precipitation

It is not possible in satellite measurements to differ between space variations and time variations short compared to the period of revolution. Balloon observations show, however, that time variations as rapid as of periods of a tenth of a second can occur in the electron flux (Winckler et al., 1962).

No systematic variation in the intensity of electron precipitation over periods as long as a day could be found by O'Brien (1964) from the Injun 1 and 3 measurements in contrast with the case for trapped electrons. Sharp et al. (1964) observed a higher nighttime than daytime precipitation flux of electrons in the energy range 0.08–24 kev—as well as of protons—during the 5 days life time of an oriented polar-orbiting satellite. They also found the energy distribution of the electrons to be harder on the dayside than on the nightside of the earth (Johnson et al., 1964). The Iowa group has recently reported the existence of a diurnal variation in the precipitated flux of electrons of energy above 40 kev at high latitudes. The maximum in the diurnal variation curve was reported to be on the dayside of the earth (Frank et al., 1964). It is not clear if the difference in the results found by various observers is due to the diurnal variation being different for different energy ranges or if it is due to some of the results being not statistically significant because of the enormous variability of the precipitation phenomenon.

(c) Pitch Angle Distribution

Figure 1 shows that over the aurora, where the precipitation is intense, the directional flux of precipitated electrons becomes equal to the flux of trapped electrons, which also is increased over the aurora. Figure 1 illustrates a general rule that has been found by O'Brien (1962a, b, 1964), namely that for electrons of $E \geq 40$ kev the pitch angle distribution approaches isotropy in the regions of intense precipitation. An isotropic pitch angle distribution in regions of strong precipitation has also been observed by Krasovskii et al. (1962) at an energy of about 10 kev. No cases have been found in which the directional flux of precipitated electrons was higher than the corresponding value for trapped electrons. The mentioned observations were made fairly close to the earth's atmosphere.

The tendency to isotropy seems to indicate that the acceleration of the electrons—if it is directed along the field lines—takes place far away from the atmosphere. This, as well as the mentioned observations of precipitation of electrons from above 1000 km in aurorae, suggest that the role of the ionosphere in the production of the energetic electrons is not important.

O'Brien (1964) found a flux upwards along the field lines, which was some 10 percent of the precipitated electron flux. He interpreted these observations as backscattering of electrons from the atmosphere.

(d) K_p—Dependence of Precipitation

The K_p dependence of the precipitation of electrons of energy above 40 kev is illustrated in

Figure 3. The average intensity was ten times higher near the auroral zone when K_p was above 4 than when it was below 4. The data material for the higher K_p range in Figure 3 does not contain data from any strong magnetic storm, so still higher values may be expected.

O'Brien (1964) found that the flux of precipitated electrons above 40 kev increased on the average by a factor of 5 for every step of K_p. A close correlation between precipitation intensity has also been observed for the 0.08–24 kev energy range by Sharp et al. (1964). O'Brien's results, obtained on Injun 3 in a low orbit, are shown in

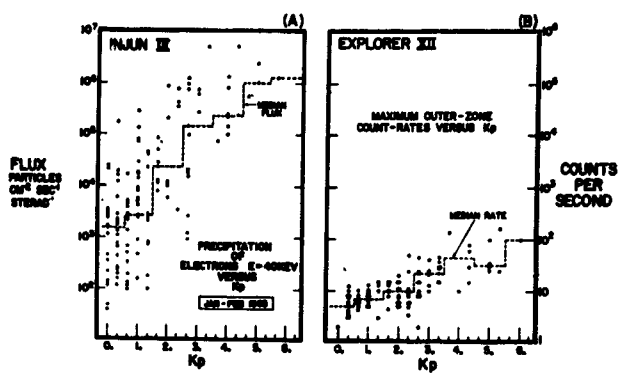


FIGURE 6.—Illustration that the flux of precipitated electrons (in A) varies more with K_p than does the omnidirectional flux (mainly of trapped electrons) in the equatorial plane (in B). Each point shows the maximum respective flux encountered on an outer-zone pass. (After O'Brien, 1964.)

Figure 6 together with the dependence of the omnidirectional flux above 40 kev as observed in the equatorial plane far from the earth by Explorer 12 (Freeman, 1963). As can be seen, the K_p dependence is very much larger close to the earth (at one end of the field line). If one assumes that both increases are due to a common acceleration mechanism it follows that it acts preferentially parallel to the geomagnetic field lines (O'Brien, 1964).

The change of omnidirectional flux above 40 kev with K_p, shown in Figure 6b, is opposite to what has mostly been observed for electrons of $E \gtrsim 2$ Mev (Arnoldy et al., 1960; Hoffman et al., 1962).

The dependence of the poleward boundary of precipitation on magnetic activity has been studied by Maehlum and O'Brien (1963). They used data for trapped electrons, but since O'Brien (1964) has shown that the precipitation has its maximum close to the poleward boundary of the trapped electrons, their results may be interpreted in terms of possible extension of precipitated electrons.

Maehlum and O'Brien (1963) found that during magnetic storms there was a very sharp boundary of the region where trapped electrons of energy above 40 kev could be observed. For this boundary, measured in L, they used the symbol L_n. It can be seen as a function of the K_p index during one geomagnetic storm in Figure 7. When K_p

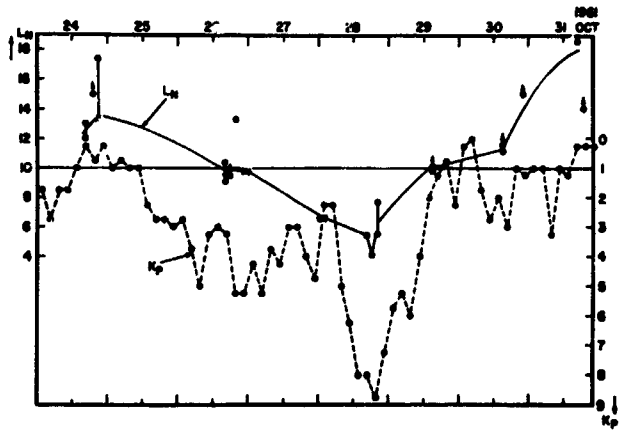


FIGURE 7.—Variation of the northern boundary of trapping with K_p during a geomagnetic storm, as observed by Injun I at a height of about 1000 km. (After Maehlum and O'Brien, 1963.)

reached its maximum value of 9, L_n had its minimum value of 4. Maehlum and O'Brien (1963) also found that the poleward boundary of strong radio wave absorption followed the L_n quite closely during the storm.

The effect of K_p on L_n is similar to the equatorward movement of the region of visual aurora during strong magnetic storms, which has been studied in the last few years in detail for some storms by Akasofu (1962, 1963a, b) and others.

(e) The Spectrum of Precipitated Electrons

Up to now only very rough measurements of the energy spectrum of the precipitated electrons have been reported. In most cases the proposed spectra have been obtained from two instruments with different energy characteristics. They are to be considered as rough equivalent spectra for

measurements made in a defined way. In addition, it has been found that the equivalent spectrum is highly variable both in space and time (cf. O'Brien et al., 1962). Nonetheless the available spectral data are of great interest at the present stage of knowledge in the field, and they even make it possible to draw some interesting conclusions about ionospheric effects of the precipitated electrons. These will be discussed in the following sections.

Most spectra that have been reported hitherto can be divided in two categories if they are expressed in an exponential form. On one hand, the e-folding value, b, in the spectrum of precipitated electrons, written in the form $\alpha \exp(-E/b)$, has been found often to be in the range 2–8 kev (McIlwain, 1960; Stilwell, 1963; Sharp et al., 1964a, b, c). This is a very steep spectrum but even steeper ones have been observed. McIlwain interpreted his rocket measurements in a strong aurora as indicating a monoenergetic flux of electrons with an energy of about 6 kev. Krasovskii et al. (1962) has observed steep spectra for precipitated electrons with a most common equivalent energy value of 14 kev.

The majority of the existing measurements have been made with Geiger tubes, for which the minimum detectable electron energy is about 40 kev. These measurements have mostly given e-folding values much higher than those mentioned above, namely, between 20 and 45 kev. (Davis et al., 1960; McDiarmid et al., 1960; O'Brien et al., 1962; Mann et al., 1963; O'Brien and Taylor, 1964). Mann et al (1963) found from measurements during only some 20 orbits that the values within this range were grouped in two classes, 25 ± 5 kev and 42 ± 3 kev.

The balloon measurements of X-rays at 30–35 km altitude also, in general, give inferred spectra of incident electrons of a fairly flat type, corresponding to e-folding energies in the range 20–45 kev (cf. e.g., Anderson and Enemark, 1960), although power law spectra often are found to fit the observations somewhat better than exponential spectra.

Recently the first direct measurements of precipitated electrons below 1 kev energy have been reported (Sharp et al., 1963; Evans et al., 1964). They showed that there is generally not very much energy flux below 1 kev. In some passages through the auroral zone Sharp et al. (1964c) found the integral fluxes above 0.18 and 10 kev to correspond to equivalent exponential spectra with e-folding values between 2 and 5 kev. On other occasions they did not find any energy flux at all below 1.5 kev. That there is not very much energy flux below 1 kev is also evident from observations of the luminosity distribution with height of aurorae. O'Brien and Taylor (1964) give an example of measurement results obtained on board Injun 3, according to which less than 0.1 percent of the auroral light originated above the satellite, that was at 250 km altitude. From this they conclude that the number of 10 ev electrons in the precipitated flux was no larger than the number of 10 kev electrons. It therefore seems probable that the spectrum, in general, does not increase very fast below 1 kev, which means that it is not of the power law type at these low energies.

Sharp et al. (1964c) found a significantly harder spectrum on the dayside than on the nightside. This observation agrees with multifrequency riometer measurements of auroral absorption by Lerfald et al. (1964), according to which the absorbing ionization is located lower down in the atmosphere in the day than in the night.

Of special interest are the statistical data on electron precipitation published recently. As mentioned earlier the precipitation is most intense near the auroral zone. O'Brien (1964) found an average flux of about 4.10^5 electrons/cm^2 sec in the precipitation cone at an invariant latitude of 65°. Injun 1 had a CdS detector, measuring electrons of energy above about 1 kev (O'Brien, 1962b). O'Brien and Taylor (1964) state that the flux of 4.10^5 electrons/cm^2 sec of energy greater than 40 kev was associated with an energy flux of 4 ergs/cm^2 sec for electrons of $E \gtrsim 1$ kev. These data should be considered as accurate to a factor of about three. Using these two values an equivalent electron spectrum in the range 1–40 kev can be deduced. It is found to be

$$n(E) = 7.8 \cdot 10^7 e^{-E/5.7} \text{ electrons/cm}^2 \text{ sec kev.}$$

For the energy values 40 and 250 kev similar average fluxes have been obtained from the Alouette measurements reported by McDiarmid et al. (1963). While the flux above 40 kev had its maximum at an invariant latitude of 65°, the

flux above 250 kev was maximal at about 60 degrees. McDiarmid et al. (1963) did not present any details for the latitudinal variation of this latter in integral flux. By applying a rough correction factor of 0.1 for the latitude variation of the average flux above 250 kev from 60 to 65 degrees invariant latitude, the following approximate equivalent spectra for the energy range 40–250 kev can be derived for two ranges of magnetic activity in the auroral zone ($\Lambda = 65°$), if isotropy is assumed for the upper hemisphere

$K_p < 4$: $n(E) = 1.2 \cdot 10^4 e^{-E/41}$ electrons/cm^2 sec kev

$K_p > 4$: $n(E) = 2.4 \cdot 10^5 e^{-E/30}$ electrons/cm^2 sec kev

(cf. Hultqvist, 1964b). The applied latitude correction is probably too large as judged from a comparison of the latitude distribution of passes in which the intensity of precipitated electrons with energies greater than 250 kev was greater than $3.2 \cdot 10^3$/cm^2 sec ster (Figure 8 of McDiarmid et al., 1963) and the corresponding diagram for

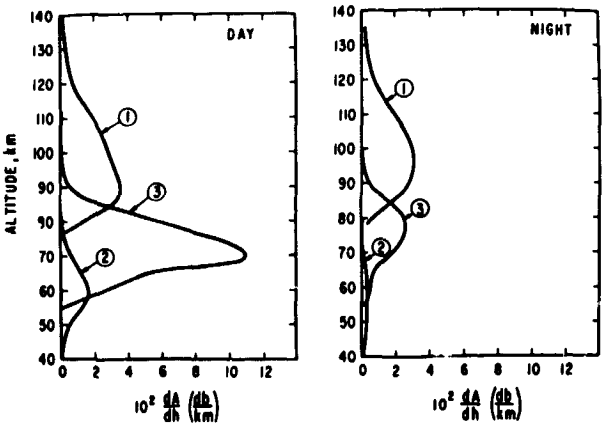


FIGURE 8.—(a) is for daytime and (b) for night. Curves in (a) and (b) show the height distribution of the absorption produced by the differential energy spectrum $N(E) = 5.10^8 e^{-E/5}$ kev electrons cm^{-2} sec^{-1} ster^{-1} kev^{-1}. The total absorption values corresponding to curves 1 amount to 1.04 db in the day and to 0.89 db at night. Curves 2 give the absorption due to the bremsstrahlung of the same electron spectrum. Total absorption in the day is 0.27 db and in the night 0.061 db. Curves 3, finally, represent the absorption distribution produced by the differential energy spectrum $N(E) = 7.10^4 e^{-E/41}$ kev electrons cm^{-2} sec^{-1} kev^{-1}, coming in along the field lines. Total daytime absorption under curves 3 is 1.9 db. The nighttime one is 0.52 db, if the height profiles of the ratio of negative ion to electron density used by Nicolet and Aikin (1960) and others are employed. 40 kev. The resulting spectra are thus too steep rather than too flat, and the absorption caused by electrons in the energy range 40–250 kev is probably larger than the values presented below.

As mentioned, the spectra of precipitated electrons are very variable. There is also an important latitude variation in the steepness, which increases with latitude (O'Brien et al. 1962; McDiarmid et al., 1963; O'Brien, 1964). On the other hand there seems to be no significant dependence of the spectrum on pitch angle or on the intensity of precipitation (O'Brien, 1964).

To obtain the equivalent spectra the averaging has been made over a large number of measured integral fluxes. With a variation of a factor of 10^5 the use of averages may be questioned. It is also questionable whether average integral fluxes are the best values to use when the interest is in the average electron density produced by the particle influx. The average electron density is then defined by the averaging process employed and this method of averaging is not identical with the averaging made in measuring the ionospheric absorption, for instance. It can, however, be shown that the average electron density produced by the precipitated electrons can be expressed in the average integral fluxes measured on the satellites. Since only two experimental integral flux values are available, one has to fit a two-parameter energy relation for the average flux to these experimental data (Hultqvist, 1964b), as has been done above.

QUANTITATIVE RELATIONS BETWEEN ELECTRON PRECIPITATION AND PHOTON EMISSION IN AURORA

Two direct measurements of the quantitative relation between precipitation and photon emission will be discussed here, namely those of McIlwain (1960) and O'Brien and Taylor (1964). It is of interest to compare their results with what is expected on theoretical grounds.

McIlwain did not use any filter, but measured the photon flux integrated over the transmission curve of the photomultiplier. If we assume that $\frac{1}{5}$ of the light was λ 3914Å, the measured photon flux corresponds to an emission rate of about 3 kilorayleigh at this wavelength. The fraction, $\frac{1}{5}$, is fairly arbitrary, but it seems reasonable (cf. e.g., Dalgarno, 1964) and is probably in error by

less than a factor of 2 and ½, respectively, for an ordinary auroral zone aurora. McIlwain's (1960) measured electron flux corresponds to an energy flux of 20 erg/cm² sec, if the spectrum obtained was extrapolated to $E=0$. Thus, the resulting electron energy flux required per unit of λ 3914Å emission rate is 7 ergs/cm² sec per *kR*.

O'Brien and Taylor (1964) reported an average λ 3914Å intensity of $2(+4/-1.5)$ *kR* at the maximum of the latitude distribution, i.e., in the auroral zone. As mentioned above, the electron measurements gave a corresponding average of the flux above 40 kev of $4\cdot10^5$ electrons/cm² sec and an energy flux for $E>1$ kev of about 4 ergs/cm² sec, which values they consider as accurate to a factor of about three. In order to get a value directly comparable with that of McIlwain one would have to extrapolate the energy spectrum from 1 kev down to $E=0$ and evaluate the total energy flux. Using the equivalent exponential spectrum, we find 5 ergs/cm² sec for $E>0$. Thus the Injun 3 results give 2.5 ergs/cm² sec per *kR*. Considering the uncertainties in measurements and the method of evaluation, this is a good agreement.

What energy flux per *kR* does one expect on the basis of existing knowledge about the emission processes? Omholt (1957, 1959), Chamberlain (1961), Rees (1963), and Dalgarno (1964), among others, have discussed this. It has been shown by Stewart (1956) that the ratio between the excitation cross section for the 3914Å band of N^+_2 and the total ionization cross section is constant, at least up to 200 ev energy, and has the value 0.02. Assuming that this value is true over the whole energy range of interest, one finds that 50 electron-ion pairs are produced for each 3914Å photon. Since the mean energy expended by fast electrons in nitrogen per electron-ion pair is 35 ev (at least for energies down to a few hundred ev; it is assumed that the figure is correct down to zero energy), we find that $50\times35=1750$ ev is dissipated per 3914Å photon. Since each photon has an energy of 3.2 ev the efficiency with which energy is converted into 3914Å radiation is $1.8\cdot10^{-3}$. When the initial energies of the fast electrons fall below perhaps 100 ev, these efficiencies must decrease sharply (Dalgarno and Griffing, 1958; Dalgarno, 1964). One *kR* of 3914Å photons corresponds to an energy influx of 2.8 ergs/cm² sec, if the efficiency figure $1.8\cdot10^{-3}$ is employed. Dalgarno (1964) has used the value $1\cdot10^{-3}$. With this conversion efficiency we find the energy flux requirement to be 5.1 ergs/cm^{-2} sec^{-1} per *kR*. The agreement between these values and the experimental ones of McIlwain (1960) and O'Brien and Taylor (1964) is better than expected, when the uncertainties in the analyses are taken into account.

The total efficiency of converting particle energy into photons of any energy is expected to be about 1 percent on theoretical basis (cf. Chamberlain, 1961). McIlwain found from his rocket measurements a value of only 0.2 percent while preliminary Injun 3 results point towards 1 percent (O'Brien and Taylor, 1964).

AURORAL ABSORPTION

Auroral absorption is in general defined on riometer records as all the absorption, generally irregularly varying, which is usually observed during magnetically disturbed conditions and which is not associated with either polar cap absorption (PCA), produced by high energy solar protons, or sudden cosmic noise absorption (SCNA), which is caused by solar ultraviolet radiation. The auroral absorption is thus defined as the remainder when one has excluded two well defined types of absorption. It is therefore not surprising that it recently has been found to contain a number of phenomena differing, for example, in the energy characteristics of the ionizing agent (Ansari, 1963). The name may, however, be motivated by the fact that there is a fairly good correlation between auroral absorption and the general level of magnetic disturbance (cf. e.g., the review by Hultqvist, 1963c).

It has long been known that some of the auroral absorption is located below the *E*-layer, i.e. in or below the *D*-layer. Such evidence arises, for instance, from ionosondes, which frequently are blacked out during magnetic storms. This shows that all the radio energy is absorbed between the earth's surface and the *E*-layer. It has not been clear, however, whether all the absorption, for sufficiency high frequency cosmic noise, takes place below the *E*-layer or whether there may be as much or even more absorption at *E*-layer heights. In fact, there are reasons to believe that the absorption produced by those electrons which

are the cause of the visible aurora should be located mainly up in the altitude range where the aurora occurs. Evidence has recently been presented against the hypothesis of Chapman and Little (1957) that the main part of the radiowave absorption is produced below 90 km by X-rays from the primary auroral electrons. The absorption due to primary electrons of an energy spectrum as measured in visual aurora by McIlwain (1960), is probably at least an order of magnitude greater than that due to the bremsstrahlung X-rays the electrons give rise to (Ansari, 1963; Hultqvist, 1963, 1964; Brown, 1964).

The height of the main part of the auroral absorption became an important parameter for the understanding of the electron reactions in the lower ionosphere when it was found that there is a much smaller influence of the sunlight on the absorption value than was expected on the basis of existing models (cf. e.g., the review by Hultqvist, 1963c). A very small sunlight effect would be expected if all auroral absorption took place up in the height interval of the ordinary visible aurora. But such a height distribution would be in conflict with a large body of other experimental observations (cf. e.g., Hultqvist, 1963a, b, c). It would appear that the satellite measurements reported by Mann et al. (1963), McDiarmid et al. (1963), O'Brien (1964) and O'Brien and Taylor (1964) give a solution of this problem.

An important satellite observation in this respect is that of Mann et al. (1963) that there occurs frequently beside the soft type of spectrum of the precipitated electrons, observed in visible aurora by McIlwain (1960), a harder type with the e-folding energy between 25 and 45 kev. The averages of the extensive statistical satellite data of McDiarmid et al. (1963) and O'Brien (1962b, 1964) correspond to equivalent spectra given on page 11. Thus the e-folding energy is about McIlwain's value for the low energy range and is appreciably higher in the higher range.

Having these average spectra it is possible to calculate how much absorption is produced by the electrons below and above 40 kev energy. This has been done by Hultqvist (1964a, b). Figure 8 shows the height distribution of absorption per km due to the following spectra (Hultqvist, 1964a)

Curves No. 1: $5 \cdot 10^8 e^{-E/5}$ electrons/cm² sec ster kev corresponding to a very strong visual aurora of international brightness coefficient between III and IV;

Curves No. 2: Due to the bremsstrahlung X-rays produced by the electrons giving rise to curve No. 1;

Curves No. 3: $7 \cdot 10^4 e^{-E/41}$ electrons/cm² sec ster kev which was observed over the auroral zone by Mann et al. (1963).

It can be seen from Figure 8 that although curve No. 1 corresponds to a very strong aurora, the riometer absorption due to the hard electron spectrum (curve 3), which has its peak as low as 70 km altitude, is larger. The total absorption for daytime was found to be 0.9 db for the soft electron spectrum and 1.9 db for the hard one. Even if the uncertainty in the computed absolute absorption values is fairly high (cf. Hultqvist, 1964a) it is probably less than a factor of three for the relative values due to the various spectra.

The profiles shown in Figure 8 can, with proper absorption scale, be used for estimating the absorption caused by the average precipitation spectra observed by the Injun and Alouette satellites. The e-folding value in the range 1–40 kev was found to be 5.7 kev (see page 11). This is very close to McIlwain's (1960) value 5.0 kev. In fact, the spectrum $2.4 \cdot 10^8 e^{-E/5}$ electrons/cm² sec kev contains the same number of electrons above 40 kev as the average observed by Injun 3 ($4 \cdot 10^5$ electrons/cm² sec; O'Brien and Taylor, 1964). It corresponds to a total energy flux above 1 kev about twice as high as the average value (4 ergs/cm² sec) given by O'Brien and Taylor (1964). However, the corresponding energy flux per kilorayleigh is about 5 ergs/cm² sec which is equal to the average of the values found by McIlwain (1960) and O'Brien and Taylor (1964).

Using the equivalent spectrum $2.4 \cdot 10^8 e^{-E/5}$ electrons/cm² sec kev for the 1–40 key range and the $1.2 \cdot 10^4 e^{-E/41}$ electrons/cm² sec kev for 40–250 kev during low magnetic activity ($K_p < 4$), the electron density profiles 3 and 1 in Figure 9 are found. The total absorption produced by these two spectra are 0.09 db and 0.32 db, respectively.

For $K_p > 4$ the average fluxes of McDiarmid et al. (1963) above 40 and 250 kev correspond to the exponential spectrum $2.5 \cdot 10^6 e^{-E/30}$ electrons/cm² sec kev. For this spectrum only an upper limit of the electron density and absorption can be obtained with the use of Figure 8. Curve 2 in

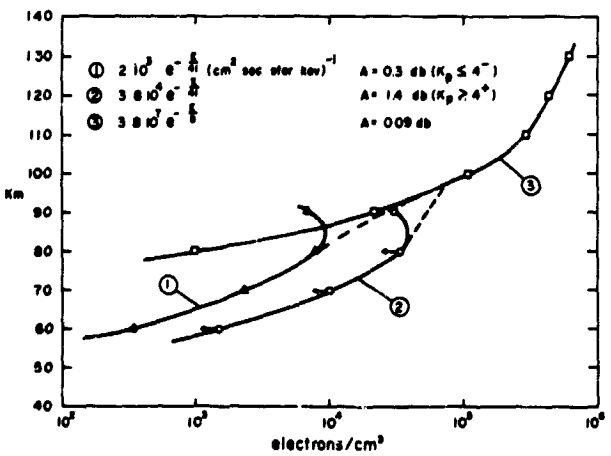


FIGURE 9.—Electron density profiles due to three different energy spectra of pecipitated electrons. The spectra as well as the total absorption, A, at 27.6 are given in the figure. Curve No. 2 gives an upper limit for the equilibrium electron density distribution produced by the spectrum $3.8 \cdot 10^4 \epsilon^{-E/30}$ (cm³ sec ster kev)$^{-1}$. A lower limit for this spectrum is curve No. 1.

Figure 9 shows this upper limit for the electron density. The corresponding upper limit of the total absorption is 1.4 db. Lower limits for the average electron density distribution and absorption are the values given for $K_p < 4$, i.e., curve 1 in Figure 9 and the corresponding absorption value of 0.32 db. The total riometer absorption due to the average electron spectrum in the energy range 40–250 kev is thus found to be about 1 db during the moderately disturbed conditions for which the observations are representative.

Two assumptions were made in the computations described above, which both tend to increase the relative importance of the absorption caused by precipitated electrons in the range 1–40 kev. One is that the equivalent average spectrum for the energy range 40-250 kev is probably less steep than the spectra used in the calculations. The other is that the equilibrium relation between electron density and the ionization rate was used. The computed absorption is therefore certainly overestimated. Since the effective recombination coefficient decreases from 60 km upwards the overestimation will be most important at greater heights. It does not seem probable that the error in the average absorption per unit height in the lowermost part due to unequilibrium would exceed a factor of two, since the rate of variation of the electron content, as seen on riometer records, mostly is slow compared with the recombination time in that part of the ionosphere. It is thus probable that the relative contribution of the absorption caused by the low energy electrons (corresponding to curve no. 3 in Figure 9) is smaller than shown by the figures given above.

Since the average absorption values produced by "soft" and "hard" electrons differ by an order of magnitude at least in the day, it seems possible to draw some conclusions on their basis.

Most of what is called auroral absorption seems to take place below 90 km. Hence the reason for the observed very small influence of sunlight on the intensity of the absorption is *not* that the absorption is located mostly above 90 km altitude, (where, according to virtually all models that have been proposed in the last few years, the density of negative ions is negligible in both day and night). One apparent interpretation of this observation is that the negative ions in fact are negligible, even in the night, down to about 60 km (as proposed by Hultqvist, 1962, 1963a, b). In any case it seems that the electron chemistry in the lowest ionosphere is quite different from what has hitherto been believed.

The virtual absence of sunlight influence on auroral absorption taking place mainly in the range 60–90 km thus leads to the conclusion that photo detachment is not important in that height interval. If this is true, the problem cannot be solved by assuming that the detachment of electrons from negative ions, in the height range mentioned, requires ultraviolet radiation, as proposed and discussed by Eriksen et al. (1960) and Reid and Leinbach (1962), and Reid (1961).

The result that the auroral absorption is produced mainly by electrons of energy above 40 kev, while the visible aurora is caused by lower energy electrons (cf. e.g., McIlwain, 1960) makes the fairly poor observed correlation between visible aurora and absorption understandable. By combining low-energy and high-energy electron spectra in various proportions, one may expect to see visible aurora practically without absorption, strong absorption without visible aurora, and all combinations in between. This is what has been found in the detailed study by Ansari (1963). There is a different local time dependence of the occurrence frequency of the low and high energy parts of the precipitated electron spectra. The auroral absorption has a pronounced maximum

in the auroral zone in the morning hours, whereas the visual aurora has its maximum in the middle of the night. So there seems to be some mechanism that accelerates the electrons on the average to higher energies over the morning side of the auroral zone.

The quantitative estimations of average absorption values presented above refer to the auroral zone. As has been pointed out earlier there is a tendency for the electron spectrum to grow steeper with increasing latitude and the conclusions drawn may not be true far inside the auroral zone. On the other hand the conclusion that auroral absorption is located mainly below 90 km is probably valid at subauroral latitudes. This is indicated also by the observed good correlation between bremsstrahlung X-rays and visual aurora at these latitudes (cf. e.g., the review by Winckler, 1962).

The deductions in this section are based on average fluxes and average spectra derived from widely varying, measured electron fluxes at two energies each, only. Conclusions should certainly be drawn from them with great care (cf. also page 12), but it seems that the order of magnitude results presented above should be fairly significant.

The height distribution of auroral absorption is quite similar to what has recently been found by means of multifrequency riometer measurements by Lerfald et al. (1964). The multifrequency absorption measurements can, however, provide height information only in the interval 35–75 km. A possible contribution from ionization up in the visible aurora in the presence of absorption also at lower altitudes cannot be measured by that technique, whereas the satellite measurements, reviewed above, clearly indicate that, on the average, most absorption is produced 90 km altitude.

The total absorption values obtained from the Injun and Alouette measurements are in quite reasonable accordance with the average absorption during the five most disturbed and five most quiet days per month, as observed over several years at College (geomagnetic latitude 64.5°) by Basler (1963). For the disturbed days he found a daily average of about 1 db in the summer and between 1 and 2 db in winter and at equinoxes—as compared to about 1 db given above for $K_p>4$. For the quiet days the daily average at College was about 0.3 db (to be compared to the value 0.3 db for $K_p<4$ above).

Figure 3 shows the fraction of passages for which the absorption would be estimated to be greater than about 0.2 db for high and low magnetic activity.

THE LOWER IONOSPHERE AT NIGHT

A number of hypothesis of corpuscular ionization of the ionosphere have been advanced to explain the existence of appreciable electron densities in the night and especially during the polar night (see Antonova and Ivanov-Kholodny, 1961). Mariani (1964) has invoked a corpuscular flux with a peak value of about 0.1 erg/cm² sec of electrons with energies of the order of 1 kev at geomagnetic latitudes between 55 and 65° to explain the correlation of maximum electron density in the *F*-region at noon with solar activity. Willmore (1964) has found evidence for a particle energy influx of the order of 20 percent of the UV photon flux from Ariel I measurements of temperatures in the *F*-region. He also found good agreement between the geographical distribution of the temperature and of particle fluxes observed by Sputnik II. Harris and Priester (1962) assume a corpuscular heat source comparable in influence to the ultraviolet solar radiation, i.e. of the order of a few ergs/cm² sec, in their model for the solar cycle variation of the upper atmosphere.

The electron density in the *F*-layer should decrease by a factor of 100 or more in 5–10 hours after sunset, at that altitude, if no ionization took place. As the observed decrease in general is only a factor of 10–20 in the altitude interval 120–200 km and 3–10 in the *F*-region, the obvious conclusion is that some nighttime ionization process exists. The need for such a process is still stronger in order to explain the fairly high electron densities observed over the central polar caps during the long polar nights. For instance, at the Pole Station (latitude 90°*S*) at midnight in the midwinter the *foF* was about 5 Mc/s in the high-solar activity year 1958 and a little above 3 Mc/s three years later. The nighttime *foF* was not very much less than the noon value in the summer. It is true that *foF* is not a good measure of the electron production rate in the upper ionosphere, especially not in the polar night, as dynamical effects within the ionosphere are very important, but

even so it can definitely be said that an ionization source must exist.

Does the observed flux of precipitated electrons contain enough energy to produce the lower ionosphere at night? It is not possible to give any accurate answer to that question now, since no extensive measurements of energy fluxes below 1 erg/cm² sec have been made in the lowest part of the electron energy spectrum of interest. Another piece of missing information is the energy flux associated with low energy protons. Krasovskii et al. (1963) have observed appreciable fluxes of positive ions with energy above 200 ev at low and medium latitudes ($\leq 49°$). The ion flux was usually about 10^8/cm² sec ster and the energy flux associated with it was probably ≤ 0.1 erg/cm² sec. It is not known whether the ions were trapped or precipitated. Above 100 km altitude not only precipitated electrons, as defined in the introduction, but also trapped ones contribute to the production of free electrons, and their contribution may even exceed that of the precipitated electrons. No observational results are yet available from which the energy dissipation into the atmosphere by protons can be evaluated.

It is obvious from Figure 9 that the average electron flux produces quite sufficient electron densities down in the *E*-layer in the auroral zone. A numerical calculation using the same spectrum as for curve 3 in Figure 9, for 200 km altitude gives an electron production rate of 1400 electrons/cm³ sec, corresponding to an equilibrium electron density of from $4 \cdot 10^5$ to $2 \cdot 10^6$/cm³, depending on the recombination coefficient used (Mitra, 1959, Van Zandt et al., 1960, or Schmerling and Grant, 1961). Therefore the corpuscular energy source is probably sufficient for production of the nighttime *F*-layer too in the auroral zone.

The electron densities shown in Figure 9 are astonishingly high. They correspond to critical frequencies higher than are normally observed in the auroral zones. They were derived, however, without taking dynamic effects into account and there may very well be uncertainties of a factor of four or so due to the inaccuracy in the experimental average particle fluxes, as well as in the atmospheric parameters used in the derivation. To this comes that the equilibrium relation has been used in computing the electron density from the ionization rate. Too large values of electron density and absorption is probably obtained in that way. On the other hand the ionization produced by electrons mirroring above 100 km has not been taken into account which counteracts the effect of the use of the equilibrium relation between electron density and ionization rate.

It is not known to this author whether the observed enormous variability of the precipitation fits into the time behaviour of the nighttime ionosphere.

What can be said about the ionosphere over the central polar cap in respect of observed particle energy dissipation? The average flux of precipitated electrons above 40 kev energy was found to be 10–100 times smaller than in the auroral zone at 76° invariant latitude by Injun 3 (O'Brien, 1964) and 100–1000 times smaller by Alouette (McDiarmid et al., 1963). Precipitation did not take place continuously well inside the auroral zones, but it varied in intensity from cosmic ray background to 10^5 electrons/cm² sec ster. Spectral information for the central polar caps is not available at the time or writing. The best that can be done to obtain an idea about the electron production rate over the polar caps from earlier presented results, seems to be to assume that the spectrum is about the same as used above for the auroral zone and to reduce the ionization rate in proportion to the difference between the two regions in counting rate found by Injun 3 and/or Alouette.

The corresponding energy flux for electrons of energy above 1 kev will thus be in the range 10^{-2} to 1 erg/cm² sec and the steady state electron densities 1/30 to 1/3 of those for curve no. 3 in Figure 9. For a 200 km electron density of 10^6 cm^{-3} in the auroral zone a corresponding electron density range of $3 \cdot 10^4 - 3 \cdot 10^5$ electrons/cm³ is obtained for the polar caps, neglecting all dynamical effects. The corresponding plasma frequency is between 1.6 and 4.9 Mc/s. These values should be compared with the *foF* value of about 3 Mc/s observed at the Pole Station during somewhat higher solar activities. At 100 km altitude the plasma frequency would be between half an Mc/s and one and a half Mc/s. Despite the roughness of this estimate it seems possible to state that the observed precipitation of electrons of energies above 1 kev must contribute significantly to the nighttime ionosphere over the polar caps and that one

cannot rule out the possibility that it is the main source.

At subauroral latitudes the average values found from the Injun 3 data (O'Brien, 1964) by making the same assumptions as for the polar caps are the following:

energy flux: $10^{-2}-10^{-1}$ erg/cm² sec,

equilibrium electron density at 200 km: $3\cdot10^{4}-10^{5}$ electrons/cm³

at 120 km: $1.5\cdot10^{4}-4.5\cdot10^{4}$ electrons/cm³

corresponding plasma frequency at 200 km: 1.6−2.8 Mc/s

at 120 km: 1.1−1.9 Mc/s.

Also in this case it seems possible to draw the same conclusion as for the polar caps: the observed precipitation of electrons of energy above 1 kev is certainly of major importance for the nighttime lower ionosphere in subauroral latitudes and it does not seem possible at the present time to exclude the possibility that it is the main source.

A problem with the precipitated-electron source of ionization is that it it is highly discontinuous in space and time. It is not known to this author if the time constants of the ionospheric processes can explain that much slower variations are observed in *foF*, for instance.

If the electron precipitation observed by the Injun and Alouette satellites is the main source of nighttime ionization, one expects a maximum for the average nighttime electron density in the auroral zone. It seems not to be clear if such a maximum exists or not. The situation is complicated by the frequent strong disturbances there, which tend to hide an average effect.

The average absorption at riometer frequencies corresponding to the given electron densities for polar cap and subauroral latitudes is below the lower measurable limit.

The electron density produced in and below the *D*-layer by the average precipitation fluxes described above is orders of magnitude greater than that due to the ordinary cosmic radiation (see e.g. Mohler, 1960) in the auroral zone and is probably of the same order of magnitude or is even larger than that due to cosmic rays in subauroral latitudes, if the height profile of the ratio of negative ion to electron densities of Nicolet and Aikin (1960) and others is valid.

The average energy flux of precipitated electrons observed by Injun 3 and Alouette meets the requirements of Mariani (1964) but is somewhat smaller than that deduced by Willmore and orders of magnitude smaller than that hypothesized by Antonova and Ivanov-Kholodny (1961) and Harris and Priester (1962).

AIRGLOW

The following questions about airglow in connection with particle precipitation are of great interest.

(1) What limits can be put on the particle energy flux in low and medium latitudes on the basis of airglow intensities?

(2) Is it possible that the observed precipitated electrons of energy above 1 kev are responsible for part of, or all airglow?

The second question has been answered with no by O'Brien (1962b) on the basis of energy considerations. The average zenith intensity of 5577Å at midlatitudes is about 250 rayleighs (Hunten et al., 1956) which requires a precipitation rate of the order of 1 erg/cm² sec. The precipitation rate observed at midlatitudes is equivalent to orders of magnitude smaller than this.

The first question has been discussed by e.g. Ivanov-Kholodny (1962), Galperin (1962), and Galgarno (1964). The N^{+}_{2} bands are produced by impact excitation. According to Roach (see Dalgarno, 1964) 60R of 3914Å emission would not occur undetected. This puts an upper limit of about 0.3 erg/cm² sec on the possible electron precipitation rate. Galperin (1962) has claimed that the observable limit can sometimes be put much lower and the corresponding limit of the energy flux has been given as $2\cdot10^{-2}$ erg/cm² sec by Dalgarno (1964).

The energy fluxes derived above from the satellite measurements reported by McDiarmid et al. (1963) and O'Brien (1964) amount to $10^{-2}-10^{-1}$ erg/cm² sec in subauroral latitudes. From this it seems that the limit given by Roach is a safe one, whereas the observed energy flux sometimes may exceed the limit given by Galperin (1962).

RELATION OF PRECIPITATED ELECTRONS TO THE OUTER RADIATION BELT

The idea about the relations between the electrons that are precipitated into the atmosphere and the trapped electrons in the radiation belts, which was prevailing one or two years ago, was roughly that the precipitated particles were dumped from the large storage of energetic trapped particles in the magnetosphere through the influence of disturbing effects caused by the solar plasma, as, for example, the magnetic disturbances. The number of trapped particles was thought to be sufficiently large to allow the precipitation rates observed to occur with only weak and perhaps independent processes of injection and acceleration required to replenish the particles in the radiation belt, the occurrence of which is necessary to be assumed in any case for understanding the existence of the trapped radiation.

This picture has been completely changed recently through the investigations by O'Brien (1962b, 1964) of the measuring results of Injun 1 and Injun 3 and also through a number of balloon studies of X-ray fluxes in the lower atmosphere (Winckler et al., 1962, Anderson, 1964).

O'Brien (1962b) estimated the lifetime of the trapped electrons in the outer radiation belt, assuming that the source was stopped but the loss mechanisms operated at the same rate as observed by Injun 1. He found that the outer zone beyond $L \sim 2$ would drain empty of electrons in a few hours. Similar average lifetimes have also been evaluated from X-ray measurements (see e.g., Winckler et al., 1962 and Anderson, 1964). Sometimes precipitation rates several orders of magnitude above the average (more than 1000 ergs/cm^2 sec) have been observed (Krasovskii et al., 1961, O'Brien and Laughlin, 1962, and Winckler et al., 1962). Thus Winckler et al. (1962) recorded an electron burst reaching 10^{11} electrons/cm^2 sec, a flux which would have used up the total energy of trapped particles in the field tube in half a second, but yet it persisted for about 100 seconds. Some efficient acceleration mechanism, which can reach full efficiency in a fraction of a second, seems required.

O'Brien (1964) has also shown that the flux of trapped radiation increases when precipitation takes place instead of diminishing, which would be expected if the trapped particles were simply dumped into the atmosphere. An acceleration mechanism seems to influence precipitated and trapped particles simultaneously.

Finally, O'Brien (1964) has demonstrated that the precipitation is highly energy dependent. There was no significant precipitation or change in the flux of trapped electrons of energy greater than 1.5 Mev observed in the middle of a strong burst of electrons of energy above 40 kev. This demonstrates that the precipitation could not be due simply to a lowering of the mirror point through a decrease in the geomagnetic field, since such a mechanism would be active over the whole spectrum.

Taken together, this new evidence clearly demonstrates that the precipitated electrons which produce aurora and ionospheric ionization are not produced by dumping trapped electrons into the atmosphere. An acceleration mechanism must be involved. This mechanism must be one that can act with full strength very quickly (in a fraction of a second) and it should not change the flux of Mev electrons more than 10 percent, when the simultaneous changes for trapped electrons of energy above 40 kev is a hundred-fold and for the precipitated electrons above 40 kev is more than three orders of magnitude. O'Brien (1964) mentioned acceleration in an electrostatic field directed along the magnetic field lines and involving a voltage drop of the order of 10 kilovolts as one possible mechanism from the precipitation-observation point of view.

The development in the last one or two years has thus in some respects brought the model of the auroral mechanism back into the situation that existed before the discovery of the Van Allen belts. These belts have been demonstrated to be probably an effect of the same acceleration mechanism that precipitates particles into the atmosphere and produces aurora, but not the source of these particles. The "leaky bucket" model has been replaced by the "splash catcher" model, in the vocabulary of O'Brien.

ACKNOWLEDGEMENT

I am grateful to L. R. Davis, D. S. Evans, and R. A. Hoffman for reading the manuscript and proposing improvements.

REFERENCES

S. I. AKASOFU (1962) J. Atm. Terr. Phys. 24, 723.

S. I. AKASOFU (1963a) J. Geophys. Res. 68, 1667.

S. I. AKASOFU (1963b) J. Atm. Terr. Phys. 25, 163.

S. I. AKASOFU and S. CHAPMAN (1962) J. Atm. Terr. Phys. 24, 785.

K. A. ANDERSON (1964) Paper presented at Lockheed Symposium on Aurora in Palo Alto, Jan. 1964.

K. A. ANDERSON and D. C. ENEMARK (1960) J. Geophys. Res. 65, 3521.

Z. A. ANSARI (1963) Univ. of Alaska, Sci. Rep. No. 4, NSF Grant No. G 14133.

L. A. ANTONOVA and G. S. IVANOV-KHOLODNY (1961) Geomagn. and Aeronomy 1, No. 2 (English translation 1, 149).

R. L. ARNOLDY, R. A. HOFFMAN and J. R. WINCKLER (1960) J. Geoph. Res. 65, 1361.

R. P. BASLER (1963) J. Geoph. Res. 68, 4665.

R. R. BROWN (1964) Arkiv för Geofysik (in print).

R. R. BROWN, K. A. ANDERSON, C. D. ANGER and D. S. EVANS (1963) J. Geophys. Res. 68, 2677.

J. W. CHAMBERLAIN (1961) Physics of the Aurora and Airglow, Academic Press, N. Y.

S. CHAPMAN and C. G. LITTLE (1957) J. Atm. Terr. Phys. 10, 20.

A. DALGARNO (1964) Ann. de Géophys., 20, 65.

A. DALGARNO and W. G. GRIFFING (1958) Proc. Roy. Soc. A248, 415.

L. R. DAVIS, O. E. BERG and L. H. MEREDITH (1960) Space Research, p. 721 (Ed. Kalmann-Bijl) North Holland Publishing Co., Amsterdam.

R. N. DEWITT (1962) J. Geophys. Res. 67, 1347.

K. W. ERICKSEN, O. HOLT, and B. LANDMARK (1960) J. Atm. Terr. Phys. 18, 78.

J. E. EVANS, R. A. JOHNSON, R. D. SHARP and J. B. REGAN (1964). Paper presented at the American Geophys. Union Meeting in Washington, D.C, April 1964.

L. A. FRANK, J. D. CRAVEN, and J. A. VAN ALLEN (1964). Paper presented at the URSI-AGU-AAS Symposium on Solar-Terrestrial Relationships, April 20–21, Wash., D.C.

J. W. FREEMAN (1963) State Univ. Iowa Publ. SUI–63–20.

YU. I. GALPERIN (1962) Bulletin (Izvestiya) Acad. Sci. USSR, Geophys. Series No. 2, 174.

T. HARRIS and W. PRIESTER (1962) J. Geophys. Res. 67, 4585.

R. A. HOFFMAN, R. L. ARNOLDY, and J. R. WINCKLER (1962) J. Geophys. Res. 67, 4543.

B. HULTQVIST (1962) Kiruna Geophysical Observatory, Sci. Rep. No. 2, Contract No. AF 61(052)–601.

B. HULTQVIST (1963a) Planet. Space Sci. 11, 371.

B. HULTQVIST (1963b) J. Atm. Terr. Phys. 25, 225.

B. HULTQVIST (1963c) Radio Astronomical and Satellite Studies of the Atmosphere, p. 163 (Ed. Aarons) North-Holland Pub. Co., Amsterdam.

B. HULTQVIST (1964a) Planet. Space Sci. (in print).

B. HULTQVIST (1964b) Planet. Space Sci. (in print).

G. S. IVANOV-KHOLDNY (1962) Geomagnetism and Aeronomy (English translation) 2, 315.

R. G. JOHNSON, J. E. Evans, R. D. SHARP and J. B. REAGAN (1964). Paper presented at the American Geophysical Union Meeting in Washington, D.C., April 1964.

V. I. KRASOVSKII, I. S. SHKLOVSKI, YU. I. GALPERIN, E. M. SVETLITSKII, YU. M. KUSHNIR, and G. A. BORDOVSKII (1962) Planet. Space. Sci. 9, 27.

V. I. KRASOVSKII, YU. I. GALPERIN, N. V. GALPERIN, N. V. JORJIO, T. M. MULARCHIK, and A. D. BOLUNOVA (1963). Paper presented at symposium in Paris.

G. M. LERFALD, C. G. LITTLE, and R. PARTHASARATHY (1964) J. Geophys. Res. (in print).

C. G. LITTLE, G. M. LERFALD and R. PARTHASARATHY (1963). Personal communication.

B. MAEHLUM and B. J. O'BRIEN (1963) J. Geophys. Res. 68, 997.

L. G. MANN, S. D. BLOOM, and H. I. WEST, JR., (1963) Space Research III, p. 447. (Ed. Priester) North Holland Publishing Co., Amsterdam.

F. MARIANI (1964) J. Atm. Sci. 20, 479.

I. B. MCDIARMID, D. C. ROSE, and E. BUDZINSKI (1961) Can. J. Phys. 39, 1888.

I. B. MCDIARMID, J. R. BURROWS, E. E. BUDZINSKI, and M. D. WILSON (1963) Can. J. Phys. 41, 2064.

C. E. MCILWAIN (1960) J. Geophys. Res. 65, 2727.

A. P. MITRA (1959) J. Geophys. Res. 64, 733.

W. F. MOLER (1960) J. Geophys. Res. 65, 1459.

M. NICOLET and A. C. AIKIN (1960) J. Geophys. Res. 65, 1469.

B. J. O'BRIEN (1962a) J. Geophys. Res. 67, 1227.

B. J. O'BRIEN (1962b) J. Geophys. Res. 67, 3687.

B. J. O'BRIEN (1963) J. Geophys. Res. 68, 989.

B. J. O'BRIEN (1964) J. Geophys. Res. 69, 13.

B. J. O'BRIEN and C. D. LAUGHLIN (1962) J. Geophys. Res. 67, 2667.

B. J. O'BRIEN, C. D. LAUGHLIN, J. A. VAN ALLEN, and L. A. FRANK (1962) J. Geophys. Res. 67, 1209.

B. J. O'BRIEN and H. TAYLOR (1964) J. Geophys. Res. 69,45.

A. OMHOLT (1957) Astrophys. J. 126, 461.

A. OMHOLT (1959) Geophys. Publikasjoner 20 (11), 1.

M. H. REES (1963) Planet. Space Sci. 11, 1209.

G. C. REID (1961) J. Geophys. Res. 66, 4071.

G. C REID and H. LEINBACH (1962) J. Atm. Terr. Phys. 13, 216.

E. R. SCHMERLING and D. GRANT (1961) Penn. State. Univ., Sci. Rep. No. 147.

R. D. SHARP, J. E. EVANS, R. G. JOHNSON and J. B. REAGAN (1963). Paper presented at American Geophys. Union Third Western Nat. Meeting. Boulder, Dec. 1963.

R. D. SHARP, J. E. EVANS, R. G. JOHNSON, and J. B. REAGAN (1963). Paper presented at American Geophys. Union Third Western Nat. Meeting. Boulder, Dec. 1963.

R. D. SHARP, J. E. EVANS, R. G. JOHNSON, and J. B. REAGAN (1964a). Paper presented at COSPAR Fifth Space Sci. Symp., Florence, May 1964.

R. D. SHARP, J. E. EVANS, W. L. IMHOP, R. G. JOHNSON, J. B. REAGAN and R. V. SMITH (1964b). Paper submitted for publication to J. Geophys. Res.

R. D. SHARP et al. (1964c). Personal communication.

D. T. STEWART (1956) Proc. Phys. Soc. A69, 437.

D. E. STILWELL (1963) State Univ. Iowa Publ., SUI 63-28.

T. E. VAN ZANDT, R. B. NORTON, and G. H. STONEBOCKER (1960). Some Ionospheric Results Obtained During the International Geophysical Year, p. 43 (Ed. Beynon), Elsevier Pub. Co., Amsterdam.

A. P. WILLMORE (1964). Personal communication.

J. R. WINCKLER (1962) J. Res. Nat. Bur. Standards, 66D, 127.

J. R. WINCKLER, P. D. BHAVSAR and K. A. ANDERSON (1962) J. Geophys. Res. 67, 3717.

N64-

ON THE HEIGHT OF AURORAL ABSORPTION, II

BENGT HULTQVIST†

Statistical data from Alouette and Injun 3 giving the average precipitation rate of electrons in the auroral zone as a function of energy, are analyzed with regard to the average radio wave absorption produced by ionization by electrons in the energy ranges 1—40 kev and 40—250 kev. From the satellite data, it is found that the higher energy range for the primary electrons is the dominating one in producing absoprtion at riometer frequencies. From this it follows that the majority of auroral absorption takes place below 90 km altitude in the auroral zone.

INTRODUCTION

The observation of a very small difference between ionospheric absorption of cosmic radio noise associated with aurora and magnetic disturbances, so called auroral absorption, before and after sunrise and sunset as compared to what was expected from existing models has made the height of auroral absorption a parameter of great importance for the understanding of the electron reactions in the lowest ionosphere (cf. e.g. Hultqvist 1963 a, b, 1964). There would be fewer difficulties in obtaining consistency between different types of experimental results if the auroral absorption could be shown to take place mainly in the height interval where the light emissions are mostly produced, i.e. in the *E*-layer. Proposals have been offered that, in fact, this is what in general occurs (Brown and Barcus, 1963). Several kinds of observational data indicate, however, that most of the auroral absorption is caused by ionization well below the *E*-layer (cf. Hultqvist, 1962, 1963 b, 1964).

Recently more observations relevant to the height of auroral absorption have been obtained. Ansari (1963) has analysed more carefully than before the correlation between auroral type of absorption and visual aurora and has found that "auroral absorption" is not one well defined type of phenomenon but contains at least two absorption phenomena different with regard to the energy characteristics of the ionizing corpuscular radiation. While some radio wave absorption is well correlated with the intensity of visual aurora, especially in the early part of the night, he found that there is one type, occurring after local midnight, that is not correlated with the light intensity. This is the type of auroral absorption which is causing the morning peak of the diurnal variation of the absorption in the auroral zone. The low light emission per absorption unit indicates that the energy of the primary electrons is much higher than for the electrons that cause the visible aurora. The increased penetrability of the higher energy electrons produces the enhancements in ionization much deeper in the atmosphere.

New extensive multiple-frequency riometer measurements of auroral absorption (Little et al., 1963, Lerfald et al., 1964) have shown that some 60 percent of 112 studied cases had the maximum in the absorption-per-km of altitude for 20 Mc/s cosmic radio noise at an altitude higher than 72 km, while for some 10 percent of the cases the maximum absorption-per-km was located at or below 60 km altitude. The multi-frequency absorption technique provides height information in the altitude range 35–75 km. It cannot tell whether there is some significant amount of absorption taking place up in the *E*-layer in the presence of absorption also in the 35–75 km interval, for instance. But even so the fact that in some 40 percent of 112 auroral absorption cases studied the peak absorption-per-km was at or below 72 km taken together with the fact that bremsstrahlung x-rays are probably negligible in producing riometer absorption (Ansari 1963, Hultqvist 1963

*Goddard Space Flight Center Document X-611-64-155, June 1964.

†Goddard Space Flight Center, NASA—National Academy of Sciences—National Research Council Senior Post—Doctoral Resident Research Associate on leave of absence from Kiruna Geophysical Observatory, Kiruna C, Sweden.

c, 1964, Brown, 1964) is enough for demonstrating the importance of higher energy electrons penetrating well below the visible aurora. The above-mentioned new results were obtained from ground observations of cosmic noise absorption.

Measurements on satellites (Mann et al., 1963) of the primary particles which cause the absorption have shown that there sometimes exists an influx of electrons with a spectrum so flat that the absorption-per-km height-profile caused by them has its maximum at about 70 km (Hultqvist, 1964). The total absorption produced by these observed flat-spectrum-electrons is higher than that caused by the steep electron spectra of McIlwain (1960) type, observed in visible aurora, even for very strong aurorae. These direct observations of the primary particles thus showed that at least sometimes most of the auroral absorption is located below 90 km altitude. However, due to the short life time of the satellites from which the measurements were made (only a few days) no good statistical information about the occurrence frequency of the various types of primary electron spectra was obtained. It is the purpose of this note to point out that the measurements by means of Alouette (McDiarmid et al., 1963) and Injun 3 (O'Brien, 1964) of precipitation of electrons into the atmosphere make it possible to attain a rough idea about how important the electrons in the energy ranges 1–40 kev and 40–250 kev, respectively, are for producing the auroral absorption, and, therefore, to provide another piece of experimental information independent of the ground observation of cosmic noise absorption about the height distribution of the auroral absorption. This information, in principle, covers the whole height interval of interest, contrary to the multi-frequency riometer measurements.

SATELLITE OBSERVATIONS OF ELECTRON PRECIPITATION IN THE AURORAL ZONE

McDiarmid et al., (1963) have reported results of electron flux measurements on board the Alouette satellite during several hundred passes through the auroral zone at an altitude of about 1025 km in the period October 1962 through January 1963. The measurements of special interest here were made by means of two Geiger-Müller tubes having electron energy thresholds of 40 and 250 kev. McDiarmid et al., found that the electron spectrum becomes progressively softer and very variable above about 59° invariant latitude. The flux of electrons with energy greater than 40 kev precipitated into the atmosphere had its maximum at an invariant latitude of 65 to 67°. The data were analyzed for two different ranges of K_p, namely <4 and >4. The average flux of precipitated electrons at the maximum in the auroral zone was found to be 3.10^5 electrons/cm^{2}sec ster for the high K_p range and 10 times less for the low K_p range. The half value width of the latitude distribution was some 8 degrees for $K_p > 4$ and about 12 degrees for $K_p < 4$.

On the average, the peak intensity of precipitated electrons with energies greater than 250 kev occurred at an invariant latitude of 60° and was approximately $1.9 \cdot 10^3$ electrons/cm^{2}sec ster and $2.6 \cdot 10^3$ electrons/cm^2 sec ster for the low and high ranges of K_p, respectively. The latitude spread was found to be smaller than for 40 kev electrons.

Because no detailed latitude distribution of the flux of 250 kev electrons is given by McDiarmid et al., we will assume that the average fluxes at an invariant latitude of 65° is down by a factor of ten from the peak values mentioned above. This latitude correction is probably too large (i.e. the correction factor 0.1 is too small) as judged from a comparison of the latitude distribution of passes in which the intensity of precipitated electrons with energies greater than 250 kev was greater than $3.2 \cdot 10^3$/cm^{2}sec ster (Fig. 8 of McDiarmid et al., 1963) and the corresponding diagram for 40 kev. The resulting spectra are thus too steep rather than too flat. The absorption caused by electrons in the energy range 40–250 kev is therefore probably larger, rather than smaller, than the values presented below.

From the two integral flux values given above for each range of magnetic activity we can derive equivalent two-parameter spectra. For $K_p < 4$ an exponential equivalent differential spectrum is found to be $n(E) = 2.10^3 e^{-E/41}$ electrons/cm sec ster kev and for $K_p > 4$: $n(E) = 3.8 \cdot 10^4 e^{-E/30}$ electrons/cm^{2}sec ster kev.

An equivalent two-parameter average differential spectrum for the primary electrons in the energy range 1–40 kev can be obtained from the results of the Injun 3 measurements presented by (O'Brien, 1964).

The data were taken in early 1963 at altitudes between 237 and 2785 km. The inclination of the orbit of Injun 3 was 70.4° and the satellite thus passed over the central polar cap over northern Canada. O'Brien gives the average integral flux for precipitated electrons with $E \geq 40$ kev as 4.10^5 electrons/cm²sec in the auroral zone. This was obtained by means of a Geiger-Müller tube. A Cds-detector on Injun 1 gave the energy flux associated with electrons of energy greater than about one kev. The integral flux of 4.10^5 electrons/cm²sec corresponded to an energy flux of about 4 ergs/cm² sec. O'Brien and Taylor (1964) state that these data can be considered as accurate to a factor of about 3. Corresponding to these two average flux values is the average equivalent exponential spectrum $n(E) = 7.8 \cdot 10^7 e^{-E/5.7}$ electrons/cm²sec kev. This spectrum is fairly close to the spectrum ($\propto e^{-E/5}$) considered in some numerical detail by Hultqvist (1964). In fact $n(E) = 2.4 \cdot 10^8 e^{-E/5}$ electrons/cm²sec kev gives the same integral electron flux for $E \geq 40$ kev as $7.8 \cdot 10^7 e^{-E/5.7}$ electrons/cm²sec kev, but the corresponding energy flux for $E \geq 1$ kev is 9 ergs/cm² sec instead of 4. In fact the energy input rate per unit of light emission rate obtained with the value 9 ergs/cm²sec is 5 ergs/cm²sec per kR, which is midway between the corresponding values of McIlwain (1960) and O'Brien (1964) (cf. Hultqvist, 1964b).

RELATIONS BETWEEN AVERAGE INTEGRAL FLUXES MEASURED IN SATELLITES AND AVERAGE ABSORPTION VALUES

The radio-wave absorption for the riometer frequency 27.6 Mc/s in the auroral zone is given for the ordinary ray by the Appleton-Hartree expression as

$$A = 0.46 \int_0^\infty \frac{N_e \nu}{3.34 \cdot 10^{16} + \nu^2} dh \quad db,$$

where N_e is the electron density and ν the electron collision frequency. The temporal structure of the electron precipitation measured by Alouette and Injun 3 is unknown. However, if the equilibrium relation between N_e and the ionization rate, $q(h)$, is used for evaluation of electron density and absorption, overly large values will certainly be obtained. Since the effective recombination coefficient decreases from 60 to 100 km altitude the overestimation will be most important in the upper part of this height interval. The relative contribution to the absorption from the lowest part of the height interval mentioned will therefore be higher than what is found below in this note. It does not seem probable that the error in the average electron density in the lowermost part of the height range 60 to 100 km due to unequilibrium would exceed a factor of two, since the rate of variation of the electron density, as seen on riometer records, generally is slow compared with the recombination time in that part of the ionosphere. As such an inaccuracy does not invalidate the conclusions drawn in this note we thus write

$$N_e(h) = \{q(h)/\alpha_{eff}\}^{1/2} \quad \text{and}$$

$$\bar{A} = 0.46 \int_0^\infty \frac{\nu}{3.34 \cdot 10^{16} + \nu^2} \overline{\left\{\frac{q(h)}{\alpha_{eff}}\right\}^{1/2}} dh,$$

where the bars indicate average values for the time function in question.

The data obtained from the above-mentioned satellite measurements above the atmosphere are the following time-average values:

$$\int_{40}^\infty \overline{n(E)} dE = \overline{N(>40)}$$

electrons/cm²sec,

$$\int_{250}^\infty \overline{n(E)} dE = \overline{N(>250)}$$

electrons/cm²sec,

and

$$\int_1^\infty \overline{En(E)} dE = \overline{\phi(>1)} \quad \text{ergs/cm²sec.}$$

Here $\overline{N(>40)} = \frac{1}{m} \sum_{i=1}^{m} N_i(>40)$,

$$\overline{n(E)} = \frac{1}{m} \sum_{i=1}^{m} n_i(E) \quad \text{and}$$

$$\overline{En(E)} = \frac{E}{m} \sum_{i=1}^{m} n_i(E) = E\overline{n(E)}.$$

$N_i(>40)$ represents a single satellite measurement of the integral flux above 40 kev energy.

The average ionization rate at height h is given by

$$\overline{q(h)} = \frac{1}{Q}\frac{d}{dh}\int_0^\infty E\overline{n(E,h)}dE \qquad \text{electrons/cm}^3\text{sec},$$

even if the electron precipitation is not homogenous over a large area, due to the effect of the geomagnetic field. Q is the average amount of energy used in producing one electron-ion pair (35 ev in molecular nitrogen) and $n(E,h)$ is the omnidirectional differential flux of electrons of energy E at altitude h. For a given particle energy, E, and angular distribution outside the atmosphere, $n(E,h)$ can be written

$$n(E,h) = n(E)\cdot F(E,h)$$

where $n(E)$ is the omnidirectional differential flux outside the atmosphere and $F(E,h)$, which contains the information about the angular distribution outside the atmosphere, gives the attenuation in the atmosphere. Thus

$$\overline{q(h)} = \frac{1}{Q}\int_0^\infty En(E)\frac{dF(E,h)}{dh}dE$$

$$= \frac{1}{Q}\int_0^\infty E\overline{n(E)}\frac{dF(E,h)}{dh}dE,$$

where the last equality is true if $\frac{dF}{dh}$ is the same for all fluxes used in the averaging i.e. if the directional distribution of the electrons outside the atmosphere does not vary from one passage to another. This is, of course, not necessarily so, but there is some experimental support for angular distribution being identical, namely isotropic, at least in the cases of intense precipitation (O'Brien, 1964), i.e. in those measuring values that contribute most to the average. It was also pointed out by Hultqvist (1964 a) that the radio wave absorption produced by an $e^{-E/b}$ differential spectrum of electrons enetering the atmosphere vertically is smaller by less than 50 percent of its value than the absorption produced by an isotropic flux of identical energy distribution but of 2 π times greater omnidirectional flux value. The variation of the directional distribution of the precipitated electrons from one precipitation event to another does therefore probably not affect the average significantly, and it is neglected in the considerations below.

When the two energy ranges about which information was obtained from Injun 3 (1–40 kev) and Alouette (40–250 kev) are considered, it can be found by numerical computations that the contribution to $\bar{q}$ from particles outside the energy ranges is negligible.

Thus we can write, with the use of the mean value theorem for integrals, for the low energy range

$$\overline{q(h)}_{1-40} = \frac{1}{Q}\int_1^\infty E\overline{n(E)}\frac{aF(E,h)}{dh}dE$$

$$= \frac{1}{Q}\left\{\frac{dF(E,h)}{dh}\right\}_{E_1}\int_1^\infty E\overline{n(E)}dE$$

$$= \frac{1}{Q}\left\{\frac{dF(E,h)}{dh}\right\}_{E_1}\overline{\phi(>1)},$$

where E_1 is a value between 1 and ∞, and for the high energy range

$$\overline{q(h)}_{40-250} = \frac{1}{Q}\int_{40}^\infty E\overline{n(E)}\frac{dF(E,h)}{dh}dE$$

$$= \frac{1}{Q}\left\{E\frac{dF(E,h)}{dh}\right\}_{E_2}\cdot\overline{N(>40)}.$$

If we, for instance, adopt the approximate formula for $F(E,h)$ of Maeda (1963) and derive a two parameter average energy distribution from $\overline{N(>40)}$ and $\phi\ (>1)$ for the energy interval 1–40 kev and from $\overline{N(>40)}$ and $\overline{N(>250)}$ for the interval 40–250 kev, simple closed expressions can be derived for $\left\{\frac{dF(E,h)}{dh}\right\}_{E_1}$ and $\left\{E\frac{dF(E,h)}{dh}\right\}_{E_2}$.

If the differential energy distribution corresponding to the experimental average integral flux values are assumed to be exponential in both energy intervals the expressions are as follows:

$$\left\{\frac{dF(E,h)}{dh}\right\}_{E_1} = -2\pi\zeta(h)\frac{e^{1/b_1}}{b_1(1+b_1)}\int_1^\infty Ee^{-E/b_1}\frac{1}{\sigma(E)}Ei\left(-\frac{x}{\sigma(E)}\right)dE \qquad (4)$$

and

$$\left\{E\frac{dF(E,h)}{dh}\right\}_{E_1} = -2\pi\zeta(h)\frac{e^{40/b_2}}{b_2}\int_{40}^{\infty} Ee^{-E/b_2}\frac{1}{\sigma(E)}Ei\left(-\frac{x}{\sigma(E)}\right)dE \tag{5}$$

b_1 is the e-folding value in the energy interval 1–40 kev and b_2 in the range 40–250 kev. $1/\sigma\ (E) = 3.18\cdot10^6E^{-2.2}$ (Maeda, 1963) and $E_i\ (-x/\sigma)$ is the exponential integral defined by $-E_i(-y) = \int_y^\infty e^{-z}z^{-1}dz$. x is the atmospheric depth in g/cm², $\zeta(h)$ is the atmospheric density, and E is everywhere measured in kev.

We thus see that $\overline{q(h)}$ can be expressed in terms of the satellite-measured averages $\overline{\phi(>1)}$ and $\overline{N(>40)}$. But the average cosmic noise absorption is given in terms of $\overline{q^{\frac{1}{2}}}$ in (1) and not in terms of $\bar{q}$.

However,

$$(\overline{q^{\frac{1}{2}}})^2 = \left\{\frac{1}{n}\sum_{i=1}^{n}q_i^{1/2}\right\}^2 = \frac{1}{n^2}\left\{\sum_{i=1}^{n}q_i + \sum_i q_i^{1/2}\sum_{j\neq i}q_j^{1/2}\right\} = \frac{1}{n^{\frac{1}{2}}}\left\{n\bar{q} + (n-1)n\bar{q}^{\frac{1}{2}}\cdot\bar{q}^{\frac{1}{2}}\right\} = \frac{1}{n}\bar{q} + \left(1-\frac{1}{n}\right)(\bar{q}^{\frac{1}{2}})^2.$$

Thus

$$\overline{q^{1/2}} = (\bar{q})^{1/2}$$

if

$$\sum_{i\neq j} q_j^{1/2} = (n-1)\overline{q^{1/2}}.$$

Numerical calculations show that $\overline{q^{\frac{1}{2}}}$ and $(\bar{q})^{\frac{1}{2}}$ do not differ by as much as 50 percent even for a small number of terms and for ratios between the various q_i as high as 10^6.

We can therefore write with sufficient accuracy for the considerations below:

$$\overline{A_{1-40}} = \frac{0.46}{\overline{Q_{1/2}}}\left\{\overline{\phi(>1)}\right\}^{1/2}\int_0^\infty \frac{\nu\alpha_{\text{eff}}^{-1/2}}{3.34\cdot10^{16}+\nu^2}\left[\left\{\frac{dF(E,h)}{dh}\right\}_{E_1}\right]^{1/2}dh \tag{6}$$

$$\overline{A_{40-250}} = \frac{0.46}{\overline{Q^{1/2}}}\left\{N(>40)\right\}^{1/2}\int_0^\infty \frac{\nu\alpha_{\text{eff}}^{-1/2}}{3.34\cdot10^{16}+\nu^2}\left[\left\{E\frac{dF(E,h)}{dh}\right\}_{E_2}\right]^{1/2}dh \tag{7}$$

where

$$\left\{\frac{dF}{dh}\right\}_{E_1}$$

and

$$\left\{E\frac{dF}{dh}\right\}_{E_2}$$

may be obtained from (4) and (5), respectively. The average absorption is thus expressed in terms of the average integral fluxes measured by the satellites.

The integrands in (6) and (7) depend only on atmospheric properties and the average flux as function of energy and can therefore be evaluated from satellite-measured average spectra. If the average spectrum is defined only in two points, as is the case with the Injun 3 and Alouette results, an assumption has to be made about the shape of the spectrum between those points. Exponential spectra are used below. The main uncertainty in the results obtained is certainly due to the limited amount of available statistical data and the enormous variation that there is in it, while the uncertainty in the detailed shape of the spectrum between the measuring points probably introduces only minor uncertainties.

ELECTRON DENSITY PROFILES AND ABSORPTION DUE TO THE EQUIVALENT AVERAGE SPECTRA

The exponential electron energy spectrum for $K_p<4$ is exactly the same as the spectrum for which Hultqvist (1964) calculated the height profile. The equilibrium electron density curve No. 1 in Fig. 1 has been obtained from Hultqvist's

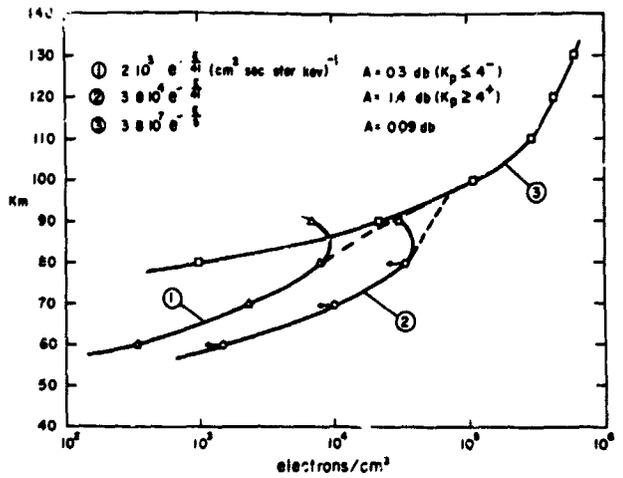


FIGURE 1.—Equilibrium electron density profiles due to three energy spectra of precipitated electrons. The spectra as well as the total absorption, A, at 27.6 Mc/s are given in the figure. Curve No. 2 gives an upper limit for the equilibrium electron density distribution produced by the spectrum $3.8 \cdot 10^4\ e^{-E/30}$ (cm² sec ster kev)$^{-1}$. A lower limit for this spectrum is curve No. 1.

(1964) results by multiplying by $\sqrt{2/70}$. Hultqvist's computations were made for vertically incident electrons of energy spectrum proportional to $\exp(-E/41)$. However, the absorption produced by $ce^{-E/41}$ electrons/cm²sec kev, entering the atmosphere vertically, is probably less than that caused by $ce^{-E/41}$ electrons/cm²sec ster kev, having isotropic directional distribution outside the atmosphere, by less than 50 percent of this higher value. The total absorption for a frequency of 27.6 Mc/s expected from the electron density curve No. 1 in Fig. 1 is 03.2 db.

An upper limit to the electron density profile and total absorption produced by an electron energy spectrum of $3.8 \cdot 10^4 e^{-E/30}$ is obtained from $3.8 \cdot 10^4 e^{E/41}$, which in turn can easily be found from the values computed by Hultqvist (1964). In this way curve No. 2 in Fig. 1 has been derived. The corresponding limiting total absorption for radio waves of frequency 27.6 Mc/s is 1.4 db. Probably the correct value for the spectrum with an e-fold value of 30 kev is close to 1 db. A lower limit for it is, of course, the value given above for $K_p < 4$ (0.32 db).

For the equivalent average spectrum for the range 1–40 kev, $2.4 \cdot 10^8 e^{-E/5}$ electrons/cm²sec kev, the numerical calculations of Hultqvist (1964) can be employed. The electron density profile No. 3 in Fig. 1 has been derived by means of them. The total absorption at 27.6 Mc/s due to this electron density profile is 0.09 db.

DISCUSSION

The increase of the pitch angle of an electron due to the magnetic forces when it passes along the field line down into the atmosphere has not been taken into account. As all measurements were made close to the earth's surface this neglection does certainly not introduce any large error. Its effect, if any, would be to increase the height of the ionization produced by the particle precipitation.

There are several simplifying assumptions and approximations made, which tend to make the deduced effect on the electrons in the energy range 40–250 kev too small relative to that of the 1–40 kev electrons. One is that Hultqvist's (1964a) numerical values for vertically penetrating electrons have been employed. As mentioned earlier, this may make the absorption too small by not more than 50 percent of the value for the case of isotropic electron angular distribution outside the atmosphere, and by still less of the value for the electrons in the loss cone. A second fact contributing to the computed absorption value for 40–250 kev electrons being too small is that the equivalent average spectrum in this energy interval is probably less steep than the spectra used in the numerical estimations. A third approximation working in the same direction is that the equilibrium relation between electron density and the ionization rate was used. This makes the computed absorption overestimated. Since the effective recombination coefficient decreases from 60 km upwards the overestimation will be most important at greater heights. It does not seem probable that the error in the average absorption per unit height in the lowermost part due to unequilibrium would exceed a factor of two, since the rate of variation of the electron content, as seen on riometer records, mostly is slow compared with the recombination time in that part of the ionosphere.

On the other hand there was an assumption made which tends to make the contribution by electrons in the 1–40 kev range too small compared to that of the 40–250 kev electrons. The average precipitated electron flux above 40 kev

energy was given by O'Brien as 4.10^5 electrons/cm²sec. It was assumed above that this omnidirectional flux was isotropic over the upper hemisphere just outisde that level where the energy loss of the electrons starts to be significant, i.e. at 200 km, say. This is probably not far from the truth, but the flux may have been limited to a somewhat smaller solid angle than 2π. Had this real angular distribution been known and taken into account the calculated absorption would have been somewhat greater.

Although only the order of magnitude of the absolute values of absorption seems to be significant, the ratio of contributions of the two energy intervals of precipitated electrons, discussed in this note, is probably uncertain by no more than a factor of three. A more accurate analysis than the one given above seems not worthwhile or possible on the basis of existing experimental data.

The electron density profiles in Fig. 1 as well as the total absorption values derived from them are those for the daytime. On the night side of the earth the electron densities may possibly be somewhat lower than those in Fig. 1 for altitudes less than 90 km due to negative ion production.

The values used as basis for the model spectra are average fluxes. The range of variation is, as mentioned, very large. O'Brien (1964) found that there was always some precipitation in the aurotal zone. The lowest value observed by him there was about $2 \cdot 10^{-4}$ of his average value quoted above. The maximum values shown in his paper were about 15 times greater than the average. To these minimum and maximum values of the ionization rate correspond extreme values of equilibrium electron densities 70 times smaller and 4 times greater than the average respectively. McDiarmid et al. (1963) showed similar variation ranges.

The use of exponential model spectra was motivated simply by the fact that numerical values were already available from Hultqvist (1964). It should be remembered that each of these spectra were obtained from only two pieces of experimental information and they are only very rough descriptions of spectra which are not known in detail. If an equivalent spectrum is to be used, as in this case, there seems to be at least as good reasons to use exponential ones as power law spectra. The observational indications of a fairly flat spectrum without any "infrared catastrophy" below a few kev (O'Brien, 1964, O'Brien and Taylor, 1964) support the use of exponential spectra in the low energy range.

The observations of both McDiarmid et al. (1963) and of O'Brien (1964) indicate a fairly strong latitude dependence of the spectral slope and the equivalent spectra used above should therefore be considered as representative only in the auroral zone.

The average total absorption values for 27.6 Mc/s given above for the various electron spectra:

(a) $n(E) = 3.8 \cdot 10^7 e^{-E/5}$ electrons/cm²sec ster kev $A = 0.090$ db

(b) $n(E) = 3.8 \cdot 10^4 e^{-E/30}$ electrons/cm²sec ster kev $(K_p > 4)$ $A < 1.4$ db

(c) $n(E) = 1.9 \cdot 10^3 e^{-E/41}$ electrons/cm²sec ster kev $(K_p < 4)$ $A = 0.32$ db,

indicate that the flat type of electron spectrum with e-fold values generally between 25–45 kev are the dominating one in producing auroral absorption in the average. The absorption due to these spectra takes place mainly between 60 and 90 km and not in the height interval where most visible auroae are located.

The range of variation of the 27.6 Mc/s absorption corresponding to the wide variations of electron fluxes reported by McDiarmid et al.(1963) and O'Brien (1964) is from the order of 0.01 db to 4 db, if smoothing due to time constants in the ionosphere is neglected, and is caused by the variation of the flux of the hard spectrum electrons. It should be remembered that there were no strong geomagnetic storms included in the material presented by McDiarmid et al. (1963).

The absorption values computed from the satellite measured electron fluxes for the two ranges of magnetic activity are in quite good accordance with the multifrequency riometer measurements mentioned in the introduction, as well as with the average absorption during the five most disturbed days and five most quiet days in each month, as observed over several years at College (geomagnetic latitude 64.5°) by Basler (1963). For the disturbed days he found a daily average of about 1 db in the summer and between 1 and 2 db in winter and at equinoxes. For the quiet days the daily average obtained from his figure 7 is about 0.3 db.

ACKNOWLEDGEMENT

I am grateful to A. C. Aikin, L. R. Davis, D. S. Evans, and R. A. Hoffman for reading the manuscript and offering criticism.

REFERENCES

Z. A. Ansari (1963) The spatial and temporal variations in high latitude cosmic noise absorption and their relation to luminous aurora. University of Alaska, Scientific Report No. 4, NSF Grant No. 14133.

R. P. Basler (1963) Radio wave absorption in the auroral ionosphere. J. Geophys. Res., 68, 4665–4681.

R. R. Brown (1964) Features of the auroral electron energy spectrum inferred from observations of ionospheric absorption. Arkiv for Geofysik (in print).

R. R. Brown and J. R. Barcus (1963) Day-night ratio for auroral absorption events associated with negative magnetic bays. J. Geophys. Res., 68, 4175–4180.

B. Hultqvist (1962) Kiruna Geophysical Observatory. Scientific Report No. 2, Contract No. AF 61(052)–601.

B. Hultqvist (1963a) Sunrise and sunset effects on cosmic noise absorption associated with aurora and magnetic storms. Planet. Space Sci. 11, 371-383.

B. Hultqvist (1963b) On the height distribution of the ratio of negative ion and electron densities in the lowest ionosphere. J. Atm. Terr. Phys. 25, 225–240.

B. Hultqvist (1963c) On the height of auroral absorption. Goddard Energetic Particles Preprint Series X-611-63-257.

B. Hultqvist (1964) On the height of auroral absorption, I. Planet Space Sci. (in print).

B. Hultqvist (1964b) Aurora and the lower ionosphere in relation to satellite observations of particle precipitation. Paper presented at the COSPAR Fifth International Space Science Symposium in Florence, May, 1964.

G. M. Lerfald, C. G. Little, and R. Parthasarathy (1964) D-region electron density profiles during auroras. J. Geophys. Res. (in print).

C. G. Little, G. M. Lerfald and R. Parthasarathy (1963). Personal communication.

L. G. Mann, S. D. Bloom and H. I. West, Jr. (1963) The electron spectrum from 90 to 1200 kev as observed on Discoverer satellites 29 and 31. Space Research III, pp. 1194–1205. North-Holland Publishing Co., Amsterdam.

J. B. McDiarmid, J. R. Burrows, E. E. Budzinski and M. D. Wilson (1963) Some average properties of the outer radiation zone at 1000 km. Can. J. Phys. 41, 2064–2079.

C. E. McIlwain (1960) Direct measurements of particles producing visible auroras. J. Geophys. Res. 65, 2727–2747.

C. E. McIlwain (1960) Direct measurements of particles producing visible auroras. J. Geophys. Res. 65, 2727–2747.

B. J. O'Brien (1964) High-latitude geophysical studies with satellite Injun 3.3. Precipitation of electrons into the atmosphere. J. Geophys. Res. 69, 13–143.

B. J. O'Brien (1964b) Paper presented at the Lockheed symposium on aurora in Palo Alto, Calif., January 1964.

B. J. O'Brein and H. Taylor (1964) High-latitude geophysical studies with satellite Injun 3. 4. Auroras and their excitation. J. Geophys. Res. 69, 45–63.

R. D. Sharp, J. E. Evans, R. G. Johnson and J. B. Regan (1963) Satellite measurements of low energy electrons in the northern auroral zone. Paper presented at the American Geophysical Union, Third Western National Meeting in Boulder, December 1963.

ON THE STRUCTURE OF A COLLISION FREE WAVE IN A PLASMA TRANSVERSE TO A MAGNETIC FIELD*

R. K. JAGGI AND A. P. STOKES

In this paper we have considered the problem of the structure of a strong collision free wave in a plasma perpendicular to its magnetic field. It is shown that conditions on the two sides of the wave are the same and are separated by a solitary pulse provided the kinetic energy of the motion at one end of the wave ($x=-\infty$, say) is less than the sum of magnetic and thermal energy of the plasma. If the kinetic energy is more than the sum of magnetic and thermal energy, no wave exists and the constant solution at $x=-\infty$ is the only solution of the equations describing the wave form.

1. INTRODUCTION

The subject of the structure of a strong collision free wave, in a plasma perpendicular to the magnetic field has been treated by Adlam and Alan[1], Burgers[2] and Hain, Lüst and Schlüter[3]. Hain, Lüst and Schlüter in particular have obtained solutions of the wave structure by solving the non-linear equations on a computer.

In this paper we have investigated mathematical properties of the equations which help us to guess the correct solution out of the many solutions obtained on a computer. As an illustrative example we shall consider a plasma with isotropic pressure. The generalization to the anisotropic case is quite straight forward. Burgers has made an extensive study of the wave structure when the plasma pressure is anisotropic and he arrives at the same general result that there is no change of state on the two sides of the wave, i.e., that no shock wave exists in a collisionless plasma perpendicular to the magnetic field.

The result of our paper is that a solitary wave only exists if the kinetic energy of motion, which produces the wave, is less than the sum of magnetic and thermal energy of the plasma; that no stationary wave exists if the kinetic energy of motion is more than the sum of magnetic and thermal energy of the plasma. In the latter case the constant solution at one end of the wave is the only solution of the differential equation (section 4).

*Goddard Space Flight Center Document X-640-64-34, February 1964.

2. EQUATIONS DESCRIBING THE WAVE FORM

Consider a stream of plasma at $x=-\infty$ with a constant velocity u along x-axis and a constant magnetic field B_1 along z-axis. The electrons and ions of the plasma will drift in the y-direction because of the magnetic field.

Our object is to obtain conditions at $x=+\infty$ and the form of the wave or waves which start from the given solution at $x=-\infty$. As a simplification we assume that the electrons and ions satisfy Maxwellian distribution function with variable number density, velocity components and the temperature. Therefore, if f_e, f_i denote the electron and ion distribution functions

$$f_e(x, v)=n_e(x)\left(\frac{m}{2\pi kT_e(x)}\right)^{3/2}\cdot$$

$$\exp\left\{-\frac{m}{2kT_e(x)}[(u-u_e(x))^2+(v-v_e(x))^2+w^2]\right\} \quad (1)$$

$$f_i(x, v)=n_i(x)\left(\frac{M}{2\pi kT_i(x)}\right)^{3/2}\cdot$$

$$\cdot\exp\left\{-\frac{M}{2kT_i(x)}[(u-u_i(x))^2+(v-v_i(x))^2+w^2]\right\} \quad (2)$$

In cartesian coordinates, assuming steady state, the Boltzmann equation can be written as

$$mu_e\frac{\partial f_i}{\partial x}-eE_x\frac{\partial f_e}{\partial u}-eE_y\frac{\partial f_e}{\partial v}-$$

$$-\frac{eB}{c}\left(v\frac{\partial f_e}{\partial u}-u\frac{\partial f_e}{\partial v}\right)=0 \quad (3)$$

for electrons, and

$$Mu\frac{\partial fi}{\partial x}+eE_x\frac{\partial fi}{\partial u}+eE_y\frac{\partial fi}{\partial v}+ +\frac{eB}{c}\left(v\frac{\partial fi}{\partial u}-u\frac{\partial fi}{\partial v}\right)=0 \quad (4)$$

for the ions. Multiplying by 1, u_e, v_e, u_e^2 and $u_e^2+v_e^2+w_e^2$ and generating moments of equation (3) we obtain

$$\frac{\partial}{\partial x}[n_e(x)u_e(x)]=0 \quad (5)$$

$$mn_eu_e\frac{du_e}{dx}+\frac{d}{dx}(n_ekT_e)+n_eeE_x+\frac{1}{c}n_eev_eB=0 \quad (6)$$

$$mn_eu_e\frac{dv_e}{dx}-n_eeE_y-\frac{1}{c}n_eeBu_e=0 \quad (7)$$

$$\frac{d}{dx}\left(u_e^2+\frac{3kT_e}{m}\right)+\frac{2eE_x}{m}+\frac{2eB}{mc}v_e=0 \quad (8)$$

$$\frac{d}{dx}u_e[mn_eu_e^2+mn_ev_e^2+5n_ekT_e]+ 2en_eu_eE_x-2en_ev_eE_y=0 \quad (9)$$

and similar equations for the ions. Let us call the similar ion equations as (5)′—(9)′. Maxwell's equations yield

$$E_y=\text{constant} \quad (10)$$

$$\frac{dE_x}{dx}=4\pi e(n_i-n_e) \quad (11)$$

$$\frac{dB}{dx}=-\frac{4\pi e}{c}(n_iv_i-n_ev_e) \quad (12)$$

To further simplify the problem, we will make the assumption that the space charge can be neglected. The characteristic length of space charge oscillations is of the order of the Debye length[4] $R_D=\left(\frac{kT}{4\pi ne^2}\right)^{1/2}$ while, as we shall see in this paper, that of magnetohydrodynamic oscillations is of the order of $c/w_p=c\sqrt{\frac{m}{4\pi Ne^2}}$. Thus assuming that the electron thermal velocity is well below the velocity of light, the space charge oscillations are of much lower scale than the oscillations we are interested in. With the assumption $E_x\simeq 0$ and $n_e=n_i=n$ (to a good approximation we then have $u_e=u_i=u$, say) equation (6), (6)′ and (8), (8)′ yield, respectively

$$\frac{d}{dx}[Mnu_i^2-n_ikT_i+nkTe]=-\frac{1}{8\pi}\frac{dB^2}{dx} \quad (13)$$

$$\frac{d}{dn}[Mu_i^2+3kTi+3kTe]=-\frac{1}{2\pi}\frac{B}{n}\frac{dB}{dn} \quad (14)$$

Equation (5) now gives

$$nu=n_1v_1 \quad (15)$$

where $\underset{x\to-\infty}{\text{Limit}}\, n(x)=n_1$ and $\underset{x\to-\infty}{\text{Limit}}\, u(x)=u_1$.

Equation (13) can be integrated and yields

$$Mnu^2+nkTi+nkTe+\frac{B^2}{8\pi}= Mn_1u_1^2+2nkT_1+\frac{B_1^2}{8\pi} \quad (16)$$

where $\underset{x\to-\infty}{\text{Limit}}\, B(x)=B_1$. Eliminating T_i+T_e from equations (16) and (14) we obtain

$$u\frac{d}{dx}\left[u\frac{dB}{dx}\right]+\frac{4\pi n_1e^2}{mc^2}u_1uB-\frac{4\pi n_1e^2}{mc^2}u_1^2B_1=0 \quad (17)$$

Dividing by $\sqrt{8\pi Mn_1u_1^2}$ and using the transformations

$$\frac{u}{u_1}=U,\ \frac{B}{\sqrt{8\pi n_1M}u_1^2}=H,$$

$$\frac{kT_{i,e}}{Mu_1^2}=\theta_{i,e},\ \beta_1^2=\frac{B_1^2}{8\pi n_1Mu_1^2} \quad (18)$$

we finally obtain the equations

$$H^2=\alpha-U-\frac{2\theta_1}{U^3} \quad (19)$$

and

$$U\frac{d}{dy}\left(U\frac{dH}{dy}\right)+UH-\beta_1=0 \quad (20)$$

where

$$\alpha=1+2\theta_1+\beta_1^2 \quad (21)$$

Equations (19), (20) satisfy the boundary conditions $H=\beta_1$, $U=1$ at $y=-\infty$.

3. A MATHEMATICAL DISCUSSION OF EQUATIONS (19), (20)

The pair $(U, H)=(1, \beta_1)$ is a constant solution of these equations. We shall study other constant solutions represented by (19), (20), but we first examine the stability of the solutions of (19), (20) in the neighborhood of $(1, \beta_1)$ and determine the nature of this critical point.

Let $U=1+U_1$, $H=\beta_1+H_1$. We then obtain from (19)

$$H_1=U_1\,F\,(H_1, U_1) \quad (22)$$

where

$$F(H_1, U_1)=\frac{2\theta_1(3+3U_1+U_1^2)-(1+U_1^3)}{(2\beta_1+H_1)(1+U_1)^3} \tag{23}$$

From (22) and (23) we obtain

$$H_1=b_1U_1+b_2U^2_1-b_3U_1^3+\cdots \tag{24}$$

where

$$b_1=F(0, 0)=\frac{6\theta_1-1}{2\beta_1} \tag{25}$$

$$b_2=F(0, 0)\frac{\partial F}{\partial H_1}(0, 0)+\frac{\partial F}{\partial U_1}(0, 0)$$

$$=-\frac{(6\theta_1-1)^2}{8\beta_1^3}-\frac{6\theta_1}{\beta_1} \tag{26}$$

etc.

Equations (20), (24) yield,

$$\frac{d^2U_1}{dy^2}+\gamma_1(U_1)\left(\frac{dU_1}{dy}\right)^2+KU_1+\gamma_2(U_1)=0 \tag{27}$$

where γ_1, γ_2 are analytic in U_1 near $U_1=0$, $\gamma_2(U_1)=0(U_1^2)$ and

$$K=1+\frac{\beta}{b_1}=\frac{1-2\beta_1^2-6\theta_1}{1-6\theta_1} \tag{28}$$

To determine the nature of the point $U=1$, we linearize equations (23), (27). The linearized form of equation (27) is

$$\frac{d^2U_1}{dy^2}+KU_1=0 \tag{29}$$

In all the discussion that follows we shall assume the flow velocity to be much higher than the thermal velocity of particles constituting the plasma, so that $6\theta_1<1$ or $1-6\theta_1>0$. The sign of k^1 is therefore determined by the expression $1-2\beta_1^2-6\theta_1$. In case I when $1-2\beta_1^2-6\theta_1>0$, k^1 is positive and therefore $U=1$ is a center in both (27) and (29) according to a known result of Lyapounov, see Malkin[5], pp 123–4, (as no odd powers of dU_1/dy occur). In case II when $1-2\beta_1^2-6\theta_1<0$, k^1 is negative; the point $U=1$ is a saddle point in both equations (27) and (29).

Let us now consider all the constant solutions represented by the equations (19), (20). According to equation (20) constant solutions satisfy the equation $HU=\beta_1$. This condition, together with equation (19) then gives the equation

$$f(U)\equiv U^4-U^3(1+2\theta_1+\beta_1^2)+\beta_1^2U+2\theta_1=0 \tag{30}$$

$f(U)$ may also be written as

$$f(U)=(U-1)[U^3-(\beta+2\theta_1)(U^2+U)-2\theta_1]$$
$$=(U-1)\,h(U) \tag{31}$$

From Descarte's rule of signs $h(U)$ cannot have more than one positive root. Also it is easy to see that it has a positive root. Let this positive root be denoted by U_o. Since $f(0)<0$, U_o will be less than or greater than 1 when $f'(1)>0$, $f'(1)<0$ respectively. We therefore again have two cases:

(I) $f'(1)=1-2\beta_1^2-6\theta_1>0$

and

(II) $1-2\beta_1^2-6\theta_1<0$.

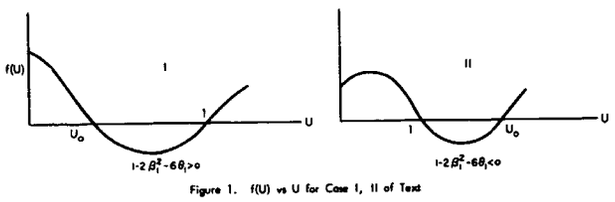


Figure 1. f(U) vs U for Case I, II of Text

FIGURE 1.—$f(U)$ vs U for Cases I, II of text.

As we require that $6\theta_1<1$, Case I may occur. Figure 1 shows a rough sketch of $f(U)$ versus U for cases I and II above.

To find the nature of the other critical point $U=U_o$, we write equation (20) as

$$\frac{d^2U}{dy^2}+\left(\frac{dU}{dy}\right)^2\left[\frac{Ug''(U)+g'(U)}{Ug'(U)}\right]+\frac{Ug(U)-\beta_1}{U^2g'(U)}=0 \tag{32}$$

where

$$H\equiv g(U)\left[a-U-\frac{2\theta_1}{U^3}\right]^{1/2} \tag{33}$$

$g'(U)$ has a zero at $U^*=(6\theta_1)^{1/4}$. It is interesting to note that this singularity of the differential equation can be quite close to 1. For example, for $\theta_1=.1$, $(6\theta_1)^{1/4}\simeq.880117$. Higher values of θ_1 can bring this singularity still closer to $U=1$.

When equation (32) is linearized about the zero U_o of $f(U)$, we obtain the equation

$$\dot{U}_1+L(U_o)U_1=0 \tag{34}$$

where $U=U_o+U_1$ and $UL(U)=1+\dfrac{2\beta_1^2}{U^2(6\theta_1-U^4)}$.

The nature of the point $U_o=1$ has already been discussed. Now let us assume in addition to

$6\theta_1<1$, that $U_o^4>6\theta_1$. We wish to establish for equation (32) that in case I when the point $U=1$ is a center the other equilibrium point $U_o<1$ is a saddle point and conversely, in case II when $U=1$ is a saddle point, the other equilibrium point $U_o>1$ is a center. By arguments similar to those given above for equation (27) and (29), this is equivalent to showing that if $h(U_o)=0$, h gvien in (31), then $U_o<1$ implies $L(U_o)<0$ and $U_o\cdot 1$ implies $L(U_o)>0$.

It can be shown that

$$Ug(U)-\beta_1=\frac{-f(U)}{U[\beta_1+Ug(U)]} \tag{35}$$

Now, consider case I. $f(U)$ has a zero at $U=U_o$ (<1) and $U=1$. In the $U,\dot{U}$ plane construct a closed loop C enclosing U_o and 1 and lying so close to the U-axis that the sign of $\dot{U}$ is determined by the last term on the left in (32).

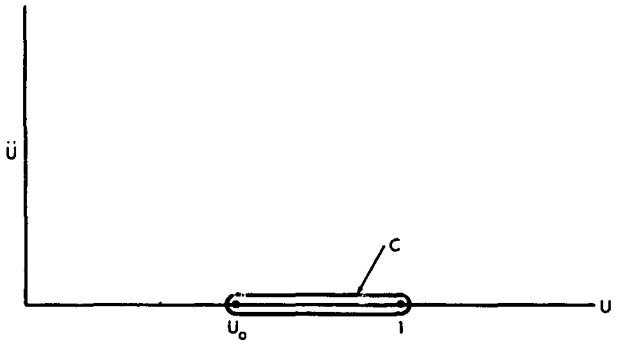


FIGURE 2.—The closed curve C enclosing the points $U=U_0$ and $U=1$ and lying close to the U axis.

(see Figure 2). Then using (35) and upon considering the changes in the sign of $f(U)$, it follows that the index of C with respect to (32) is 0[6]. As the index of a center $(U=1)$ is 1, this implies that the index of U_o is -1. If $L(U_o)\neq 0$, i.e., if U_o is an elementary critical point, then U_o of necessity must be a saddle point and $L(U_o)<0$ will hold. A similar arrangement for $U_o>1$ in case II shows that the index of U_o is 1 and therefore $L(U_o)>0$ holds (provided of course $L(U_o)\neq 0$).

Thus the problem is reduced to showing that the two polynomials $h(U)=U^3-bU^2-h_2-2\theta$, and $M(U)=U_6-6\theta_1U^2+4\theta_1-2b$ have no common positive zeros in either case I or case II. Here we have used the definition $b=2\theta_1+\beta_1^2$ in deriving the form of $M(U)$ which is the numerator of $UL(U)$.

This has not been shown in general, although for particular choices of β_1, θ_1, it is relatively straightforward. The critical case, however, is when $b+\theta_1=\frac{1}{2}$ or $2\beta_1^2+6\theta_1=1$, for then the two roots U_o and 1 become equal and, of course, in this case $M(1)=0$.

Let us suppose that $b+\theta_1$ is close to $\frac{1}{2}$, i.e., U_o is close to 1 so that $(U_o-1)^2$, $(U_o-1)^3$ are negligible in comparison with U_o-1. Now let $R=U^2$ and expand h and M about $U=1$ and $R=1$. This gives

$$h(U)=1-2(b+\theta_1)+3(1-b)(U-1)+(3-b)(U-1)^2+(U-1)^3 \tag{36}$$

and

$$M(R)=1-2(b+\theta_1)+3(1-2\theta_1)(R-1)+3(R-1)^2+(R-1)^3 \tag{37}$$

The zeros U_o, R_o $(=U^2-1)$ of h and M may be approximated by

$$U_0=1+\frac{2(b+\theta_1)-1}{3(1-b)}$$

$$R_0=1+\frac{2(b+\theta_1)-1}{3(1-2\theta_1)} \tag{38}$$

Now suppose we are in case I, that is $b+\theta_2<\frac{1}{2}$. Then as $b>2\theta_1$, we have $U_o-1<R_o-1=\bar{U}^2-1<0$; where $M(\bar{U})=0$, $\bar{U}>0$. This implies

$$1-U_o>(1-U)(1+U)>(1-U)$$

Therefore

$$U_o<\bar{U}<1$$

In case II, $b+\theta_2>\frac{1}{2}$, it follows similarly from $U_o-1>\bar{U}^2-1$, that $1<\bar{U}<U_o$. Thus the positive roots of M and R, which coincide for $b+\theta_1=\frac{1}{2}$, do separate as required, for $b+\theta_1$ near $\frac{1}{2}$. Although a proof for the general case is lacking, the following calculations may be of some interest.

Taking $\theta_1=.1$, $b=.3$, we find $(6\theta_1)^{1/4}=.880117$, $U_o\simeq.88978$ and $M(U_o)=-.1788<0$. With $\theta_1=.1$, $b=.5$, we find $U_o\simeq1.1116$ and $M(U_o)\cong.54567>0$. Thus we observe that at $b=.3$, U_o is very close to $U^*=(6\theta_1)^{1/4}$ and so b could not be decreased further and still have $U_o^4>6\theta_1$.

Thus we have seen that in case II, that is $U_o>1$, if $M(U_o)>0$, then U_o is a center. In this

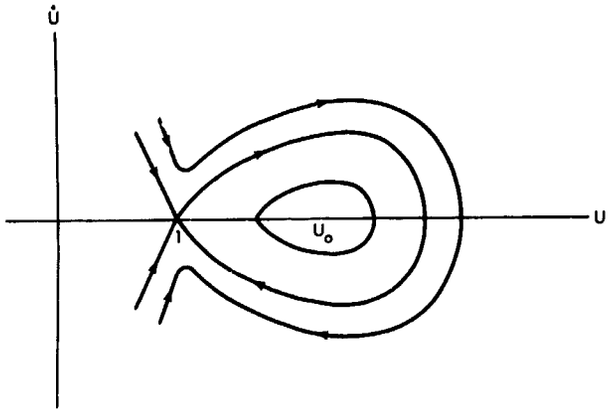


FIGURE 3.—Approximate phase plane picture of the solution for the case II when $U=1$ is a saddle point and $U=U_0$ is a center.

case the phase plane portrait of the solution is shown in figure 3. In the other case, in which U_0 is a saddle point and 1 a center, the phase plane picture is the same as figure 3 with the role of 1, U_0 reversed. As is clear from our analysis we have tried to determine the structure of the solutions of equations (19), (20) from the properties of the differential equations (27) or (32). In the next section, we shall verify our predictions by obtaining some solutions on a computer.

4. COMPUTER CALCULATIONS

Some of the results given in the previous section were verified by computing the solution of equations (19), (20) on an IBM 7094. The way to do this is to choose values of U, H (different from 1, β_1) lying on the eigen solution by using the linear equation (29). All calculations were started close to $U=1$, $H=\beta_1$. Since the solution depends critically on the position of the point of start in the phase plane we ran several cases using values slightly different from those obtained from the linear equation (29). For example, for $\beta_1=.81$, $\theta_1=.05$, we obtained the solution using (a) $U=1.0001$, $\frac{dU}{dy}=.00001$ and (b) $U=1.0001$,

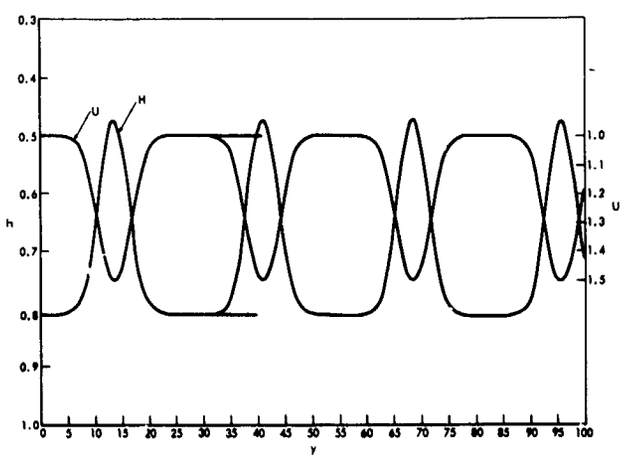


FIGURE 4a.—Computed solution of equations (19), (20) for $\beta_1=.81, \theta_1=.05$, using perturbed values $U=1.0001$, $\frac{dU}{dy}=100$.

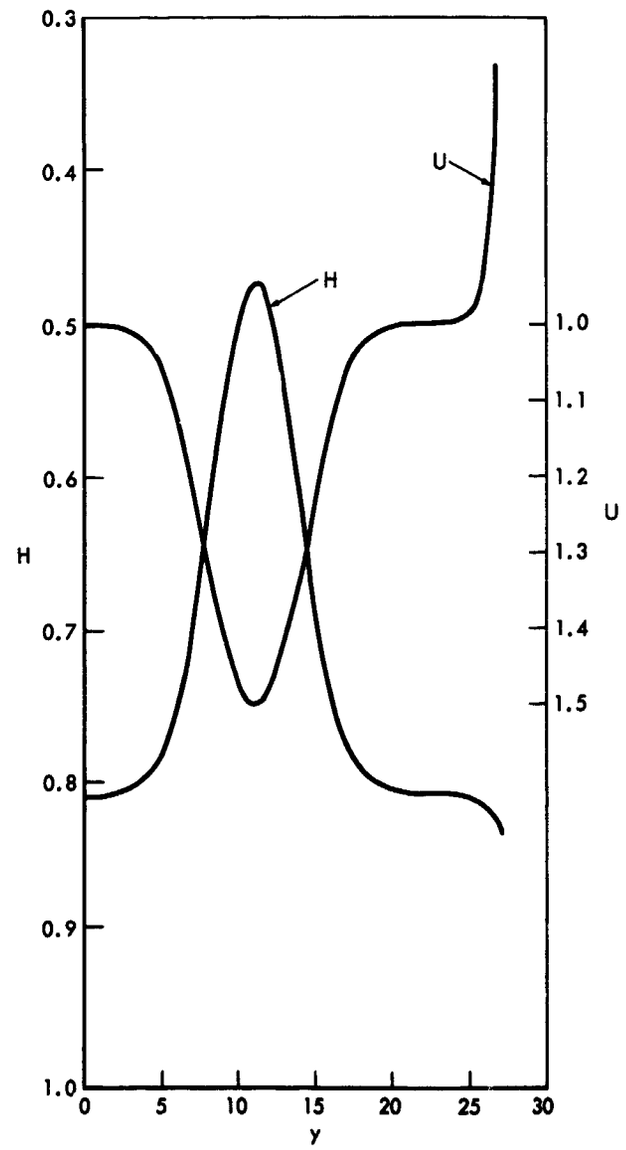


FIGURE 4b.—Computed solution of equations (19), (20) for $\beta_1=.81, \theta_1=.05$, using perturbed values $U=1.0001$, $\frac{dU}{dy}=.001$.

$\frac{dU}{dy}=.001$ while from the linear equation we obtain a value of $\frac{dU}{dy}$ which lies between .00001, .001. Figures 4a and 4b give the solution for the cases (a), (b) above. In the case (b) the solution becomes unstable because we happen to use a value of $\frac{dU}{dy}$ which lies outside the oval of figure 3. In the case (a), we obtain a solution which almost returns back to its initial value and repeats itself. (The repetitions are evidently the cause of the computational error, which if decreased will show that the solution is a solitary pulse. Since only two branches of the curve approach $U=1$, $U=0$ the solitary pulse is the required solution of the wave form.)

We now turn to a calculation of case I considered in section 3. We used $\beta_1=.1$, $\theta_1=.1$. Taking a point away from $U=1$, $\frac{dU}{dy}=0$, we obtained a solution consisting of oscillations having $U=1$ $\frac{dU}{yd}=0$ for its center in accord with the fact that $U=1$, $\frac{dU}{dy}=0$ is a center. We see therefore that in case (I) even the solitary wave does not exist.

We have not made any computations about the point $U=U_0$ as this point is of no interest to us. This corresponds to a solution lying just within the oval pictured in figure 3. A more accurate choice of initial conditions would show that the boundary of this oval corresponds to a solitary pulse, that is, the solution forming the boundary of the oval leaves $U=1$, $U=0$ at $y=-\infty$, and returns to this point at $y=+\infty$.

ACKNOWLEDGEMENTS

One of us (R.K.J.) is grateful to the National Academy of Sciences, National Research Council for the award of the fellowship during which this work was done. Part of this work (R.K.J.) has been supported by AEC under a subcontract 2220 with Oak Ridge National Laboratory.

REFERENCES

1. J. H. Adlam and J. E. Allen, Phil. Mag. 3, 448 (1958)
2. J. M. Burgers, Rev. Mod. Phys. 32, 868 (1960)
3. K. Hain, R. Lust and A. Schluter, Rev. Mod. Phys. 32 967 (1960)
4. O. W. Greenberg and Y. M. Treve, Phys. Fluids 5, 769 (1960)
5. I. G. Malkin, "Theory of Stability of Motion." AEC-tr-3352 (Russian Edition 1952)
6. V. V. Nemytskii and V. V. Stepanov, "Quantitative Theory of Differential Equations," Princeton University Press, 1960 (pp 125–130)

N66 37961

EVIDENCE FOR RAPID MOTION OF THE OUTER BOUNDARY OF THE MAGNETOSPHERE*

ANDREI KONRADI AND RICHARD L. KAUFMANN†

A sudden increase of the particle intensity and of the magnetic field strength on September 25, 1961 at 17:31 U.T. was detected by the ion-electron detector and the magnetometer on board Explorer XII outside the earth's magnetosphere. From the physical extent of the region and the energy of the trapped protons the temporal nature of the event was established. Similarities between particle fluxes as well as between magnetic fields during the increase and within the outer belt make it probable that the increase represents an impulse-like extension of at least a portion of the outer belt. A deduced lower limit on the outward radial velocity of the extension is 156 km/sec. Crude calculations show that either the proton density near the outer edge is less than 100 cm^{-3} or the temperature is greater than 15 ev.

INTRODUCTION

The Explorer XII satellite was well suited for the study of the outer boundary of the trapping region. With an apogee of over 13 R_e (earth's radii) it usually penetrated the boundary on the sunward side twice during each orbit. The position of the boundary ranged from about 8 R_e to beyond the apogee of the satellite with an average value of somewhere about 10 R_e (Cahill and Amazeen, 1963).

Data from the ion-electron detector flown by Davis and Williamson and from the magnetometer flown by Cahill on board Explorer XII has yielded more than 100 transitions of the outer boundary.

For our purposes a transition shall be defined to occur when on an inbound pass the output current of the photomultiplier, measured in the mode which is most sensitive to the incident energy flux, rises significantly above the dark current (usually at least by an order of magnitude) in a time interval of the order of minutes corresponding to a radial distance of about one hundred kilometers covered by the satellite. The reverse is of course true for an outbound pass (Davis and Williamson, 1963). Similarly, a transition of the boundary is indicated by the magnetometer when there is usually an abrupt change in either both the magnitude and direction of the magnetic field, or just a change in the direction of the magnetic field. Several typical transitions are shown in Figures 1a and 1b. As can be seen from Figure 1a,

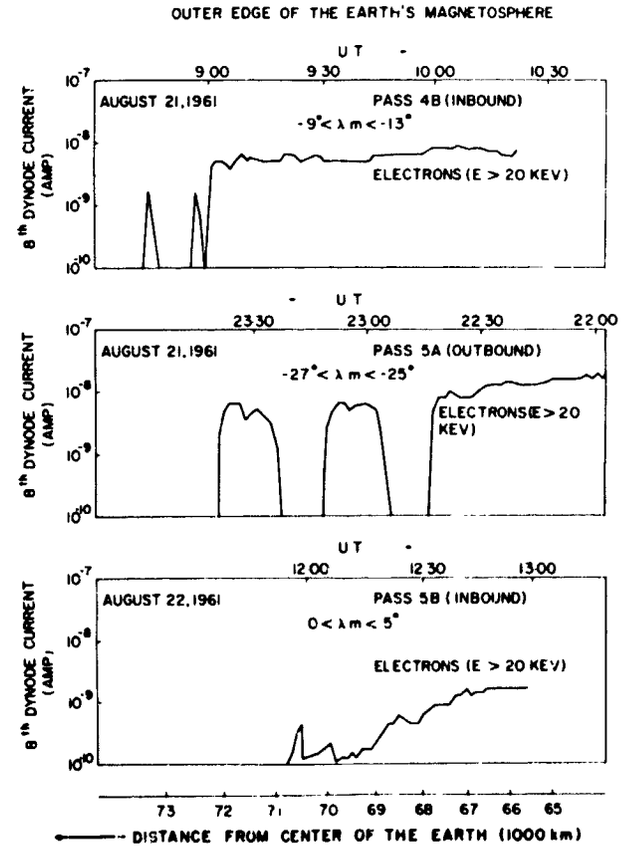


FIGURE 1a.—Three samples of Explorer XII transitions of the outer boundary of the trapping region as seen by the ion-electron detector. The 8th dynode current is a measure of the electron intensity for electrons with 20 kev < E < 1 Mev.

*Published as *Goddard Space Flight Center Document X-611-64-297, October 1964.*

†University of New Hampshire, Durham, New Hampshire.

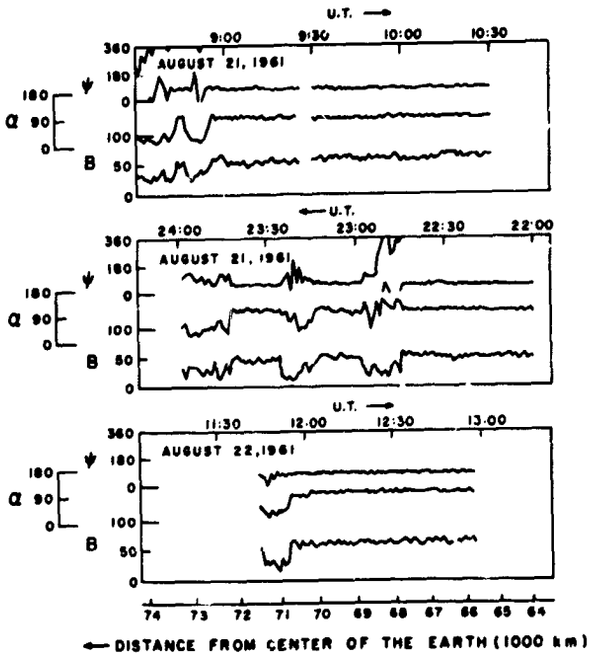


FIGURE 1b.—Magnetic field directions and intensity for the same transitions as in Figure 1a. The strength is measured in gammas and the directions are given in the satellite centered spherical coordinates α and ψ.

at times the satellite passes through regions of enhanced particle intensity which are separated from the boundary by stretches where the particle intensity falls below the threshold of detectability. Unfortunately, in most cases it is impossible to separate temporal and spatial phenomena on the basis of only particle data. One could argue equally well that the intensity represent actually a part of the outer belt that for some reason is separated from the rest of the belt by particle free regions, or that they represent an expansion and contraction of the trapping region in such a way that the satellite crosses the boundary several times during one orbit. Figure 1b shows the magnitude and direction of the magnetic field during the same transitions. The directions of the magnetic field are given in satellite centered spherical coordinates (Cahill and Amazeen 1963). It should be noted that in analogy to particle intensities during the transition of the boundary the magnetic field sometimes shows the same direction and magnitude in regions separated from the magnetosphere as were observed in the trapping region. Using magnetic field data alone one again cannot distinguish between temporal and spatial phenomena.

In the analysis of the data thus far, one case was found in which a combination of magnetic field data and particle data can resolve this question.

The present paper shall be concerned with a sudden increase of the particle flux on September 25, 1961 at 17:31 U.T. at a radial distance of 77,465 km as seen by the ion-electron detector and changes in the magnetic field as seen by the magnetometer on board Explorer XII.

INSTRUMENTATION

A brief review of some of the features of the ion-electron detector pertinent to the following discussion will be given here. A detailed description of the apparatus will be published elsewhere (Davis and Williamson, to be published).

The detector is a scintillation counter consisting of a photomultiplier tube on the face of which is deposited a 5 mg/cm² thick layer of crystalline ZnS covered by a 1000 A° thick layer of Al. It has a geometric factor of 5.4×10^{-3} cm² ster. A collimator allows particles to enter from a viewing cone with a half angle of 11°.

A wheel driven by a stepping motor introduces varying thicknesses of Ni absorber between the collimator and the phosphor. Three positions on the wheel carry Au discs which scatter the incident flux entering through an alternate collimator into the phosphor.

There are three modes of operation:

1. Pulse output from the anode measures protons between the energies of 100 kev and 5 Mev.
2. 8th dynode current measures the total incident energy flux from both protons and electrons.
3. Electrons can be selected preferentially by scattering an incident beam of both protons and electrons from an Au disc. In this mode of operation the 8th dynode current measures the total incident energy flux due to electrons alone.

In all three modes of operation the lower energy cutoff can be raised by introducing varying thicknesses of Ni absorber.

Throughout the following discussion we shall make the assumption that for different absorbers we are measuring the integral proton flux above some energy for which the detector efficiency is 50 percent. The situation is more complicated for electrons because of the shape of the sensitivity

curves and the necessity to use approximation methods to obtain electron spectrums.

The magnetic field data was obtained from the magnetometer flown by L. J. Cahill of the University of New Hampshire and its operation had been described earlier (Cahill and Amazeen, 1963).

ANALYSIS

The inbound pass of September 25, 1961 had a geomagnetic latitude of less than 5° for radial distances of more than 40,000 km. Figure 2 shows intensities of low energy protons and the value of the photomultiplier 8th dynode current obtained with 5μ inch Ni absorber when the detector was looking at particles with local pitch angles of around 90°. Also shown are the intensity and direction of the magnetic field in the satellite coordinate system.

The electrometer data represents instantaneous values of the 8th dynode current sampled at intervals of 83 seconds. To gain some physical idea of the electron intensities involved, a crude analysis shows that at 19:45 U.T. the differential electron spectrum could be approximated by a curve of the form $E^{-\gamma}$ where E is the electron energy and $\gamma = 3.7$. The integral intensity of electrons for $E > 20$ kev is 4×10^6 electrons (cm^2 sec ster)$^{-1}$. This number is accurate to better than a factor of 2.

The plot of the proton data was made for particles with energies $E > 140$ kev and $E > 200$ kev. The proton intensities are obtained from data accumulated over a time period of 0.3 seconds. During that time the satellite turns about its axis through an angle of about 58°. Thus the 90° pitch angle intensity is an average value covering also other than locally mirroring particles.

The magnetometer data corresponds to an average of ten successive telemetry samples taken at intervals of 50 km. The field strength is expressed in gammas, and the direction is indicated in spherical satellite centered coordinates mentioned earlier.

The graph in Figure 2 is divided into three parts: Region I represents the proton and current intensities in the outer belt. Region II is characteristic of the space outside the trapping region. In this case the current output of the detector consists only of the dark current and the pulse output merges into the omnidirectional background. Region III shows a sudden increase in the electron and proton intensity at 17:31 U.T. which lasted for about one minute.

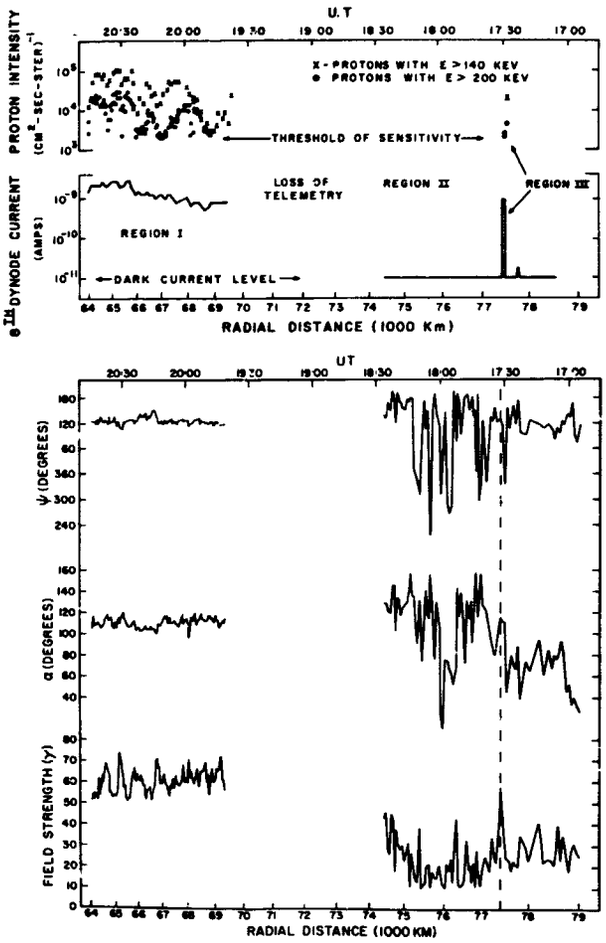


FIGURE 2.—Inbound pass of September 25, 1961. The upper two curves represent the unidirectional proton intensity for particles with 90° local pitch angle and the 8th dynode current corresponding to electrons with E > 20 kev and 90° pitch angle respectively. The lower three curves show the strength and directions of the magnetic field. Angles α and ψ were previously defined by Cahill and Amazeen (1963). The left side is a typical response for the trapping region (Region I). The right side (Region II) shows the usual response of the detectors outside the magnetosphere. For protons the threshold of sensitivity is the count rate corresponding to an omnidirectional flux of penetrating particles. Region III is the sudden increase in particle and magnetic field intensity which is the subject of the present study.

Similarly the magnetic field in Region I is the rather constant field encountered in the outer belt. Region II is typical of the field outside the boundary of the magnetosphere. The field is rather disorganized and shows many changes in direction.

The dashed line indicates the magnetic field in Region III. During the transition of Region III the magnetic field agreed in magnitude and in direction with the field observed in Region I. Before and after the sudden increase the magnetic field had an average value of 25 gamma. During the transition of Region III the field rose to 61 gamma. Aside from this single increase there is nothing unusual happening in the behavior of the field outside the magnetospheric boundary. It should be emphasized that in this context the word "region" is used only with reference to the graph and does not necessarily imply any physical structure in space. The increase is real and not due to faulty data processing because it appears on two completely independent data tapes acquired at Fort Myers, Florida and at South Point, Hawaii. This increase in particle and field intensity will from now on be referred to as Region III.

While the satellite was in Region III it was possible to obtain 15 data points for protons with $E > 140$ kev and 45 data points for protons with $E > 200$ kev. Figures 3 and 4 present the pitch

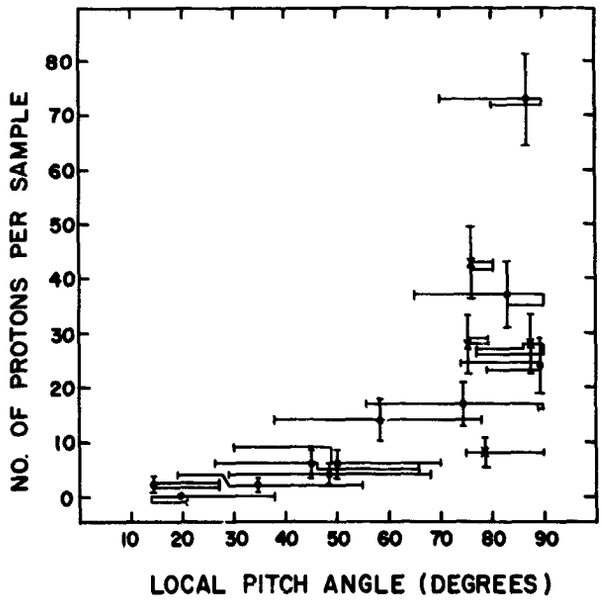


FIGURE 3.—Pitch angle distribution of protons with E>140 kev found in Region III. The vertical bars represent Poisson standard deviations. The uncertainties due to the digitization of the analog data are about half as large again. The horizontal bars represent the range of angles over which the samples were collected. The assigned p.a. corresponds to the center of the accumulation period.

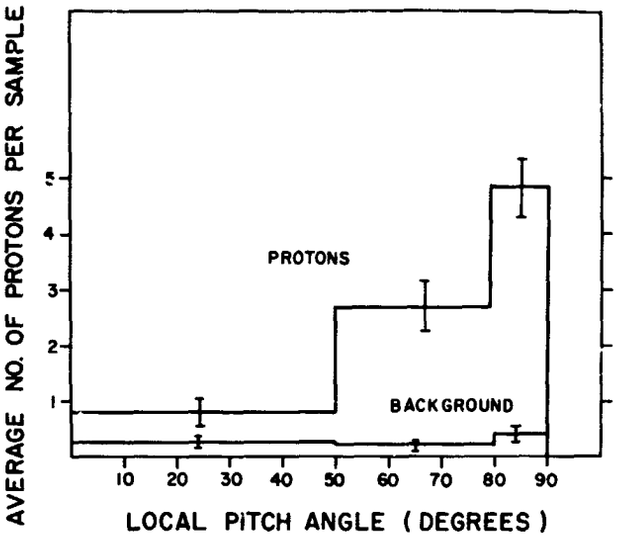


FIGURE 4.—Pitch angle distribution of protons with E>200 kev. The data points were lumped together into three groups each containing 15 points. The vertical bars represent Poisson deviations. The background count obtained with a thick plug in front of the detector aperture is given below.

angle distributions for these two cutoff energies. In both cases the pitch angles were assigned from the knowledge of the direction of the magnetic field as seen by the magnetometer. The vertical bars represent Poisson standard deviations only. The uncertainty due to the digitization of the analog data tapes is about half as large again. In Figure 3 the horizontal bars represent the ranges of pitch angles which were scanned during the accumulation period. The pitch angle assigned to each point is the value seen at the center of the accumulation period. The points represented by crosses actually have pitch angles which are supplements of the indicated pitch angles. Figure 4 is a similar diagram. In this case the number of counts per data point was small so that data points were lumped together into three groups each containing 15 points. The bars again represent Poisson standard deviations. For contrast the omnidirectional background count obtained with a thick plug in front of the detector is plotted below.

The shape of the pitch angle distributions is quite similar to those obtained for trapped particles within the magnetosphere in that the maximum intensity is found at 90° local pitch angle (Davis and Williamson, 1963). By contrast, solar protons during a sudden commencement have been observed to have maximum intensities in

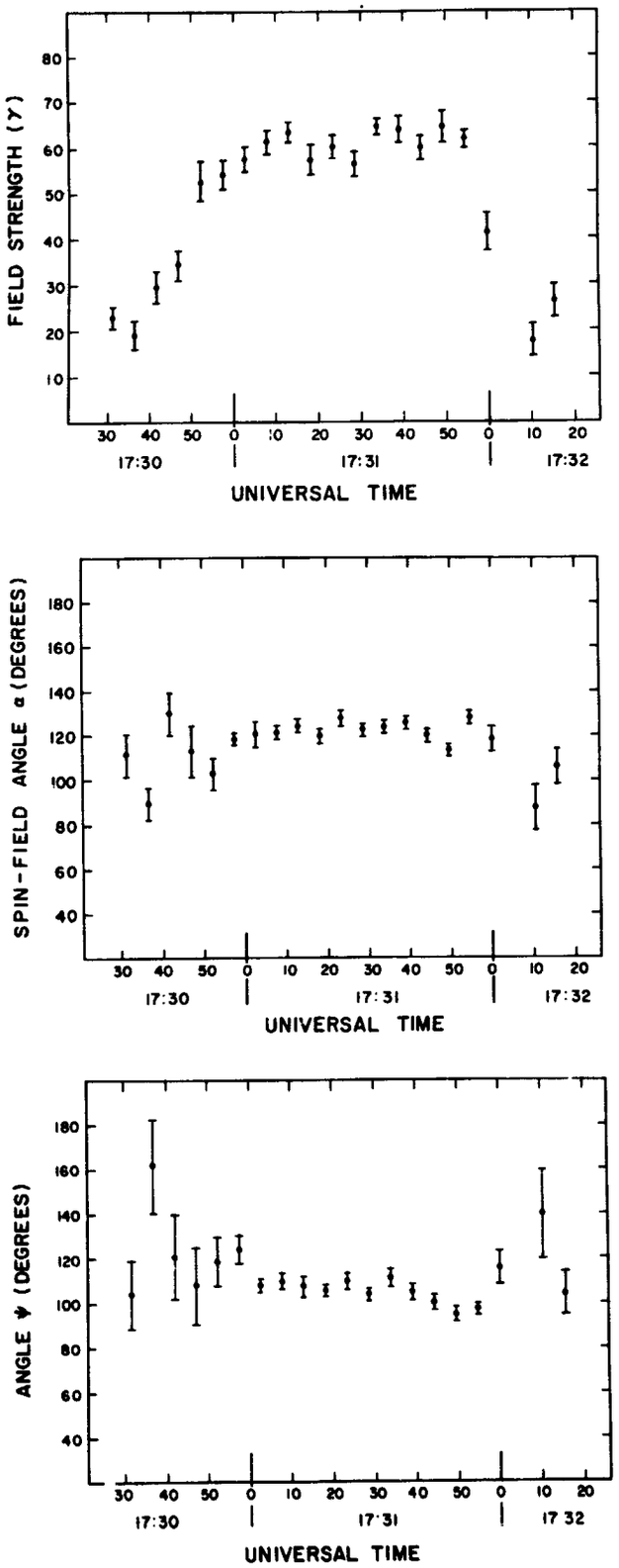


FIGURE 5.—Intensity and directions of the magnetic field in Region III. The vertical bars represent standard deviations of the mean of 16 data points each. The angles α and ψ are identical to those defined by Cahill.

directions other than 90° to the field lines (Hoffman, et. at., 1962).

Figure 5 shows the magnetic field intensity and directions during the transition of Region III. The vertical bars represent standard deviations of the mean of 16 data points each. The scatter in the original data is mostly due to the digitization of the magnetometer analog output voltage. The standard deviations increase as the field decreases in magnitude. α is the angle between the spin axis of the satellite and the magnetic field. ψ is the angle between the spin-field plane and the spin-sun plane. They are identical to those defined earlier by Cahill and Amazeen (1963). Before and after the transition of Region III the magnitude of the field was about 25 gamma. During the transition the magnitude rose to 60 gamma. There was no drastic change in either α or ψ.

Assuming for the moment that Region III is stationary an estimate of its maximum width can be made by finding the time when the particle flux was zero on the step corresponding to the lowest energy cutoff just before and after the transition of Region III and multiplying it by the velocity of the satellite. This gives the region a maximum width of 200 km. A more precise value obtained from the magnetometer data is 83 km corresponding to a width of 56 km in the radial direction from the center of the earth. It should be pointed out that the angle between the direction of the magnetic field and the velocity vector of the satellite is about 60°. This means that the angle between the velocity vector and the plane of gyration of protons with 90° pitch angle is 30°. Therefore, the width of Region III in the plane containing protons with 90° pitch angles must be close to that of Region III as measured along the velocity vector.

Table 1 shows the cyclotron radii of protons with several energies corresponding to the magnetic field encountered within Region III. It is clear from the Table that no 200 kev or 140 kev protons with pitch angles close to 90° could be trapped within a stationary Region III where the magnetic field strength is about 60 gamma. The efficiency of the detector for 310 ev protons is zero. Thus the conclusion must be drawn that Region III is a temporal rather than a spatial phenomenon.

TABLE 1.—*Cyclotron Radii of Protons*

Magnetic field strength H = 60 gamma

Proton energy	Cyclotron radius
200 kev	1075 km
140 kev	899 km
310 kev	42 km

The origin of Region III can only be conjectured. In what follows we shall try to produce convincing evidence that Region III represents a temporal extension of the outer belt.

Comparison between Region I and Region III shows that the intensities of both the 8th dynode current and of the protons are comparable in several places for particles with 90° pitch angle. The order of magnitude fluctuations of the proton flux and the somewhat smaller fluctuations of the 8th dynode current make it difficult to perform any more detailed analysis. Magnetometer data, however, provide a more quantitative comparison.

For reasons of clarity, we shall transform the angles α and ψ into the geomagnetic coordinate system.

To make a comparison with the directions of the geomagnetic field within the radiation belt, the direction of the field was obtained from experimental data at three points shown in Table 2. Region Ia occurs just after the recovery of the telemetry within the belt. Region Ib was selected to have the same increment of distance as exists between III and Ia.

Several theoretical models indicate that in the absence of a ring current and close to the equatorial plane the boundary of the magnetosphere is almost circular in the meridian plane on the day side of the earth. (Beard, 1960; Beard, 1962; Beard and Jenkins, 1962; Davis and Beard, 1962; Mead and Beard, 1964; Spreiter and Briggs, 1962; Spreiter and Alksne, 1962; Spreiter and Hyett, 1963). Assuming that this holds true the direction of the magnetic field was calculated just within the boundary of the magnetosphere at +4° geomagnetic latitude and at 10:24 hours local time.

TABLE 2

Time (U.T.)	Region	Radial distance (earth radii)
17:31	III	12.135
19:42	Ia	10.877
21:12	Ib	9.618

The results of the field calculations and observations are given in Table 3.

For comparison the satellite and sun positions were also included. A diagram of the results is given in Figure 6, where a polar projection is used to represent the field directions. To make possible the comparison between magnetic fields as observed in different regions we re-define the zero magnetic meridian to lie in the plane formed by the earth's magnetic polar axis and the radius vector of the satellite position. In this coordinate

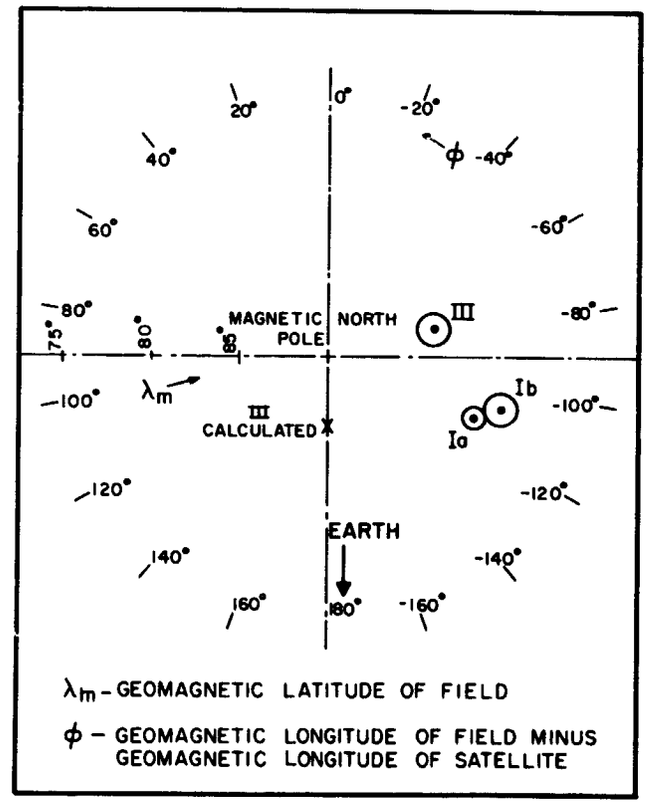


FIGURE 6.—Polar plot showing the directions of the magnetic field in Regions Ia, Ib, and III. Also shown is the direction of the magnetic field as calculated under assumption described in the text. The azimuthal variable is the difference between the geomagnetic longitudes of the field and of the satellite.

TABLE 3.—*Directions of the Magnetic Field*

Region	Time	Field(γ)	ψ (deg.)	α (deg.)	Field Geomagnetic Coord.		
					Lat. (deg.)	Long. (deg.)	ϕ (deg.)
III calc.	17:31		113.45	126.4	86.06	139.81	180.00
III obs.	17:31	61.1±0.8	105.0±1.0	122.4±0.8	83.92	−115.50	− 75.69
Ia obs.	19:41	71.5±0.7	116.3±0.7	120.4±0.6	81.11	179.82	−112.53
Ib obs.	21:12	73.5±1.1	119.5±0.9	118.2±0.9	79.92	116.54	−107.12

Positions of the Sun and Satellite

Region	Time	Satellite Geom. Coord.		Sun Geom. Coord.	
		Lat. (deg.)	Long. (deg.)	Lat. (deg.)	Long. (deg.)
III	17:31	3.94	−39.81	10.38	−16.08
Ia	19:41	2.61	−67.65	6.82	−48.77
Ib	21:12	1.77	−86.34	2.79	−71.24

system the azimuthal variable of the field direction, ϕ. is the angle whose magnitude is equal to the geomagnetic longitude of the field minus the geomagnetic longitude of the satellite position.

The points represent the directions of the magnetic field as observed in Regions III, Ia, Ib, and as calculated in Region III. The circles about the points correspond to standard deviations due to scatter of the raw data points. The lack of overlap of the data is evidence for slight temporal variations of the magnetic field. This was confirmed by taking several adjacent sets of data points in both Regions Ia and Ib. The results show a clustering of points with an average distance between the points of about two degrees.

The observed directions of the magnetic field seem to be out of phase with respect to the satellite by about −90° in the azimuthal direction rather than the expected 180° if the field is calculated under assumptions stated before.

This difference does not seem significant. Even at 20,000 and 30,000 kilometers field measurements disagree by about 7 degrees of arc with a theoretical model based on the Jensen and Cain coefficients and the Mead cavity field. (*Jensen and Cain*, 1962; *Mead*, 1964). This disagreement may be due to slight uncertainties in the calibration of the instruments or an error in the assumed direction of the satellite spin axis, or both. The problem is currently being investigated.

The significance of these results is that in spite of several systematic uncertainties in the absolute measurement of the magnetic field and other effects such as the change in the satellite's geomagnetic latitude from +3.94° at 17:31 U.T. to 1.77° at 21:12 U.T., and the unknown spatial and temporal changes in the direction of the magnetic field, the observed field directions in Regions III, Ia and Ib can be fitted into a cone with a half angle of 3°.

This means that the magnetic field in Region III points almost in the same direction as the field in the trapping region and it is, therefore, reasonable to assume that Region III is just a temporary extension of the trapping region.

It is also possible to put a lower limit on the speed with which the trapping region overtook the satellite. The initial rise in the intensity of the

magnetic field occurred at 17:30:52.9 U.T. At 17:31:04.3 U.T. the detector saw 140 kev protons. During the 5.2 sec. counting period the detector could see protons with 90° pitch angle. They were coming from directions shown in Figure 7.

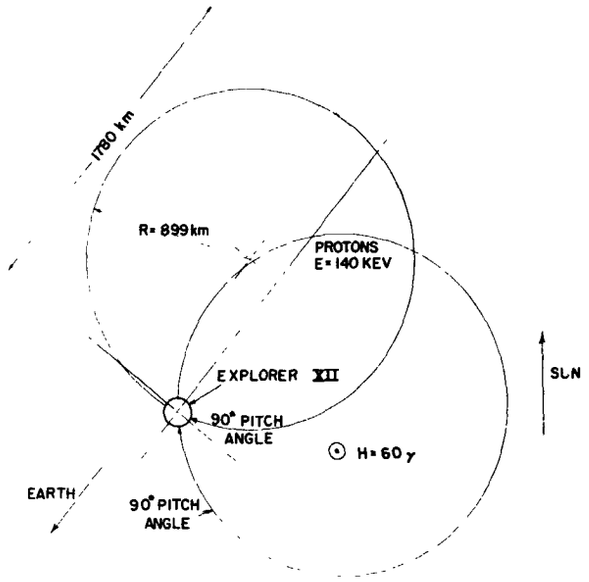


FIGURE 7.—Orbits of trapped protons with 90° pitch angle which were observed in Region III. At the time when the trapped protons were observed the boundary of the magnetosphere must have been at least at a distance of 1780 km in the radial direction away from the satellite. The interval between the first observation of the increased magnetic field and the observation of 140 kev protons was 11.4 sec. This puts a lower limit on the velocity of advance of the magnetosphere of 156 km/sec.

Since these protons were trapped their orbit could not extend beyond the boundary of the magnetic field. Thus we are led to the conclusion that during the interval of 11.4 seconds between the increase in the field and the appearance of 140 kev trapped protons the boundary must have moved through a radial distance of at least 1780 km. This puts a lower limit on its radial velocity of advance equal to 156 km/sec. A similar analysis can be performed in trying to estimate the speed of recession of the magnetosphere. The time interval between the instant when 200 kev locally mirroring protons were seen and the instant when the field strength dropped to about 25 gamma was 29 seconds. To preserve closed orbits of trapped protons, the magnetosphere must have moved in that interval at least through 2130 km. This puts a lower limit on the radial velocity of recession equal to 75 km/sec.

A survey of ground magnetograms was made. No identifiable features could be found in the records from low latitude stations. Large regular pulsation events were observed in two out of approximately twenty high latitude magnetograms which were surveyed. At Point Barrow (geomagnetic latitude 68.6°N, geomagnetic longitude 241.2) the field vector began rotating counterclockwise, as seen by an observer looking down at the earth, at about 17:31 U.T., or 0700 local time. The peak to peak amplitude was about 150 gammas and the period was 4 to 5 minutes. At Baker Lake (geomagnetic latitude 73.8°N, geomagnetic longitude 315.2°) the peak amplitude was again about 150 gammas and the period was 4 to 5 minutes, but here the wave had a roughly linear polarization and began about 17:27 U.T., or 11:00 local time.

These ground records suggest that the observed rapid expansion and probable compression of the magnetosphere were associated with the excitation of a hydromagnetic disturbance in a thin shell of the outer magnetosphere. We do not, however, believe that the nearly simultaneous observation of one event on the satellite and of an event at two out of twenty ground stations can be interpreted as conclusive evidence that the two phenomena are related.

INTERPRETATIONS

We have argued that at 17:31 U.T. the magnetosphere expanded past Explorer XII at a velocity of at least 156 km/sec. Even if the external pressure dropped to zero, the plasma in the magnetosphere would not expand at a velocity greater than

$$V = \sqrt{\frac{B^2}{4\pi mn} + \frac{2c^2}{\gamma - 1}}$$

where n is the number of protons per cm^3, c is the speed of sound, and we will use B equal to 60 gammas, m equal to the mass of a proton, and γ equal to 5/3. The speed of sound in an ideal gas is given by

$$c^2 = \frac{\gamma RT}{\mu}$$

where R is the gas constant and μ is the molecular weight of the gas. Using these expressions, the

expansion velocity is given by where T is the

$$V^2 = \frac{1.71 \times 10^6}{n} + 482T$$

temperature in electron volts and V is in km/sec.

If the magnetic energy density far exceeds the thermal energy density, the first term will dominate and an expansion velocity of 156 km/sec implies a number density of less than 70 protons per cm^3. If the magnetic field can be neglected, the second term dominates and implies a proton temperature of at least 50 ev. Several other possible combinations are listed in Table 4 where it can be seen that either the number density must be less than about 100 protons per cm^3 or the temperature must be greater than about 15 ev near the outer edge of the magnetosphere.

TABLE 4

upper limit on n(cm^{-3})	∞	341	174	117	88	70
lower limit on T(ev)	50	40	30	20	10	0

After a time interval of 69 seconds a second boundary was observed. Several interpretations are possible. The observations could be caused by an expansion and subsequent compression of the entire magnetosphere, or a small portion of the magnetosphere; they could be caused by the expansion of a small portion of the magnetosphere which breaks off and continues to move outwards; they could be caused by a wave propagating in the surface of the magnetosphere. We cannot presently determine the cause of the observation.

CONCLUSIONS

On September 25, 1961 at 17:31 U.T. Explorer XII passed through a region of enhanced proton and electron intensity which lasted for 1 minute and extended over an apparent distance of 83 km. At the same time the on board magnetometer registered an increase in the magnetic field.

Analysis of the data shows that the detected protons were trapped in the magnetic field. At the same time their cyclotron radii were much larger than the apparent width of the region. This fact points out the temporal character of the region. Similarities between the particle intensities and the strength and directions of the magnetic field in the region and in the outer belt suggest that the region comprises an impulse-like extension of the magnetosphere.

From knowing the time of the increase and the intensity of the magnetic field as well as the time when the 140 kev trapped protons were detected, a lower limit on the outward radial velocity of the magnetospheric boundary can be determined to be 156km/sec. Similarly the lower limit on the inward radial velocity of the magnetospheric boundary has been determined to be 75 km/sec.

The conclusions that can be drawn from our crude calculations are that either the proton density near the outer edge of the magnetosphere is less than about 100 per cm^3 or the temperature is greater than about 15 ev.

ACKNOWLEDGMENTS

We wish to express our appreciation to Dr. L. J. Cahill of the University of New Hampshire and to Mr. L. R. Davis and Mr. J. M. Williamson of the Goddard Space Flight Center for the use of their data from detectors flown on board Explorer XII. We also would like to thank Dr. Cahill and Mr. Davis for many helpful discussions.

REFERENCES

1. BEARD, D. B., The Interaction of the Terrestrial Magnetic Field with the Solar Corpuscular Radiation, *J. Geophys. Res.*, **65**, 3559-3568, 1960.
2. BEARD, D. B., The Interaction of the Terrestrial Magnetic Field with the Solar Corpuscular Radiation, 2, second-order approximation, *J. Geophys. Res.*, **67**, 477–483, 1962.
3. BEARD, D. B., and E. B. JENKINS, Correction to the Second Approximation Calculation of the Geomagnetic Field, Solar Wind Interface, *J. Geophys. Res.*, **67**, 4895–4896, 1962(1).
4. CAHILL, L. J., and P. G. AMAZEEN, The Boundary of the Geomagnetic Field, *J. Geophys. Res.*, **68**, 1835–1843, 1963.
5. DAVIS, L., JR., and D. B. BEARD, A Correction to the Approximate Condition for Locating the Boundary Between a Magnetic Field and a Plasma Field, *J. Geophys. Res.*, **67**, 4505–4507, 1962(1).
6. DAVIS, L. R., and J. M. WILLIAMSON, Low Energy Trapped Protons, *Space Research*, Proc. Intn'l Space Sci. Symp., 3rd, Washington, D.C., 1962, North-Holland Publishing Co., Amsterdam, 1963.
7. DAVIS, L. R., and J. M. WILLIAMSON, Ion-Electron Detector, *J. Geophys. Res.*, (to be published).
8. HOFFMAN, R. A., L. R. DAVIS, and J. M. WILLIAMSON, Protons of 0.1 to 5 Mev and Electrons of 20 Kev at 12 Earth Radii During Sudden Commencement on September 30, 1961, *J. Geophys. Res.*, **67**, 5001–5005, 1962.

9. JENSEN, D. C., and J. C. CAIN, An Interim Geomagnetic Field (abstract), *J. Geophys. Res.*, **67**, 3568–3569, 1962.
10. MEAD, G. D., The Deformation of the Geomagnetic Field by the Solar Wind, *J. Geophys. Res.*, **69**, 1181–1195, 1964.
11. MEAD, G. D., and D. B. BEARD, The Shape of the Geomagnetic Field Solar-Wind Boundary, *J. Geophys. Res.*, **69**, 1169–1179, 1964.
12. SPREITER, J. R., and A. Y. ALKSNE, On the Effect of a Ring Current on the Terminal Shape of the Geomagnetic Field, *J. Geophys. Res.*, **67**, 2193–2205, 1962.
13. SPREITER, J. R., and B. R. BRIGGS, Theoretical Determination of the Form of the Boundary of the Solar Corpuscular Stream Produced by Interaction With the Magnetic Dipole Field of the Earth, *J. Geophys. Res.*, **67**, 37–51, 1962.
14. SPREITER, J. R., and B. J. HYETT, The Effect of a Uniform External Pressure on the Boundary of the Geomagnetic Field in a Steady Wind, *J. Geophys. Res.*, **68**, 1631–1642, 1963.

N65-18274

MEASUREMENT OF LOW ENERGY PRIMARY COSMIC RAY PROTONS ON IMP-1 SATELLITE*

F. B. McDONALD AND G. H. LUDWIG

The first precise determination of the intensity and energy spectra of primary cosmic ray protons in the 15–75 MeV interval has been made with a three element energy vs. energy-loss telescope aboard the IMP–1 (Explorer XVIII) satellite. This spacecraft had an apogee of 193,000 km and only data obtained well beyond the effects of the earth's magnetosphere are considered. The measurements reported here cover the time interval 8 December 1963 to 6 May 1964 and are considered to be representative of the period just prior to solar minimum. The proton intensity in the range 15–75 MeV was observed to be 19 proton/m^2-sec-ster or approximately 1 percent of the total primary cosmic ray intensity and to exhibit a steeply falling energy spectrum toward lower energies, decreasing by a factor of 5 over this interval. One point for helium was obtained in the range 65–75 MeV/nuc.

The E vs dE/dx telescope (Figure 1a) provides a means of studying this low energy component in the presence of higher energy cosmic rays.[1] Because of the background and low flux in this energy interval, significant measurements cannot be obtained by balloon or rocket techniques but are dependent on satellite measurements. The principal of operation of the telescope is shown in Figure 1. For each particle which traverses the ΔE counter and stops in the E crystal (as defined by the plastic scintillator anticoincidence cup) measurements of ΔE and E—ΔE are made by 512 channel pulse height analyzers and transmitted over the satellite telemetry systems. As shown in Figure 1b, this provides a measure of mass, charge and energy resolution for $Z = 1, 2$ particles. The electron data are discussed in an accompanying paper. Saturation effects prevented the alpha particles data from being extended to lower energies. This does not affect the H or e data. Mass histograms have been obtained in the proton region for six energy intervals of about 10 MeV

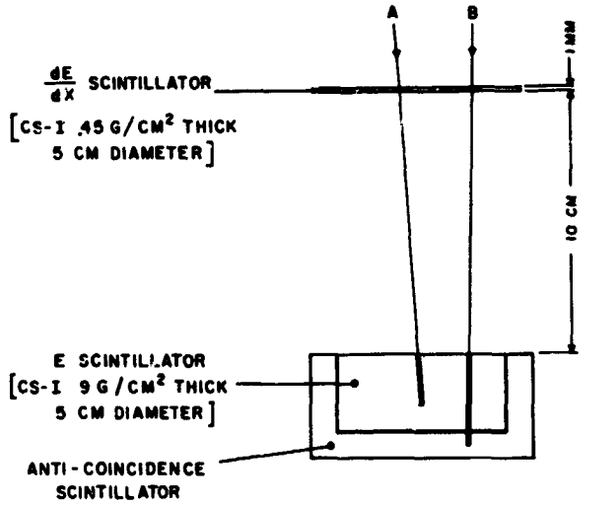


FIGURE 1a.—Schematic drawing of E vs dE/dx telescope. Particles which stop in the lower scintillator (A) are accepted for analysis while those which enter the anti-coincidence cup (B) are rejected.

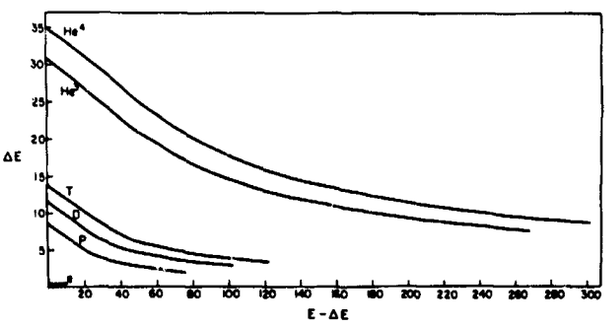


FIGURE 1b.—Mass and energy response of E[~(E-ΔE)] and dE/dx (~ΔE) telescope.

*Published as *Goddard Space Flight Center Document X-611-64-363*, November 1964.

each by summing the data in 11 channels constructed parallel to and centered about the proton line of Figure 1b. These six energy intervals for H are shown in Figure 2. The clean resolution of

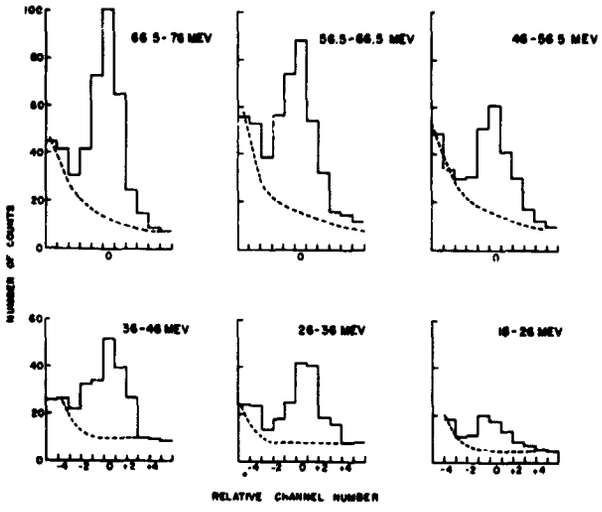


Figure 2—Mass histogram for 6 energy intervals in proton region.

the proton distribution at all energies is striking. The background correction (illustrated by the dashed line of Figure 2) has been applied using the experimental proton distribution obtained from the small solar cosmic ray event of 16 March 1964. The rigidity and energy spectra obtained from the mass histogram data of Figure 2 along with the single α point is shown in Figure 3.

To search for long term variations, the data were divided into two intervals covering the period 8 December 1963—14 March 1964 and 22 March 1964—6 May 1964. Except at the lowest energy point, the data are consistent with a change of less than 10 percent in the proton flux in the 15–75 MeV region between these two periods. The increase on the low energy point may be due to an increase in the galactic flux or, most probably, to the small solar proton event of 16 March 1964. We cannot rule out large short time variations since the long times required to obtain meaningful statistics preclude the possibility of observing variations occurring over a shorter time scale. Proton and helium data obtained at higher energies from the balloon data of Balasubrahmanyan and McDonald,[2] and Fichtel et. al.[3] are shown for comparison. It is observed that at this period of the solar cycle just prior to solar minimum there is a sharp decrease is a sharp decrease in both spectra toward lower energies that extends to the lowest observed energy.

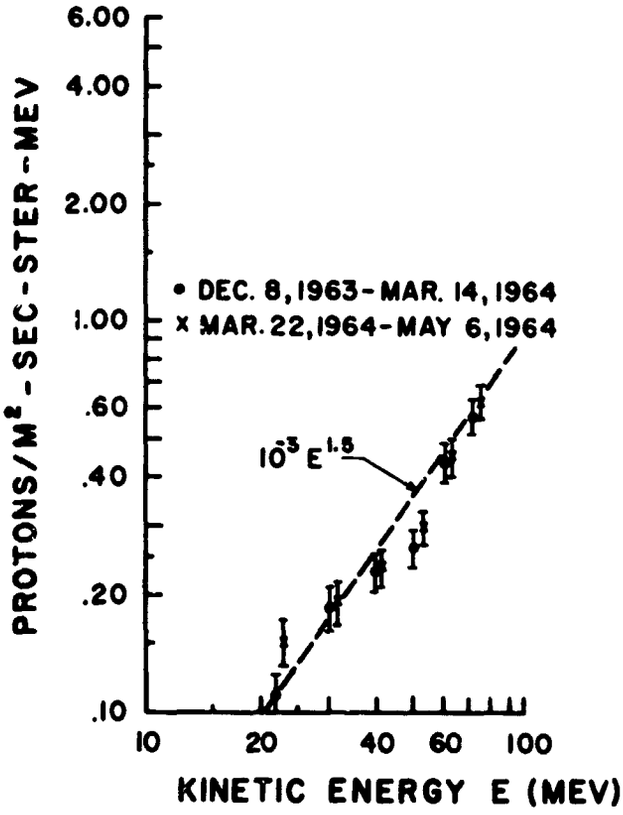


FIGURE 3a.—Cosmic ray proton energy spectra for two different time periods. The dotted line indicates an empirical fit to the data of the form $10^{-3}E^{1.5}$.

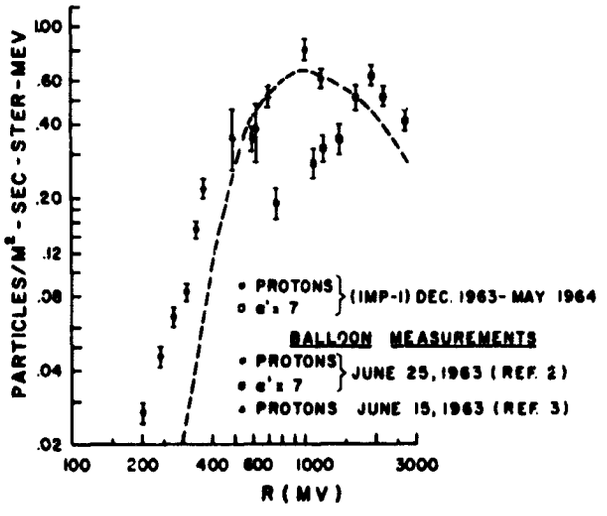


FIGURE 3b.—IMP proton and α rigidity spectra. Shown for comparison is the balloon data of Fichtel et. al.[3] and Balasubrahmanyan et. al.[2].

The flux of deuterium could not be resolved from the background and an upper limit of 8 percent of the proton flux over the corresponding energy interval was placed on its abundance.

During the last solar minimum an arctic latitude survey by Meredith, et. al.[4] using a rocket borne single Geiger counter had indicated the cosmic ray intensity did not increase strongly at low energies. The ionization chamber data of Neher [5] at balloon altitude suggested a strong variation in the low energy range >100 MeV anti-correlated with solar activity.

The only previous satellite measurements in this region were made by Stone [6] in late 1961 on a polar orbiting satellite. This study was based on 12 counts in the energy interval 11–125 MeV. They were consistent with a flat energy spectrum and served to place an upper limit of about 1 proton/m^2-sec-MeV on the proton flux since an active anticoincidence device was not used. The data in Figure 3 show the highest energy point of the IMP data reported here is in general agreement with the lowest energy point obtained in the June 1963 balloon flight by Balasubrahmanyan and McDonald [2] and Fichtel,[3] et. al. It is also consistent with the upper limit of $dJ/dE<.49$ protons/m^2-sec-MeV obtained on a high latitude flight of Brunstein in 1962 [7] and with the lowest energy points obtained by Freier and Waddington[8] and Ornes and Webber [9] from 1963 balloon flights. Further confirmation of the splitting of the normalized low rigidity spectra is also obtained.

The proton and alpha spectra as observed in interplanetary space result from a superposition of three processes—initial acceleration, diffusion through the galaxy and solar modulation. To illustrate these processes assume

(1) Energy spectra at injection are of the form

$$\mu(E)=\frac{10^4}{w(1+E)^{2.5}} \text{ protons/M}^2\text{-ster-Bev}$$

(2) Traversal of 2.5 gm/cm^2 of hydrogen before reaching the solar system; and

(3) The solar modulation is of the form

$$\frac{dj}{dE}=\frac{dj_0}{dE}\exp\left(-\frac{\text{const}}{w}\right)$$

where $\mu(E)$ is the density of cosmic rays with kinetic energy E, and dj/dE is the observed differential energy spectrum while dj_0/dE is the unmodulated or stellar spectrum and w is the particle velocity. Assumption (1) is based on an extrapolation to low energies of observations above 4 Bev [10]. The 2.5 gm/cm^2 of H is implied by the relative abundance of *Li*, *Be*, and *B* in the primary beam [11] and the solar modulation function is a diffusion model proposed by Parker [12]. In this model an equilibrium state is established between diffusion inward through irregular magnetic fields and removal by outward convection. The solar modulation effects have been obtained by normalizing at 200 MeV. These calculations are shown by the dotted curve of Figure 3b. While it can be concluded that either the injection spectra are steeper than a simple power law (as one might hypothesize from solar cosmic ray studies) or that the effects of solar modulation are not as strong at low energies (in agreement with the lack of significant long term time variations), the primary effect of these admittedly crude assumptions is to show that indeed the production of the observed steep proton energy spectrum is not unexpected on the basis of other cosmic ray observations. It is to be expected that further observations over the period of solar minimum coupled with low energy measurements of multiply charged primaries will permit more accurate determinations of the source spectra and the solar modulation mechanism.

The authors wish to thank S. Paull, C. A. Cancro and P. J. Janniche of the Goddard Space Flight Center for designing and testing the spacecraft electronics for the IMP experiment, and to A. P. Flanick and D. Clark of Goddard for testing and calibrating the detector system. The support of Paul Butler and the IMP project team is greatefully acknowledged.

REFERENCES

1. G. H. Ludwig and F. B. McDonald, Proceedings of the International Symposium on Nuclear Electronics, Paris 195, (1963).
2. V. K. Balasubrahmanyan and F. B. McDonald, Journal of Geophysical Research 69, 3289, 1964.
3. C. E. Fichtel, D. E. Guss, D. A. Kniffen and K. A. Neelakantan, Journal of Geophysical Research 69, 4293, 1964.
4. L. H. Meredith, J. A. Van Allen and M. B. Gottlieb, Physical Review 99, (198), 1955.
5. H. V. Neher, Physical Review 107, 588, 1957.
6. E. C. Stone, Journal of Geophysical Research 69, 3939, 1964.

7. K. A. BRUNSTEIN, Physical Review **133**, B(1520), 1964.
8. P. S. FREIER and C. J. WADDINGTON, Physical Review Letters **13**, 108 (1964).
9. J. ORMES and W. R. WEBBER, Physical Review Letters **13**, 106 (1964).
10. F. B. McDONALD, Physical Review **109**, 1370 (1958).
11. G. D. BADHWAR, R. R. DANIEL and B. VIJAYALAKSHMI, Prog. Theor. Phys. **28**, 607 (1962).
12. E. N. PARKER, Interplanet Dynamical Processes (Interscience, 1963) P. 199.

N65-37959

SECULAR AND NON-SECULAR BEHAVIOR FOR THE COLD PLASMA EQUATIONS*

N66 37962

DAVID MONTGOMERY AND DEREK A. TIDMAN†

The origin of "secular" behavior for the nonlinear cold electron plasma equations is studied. The equations involved are closely related to the Klein-Gordon equation with a small nonlinear term. A method is developed for arriving at perturbation theoretic solutions of this equation, and the method is then applied to the case of the higher order effects of an electromagnetic wave propagating in the cold electron plasma. An explicit expression for the second order frequency shift is calculated.

INTRODUCTION

When attempts are made to calculate the behavior of nonlinear mechanical systems in perturbation theory which goes beyond the linearized approximation, the difficulty of "secular" behavior often (though by no means always) appears. The higher orders contain, in addition to trigonometric terms, time-proportional or "secular" terms. The unbounded character of these terms soon invalidates the perturbation theoretic assumptions of smallness on which they were derived.

A systematic program for doing a type of perturbation theory that is free of secular terms in the case of the harmonic oscillator equation with a "small" nonlinear term was given some time ago by Krylov and Bogolyubov (Reference 1), and later refined and mathematically justified by Bogolyubov and Mitropolskii (Reference 2).** Recently, considerable interest has arisen in modifying these techniques to deal with partial differential and differentio-integral equations, especially in connection with the work of Frieman (Reference 3) and Sandri (Reference 4). Earlier calculations, which are more closely related to this work, were made by Jackson (Reference 5) and Sturrock (Reference 6).‡

The system treated here, a partial differential equation, is considerably simpler than those described in References 3 and 4. It is nonetheless believed to be of value in that it illuminates, in a relatively uncluttered way, some new features which emerge as a consequence of its being a partial differential, rather than an ordinary differential, system.

The equation treated by Bogolyubov and Mitropolskii is

$$\left(\frac{d^2}{dt^2} + \omega^2\right) x(t) = \epsilon F\left(x, \frac{dx}{dt}\right) \qquad (1)$$

where ϵ is a formal expansion parameter (eventually to be set equal to one) written to indicate the relative "smallness" of the right hand side; F(x, dx/dt) is a known nonlinear functional of x and dx/dt; and ω^2 is a constant.

We shall examine the equation

$$\left(\frac{\partial^2}{\partial t^2} - c^2 \frac{\partial^2}{\partial x^2} + \lambda^2\right) f(x,t) = \epsilon F\left(f, \frac{\partial f}{\partial t}, \frac{\partial f}{\partial x}\right) \qquad (2)$$

where ϵ is again the formal expansion parameter; F is again a known function (which depends on the problem under consideration); and c^2 and λ^2 are non-negative real constants.

The immediate motivation for studying Equation 2 was that the equations of motion for a "cold" electron plasma in a uniform positive background may be put in this form, component by component. But it also has the value of being sufficiently simple that explicit zeroth order

*Published as *NASA Technical Note D-2305*, June 1964; also published in *Physics of Fluids*, 7(2); 242-249, February 1964.

†Both authors are affiliated with Goddard Space Flight Center and the University of Maryland, College Park, Maryland.

**A quite similar technique was put forward independently by M. V. Lighthill: "A Technique for Rendering Approximate Solutions to Physical Problems Uniformly Valid," *Phil. Mag. Series* 7, 40(311):1179-1201, December 1949.

‡See also Dolph, C. L., "A Unified Theory of the Nonlinear Oscillations of a Cold Plasma," *J. Math. Anal. Appl.*, 5(1):94-118, August 1962.

($\epsilon = 0$) solutions of it can be constructed which *both do and do not* lead to secular behavior when a straightforward attempt is made to extend them to higher order in ϵ. This simple example makes clear one difference between Equation 1 and an equation of the type 2. In general, for a given F, any $\epsilon = 0$ solution of Equation 1 will give rise to secularity when extended to higher order, whereas the conditions under which Equation 2 gives rise to secular behavior beyond the first order *depend sensitively upon which* $\epsilon = 0$ *solution it is that one wishes to extend to higher order in* ϵ, *even for a given* F.

We shall first give an example of both secular and non-secular behavior for Equation 2, and show how the difficulties may be remedied in the secular case. Then we shall give a simple physical example, that of the second order behavior of a nonlinear electromagnetic wave in a "cold" electron plasma.

THE KLEIN-GORDON EQUATION WITH A SMALL NONLINEAR TERM

A Single Monochromatic Wave

We shall first consider perturbations about the following $\epsilon = 0$ solution to Equation 2:

$$f = a \cos (K_0 x - \omega_0 t + \phi) \tag{3}$$

where

$$\omega_0^2 = c^2 K_0^2 + \lambda^2 \tag{4}$$

and, for $\epsilon = 0$, the quantities K_0, ω_0, a, and ϕ are constants.

This happens to be a situation for which secularity arises, so we shall set up the necessary formalism for handling it from the beginning. Following Reference 2, we seek a solution to Equation 2 of the form

$$f = a \cos \psi + \epsilon u_1(a, \psi) + \epsilon^2 u_2(a, \psi) + \cdots , \tag{5}$$

where the amplitude a is now determined as a "slowly-varying" function of x and t by the relations

$$\frac{\partial a}{\partial t} = \epsilon A_1(a) + \epsilon^2 A_2(a) + \cdots , \tag{6a}$$

$$\frac{\partial a}{\partial x} = \epsilon D_1(a) + \epsilon^2 D_2(a) + \cdots . \tag{6b}$$

Here, ψ is a new "phase" variable, to be chosen to coincide with the phase of Equation 3 for $\epsilon = 0$:

$$\frac{\partial \psi}{\partial t} = -\omega_0 + \epsilon B_1(a) + \epsilon^2 B_2(a) + \cdots , \tag{7a}$$

$$\frac{\partial \psi}{\partial x} = K_0 + \epsilon C_1(a) + \epsilon^2 C_2(a) + \cdots . \tag{7b}$$

The (as yet undetermined) functions A_1, A_2, $\cdots$, B_1, B_2, $\cdots$, C_1, C_2, $\cdots$, D_1, D_2, $\cdots$, are to be *chosen* so as to render the solution (5) free from secular terms. Here, "secular" must be interpreted to mean ψ-proportional; i.e., solutions can break down because of linear growth in x as well as in t. The functions u_1, u_2, $\cdots$, are to be *peroidic* in ψ.

The program is to express the various terms in Equation 2, by means of the relations (5), (6), and (7), in terms of the u's, A's, B's, C's and D's, as functions of a and ψ. For brevity's sake, we shall first only go to $0(\epsilon)$, though in principle the method may be carried to any order in ϵ. Generally, the algebra becomes prohibitive beyond $0(\epsilon^2)$. Hereafter, the notation "+ $\cdots$" will mean "of higher order in ϵ".

The result of differentiating f with respect to x and t and using (6) and (7) is:

$$\begin{aligned} \frac{\partial^2 f}{\partial t^2} - c^2 \frac{\partial^2 f}{\partial x^2} + \lambda^2 f &= a \cos \psi \left(-\omega_0^2 + c^2 K_0^2 + \lambda^2\right) \\ &+ \epsilon \left\{ \lambda^2 \left(\frac{\partial^2 u_1}{\partial \psi^2} + u_1 \right) + 2\left(\omega_0 A_1 + c^2 K_0 D_1\right) \sin \psi \right. \\ &\left. + 2a\left(\omega_0 B_1 + c^2 K_0 C_1\right) \cos \psi \right\} \\ &+ 0\left(\epsilon^2\right) . \end{aligned} \tag{8}$$

The zeroth order part of (8) vanishes identically, by virtue of (4). The coefficient of ϵ is to be equated to $F(f, \partial f/\partial t, \partial f/\partial x)$, with the arguments replaced by their zeroth order values:

$$f \to a \cos \psi ,$$

$$\partial f/\partial t \to \omega_0 a \sin \psi ,$$

$$\partial f/\partial x \to -K_0 a \sin \psi .$$

For the purpose of finding u_1, it is most convenient to write F as a Fourier series in ψ:

$$F(a\cos\psi,\ \omega_0 a\sin\psi,\ -K_0 a\sin\psi)$$

$$= g_0(a) + \sum_{n=1}^{\infty}\left(g_n(a)\cos n\psi + f_n(a)\sin n\psi\right) . \quad (9)$$

The $f_n(a)$ and $g_n(a)$ are *known* functions of the amplitude a, determined by the functional form of F. We shall make the assumption—always satisfied in practice—that the $f_n(a)$ and $g_n(a)$ go to zero, as a goes to zero, at least as fast as 0 (a^2).

The equation for u_1 may be written as:

$$\lambda^2\left(\frac{\partial^2}{\partial\psi^2}+1\right)u_1 = g_0(a)$$

$$+\left[f_1(a) - 2\omega_0 A_1 - 2c^2K_0D_1\right]\sin\psi$$

$$+\left[g_1(a) - 2a\omega_0 B_1 - 2ac^2K_0C_1\right]\cos\psi$$

$$+\sum_{n=2}^{\infty}\left(g_n(a)\cos n\psi + f_n(a)\sin n\psi\right) . \quad (10)$$

This is, effectively, an ordinary differential equation in ψ of a standard type. Its solution contains terms proportional to $\psi\sin\psi$ and $\psi\cos\psi$ *unless* the coefficients of $\sin\psi$ and $\cos\psi$ on the right hand side of (10) vanish. *If* they vanish, then the general solution may be found in the form

$$u_1 = v_0(a) + \sum_{n=1}^{\infty}\left(v_n(a)\cos n\psi + w_n(a)\sin n\psi\right) . \quad (11)$$

Our objective is to find a solution free from ψ-proportional terms. This impels us to choose A_1, B_1, C_1, D_1, so that

$$2\left(\omega_0 A_1 + c^2K_0D_1\right) = f_1(a) , \quad (12a)$$

$$2a\left(\omega_0 B_1 + c^2K_0C_1\right) = g_1(a) .$$

Two more relations may be deduced from the conditions

$$\frac{\partial^2\psi}{\partial t\,\partial x} = \frac{\partial^2\psi}{\partial x\,\partial t}$$

and

$$\frac{\partial^2 a}{\partial t\,\partial x} = \frac{\partial^2 a}{\partial x\,\partial t} .$$

Reference to (6) and (7) shows that, to lowest significant order, these become

$$A_1\frac{dC_1}{da} = D_1\frac{dB_1}{da} , \quad (13a)$$

$$A_1\frac{dD_1}{da} = D_1\frac{dA_1}{da} . \quad (13b)$$

Equations 12 and 13 are four relations for the four unknowns A_1, B_1, C_1, D_1, and once they have been obtained, (10) can be solved in a straightforward way by means of (11), and u_1 will contain no ψ-proportional terms. The function B_1 can be interpreted as a "frequency shift" and C_1 as a "wave number shift."

However, it is impossible to solve (12) and (13) completely without specifying the physical problem in more detail. For instance, if we wish to work a *boundary value* problem in which f is required to oscillate sinusoidally at a given x for all t, we do not expect a or ψ to vary with t, and B_1 and A_1 may be set equal to zero. This leaves

$$C_1 = \frac{g_1(a)}{2c^2K_0a} \quad (14a)$$

as the "wave number shift" and

$$D_1 = \frac{f_1(a)}{2c^2K_0} . \quad (14b)$$

If we are interested in an *initial value* problem in which a pure sine wave is given for all x at t = 0, we may correspondingly set C_1 and D_1 equal to zero and solve for A_1 and B_1. Other choices are necessitated by still other problems. (Note in passing that in both cases, the corrections approach zero as a approaches zero, as they must in order to make physical sense.

All this has been for a completely general F, imagined to contain all harmonics in ψ. In practice, it often happens that F contains only a few harmonics. Observe that secularity may *not* arise for some forms of the nonlinear term. Thus, if F is proportional to $(f)^2$ or $f\partial f/\partial x$, say, then $f_1(a)$ and $g_1(a)$ will both vanish; and in this order, we may set A_1, B_1, C_1, D_1, all identically zero, which is equivalent to a completely straightforward kind of perturbation theory. On the other hand, if we were to have F proportional to $(f)^3$, $f_1(a)$ is still zero, but $g_1(a) = 3a^3/4$ and straightforward perturbation theory no longer works. The occurrence of secularity thus depends, in a given order, on the particular form of the nonlinear term. In the following sub-section, we shall show that, for a given F secularity likewise may or may not appear, depending on the choice of the $\epsilon = 0$ solution.

To close this discussion, we give the full solution for $u_1(a,\psi)$ of Equation 11. If we make the (arbitrary) choice that all the first harmonic shall be collected in lowest order, we may choose $v_1(a) = w_1(a) = 0$. Then for $n \neq 1$,

$$\left.\begin{aligned} v_n(a) &= \frac{g_n(a)}{\lambda^2(-n^2+1)} , \\ w_n(a) &= \frac{f_n(a)}{\lambda^2(-n^2+1)} , \end{aligned}\right\} \qquad (15a)$$

$$\begin{pmatrix} f_n(a) \\ g_n(a) \end{pmatrix} = \frac{1}{\pi} \int_0^{2\pi} \begin{pmatrix} \sin n\psi \\ \cos n\psi \end{pmatrix} F\left(a\cos\psi,\ \omega_0 a \sin\psi,\ -K_0 a \sin\psi\right) d\psi . \qquad (15b)$$

When the phenomenon of secularity occurs, one must go to one higher order in ϵ to get corrections to f which are uniformly valid through terms of $0(\epsilon)$ for changes in x and t of $0(1/\epsilon)$, and so the solution (15) is not of great interest by itself (Reference 2). However, the expressions such as (14) for frequency and wave number shifts *are* accurate to $0(\epsilon)$, and are usually more accessible to measurement than the expression for u_1, in any case.

Two Coupled Monochromatic Waves of Differing Frequencies

We now proceed to the case in which Equation 2 has the $\epsilon = 0$ solution

$$f = a\left[\cos\left(K_1 x - \omega_1 t + \phi_1\right) + \cos\left(K_2 x - \omega_2 t + \phi_2\right)\right] , \qquad (16)$$

where

$$\omega_1^2 - c^2K_1^2 - \lambda^2 = \omega_2^2 - c^2K_2^2 - \lambda^2 = 0 , \qquad (17)$$

but where there is not necessarily any other relationship between ω_1, K_1 and ω_2, K_2. If we were to do a straightforward perturbation theory, assuming

$$f = f^{(0)} + \epsilon f^{(1)} + \epsilon^2 f^{(2)} + \cdots , \qquad (18)$$

with $f^{(0)}$ given by (16), it is clear that we can always find a secularity-free perturbation theoretic solution to (2) of the form

$$f^{(n)} = \sum_{r,s=-\infty}^{\infty} f_{rs}^{(n)}(a) \exp i\left[r\left(K_1 x - \omega_1 t + \phi_1\right) + s\left(K_2 x - \omega_2 t + \phi_2\right)\right] \qquad (19)$$

except when there happen to exist two integers $\bar{r}$, $\bar{s}$ such that

$$\lambda^2\left(-\bar{r}^2 - \bar{s}^2 + 1\right) - 2\bar{r}\bar{s}\left(\omega_1\omega_2 - c^2K_1K_2\right) = 0 . \qquad (20)$$

When (20) is fulfilled $f^{(1)}$, like the u_1 of Equation 10, will in general contain terms proportional to the phases, as well as trigonometric terms.

For a randomly selected pair of values ω_1, K_1 and ω_2, K_2, there will exist no integers $\bar{r}$, $\bar{s}$ which

fulfill (20), and there is no necessity for the Bogolyubov techniques. When, on the other hand, two such integers $\tilde{r}$, $\tilde{s}$, do exist [the relation (20) will also hold for $-\tilde{r}$, $-\tilde{s}$, of course], then the Bogolyubov techniques are called for. One method of introducing them is the following, though it is not the only method.* Seek a solution of the form

$$f = a(\cos\psi_1 + \cos\psi_2) + \epsilon u_1(a, \psi_1, \psi_2) + \cdots$$

where now the amplitude a and the phases ψ_1 and ψ_2 develop according to:

$$\left.\begin{aligned}
\frac{\partial a}{\partial t} &= \epsilon A(a, \psi_1, \psi_2) + \cdots, \\
\frac{\partial a}{\partial x} &= \epsilon B(a, \psi_1, \psi_2) + \cdots, \\
\frac{\partial \psi_{1,2}}{\partial t} &= -\omega_{1,2} + \epsilon C_{1,2}(a, \psi_1, \psi_2) + \cdots, \\
\frac{\partial \psi_{1,2}}{\partial x} &= K_{1,2} + \epsilon D_{1,2}(a, \psi_1, \psi_2) + \cdots.
\end{aligned}\right\} \quad (21)$$

The left hand side of (2) may again be computed, by using the relations (21), to give

$$\begin{aligned}
\frac{\partial^2 f}{\partial t^2} - c^2\frac{\partial^2 f}{\partial x^2} + \lambda^2 f = {} & a\left[\left(-\omega_1^2 + c^2K_1^2 + \lambda^2\right)\cos\psi_1 + \left(-\omega_2^2 + c^2K_2^2 + \lambda^2\right)\cos\psi_2\right] \\
& + \epsilon\Bigg\{\sin\psi_1\left[2\omega_1 A + a\omega_1\frac{\partial C_1}{\partial\psi_1} + a\omega_2\frac{\partial C_1}{\partial\psi_2}\right. \\
& \left. + 2c^2K_1B + ac^2K_1\frac{\partial D_1}{\partial\psi_1} + ac^2K_2\frac{\partial D_1}{\partial\psi_2}\right] + \sin\psi_2\left[2\omega_2 A\right. \\
& \left. + a\omega_1\frac{\partial C_2}{\partial\psi_1} + a\omega_2\frac{\partial C_2}{\partial\psi_2} + 2c^2K_2B + ac^2K_1\frac{\partial D_2}{\partial\psi_1} + ac^2K_2\frac{\partial D_2}{\partial\psi_2}\right] \\
& + \cos\psi_1\left[2a\omega_1C_1 - \omega_1\frac{\partial A}{\partial\psi_1} - \omega_2\frac{\partial A}{\partial\psi_2} + 2ac^2K_1D_1 - c^2K_1\frac{\partial B}{\partial\psi_1}\right. \\
& \left. - c^2K_2\frac{\partial B}{\partial\psi_2}\right] + \cos\psi_2\left[2a\omega_2C_2 - \omega_1\frac{\partial A}{\partial\psi_1} - \omega_2\frac{\partial A}{\partial\psi_2} + 2c^2aK_2D_2\right. \\
& \left. - c^2K_1\frac{\partial B}{\partial\psi_1} - c^2K_2\frac{\partial B}{\partial\psi_2}\right]\Bigg\} + \epsilon\left\{\lambda^2\left[\frac{\partial^2 u_1}{\partial\psi_1^2} + \frac{\partial^2 u_1}{\partial\psi_2^2} + u_1\right]\right. \\
& \left. + 2\left[\omega_1\omega_2 - c^2K_1K_2\right]\frac{\partial^2 u_1}{\partial\psi_1\partial\psi_2}\right\} + 0(\epsilon^2) \qquad (22)
\end{aligned}$$

The zeroth order part of (22) vanishes identically. The first order part now must be equated to $\epsilon F[a(\cos\psi_1+\cos\psi_2), a(\omega_1\sin\psi_1+\omega_2\sin\psi_2), -a(K_1\sin\psi_1+K_2\sin\psi_2)]$. It is most convenient to write this as a complex Fourier series,

$$F = \sum_{m,n=-\infty}^{+\infty} F_{mn}(a)\, e^{i(m\psi_1+n\psi_2)} \qquad (23)$$

where $F_{mn}^* = F_{-m-n}$, since we deal only with real quantities. Calling everything in the first set of braces in Equation 22 "$G(a, A, B, C_{1,2}, D_{1,2})$," we may finally write the equation for u_1 as

$$\lambda^2\left(\frac{\partial^2 u_1}{\partial\psi_1^2} + \frac{\partial^2 u_1}{\partial\psi_2^2} + u_1\right) + 2\left(\omega_1\omega_2 - c^2K_1K_2\right)\frac{\partial^2 u_1}{\partial\psi_1\partial\psi_2} = -G(a, A, B, C_{1,2}, D_{1,2}) + \sum_{m,n=-\infty}^{+\infty} F_{mn}(a)e^{i(m\psi_1+n\psi_2)}. \qquad (24)$$

*For simplicity, we assume only one independent pair of such integers, $\tilde{r}$, $\tilde{s}$, exists.

This equation, analogously to (10), has a readily obtainable secularity-free solution of the form

$$u_1(\psi_1, \psi_2, a) = \sum_{m,n=-\infty}^{+\infty} u_{mn}(a) e^{i(m\psi_1 + n\psi_2)} \tag{25}$$

if and only if the functions A, B, $C_{1,2}$, $D_{1,2}$ *are chosen to depend on* ψ_1 *and* ψ_2 *in such a manner that:*

$$\oint d\psi_1 \oint d\psi_2 \, e^{-i(\vec{r}\psi_1 + \vec{s}\psi_2)} G(a, A, B, C_{1,2}, D_{1,2}) - F_{\vec{r}\,\vec{s}}(a) = 0 . \tag{26}$$

Since (26) involves complex numbers, it really amounts to two real equations upon A, B, $C_{1,2}$, $D_{1,2}$, which are linear partial differential equations with periodic coefficients.* We need not write them down in full detail, since they are most cumbersome, and not of interest for our purposes here.

From the requirements that

$$\partial^2 a/\partial x\,\partial t = \partial^2 a/\partial t\,\partial x , \quad \partial^2 \psi_{1,2}/\partial x\,\partial t = \partial^2 \psi_{1,2}/\partial t\,\partial x ,$$

three additional conditions may be adduced. They are:

$$K_1 \frac{\partial A}{\partial \psi_1} + K_2 \frac{\partial A}{\partial \psi_2} = -\left(\omega_1 \frac{\partial B}{\partial \psi_1} + \omega_2 \frac{\partial B}{\partial \psi_2}\right) , \tag{27a}$$

$$K_1 \frac{\partial D_1}{\partial \psi_1} + K_2 \frac{\partial D_1}{\partial \psi_2} = -\left(\omega_1 \frac{\partial C_1}{\partial \psi_1} + \omega_2 \frac{\partial C_1}{\partial \psi_2}\right) , \tag{27b}$$

$$K_1 \frac{\partial D_2}{\partial \psi_1} + K_2 \frac{\partial D_2}{\partial \psi_2} = -\left(\omega_1 \frac{\partial C_2}{\partial \psi_1} + \omega_2 \frac{\partial C_2}{\partial \psi_2}\right) \tag{27c}$$

Satisfaction of these five relations on the six quantities A, B, $C_{1,2}$, $D_{1,2}$, then, will guarantee that the solution shall be secularity-free. It is clear that considerable latitude is left to choose the functions conveniently for whatever problem is under consideration. The point to be emphasized, however, is that secularity is by no means inherent in the equation to be studied, or in the form of the interaction term. Nor does its origin necessarily have a simple physical interpretation as it does for Equation 1, where the presence of secular terms is simply interpreted as a *resonance* between one of the frequencies present in the nonlinear coupling term and one of the natural frequencies of the system. It would take considerably more insight to apprehend the physical meaning of Equation 20, or any connection it might have with the secularity condition for Equation 10.

*We may consider A, B, · · · · as being expressed as Fourier series in ψ_1, ψ_2, in which case (26) leads to algebraic conditions on the Fourier coefficients.

THE COLD ELECTRON PLASMA

Suppose we consider a cold (no thermal motions) electron plasma of equilibrium number density n_0, moving in a uniform positive background, assumed immobile. If we make a perturbation about a uniform, field-free equilibrium, the appropriate variables for describing the system are:

$\vec{v}$ = electron velocity,

$-e(n_0 + n(\vec{x}, t))$ = electron charge density,

$+en_0$ = positive background charge density,

$\vec{E}$, $\vec{B}$ = electric and magnetic fields.

All these will be treated as perturbations—i.e., as first order in the amplitude—except for n_0.

The dynamical equations are well known:

$$\frac{\partial n}{\partial t} + \frac{\partial}{\partial \vec{x}} \cdot (n\vec{v}) + \frac{\partial}{\partial \vec{x}} \cdot (n_0 \vec{v}) = 0 , \tag{28a}$$

$$\frac{\partial \vec{v}}{\partial t} + \vec{v} \cdot \frac{\partial}{\partial \vec{x}} \vec{v} = -\frac{e}{m}\left(\vec{E} + \frac{1}{c}\vec{v} \times B\right) , \tag{28b}$$

$$\frac{1}{c}\frac{\partial \vec{B}}{\partial t} = -\frac{\partial}{\partial \vec{x}} \times \vec{E} , \tag{28c}$$

$$\frac{1}{c}\frac{\partial \vec{E}}{\partial t} = \frac{\partial}{\partial \vec{x}} \times \vec{B} + \frac{4\pi e}{c}(n_0 + n)\vec{v} , \tag{28d}$$

$$\frac{\partial}{\partial \vec{x}} \cdot \vec{E} = -4\pi en \,, \tag{29a}$$

$$\frac{\partial}{\partial \vec{x}} \cdot \vec{B} = 0 \,. \tag{29b}$$

Equations 29 can be regarded as initial conditions; once fulfilled, they are preserved by Equations 28.

A small amount of algebraic juggling shows that Equations 28 lead to

$$\frac{\partial^2}{\partial t^2}\vec{E} + c^2 \frac{\partial}{\partial \vec{x}} \times \left(\frac{\partial}{\partial \vec{x}} \times \vec{E}\right) + \omega_{pe}^2 \vec{E} = -\epsilon\, 4\pi e\left[n_0\left(\vec{v} \cdot \frac{\partial}{\partial \vec{x}}\vec{v} + \frac{e}{mc}\vec{v} \times \vec{B}\right) - \frac{\partial}{\partial t}(n\vec{v})\right], \tag{30a}$$

$$\frac{\partial^2}{\partial t^2}\frac{\partial \vec{B}}{\partial t} + c^2 \frac{\partial}{\partial \vec{x}} \times \left(\frac{\partial}{\partial \vec{x}} \times \frac{\partial \vec{B}}{\partial t}\right) + \omega_{pe}^2 \frac{\partial \vec{B}}{\partial t} = \epsilon\left[4\pi e n_0 c \frac{\partial}{\partial \vec{x}} \times \left(\vec{v} \cdot \frac{\partial}{\partial \vec{x}}\vec{v} + \frac{e\vec{v}}{mc} \times \vec{B}\right) - 4\pi ec \frac{\partial}{\partial \vec{x}} \times \frac{\partial}{\partial t}(n\vec{v})\right]. \tag{30b}$$

The formal expansion parameter ϵ has been written on the right hand side of Equations 30 only to remind us that these terms are "small" in the sense of being second order in the amplitude. The quantity ω_{pe}^2 is $4\pi n_0 e^2/m$, the plasma frequency.

If we now restrict ourselves to disturbances which are function of only one spatial dimension (x, say), we may write the expression (30a) in the form

$$\left(\frac{\partial^2}{\partial t^2} - c^2\frac{\partial^2}{\partial x^2} + \omega_{pe}^2 + c^2\delta_{1,i}\frac{\partial^2}{\partial x^2}\right)E_i = \epsilon \mathfrak{F}_i \,, \tag{31}$$

where $\delta_{1,i} = 0$ unless $i = 1$, and $\delta_{1,1} = 1$; $\mathfrak{F}_i$ represents the i^{th} component of the right hand side of (30a). Since (30b) is just the curl of (30a), it is not necessary to write it in the form of (31). It matters not at all that we have not bothered to express the right hand side of (30a) as a function of $\vec{E}$ alone, since we always need the values of the quantities which enter in $\mathfrak{F}_i$ to one lower order than those under consideration on the left hand side of (30a); and those will be shown to be obtainable in each order directly from (28).

The $\epsilon = 0$ solution to (31) is just the standard cold-plasma set of field-free normal modes, which is well understood (Reference 7). It will be apparent that each component of (31) has the form of Equation 2 and is therefore immediately susceptible to the methods already discussed. Equation 31 represents a generalization of the equation treated by Jackson (Reference 5), which, however, does not exhibit all the features of (31) because of the absence of x-derivatives on the left hand side in the purely electrostatic case, and of the fact that the x and t dependences separate.

The number of possibilities from (31) is very large, due to the wide range of choices for the $\epsilon = 0$ solution. The simplest case—beyond that of a pure electrostatic oscillation (Reference 5)—is that of a pure transverse, linearly polarized, electromagnetic wave. We shall extend this solution to the next two orders above the linear approximation in the following paragraphs.

Calling our base vectors $\hat{e}_x$, $\hat{e}_y$, $\hat{e}_z$, the appropriate zeroth order solution is

$$\begin{aligned} \vec{E} &= E_0 \hat{e}_y \cos(k_0 x - \omega_0 t) \,, \\ \vec{B} &= \frac{ck_0}{\omega_0} E_0 \hat{e}_z \cos(k_0 x - \omega_0 t) \\ \vec{v} &= \frac{eE_0}{m\omega_0} \hat{e}_y \sin(k_0 x - \omega_0 t) \\ n &= 0 \end{aligned} \tag{32}$$

where

$$\omega_0^2 = \omega_{pe}^2 + c^2 k_0^2 \tag{33}$$

This zeroth order solution for $\vec{E}$ is conveniently written:

$$E_i = a\left[\phi_i e^{i(k_0 x - \omega_0 t)} + \phi_i^* e^{-i(k_0 x - \omega_0 t)}\right]$$

where ϕ is a column vector

$$\phi = \begin{pmatrix} 0 \\ \frac{1}{2} \\ 0 \end{pmatrix} \tag{34}$$

and $a = E_0$ is the amplitude.

The results of Section II suggest that we seek a solution of the form

$$E_i = a\left(\phi_i e^{i\psi} + \phi_i^* e^{-i\psi}\right) + \epsilon u_i^{(1)}(a,\psi) + \epsilon^2 u_i^{(2)}(a,\psi) + \cdots \tag{35}$$

where now the amplitude a and phase ψ are to vary in x and t according to

$$\left.\begin{aligned} \frac{\partial a}{\partial t} &= \epsilon A_1(a) + \epsilon^2 A_2(a) + \cdots , \\ \frac{\partial a}{\partial x} &= \epsilon D_1(a) + \epsilon^2 D_2(a) + \cdots , \\ \frac{\partial \psi}{\partial t} &= -\omega_0 + \epsilon B_1(a) + \epsilon^2 B_2(a) + \cdots , \\ \frac{\partial \psi}{\partial x} &= k_0 + \epsilon C_1(a) + \epsilon^2 C_2(a) + \cdots , \end{aligned}\right\} \tag{36}$$

with the functions $A_1, B_1, \cdots$, as yet undetermined.

The left hand side of (31), expressed in terms of these functions, becomes, for i = 1,

$$\left(\frac{\partial^2}{\partial t^2} + \omega_{pe}^2\right) E_1 = \epsilon\left\{\omega_0^2 \frac{\partial^2 u_1^{(1)}}{\partial \psi^2} + \omega_{pe}^2 u_1^{(1)} + 2\left(-i\omega_0 A_1 + a\omega_0 B_1\right)\phi_1 e^{i\psi} + 2\left(i\omega_0 A_1 + a\omega_0 B_1\right)\phi_1^* e^{-i\psi}\right\} + 0(\epsilon^2) ; \tag{37}$$

and for i = 2 or 3,

$$\left(\frac{\partial^2}{\partial t^2} - c^2\frac{\partial^2}{\partial x^2} + \omega_{pe}^2\right) E_i = \epsilon\left\{\omega_{pe}^2\left(\frac{\partial^2 u_i^{(1)}}{\partial \psi^2} + u_i^{(1)}\right) + 2\left(-i\omega_0 A_1 + a\omega_0 B_1 - ic^2 k_0 D_1 + ac^2 C_1 k_0\right)\phi_i e^{i\psi} + 2\left(i\omega_0 A_1 + a\omega_0 B_1 + ic^2 k_0 D_1 + ac^2 C_1 k_0\right)\phi_i^* e^{-i\psi}\right\} + 0(\epsilon^2) . \tag{38}$$

These three expressions are to be equated to the three components of the right hand sides of (31), with the quantities $\vec{v}$, $\vec{E}$, $\vec{B}$, n replaced by their values (32), and with the substitution $E_0 \to a$. The result can be conveniently written as a Fourier series

$$\mathfrak{F}_i = \sum_{m=-\infty}^{+\infty} \Phi_i(m,a)\, e^{im\psi} , \tag{39}$$

where the Φ_i (m, a) are *known* functionals of a.

Suppose we now consider whether secular terms can arise. We seek the solution for $u_i^{(1)}$ in the form

$$u_i^{(1)} = \sum_{m=-\infty}^{\infty} u_i^{(1)}(m,a)\, e^{im\psi} \tag{40}$$

The equation for $u_i^{(1)}$ then becomes

$$\sum_{m=-\infty}^{\infty}\left[\left(-m^2+1\right)\omega_{pe}^2 + m^2\delta_{1,i}\left(\omega_{pe}^2 - \omega_0^2\right)\right] u_i^{(1)}(m,a)\, e^{im\psi} = -2\left(-i\omega_0 A_1 + a\omega_0 B_1 - ic^2 k_0 D_1 + ac^2 C_1 k_0\right)\phi_i e^{i\psi} - 2\left(i\omega_0 A_1 + a\omega_0 B_1 + ic^2 k_0 D_1 + ac^2 C_1 k_0\right)\phi_i^* e^{-i\psi} + \sum_{m=-\infty}^{\infty}\Phi_i(m,a)\, e^{im\psi} , \tag{41}$$

(we have made use of $\phi_1 = 0$).

We may always solve (41) for $u_i^{(1)}$ (m, a) with the absence of ψ-proportional terms, except when a coefficient of $u_i^{(1)}$ (m, a,) on the left hand side of (41) vanishes for some m. Since $\omega_0^2 > \omega_{pe}^2$ it is clear that this can only occur when $m = \pm 1$, and only in the components i = 2 or 3. However, a term-by-term inspection of Φ_i (m, a) reveals that for i = 2 or 3, the quantities Φ_i (m, a) vanish identically for the choice (33) of the zeroth order values. Thus a result emerges which would have been hard to guess from the original set of equations

(28). Secular behavior cannot arise in second order for this case, and we may set $A_1 = B_1 = C_1 = D_1 = 0$, and solve in a completely straightforward way for $u_i^{(1)}(m, a)$:

$$u_i^{(1)}(m,a) = \frac{\Phi_i(m,a)}{\left(-m^2+1\right)\omega_{oe}^2 + m^2\delta_{1,i}\left(\omega_{pe}^2 - \omega_0^2\right)}, \quad m \neq \pm 1,$$

$$u_i^{(1)}(\pm 1, a) = 0. \tag{42}$$

The calculation of the coefficients of the Fourier coefficients is a matter of simple algebra, and we may write down in full the first correction to the linear solution (32):

$$\vec{u}^{(1)} = \vec{E}^{(1)} = E_0^2 \frac{ek_0\omega_{pe}^2}{2m\omega_0^2} \frac{\sin 2\left(k_0x - \omega_0 t\right)}{\left(3\omega_{pe}^2 + 4c^2k_0^2\right)} \hat{e}_x ,$$

$$\vec{B}^{(1)} = 0 ,$$

$$n^{(1)} = -E_0^2 \frac{e^2n_0k_0^2}{m^2\omega_0^2} \frac{\cos 2\left(k_0x - \omega_0 t\right)}{\left(3\omega_{pe}^2 + 4c^2k_0^2\right)} ,$$

$$\vec{v}^{(1)} = -E_0^2 \frac{e^2k_0}{m^2\omega_0} \frac{\cos 2\left(k_0x - \omega_0 t\right)}{\left(3\omega_{pe}^2 + 4c^2k_0^2\right)} \hat{e}_x \tag{43}$$

The only qualitatively new feature which shows up in $0(\epsilon)$, then, is a *longitudinal* electric field. It is only in $0(\epsilon^2)$ that secularity manifests itself and it becomes necessary to use the Bogolyubov methods. Making use of the $0(\epsilon)$ solution we have just computed, we may write down the $0(\epsilon^2)$ part of the left hand side of (31) in terms of the A's, B's, etc. We shall need:

$$\left\{\frac{\partial^2 E_i}{\partial t^2} - c^2\frac{\partial^2 E_i}{\partial x^2} + \omega_{pe}^2 E_i\right\}_{0(\epsilon^2)_{part}}$$

$$= \epsilon^2\Big[\left(\omega_0^2 - c^2k_0^2\right)\frac{\partial^2 u_i^{(2)}}{\partial\psi^2} + \omega_{pe}^2 u_i^{(2)}$$

$$+ \left(-2i\omega_0A_2 + 2a\omega_0B_2 - 2ik_0c^2D_2 + 2ak_0C_2\right)\phi_i e^{i\psi}$$

$$+ \left(2i\omega_0A_2 + 2a\omega_0B_2 + 2ik_0c^2D_2 + 2ac^2k_0C_2\right)\phi_i^* e^{-i\psi}\Big]. \tag{44}$$

It is also a simple matter to compute the $0(\epsilon^2)$ part of the right hand side of (31) from (43) and (32). The result expressed in terms of a and ψ, is

$$\mathcal{F}_i\Big]_{0(\epsilon^2)_{part}} = \frac{\omega_{pe}^2k_0^2e^2a^3\delta_{i,2}}{4\omega_0^2m^2\left(3\omega_{pe}^2 + 4c^2k_0^2\right)}\left\{3\left(e^{3i\psi} + e^{-3i\psi}\right) - \left(e^{i\psi} + e^{-i\psi}\right)\right\}. \tag{45}$$

Since the only nonvanishing component of (45) is in the 2-direction (y-direction), the only nonvanishing component of $u_i^{(2)}$ will also be $u_2^{(2)}$. The equation to be solved in the $0(\epsilon^2)$ approximation is, therefore:

$$\left(\omega_0^2 - c^2k_0^2\right)\frac{\partial^2 u_2^{(2)}}{\partial\psi^2} + \omega_{pe}^2u_2^{(2)} =$$

$$-\left(-2i\omega_0A_2 + 2a\omega_0B_2 - 2ik_0c^2D_2 + 2ak_0c^2C_2\right)\phi_2 e^{i\psi}$$

$$-\left(2i\omega_0A_2 + 2a\omega_0B_2 + 2ik_0c^2D_2 + 2ac^2k_0C_2\right)\phi_2^* e^{-i\psi}$$

$$+\frac{\omega_{pe}^2k_0^2e^2a^3}{4\omega_0^2m^2\left(3\omega_{pe}^2 + 4c^2k_0^2\right)}\left[3\left(e^{3i\psi} + e^{-3i\psi}\right) - \left(e^{i\psi} + e^{-i\psi}\right)\right]. \tag{46}$$

We may find a solution to (46) of the form

$$u_2^{(2)}(a,\psi) = \sum_{m=-\infty}^{\infty} u_2^{(2)}(a,m)e^{im\psi} . \tag{47}$$

which is free of ψ-proportional terms, if and only if the coefficients of $e^{i\psi}$ and $e^{-i\psi}$ on the right hand side of (46) vanish. This gives us two conditions on A_2, B_2, C_2, D_2, which are algebraic relations. Two more, analogous to Equations 13, are given by the requirements that

$$\frac{\partial^2 a}{\partial t\partial x} = \frac{\partial^2 a}{\partial x\partial t}$$

and

$$\frac{\partial^2\psi}{\partial t\,\partial x} = \frac{\partial^2\psi}{\partial x\,\partial t} \ ;$$

they are:

$$D_2 \frac{dA_2}{da} = A_2 \frac{dD_2}{da} \ , \qquad (48a)$$

$$D_2 \frac{dB_2}{da} = A_2 \frac{dC_2}{da} \ . \qquad (48b)$$

Let us now specialize the problem to one in which the spatial periodicity and amplitude are given and we are to calculate the "frequency shift." This means setting $C_2 = D_2 = 0$. The condition that the coefficients of $e^{i\psi}$ and $e^{-i\psi}$ on the right hand side of (46) vanish reduces to the equation

$$(i\omega_0 A_2 - a\omega_0 B_2) = \frac{\omega_{pe}^2 k_0^2 a^3}{4m^2\omega_0^2\left(3\omega_{pe}^2 + 4c^2k_0^2\right)} \qquad (49)$$

and its complex conjugate relation. We may therefore find $A_2 = 0$, and the frequency shift $\Delta\omega$ becomes (substituting E_0 for the amplitude a):

$$\Delta\omega = -B_2 = \frac{e^2k_0^2\,\omega_{pe}^2}{4m^2\omega_0^3}\,\frac{E_0^2}{\left(3\omega_{pe}^2 + 4c^2k_0^2\right)} \qquad (50)$$

It does not seem worthwhile to write down the $0(\epsilon^2)$ corrections to $\vec{E}$, $\vec{B}$, $\vec{v}$, and n, though it would be easy to do so. In any experiment one might imagine, Equation 50 would probably be the easiest quantity to measure.

It is possible to calculate the frequency shift of a *standing* electromagnetic wave of given periodicity as well. One assumes a linearly polarized standing wave solution for (31) in lowest order and writes

$$E_i = a_i \sin k_0 x \cos\psi + \epsilon u_i^{(1)}(a, \psi, x) + \epsilon^2 u_i^{(2)}(a, \psi, x) + \cdots$$

Then, only the t-dependence of the phase variable and amplitude are expanded:

$$\frac{da}{dt} = \epsilon A_1(a) + \epsilon^2 A_2(a) + \cdots$$

$$\frac{d\psi}{dt} = \omega_0 + \epsilon B_1(a) + \epsilon^2 B_2(a) + \cdots$$

In lowest significant order, the frequency shift turns out to be exactly one eighth of the result given in Equation (50).

DISCUSSION

We have given a technique for obtaining uniformly valid, perturbation theoretic solutions to the Klein-Gordon equation with a small nonlinear term. The method has been applied to calculate the second order frequency shift of an electromagnetic wave in a cold electron plasma. The smallness of the expression (50) for attainable parameters is an indication of just how good an approximation the linear theory is at these frequencies.

It is not to be inferred, however, that the method adapts itself readily to all partial differential equations. For instance, the reader can easily convince himself that it fails for a nonlinear sound wave described by the Euler equations. (It appears to fail in all situations for which the $\epsilon = 0$ equation is a wave equation, $(\partial^2/\partial t^2 - c^2\partial^2/\partial x^2)f = 0$). The real reason is that, due to a steepening of the exact nonlinear wave front—obtainable from the Riemann invariants—a vertical tangent develops after a time of 0(1/ the amplitude). This destroys any regularity properties which may have existed in the original wave profile. Since the exact solution does not remain "close" to the $\epsilon = 0$ solution in *any* sense, after a time of order $1/\epsilon$, it is not surprising that perturbation theory is of little use beyond this time.

REFERENCES

1. Krylov, N. M., and Bogolyubov, N. N., "Introduction to Non-Linear Mechanics," (Translated from two Russian monographs by S. Lefschetz), Princeton: Princeton University Press, 1943.
2. Bogolyubov, N., and Mitropolskii, Y. A., "Asymptotic Methods in the Theory of Non-Linear Oscillations," 2nd ed. (Translated from Russian), New York: Gordon and Breach Science Publishers, 1961.
3. Frieman, E. A., "On a New Method in the Theory of Irreversible Processes," *J. Math. Phys.*, 4:410-418, March 1963.
4. Sanderi, G., "The Foundations of Nonequilibrium Statistical Mechanics, I and II," Ann. Physics (N Y), 24, 332, 380 (1963).
5. Jackson, E. A., "Nonlinear Oscillations in a Cold Plasma," *Phys. Fluids*, 3:831-833, September-October 1960.
6. Sturrock, P. A., "Non-Linear Effects in Electron Plasmas," *Roy. Soc. Proc.*, A242:277-299, November 5, 1957.
7. Bernstein, I. B., and Trehan, S. K., "Plasma Oscillations, I." *Nucl. Fission*, 1:3-41, 1960.

INITIAL RESULTS OF THE IMP–I MAGNETIC FIELD EXPERIMENT*

NORMAN F. NESS, CLELL S. SCEARCE AND JOSEPH B. SEEK

The Interplanetary Monitoring Platform, IMP–I, or Explorer XVIII launched on November 27, 1963 has provided the first accurate measurements of interplanetary magnetic fields. The initial apogee of the satellite was located 197,616 kilometers on the Sunlit side of the Earth, with an apogee-Earth-Sun angle of 33°. Detailed measurements of the interplanetary magnetic fields and the interaction of the solar wind with the geomagnetic field have been performed. This paper presents the initial results of the magnetic field experiement.

The strength of the interplanetary magnetic field is found to vary between 4 and 7 gammas with extreme values as low as 1 and as high as 10 gammas. The magnitude, however, is extremely stable over time scales of hours, although direction changes are significant. The general configuration of the interplanetary magnetic field is slightly below the plane of the ecliptic and approximately along the streaming angle predicted for a steady state solar wind. A significant feature of the magnetic field measurements is the discovery of fields pointed diametrically opposite the streaming angle indicating filamentary structure of interplanetary fields. Associated with the fields of opposite direction are null surfaces between the filaments and in the overall field structure.

The complex interaction of the solar wind and geomagnetic field shows a variety of significant magnetic field fluctuations and transition characteristics. The discovery of the collisionless magnetohydrodynamic shock wave at 13.4 R_e at the stagnation point associated with the super Alfvenic flow of solar plasma is one of the major results of this experiment. Details of the fluctuations are discussed as well as the gross structure and shape of the magnetospheric surface (10.2 R_e at the subsolar point) and the shock wave from the subsolar point to the nighttime geomagnetic tail. The transition region between the shock wave and the magnetopause is one of high turbulence in the magnetic field.

A unique aspect of the magnetic field data is the detection of the magnetohydrodynamic wake of the Moon during the fifth orbit when the satellite was eclipsed by the Moon's magnetosphere while in interplanetary space. The implications of this experimental discovery are discussed in view of other pertinent geophysical and astrophysical data.

1.0 INTRODUCTION

The Interplanetary Monitoring Platform, IMP–I, or Explorer XVIII, was launched from the Atlantic Missile Range, Cape Kennedy, Florida on November 27, 1963 at 0230:01.07 Universal Time. The 63 kilogram spacecraft shown in Figure 1. was placed into a highly elliptical orbit by a three stage Thor-Delta vehicle with an initial orbital period of 93.5 hours, inclination to the Earth's equator of 33.33° and eccentricity of 0.937. The initial apogee was on the Sunlit side of the Earth at an angle to the Sun of 32.7° and at a geocentric distance of 197,616 kilometers (initial perigee = 192 kilometers). The spacecraft was spin stabilized at an initial rate of 22.27 rpm with an initial spin axis-satellite-Sun angle measured by the spacecraft optical aspect sensor of 111°. The right ascension and declination of the spacecraft spin axis were initially 116.6° and −23.5° respectively. An important physical force acting on the dynamically spin stabilized satellite and its solar paddles is solar radiation pressure. This force, or more properly, the torque effects associated with this force have led to both a gradual increase of the satellite spin rate and a precessional motion of the spin axis in celestial inertial coordinates. When the spacecraft aspect to the Sun decreases below 90° the net effect of the solar radiation

*Published as *Goddard Space Flight Center Document X-612-64-95*, April 1964; also published in *Journal of Geophysical Research*, 69(17):-3531-3569 September 1, 1964.

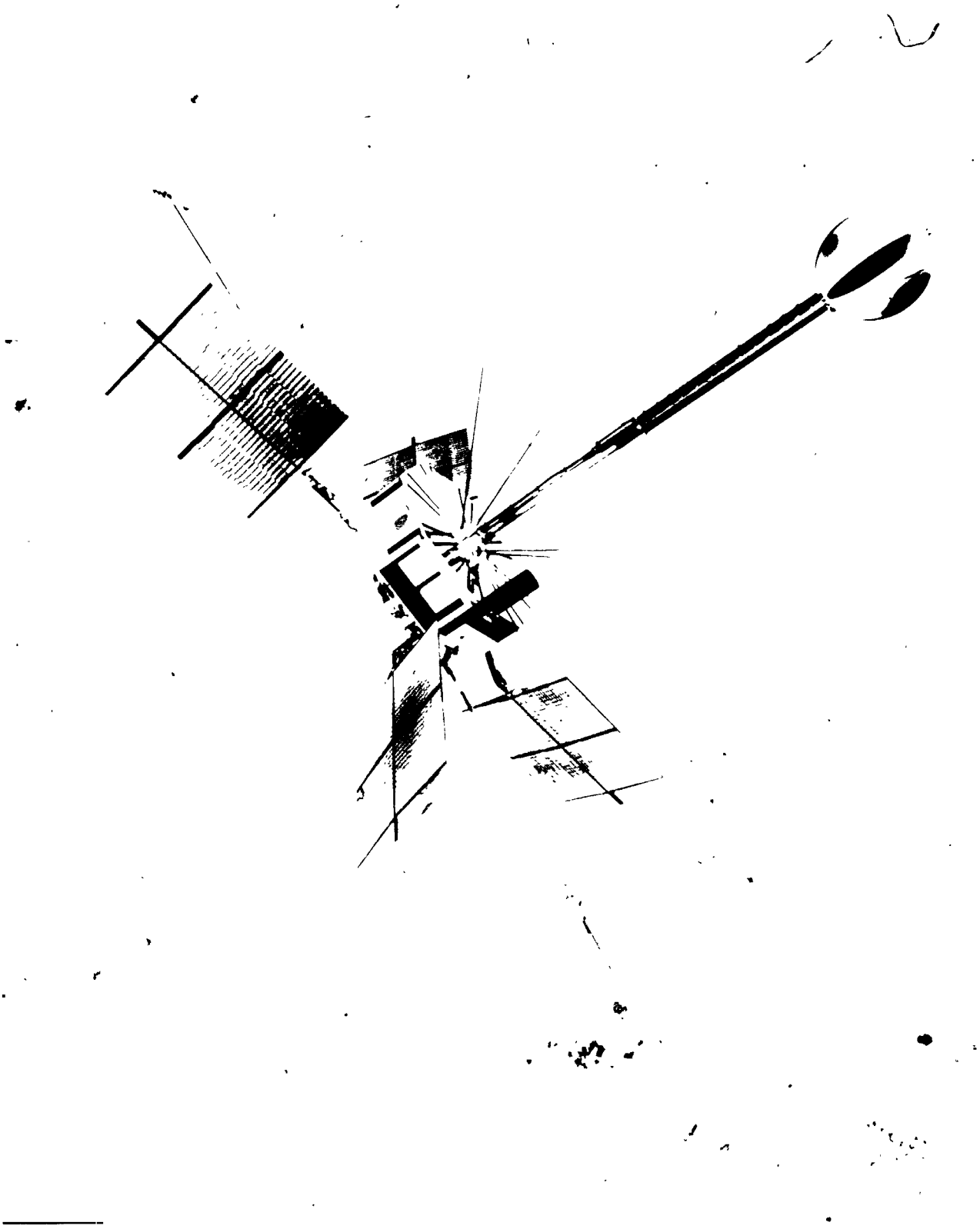


FIGURE 1.—IMP-I spacecraft in flight configuration with all appendages (solar paddles, fluxgate magnetometer booms and Rubidium vapor magnetometer tube) fully deployed.

pressure torque will be to decrease the spin rate. After 70 days in orbit the spin rate was 24.30 rpm at an aspect angle of 125° to the Sun with the location of the spin axis given by a right ascension of 111.1° and declination of −30.6°. The maximum solar aspect was 129° after 47 days in orbit. The lifetime of the satellite is expected to be at least 160 days and may well be 360 days or longer.

It is the purpose of this paper to discuss primarily the results of the magnetic field experiment on board the IMP–I spacecraft. At this early date approximately 70 days of data collected by the spacecraft has been acquired, processed and analyzed. It is important to note that our initial impressions of the significance of certain features of the data and their interpretations may be changed as the quantity of information returned from the satellite is increased and as our overall statistics and characteristics of the magnetic fields and associated phenomena become better understood.

The primary mission of the magnetic field experiment was four-fold; the investigation of:

1. The *interplanetary magnetic field*, its magnitude, direction and temporal variations as correlated with solar and terrestrial phenomena,
2. The *transition region* associated with the interaction of the solar wind with the geomagnetic field which leads to the study of two characteristic boundary surfaces, namely:
3. The *collisionless magnetohydrodynamic shock wave* surface separating the undisturbed interplanetary medium from the transition region, and
4. The *magnetopause boundary* separating the transition region from the magnetosphere, that region of space containing the geomagnetic field and encompassing the Earth in the classical concept of the Chapman-Ferraro geomagnetic cavity.

These four major phenomena have been investigated in considerable observational detail thus far by this experiment but shall be treated from a summary viewpoint at this time. Future publications concerning the IMP–I magnetic field experiment will treat substantially more detailed results of the magnetic fields in space and their solar-terrestrial correlations and also the correlations with other sensing elements carried on board the same spacecraft measuring importantly related phenomena such as the low energy proton flux "solar wind" and the higher energy proton-electron particle fluxes.

An additional feature investigated by the spacecraft has been associated with the magnetohydrodynamic wake of the Moon, as it interacts with the solar wind. This was an unexpected event during December, 1963 which placed the satellite in an approximately lunar eclipsed position with respect to solar plasma propagation from the general direction of the Sun. An even more fortuitous event studied during the lifetime of the satellite occurred near the apogee of orbit No. 2 during which time a geomagnetic sudden commencement storm occurred on December 2, 1963 at 2117 UT. This was the fifth in a series of recurrent M-region storms and such events frequently appear as a sudden commencement.

A clearly unique event was observed to occur in the interplanetary magnetic field data three minutes prior to the terrestrial magnetic field event.

2.0 HISTORICAL BACKGROUND

The existence, general description and temporal behavior of the interplanetary magnetic field have been deduced in the past from a variety of terrestrial observations notably the analysis of energetic particle trajectories impacting the Earth energetic particle trajectories impacting the Earth. An analysis by McCracken (1962) is indicative of the most recent efforts in this respect. Direct measurements in space by Pioneer V (Coleman, Davis and Sonett, 1960) with a search coil magnetometer were initially interpreted as consistent with a steady field component of 2.5 gammas normal to the ecliptic plane. In addition it was also reported that fluctuations as large as 50 gammas were detected during times of magnetic disturbance (Coleman, Sonett and Davis, 1961). Recent efforts to re-analyze and re-interpret the Pioneer V data by Greenstadt (1963, 1964) have yielded results inconsistent with the initial findings in that the fields are now found to have been much nearer to being in the plane of the ecliptic with a magnitude of 5–10 gammas.

A steady field normal to the plane of the ecliptic is inconsistent with a number of models of the interplanetary magnetic field in which solar magnetic lines of force are stretched out away from the Sun by the highly conducting streaming solar plasma (Parker, 1958a). The low energy proton flux detected by Explorer X (Bonetti, Bridge, Lazarus, Rossi and Scherb, 1963) was directed at all times away from the sun but fluctuated in magnitude and energy spectra as a function of time. Measurements by the Mariner spacecraft (Neugebauer and Snyder, 1962; Snyder and Neugebauer, 1964) have indicated that the flux of low energy plasma from the Sun presents a spectrum with considerable temporal and energy structure but in general is in agreement with the extrapolated results of Explorer X.

Measurements of the interplanetary magnetic field by Explorer X (Heppner, Ness, Skillman and Scearce, 1963) were distorted by a strong interaction between the streaming plasma and the magnetic field of the Earth which led to the formation within the plasma stream of a cavity containing the geomagnetic field. The magnetometer measurements on the Mariner spacecraft (Coleman, Davis, Smith and Sonett, 1962) are incomplete for total vector information, because of unknown spacecraft magnetic fields and do not provide accurate vector data on the interplanetary magnetic field. The data is consistent with an interplanetary field in the plane of the ecliptic with a strength of approximately 5 gammas normal to a Sun-satellite direction. The magnitude of this field component compares somewhat favorably with Pioneer V but the direction is different by 90°. Recent data on galactic fields (Davies, Verschuor and Wild, 1962) indicate that they are on the order of 0.5 gammas and may be up to 2.5 gammas in interstellar space.

In summary the interplanetary magnetic field has been estimated by various methods to be approximately 1–10 gammas average value, although steady periods might be infrequent. The variability of the interplanetary magnetic field and the extreme variations expected during solar disturbances indicated a wide dynamic range was required as well as precise and accurate measurements in order to achieve the desired scientific objectives of this experiment.

3.0 INSTRUMENTATION

The IMP–I satellite experiment complement included two types of magnetic field instruments: A Rubidium 87 vapor magnetometer and two monoaxial fluxgate magnetometers. The purpose of the Rubidium vapor magnetometer was to allow extremely accurate measurements of the interplanetary magnetic field to be performed and also to provide a wide dynamic range for magnetic field measurements within the magnetosphere and near its boundary, the magnetopause. The two monoaxial fluxgate magnetometers were employed to delineate precisely the vector characteristics of the interplanetary magnetic field. Their zero levels were calibrated inflight on the spacecraft through a combined use of the Rubidium vapor magnetometer data. The dynamic range of the fluxgates is ± 40 gammas with a sensitivity of $\pm \frac{1}{4}$ gamma while the dynamic range of the Rubidium vapor magnetometer is 3–500 gammas with a sensitivity dependent upon the magnitude of the field; the 0.1 percent precision of the ground digitization system leading to maximum errors of ± 0.5 gammas at the upper limit.

The Rubidium vapor magnetometer measures the absolute scalar intensity of the magnetic field by measuring the Zeeman splitting of the sublevels in the ground state of the atom (Bloom, 1962). For the Rubidium 87 isotope the separation between adjacent levels is 6.99592 cycles per sec. per gamma. In order to detect these separations, a technique known as optical pumping is employed to selectively populate one of these sub-levels. As the vapor in the absorption cell is "pumped" by a spectral lamp it absorbs the light until a final state is reached in which the cell becomes transparent. The application of a weak AC magnetic field at the frequency corresponding to the separation of the Zeeman sublevels (which is the Larmor frequency) then results in the redistribution of the population to all the levels and hence re-absorption of the incident light. It is possible to couple the modulated light produced and subsequently detected by a silicon photocell to the coil producing the AC magnetic field. Such a feed-back loop with the appropriate phasing of signals operates the magnetometer as an atomic self-oscillator with the resonant frequency being the Larmor frequency.

A "flashlight" configuration (Ruddock, 1961) of one lamp and absorption cell has been employed in the particular configuration on the IMP spacecraft (see Figure 2).

Although the Larmor frequency is dependent only upon the magnetic field intensity the inherent signal-to-noise ratio depends upon the angle (α) between the optical axis of the system (along the spin axis in IMP–I) and the magnetic field as approximately sin 2α. This leads to the formation of two null regions of field orientation, polar and equatorial relative to the spin axis, in which

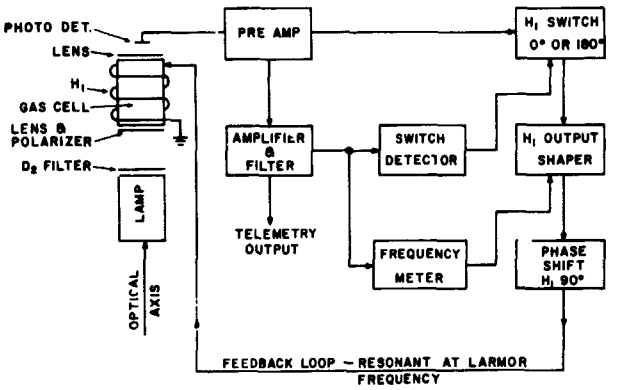


FIGURE 2.—Functional block diagram of single absorption cell self-resonant Rubidium vapor magnetometer instrumented for IMP–I.

the magnetometer will not self-oscillate with sufficient signal-to-noise ratio for accurate quantization by the ground digitization system. The fluxgate magnetometers however do measure the magnetic field vector at an extended range (see below) when in the null regions of the Rubidium vapor magnetometer and thus provide complete coverage of the magnetic field in interplanetary space over the range 0.25 to 40 gammas or more.

It is possible to convert the scalar Rb^{87} magnetometer to a vector instrument by application of a sequence of known triaxial bias magnetic fields. On a spin stabilized spacecraft, the rotation of the structure reduces the complexity of the bias coil system so that only one set of coils and one bias vector field is required. By a proper choice of the orientation and magnitude of the bias field the instrument will measure vector magnetic fields with an accuracy dependent upon the unknown vector field magnitude and orientation. An additional purpose of the bias coil system is to provide a means for in-flight calibration of the fluxgate magnetometer zero field levels since it is possible to measure the unknown vector field simultaneously with two vector instruments.

Let F = magnitude of unknown magnetic field

α = polar angle of F to spin axis

B = magnitude of bias magnetic field

θ = polar angle of B to spin axis

R = magnitude of resultant field measured

ω = spin frequency of spacecraft

Then:

$$R = \sqrt{F^2+B^2+2FB\cos\alpha\cos\theta \overline{+2FB\sin\alpha\sin\theta\cos\omega t}} \quad (3.1)$$

The analysis of $R(t)$ will yield values of F and α assuming B and θ are known.

A major problem associated with the utilization of a Rubidium vapor magnetometer in satellite applications is the necessity to maintain both the lamp used for optical pumping and resonance observation as well as the absorption cell employed at optimum temperatures. These unfortunately range over only limited values and for the gas cell an optimum temperature is approximately $42.5° \pm 50°$C. The temperature at the lamp bulb base support structure is optimumly $110° \pm 10°$C. The development of the IMP Rubidium 87 vapor magnetometer was predicated upon the necessity to include active thermal control circuits and heating coils on the lamp and gas cell. Because of the difference in optimum temperatures required it was necessary to operate three completely independent temperature control systems: one for the lamp and two for the gas cell. Figure 3. illustrates the results of the active thermal control system on the spectral lamp from telemetered temperatures of the lamp bulb. It is seen that shortly after launch the temperature of the bulb begins to rise until it reaches the critical value of 120°C at which time the heat source is turned off. When the temperature reaches the lower limit of 100°C the active thermal control once again begins to heat the bulb to raise its temperature and this cycle

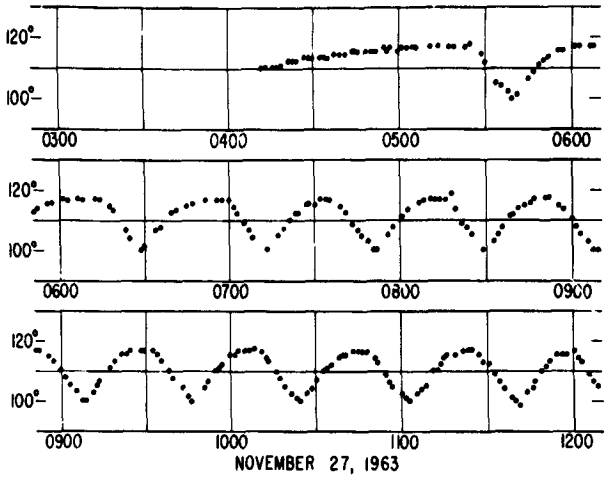


FIGURE 3.—Representative cycling of active thermal control on spectral lamp of Rubidium[87] vapor magnetometer following launch of IMP–I.

has continued since launch stabilizing at a periodicity of approximately 32 minutes. This is indicative of the thermal time constants of the lamp bulb housing and the radiative properties of the support structures enclosing the Rubidium spectral lamp. The temperature of the gas cell has always been above its lower limit and has not cycled thus far. It maintains temperature at approximately 47°C being heated by the thermal losses from the physically adjacent spectral lamp. The IMP–I Rubidium magnetometer is completely transistorized. The additional power required for controlling the environmental temperatures of these elements is a fraction of the total power required by the system (approximately 4.5 watts).

The characteristics of the fluxgate magnetometers are considerably less sensitive to temperature and their operation is proper and within the specified limits over a very wide temperature range. Thus no measurement of the temperatures of the fluxgates was required and their operation has proceeded normally. Standard calibrations on the sensitivity of the fluxgates are conducted in-flight and have been within the desired range. The operation of the instrumentation for the magnetic field experiment on the IMP–I spacecraft has been excellent since date of launch and no failures have been experienced to date.

In order to provide a uniform environment for the Rubidium magnetometer system, a spherical enclosure with a 13 inch diameter encompasses the lamp and gas cell mounted on top of the spacecraft (see Figure 1.). This sphere, appropriately surface coated for passive thermal control, provides a uniform temperature regime in which thermal equilibrium of the sensing elements can be maintained with minimum power. In addition this sphere supports a set of McKeehan (1936) coils for generating the known bias fields. The strength of the magnetic bias field employed on the IMP–I spacecraft was 20 gammas at an angle of 54° 45′ to the spin axis. These values were chosen to optimize the directional sensitivity of the resulting vector magnetometer for interplanetary magnetic fields. The output frequency from the magnetometer was used to directly phase modulate the 136 megacycle carrier of the satellite telemetry system. Subsequent digitization of the signal on the ground via a period count procedure was employed using a multiplied reference frequency recorded on the raw analog data tape as the counter. This eliminated to first order the wow and flutter errors associated with record and playback of data.

The fluxgate magnetometer measures the relative magnetic field intensity along the axis of the sensing element. The fluxgate sensor consists essentially of a saturable magnetic core which is driven at a high rate (10Kc) from positive to negative saturation by a solenoidal drive coil. Any second harmonic signal generated is due to the presence either of an ambient field component along the axis of the element or to permanent magnetization of the core material. The voltage output (0 to +˜) represents the discriminated second harmonic output which is calibrated to yield the magnitude of the field component parallel to the sensor axis while the phase indicates the direction, parallel or antiparallel. Figure 4. presents a block diagram of the fluxgate magnetometers used on IMP–I.

The PFM (pulsed frequency modulation) telemetry system employed on IMP–I encoded the fluxgate data by applying it to a voltage controlled oscillator whose frequency output, 333 cps to 938 cps, modulated the 136 mc. carrier. The fluxgate signals were digitized on the ground through the combined use of a contiguous set of narrow bandpass filters specifically designed for the PFM telemetry scheme and commonly referred to as a

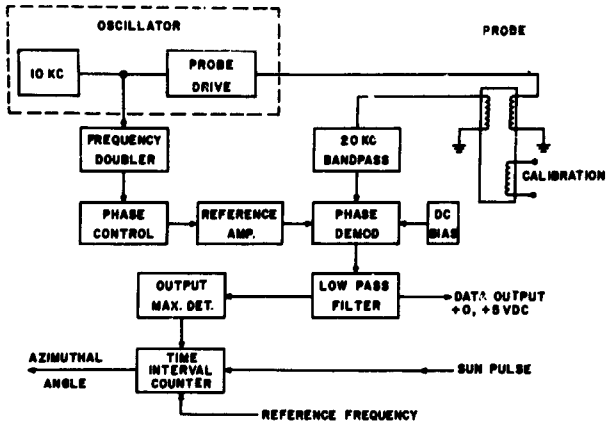


FIGURE 4.—Functional block diagram of monoaxial fluxgate magnetometer and on-board data processor of azimuthal angle instrumented for IMP-I.

"comb filter." This led to a precision in the digitization of ± 0.4 gamma associated with the precision of the comb filters (±1 percent of full scale range).

4.0 SPACECRAFT STRUCTURE AND MAGNETIC CLEANLINESS

In order to insure the success of the magnetic field experiments on the IMP spacecraft it was absolutely essential that permanent or induced magnetic fields associated with the structure and/or the electronics and instrumentation be kept to an absolute minimum. In the design and fabrication of all sub-systems, components that were highly magnetic were replaced by non-magnetic equivalents and all circuit leads and welding material were chosen to be non-magnetic or minimally magnetic. Considerable care was taken in the design and wiring of the spacecraft harness and power distribution system so that no net circulating currents and/or ground loops existed. In some instances this necessitated complex wiring techniques between and within subsystems. In addition, each solar cell module (8 x 15 series—parallel individual cell arrangement) was back wired within the paddle structure for compensation so that cell and module degradation did not change the compensating currents required. All power leads between subsystems and within subsystems at current levels above 20 ma were bifilar or twisted pair wound.

Extensive tests of individual subsystems and illuminated solar paddles at the Goddard Space Flight Center Magnetic Test Facility and finally the entire spacecraft in flight configuration at the Naval Ordnance Laboratory Magnetic Ship Modeling Facility in White Oak, Maryland and the U.S. Coast and Geodetic Survey, Fredericksburg Magnetic Observatory in Corbin, Virginia provided confidence that minimum magnetic fields would be enjoyed on the satellite after launch. Mapping of the actual spacecraft magnetic fields due to permanent, induced and stray magnetic fields indicated contamination levels at the fluxgate sensors of less than 0.6 gamma and contamination levels at the Rubidium vapor magnetometer of less than 1 gamma. These limits were set by the natural ambient magnetic field variations at the two facilities since it was not possible to completely cancel out the normal geomagnetic field fluctuations. In addition extensive tests were performed on the effects of vibration of ferro-magnetic materials in the Earth's magnetic field to evaluate the potential "perming" which might result from the vibration of the satellite during the launch maneuver. These tests indicated contamination fields which were significantly less than the sensitivity of the experiment. The in-flight measurements (see Section 5.0) show that the spacecraft fields actually achieved by the IMP-I were less than ±0.25 gamma representing one of the "cleanest" satellites carrying magnetometers launched to date.

Figure 5. illustrates a schematic cross section of the IMP satellite showing the geometrical

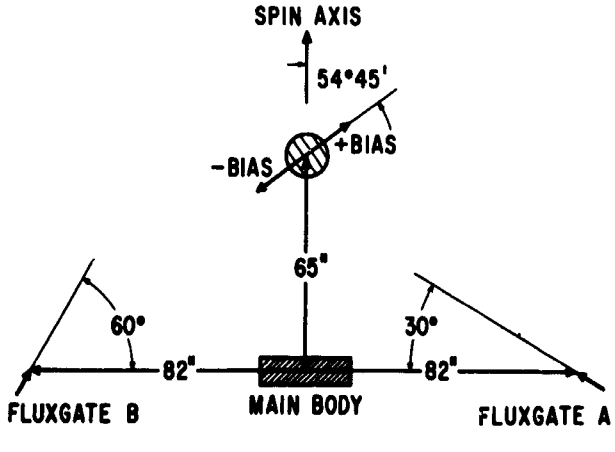


FIGURE 5.—Schematic cross section of the IMP Satellite illustrating location and orientation of sensors for magnetic field measurements.

configuration and positional location of the three magnetic field sensing elements. In spite of all the precautions taken to restrict spacecraft magnetic fields to an absolute minimum it was still necessary to place the magnetometer sensors at as remote a distance from the spacecraft as possible. The Rubidium vapor magnetometer was contained within the spherical enclosure mounted on top of a telescoping two section boom. This was erected at the same time the spacecraft was separated from the third stage ABL–258 rocket motor during the launch maneuver. This placed the Rb^{87} magnetometer resonance gas cell at a distance of 1.65 meters (65″) from the center of the spacecraft. The two monoaxial fluxgate magnetometers were mounted at the extremities of double sectioned folding booms. These were also erected at 3rd stage separation and placed the sensors 2.1 meters (82″) from the center of the spacecraft. The use of two booms was part of the experiment design to yield a spacecraft structure that was both statically and dynamically stable. Their axes were oriented so as to both lie in the same meridian plane containing the spin axis of the satellite but at different angles to the spin axis. The purpose of the two angles was to allow the sampling of magnetic fields larger than 40 gammas by either one of the two sensors depending upon the orientation of the unknown field to the spin axis. This is illustrated in Figure 6. in which the allowable magnitude as a function of polar angle (α) is shown for the θ angles $\theta_A = 30°$ and $\theta_B = 60°$ chosen for the two sensors. It is seen that by use of the two angles whose sum equals 90° the sensors tend to compliment each other in their capability to accurately and in an undistorted manner measure fields of intensities greater than 40 gammas. It was the intent of the experiment by these means to extend the effective range of the fluxgate magnetometers beyond 40 gammas and thereby compliment the Rubidium vapor magnetometer for field orientations that were nearly parallel or perpendicular to the spin axis of the spacecraft because of its signal-to-noise degradation at these orientations.

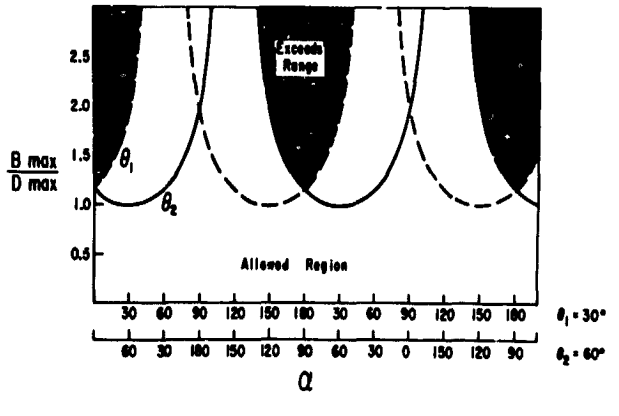


FIGURE 6.—Dynamic range of IMP-I fluxgate magnetometers as a function of sensor and field orientation. Allowed region is defined to be such that the detected output of the sensor is linearly related to the component of the field parallel to the sensor axis.

5.0 DATA SAMPLING AND ANALYSIS PROCEDURES

Magnetic field data is transmitted by the satellite sequentially in a format which time multiplexes the magnetic field information with the other scientific sensors and spacecraft performance parameters in a predetermined time shared manner. Figure 7. illustrates on a linear time scale four sequences of 16 frames containing fluxgate and Rubidium vapor magnetometer data extending over a time interval of 327.65 seconds. Each normal telemetry sequence is 81.9144 seconds in length and contains four samples of real time fluxgate magnetometer data consisting of continuous transmissions for 4.8 seconds at intervals of 20.5 seconds. Four such samples per sequence are shared between the two fluxgates,

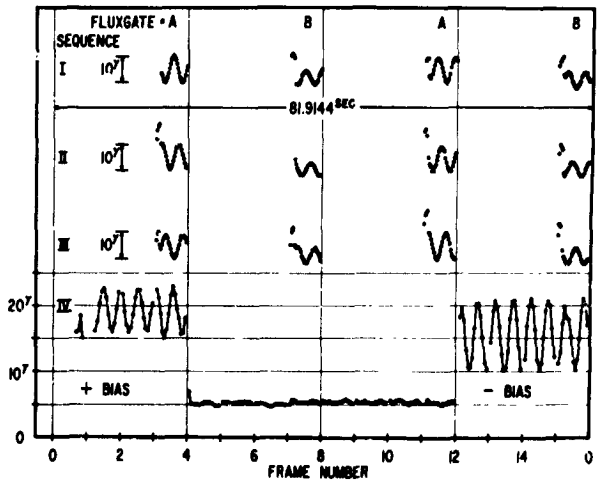


FIGURE 7.—Sample telemetry data format for IMP-I taken at a distance of 100,000 kilometers illustrating inerlacing of fluxgate and Rubidium vapor magnetometer data.

A and B, previously referred to in the description of the location of sensors on board the spacecraft. This procedure is repeated for the next two sequences in identical fashion. Every fourth sequence transmitted, however, is uniquely different from the previous three in that it contains only Rubidium vapor magnetometer data continuously transmitted for a period of 81.91 seconds with gaps associated with the synchronization time channels required by the PFM telemetry format every 5.12 seconds.

Figure 7. illustrates the effects due to the different θ angles of the monoaxial fluxgate sensors as indicated by the difference in amplitude of the spin modulated sine wave for fluxgates A and B. In addition during the first few samples of each fluxgate a sensitivity calibration is performed which linearly adds a field of 10 gammas to the measured field. By measuring the difference in the recorded fields, and taking into account the rotation of the spacecraft, the verification of the sensor's sentitivities can be performed.

The programming of the Rubidium vapor magnetometer bias coil system is also evident. During the first four frames of the 4th sequence the Rubidium bias field is in a positive direction pointing forward relative to the spin axis. During the next eight frames the bias coil is programmed off and during the last four frames the current through the bias coils is directed in the reverse sense so as to yield a bias vector which points aft of the spacecraft.

The design of the experiment was contingent upon a measurement of interplanetary magnetic fields ranging between 1 and 10 gammas. Utilization of a bias magnetic field of 20 gammas and the illustrated programming sequence allows a unique and a complete determination of the vector magnetic field characteristics through the use of mathematical procedures to analyze the resultant fields given in equation 3.1. This data was obtained at an elevation of approximately 10^5 kilometers during the first month's operation of the satellite and is representative of the general characteristics of magnetic field data obtained in interplanetary space on the IMP-I spacecraft.

A principal problem with the use of fluxgate magnetometers is the fact that they are relative devices whose absolute zeros can only be calibrated by physically reversing the sensor in a stationary magnetic field. Although the spacecraft rotates about its spin axis and thereby allows a unique determination of the magnetic field component perpendicular to the spin axis $F_\perp$, the component parallel to the spin axis, $F_\parallel$, is not capable of unique and absolute determination from the fluxgate data alone. Thus it has been necessary to calibrate the fluxgate magnetometer zero levels through the use of the absolute Rubidium vapor magnetometer data while in flight.

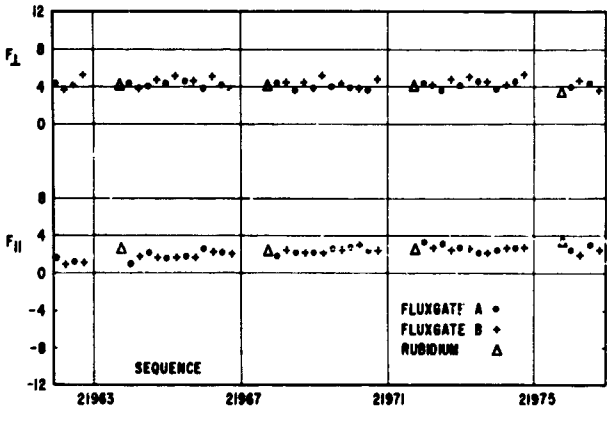


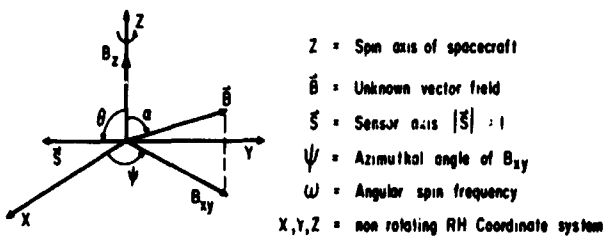
FIGURE 8.—Results of calibration of fluxgate magnetometer zero levels with Rubidium vapor magnetometer data obtained while IMP-I was in orbit.

Figure 8. illustrates the comparison of the fluxgate magnetometer data and the Rubidium vapor magnetometer data when viewed in a coordinate system in which magnetic field data is represented by perpendicular and parallel component values.

In this representation the magnetometer data from Fluxgate A has been used without any adjustment of its zero level. The data from Fluxgate B has been adjusted taking into account a zero offset of -2.1 gammas. Throughout the 70 day life of the satellite thus far, the fluxgate magnetometer data and the Rubidium vapor magnetometer data have compared very well with the zero offset for Fuxgate A remaining constant at 0 ± 0.25 gammas which is equal to the sensitivity of the analysis and the noise level of the instrument. The zero offset for Fluxgate B has changed from approximately -2.1 gammas to -2.5 gammas. The fluxgate magnetometers were manufactured by the Schonstedt Instrument Corporation, Silver Spring, Maryland. The Rubidium spectral lamp and absorption cell were

manufactured by Varian Associates, Palo Alto, California. The ancillary electronics for Fluxgate A and the complete electronics for the Rubidium vapor magnetometer were fabricated at Goddard Space Flight Center.

A natural coordinate system to use in the initial reduction of magnetic field data on a spin stabilized spacecraft is associated with the spin axis of the satellite and the satellite-Sun direction. A payload coordinate system is defined in Figure 9. in which the Z-axis is coincident with the spin axis of the spacecraft, the X-Z plane contains the satellite-Sun vector and the Y-axis is chosen to form a right-handed coordinate system. The analysis of the monoaxial fluxgate data has proceeded through the use of numerical filters designed to extract from the spin modulated magnetometer signal the first harmonic and the first and second time derivatives of the magnetometer signal at a particular instant of time. As illustrated in Figure 9. it is possible to combine the first and second time derivatives to determine both the component of the magnetic field perpendicular to the spin axis $B_{xy}=B \sin \alpha$ and also the azimuthal angle of the magnetic field ψ. These two quantities and the first harmonic then determine the component of the magnetic field parallel to the spin axis $=B \cos \alpha$. These three quantities: parallel, perpendicular and azimuthal angle completely specify the vector magnetic field $\vec{B}$.



$$\text{Detector Output} = D = \vec{B}\cdot\vec{S} = B\cos\alpha\cos\theta + B\sin\alpha\sin\theta\cos(\omega t-\psi)$$

$$\frac{\partial D}{\partial t} = -\omega B_{xy}\sin\theta\sin(\omega t-\psi)$$

$$\frac{\partial^2 D}{\partial t^2} = -\omega^2 B_{xy}\sin\theta\cos(\omega t-\psi)$$

$$B_{xy}^2 = \left[\left(\frac{\partial D}{\partial t}\right)^2\omega^2 + \left(\frac{\partial^2 D}{\partial t^2}\right)^2\right]\Big/\omega^4\sin^2\theta$$

$$\psi = \tan^{-1}\left[\omega\frac{\partial D}{\partial t}\Big/\frac{\partial^2 D}{\partial t^2}\right]$$

FIGURE 9.—Payload coordinate system and analysis procedure to obtain vector fields with monoaxial fluxgate magnetometers on a spin stabilized spacecraft

A similar procedure is used to deduce the vector fields from the Rubidium magnetometer data.

A limited amount of on-board processing of the Fluxgate A magnetometer data was performed on IMP-I as shown in the block diagram of Figure 4. A signal peak detector was developed to provide a pulse in time coincident with the maximum field value which occurred each spin period. A counter using the optical-aspect Sun pulse to open a gate and start a 100 cps count rate was terminated by this peak pulse. By this means an on-board measure of the azimuthal angle ψ was made once each telemetry sequence and provided an accurate check on the numerical procedures used to *de*-modulate the spin modulated magnetometer data. Angle accuracies are more precisely determined for azimuth because of this and are accurate to $\pm 2°$. The angular error of the polar angle α is dependent upon direct measurements of $B_\perp$ and $B_\parallel$ and thus can be large for certain ratios of $B_\perp$ and $B_\parallel$. This angular error is not significant in the transformation of the data from the initial payload coordinate system to another coordinate system by using a rotation matrix since one uses $B_\perp \sin\psi$, $B_\perp \cos\psi$ and $B_\parallel$ directly. The uncertainties in the field values and azimuth angle combine so that an estimate of the general directional accuracy of $\pm 5°$ is conservative.

Each fluxgate magnetic field vector measurement is obtained from a sample of data 4.8 seconds in length containing 30 discrete samples of the magnetic field with a sensitivity of 0.4 gammas. The bandpass of the magnetometer sensors is flat with negligible phase shift from 0 to 5 cps and falls off at 6 db per decade for higher frequencies. The analysis procedure employed to obtain a vector measurement from the sampling of the magnetic field limits the final information band width in the analysis to approximately 0 to 0.1 cycles per second. As is shown in Figure 7. the field samples from the fluxgates are obtained once every twenty seconds.

It is important to note from Figure 7. that during the interval when the Rubidium vapor magnetometer bias coil was off the spin modulation observed on the total magnetic field output is less than ± 0.25 gamma. During this particular time interval the magnetometer is sensitive to contamination fields which are perpendicular to the spin axis of the satellite at the position of the

Rubidium vapor sensor. The fluxgates always measure the component of the magnetic field perpendicular to the spacecraft spin axis accurately, independent of spacecraft contamination fields and zero offsets of the sensors. The agreement of the fluxgate and *Rb* magnetometer fields (F.) shown in Figure 8. combined with the absence of spin modulation in the Rb^{87} data gives assurance that the magnetic field data being discussed is accurate to ± 0.25 gammas and is not affected by spacecraft fields.

6.0 COORDINATE SYSTEM FOR RESULTS

The primary mission of the IMP magnetic field experiment was the delineation of the characteristics and statics of the interplanetary magnetic field. However, the particular satellite chosen for these measurements was Earth-orbiting so that all data collected was obtained in the near vicinity of the Earth. A geocentric solar ecliptic coordinate system has been defined which reflects both the interplanetary nature of the data and takes into account the motion of the Earth around the Sun. This is identical to the coordinate system used previously in the Explorer X data analysis and interpretation (Heppner et al., 1963). In this coordinate system shown in Figure 10. the origin is at the center of the Earth, the X_{se} axis points to the Sun, the Z_{se} axis is normal to the ecliptic plane and Y_{se} is chosen to form a right-handed coordinate system. These coordinates have proven extremely successful in the analysis and interpretation of the magnetic field data.

The vector magnetic field will be presented as a magnitude F and two angles: θ represents the latitude of the magnetic field being positive above the plane of the ecliptic and negative below and ϕ the longitude, being 0° when it points to the Sun and 180° when it points away from the Sun. As will become evident from studying the results of this experiment, the choice of a magnitude and two angle representation of the vector magnetic field appears to have more clearly elucidated its unique characteristics rather than the alternate representation using three components of the magnetic field. It should be noted that the coordinate system is not stationary in inertial coordinates but rotates about two principal axes and maintains its orientation as shown in Figure 10. at all times. Viewed from a non-rotating

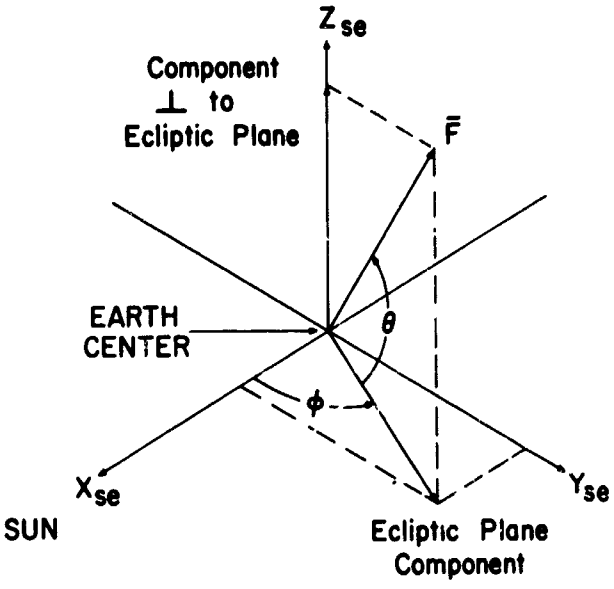


FIGURE 10.—Solar ecliptic coordinate system and associated parameter nomenclature and definitions.

solar-centric position the coordinate system moves around the Sun once each year following the orbital motion of the Earth.

In addition to its utility in the investigation of the interplanetary magnetic fields, the solar ecliptic coordinate system is also useful for studying the magnetic field phenomena associated with the interaction of the solar wind with the geomagnetic field. Since this interaction leads to the formation of a geomagnetic cavity which is strongly solar oriented the coordinate system chosen is particularly meaningful in presenting the magnetic field data in a frame of reference which illustrates the intimate solar-terrestrial phenomena being studied.

The magnetic field results herein presented represent the time average of the 12 fluxgate measurements occurring in the telemetry transmission in sequences I, II and III, yielding a sampling interval of 5.46 minutes. The solar ecliptic components for these 12 measurements are averaged by simple addition and no special numerical filters are employed. For the X_{se} component, if X^i_{se} represents the i^{th} value of the magnetic field in the sample set possible:

$$\overline{X}_{se} = \frac{1}{N} \sum_{i=1}^{N} X^i_{se} \qquad (6.1)$$

where N is less than or equal to 12 depending

upon the percentage of missing data points in the original 30 point data samples. If the number of missing points due to either digitization errors or transmission errors is greater than 10 percent, the corresponding vector field sample is not used in the analysis to determine the 5.46 minute average.

These average values for the three components are used to compute the magnitude $\bar{F}$ and the two angles θ and ϕ. In addition the average value for the components is used in computing estimates of the variance of the magnetic field for the 12 sample data set. The variance is defined as

$$\delta Y_{se}=\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(X_{se}^{i}-\bar{X}_{se}\right)^{2}} \qquad (6.2)$$

for the X_{se} component and similarly for Y_{se} and Z_{se}. The Rubidium vapor magnetometer data has not been included in this field averaging initially because its directional sensitivity for magnetic fields is not as high as the fluxgate magnetometers.

In summary, the magnetic field data is represented by the following six parameters sampled at 5.46 minute intervals:

1. $\bar{F}$ = the magnitude of the average components
$$=(\bar{X}_{se})^2+(\bar{Y}_{se})^2+(\bar{Z}_{se})^2$$
2. θ = the latitude of the field direction
$$=\tan^{-1}\{\bar{Z}_{se}/[(\bar{X}_{se})^2+(\bar{Y}_{se})^2]^{1/2}\}$$
3. ϕ = the longitude of the field measured in the plane of the ecliptic
$$=\tan^{-1}[\bar{Y}_{se}/\bar{X}_{se}] \qquad (6.3)$$
4. δX_{se} = variance of the field in the X_{se} direction
5. δY_{se} = variance of the field in the Y_{se} direction
6. δZ_{se} = variance of the field in the Z_{se} direction

7.0 ORBIT CHARACTERISTICS

The orbit of the IMP–I is highly eccentric with an apogee of approximately 31.7 R_e (Earth radii).

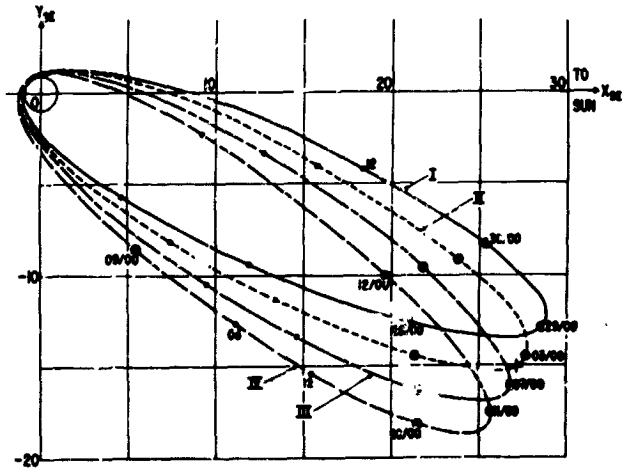


FIGURE 11.—Projection on solar ecliptic XY plane of orbits 1 through 4. Cross indicates location of satellite at time of December 2 sudden commencement geomagnetic storm.

The projection of the first four orbits as a function of time on the plane of the ecliptic or X-Y solar ecliptic plane is shown in Figure 11. The figures placed on the projection of the trajectory reference the date of the month and/or the time of the day measured in hours (UT) at which the satellite is located at the particular position shown. The location of the satellite at the time of the sudden commencement geomagnetic storm on December 2 is indicated by a cross on orbit No. 2. The data from this particular orbit will be presented later and discussed in detail with regard to the propagation of magnetic storms in interplanetary space.

A unique feature of highly eccentric orbits is related to the basic dynamics of satellites orbiting the central force field of the Earth. At apogee the satellite moves very slowly in longitude so that with IMP–I the spacecraft is located for 66 percent of its orbital period beyond 20 R_e. Thus the measurements of properties of interplanetary space is enhanced with respect to total time of possible observations. However, because of the high eccentricity of the IMP–I orbit it is severely affected by the gravitational field of the moon. The selection of an appropriate "launch window" for the IMP must take into account the multiplicity of spacecraft requirements for favorable solar aspect for power and thermal control as well as lifetime in orbit before lunar perturbations lead to a perigee which intersects the Earth's surface.

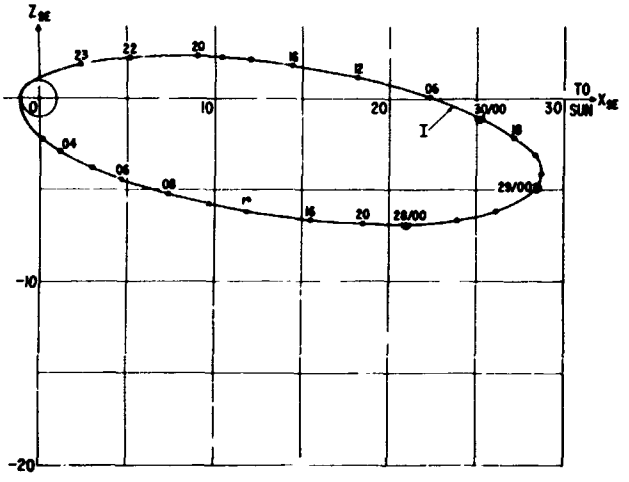


FIGURE 12.—Projection on solar ecliptic XZ plane of orbit No. 1

IMP–I has a lifetime on this basis of approximately two years.

The highly elliptical orbit does not place the satellite far above nor below the ecliptic plane as seen in Figure 12. Here a projection of orbit No. 1 is made on a plane perpendicular to the ecliptic plane passing through the Earth and the Sun, which is the X-Z plane in solar ecliptic coordinates. As time has progressed since launch the projection becomes foreshortened on this particular plane as the Earth moves about the Sun until the apogee eventually occurs behind the Earth, 160 days after launch. Figure 13. illustrates the predicted variation of the apogee-Earth-Sun angle as a function of lifetime of the satellite.

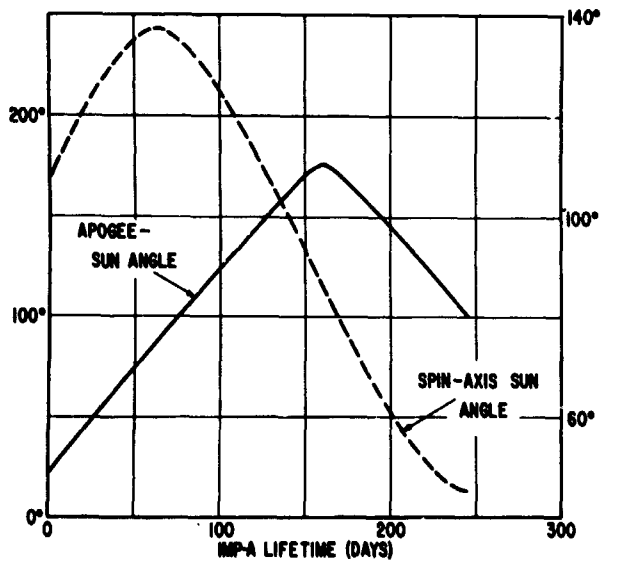


FIGURE 13.—Predicted variation of apogee Sun-angle and solar aspect angle for IMP–I.

The fact that the apogee-Earth-Sun angle never is 180° or 0° is due to the inclination of the line of apsides to the ecliptic plane. The magnitude of the inclination is determined from this figure to be 5°. The spacecraft spin axis-Sun angle shown is important for certain of the directional sensors on board the spacecraft such as the two plasma probes, whose response is principally associated with solar originating and radially directed plasma flux. The solar aspect angle is also important in determining the effects of solar radiation pressure on the spin period of the satellite since the four solar paddles were all pitched in the same sense and at the same angle and a spin up or spin down of the spacecraft occurs dependent upon whether the Sun is below or above the equatorial plane of the satellite.

8.0 MAGNETIC FIELD RESULTS

A selection of one particular orbit of the available and already analyzed 19 orbits of the IMP satellite has been made to illustrate the characteristics of the data obtained and the format in which the data will be presented. The initial studies reported herein are preliminary only in the sense that detailed correlations of magnetic fields with other sensors on board and certain other solar and terrestrial phenomena have yet to be completed. The data themselves are final as no other changes, corrections or calibrations are necessary since all have been included in this presentation.

However, as the data are further studied certain characteristic features may become more evident than they are at the present time. The study of long term statistics of the interplanetary fields and the quantitative descriptions of the shock wave and magnetopause boundaries statistically are subjects for future studies.

The magnetic field data obtained on an outbound portion of orbit No. 11 on January 5, 1964 is presented in Figure 14. As previously indicated the data is presented in a magnitude and two angle representation. The RMSE or variance of the solar ecliptic components is indicated in the lower three curves. The numerical values at the bottom refer to UT (hours) on the date

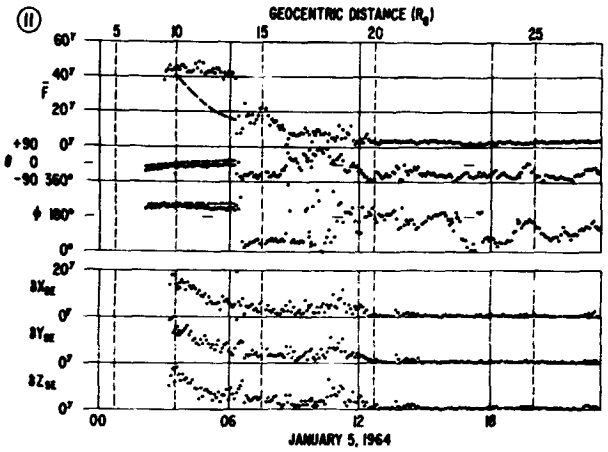


FIGURE 14.—Magnetic field data from orbit No. 11 January 5, 1964, illustrating outbound traversal of magnetopause boundary at 13.6R_e and shock wave boundary at 19.7R_e.

indicated while the values at the top indicate the geocentric distance of the satellite in Earth radii. By inspection of the data and reference to previous results on Explorer X (Heppner, Ness, Skillman and Scearce, (1963)), and Explorer XII, (Cahill and Amazeen (1963); Freeman, Van Allen and Cahill (1963)) it is clear that the magnetopause boundary has been traversed by the IMP–I satellite at a distance of 13.6 R_e at 0630 on January 5, 1964.

As the satellite moved away from the Earth the measured magnetic field began to increase above the predicted value at about 10 R_e, and proceeded to increase to more than twice that expected on the basis of the Finch and Leaton (1957) coefficients for spherical harmonic expansion of the Earth's terrestrial fields. In addition there were slight but regular angular deviations of the magnetic field increasing in importance until at 13.6 R_e there was an abrupt change in the character of the magnetic field in both magnitude and direction. The direction of the magnetic field appears to reverse by 180° while the magnitude simultaneously and just as suddenly decreases to less than 50 percent of its original value. The magnetic field then begins to fluctuate appreciably in its average magnitude and direction as the satellite continues moving further away. Although the direction of the field remains roughly stable pointing generally back toward the Sun and below the ecliptic plane, the magnetic field continues to display large deviations in the three solar ecliptic components which indicate a turbulent characteristic until a distance of 19.7 R_e is achieved. At this point in space the magnetic field becomes stable both in magnitude and direction as evidenced in the F, θ, ϕ presentation but shown much more dramatically in the presentation of the variance results. This sudden decrease in the variance from several gammas to less than 0.4 gammas is a consistently remarkable feature of the IMP–I data. This value is equivalent to the noise level of the instrument and the analysis procedure. For the next twelve hours the magnetic field is approximately 4 to 5 gammas in magnitude and pointed about 30 to 40 degrees below the ecliptic plane while directed either away or generally towards the Sun. The change in character and quantitative values describing the field at 19.7 R_e is identified as the collisionless shock wave suggested by Zhigulev (1959), Axford (1962), Gold (1962), Kellogg (1962) and Rossi (1963), associated with the standing bow wave due to the super Alfvénic streaming velocity of the solar wind impacting the geomagnetic field. Theoretical aspects of these data will be discussed later in this paper.

Continuing through orbit No. 11, Figure 16. illustrates data taken on January 6, 1964 immediately following that shown in Figure 15. In this case it is seen that the interplanetary field remains very steady at a level of 4 to 7 gammas and in general is below the plane of the ecliptic by 10° to 20° although the sense of the field is to point either away from the Sun at approximately the streaming angle associated with steady state models of coronal expansion or towards the Sun

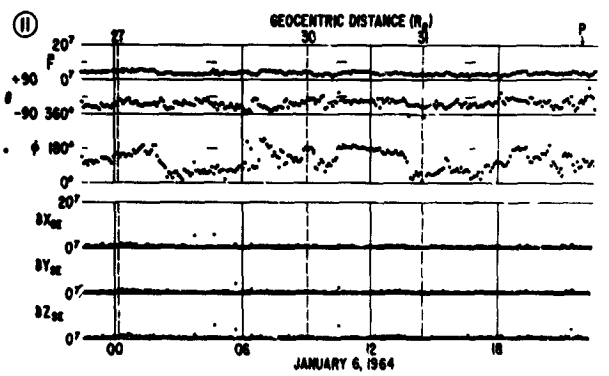


FIGURE 15.—Magnetic field data from orbit No. 11 January 6, 1964, illustrating interplanetary magnetic field.

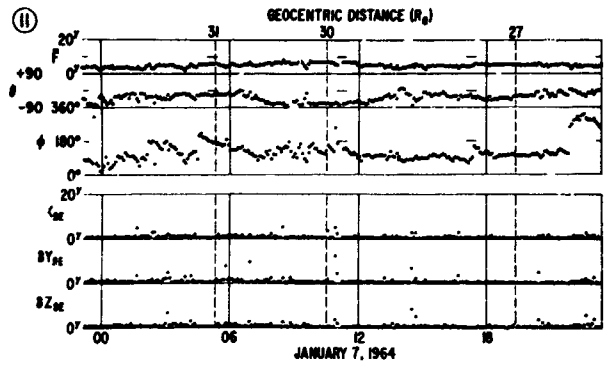


FIGURE 16.—Magnetic field data from orbit No. 11 January 7, 1964, illustrating interplanetary magnetic field and null surface separating filaments.

at the streaming angle. A significant feature in Figure 16. is the occurrence of a reversal of field direction at 2300 UT on January 7. At the same time as the field changed from being along the streaming angle to 180° opposite the streaming angle, the magnitude abruptly decreased to 10 minutes. Detailed inspection of the data at this time shows that the field went to zero as the field changed directions indicative of a *neutral surface* in interplanetary space.

The variance of the field remains quite low and clearly indicates a relatively stable magnetic field configuration somewhat in contrast to those results previously reported on Mariner 2 by Coleman et al., (1962). Figures 17. continues the presentation of data from orbit No. 11 on January 7, 1964 and supports the general statements of the stability of magnitude of the interplanetary field

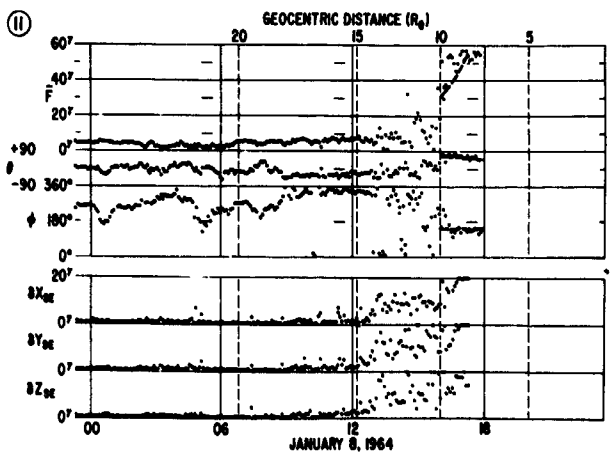


FIGURE 17.—Magnetic field data from orbit No. 11 January 8, 1964, illustrating inbound traversal of magnetopause boundary at 9.7R_e and shock wave boundary at 16.0R_e.

between 4 and 7 gammas while the directional changes in general are slow, on time scales of hours to several hours, and average directions correspond to pointing towards and away from the Sun at the streaming angle.

The inbound pass on orbit No. 11 is shown in Figure 17. presenting the magnetic field data obtained on January 8, 1964. The presentation of data is such that time always increases from left to right and the satellite is identified as being either on the inbound or outbound portion of its orbit by the sense in which the geocentric distance increases. Figure 17. illustrates two more traversals of the boundaries characteristically found throughout the lifetime of the IMP satellite. The shock wave identified as before on the basis of the increased turbulence in the magnetic field shows in this sample as an abrupt and very distinct increase in magnetic field strength at 16.0 R_e. The magnetopause boundary at which the field jumps rapidly to very large values more than twice those theoretically predicted, at the same time as its direction abruptly changes, is found to occur at 9.7 R_e. These data very briefly summarize the salient features of the 19 orbits which have been studied to date on the IMP–I satellite.

The general characteristics of all the orbital passes have been as shown in orbit No. 11, namely: traversals of two characteristic boundaries twice in each orbit. As the satellite orbit has progressed around from the Sun-Earth line the distance at which the magnetopause boundary and the shock wave boundary have been found to occur has changed. These changes have been correlated with a number of phenomena and will be discussed shortly. However, the most unique aspect of the boundary crossings in general has been their consistent and readily identifiable occurrence on all of the orbits.

The most important feature of the interplanetary magnetic field results has been in general the finding that the interplanetary magnetic field magnitude is quite steady and varies only over fractions of a gamma on the relatively long time scale of hours. This stability of the field magnitude but the variability of the direction may lend support to the studies by Dessler and Walters (1964) and Walters (1964) as to the effects of the streaming magnetized plasma on terrestrial magnetic activity and magnetospheric geometry.

The correlation between directional changes of the interplanetary magnetic field and the associated terrestrial activity as indicated by various indices, including K_p and A_p, is currently being investigated in detail. Previous studies experimentally by Snyder, Neugebauer and Rao (1963), Greenstadt (1964) and Maer and Dessler (1964) have not been completely consistent although they have generally indicated a definite correlation between K_p or (A_p) and solar wind velocity. These results are in some contradiction to the theoretical suggestions by Dessler and Fejer (1963) which deduce a correlation between K_p and changes in the solar wind velocity. It is important to note that the time variations of the interplanetary magnetic field observed reflect mainly the structural variations within the interplanetary medium. This is because the proper frame of reference moves with the bulk velocity of the solar plasma.

Although the statistics are not as yet complete it appears that in general there is a preference for the interplanetary magnetic field over this portion of time in the solar cycle to lie generally below the plane of the ecliptic by approximately 10° to 20°. At the same time the projected vector on the ecliptic plane appears to have three preferred orientations to which it returns and near which it is observed most of the time. In order of duration these three directions are:

1. Anti-solar,
2. At the streaming angle associated with the theoretical models of the solar wind,
3. At the anti-streaming angle, meaning exactly 180° opposite to the streaming angle theoretically predicted.

These results of the IMP-I magnetic field experiment suggest very strongly the existence of filamentary structure in the interplanetary medium associated with unique and separate sources of solar associated magnetic fields. These fields are stretched out away from the Sun by the highly ionized streaming solar plasma as modeled and discussed by Parker (1958). In between these streaming filamentary elements evidence has been found for magnetic field magnitudes decreasing abruptly to zero so as to essentially be in consonance with the idea of a null surface separating regions of interplanetary magnetic fields which are oppositely directed. Throughout the lifetime of the satellite thus far approximately 12 such readily identifiable filamentary structures have been observed.

A comparison of the observed streaming angle with those predicted by steady models assuming an axially symmetrical expanding solar plasma flux is shown in Figure 18. The observed interplanetary magnetic field is close to $\phi = 130$ to 150°

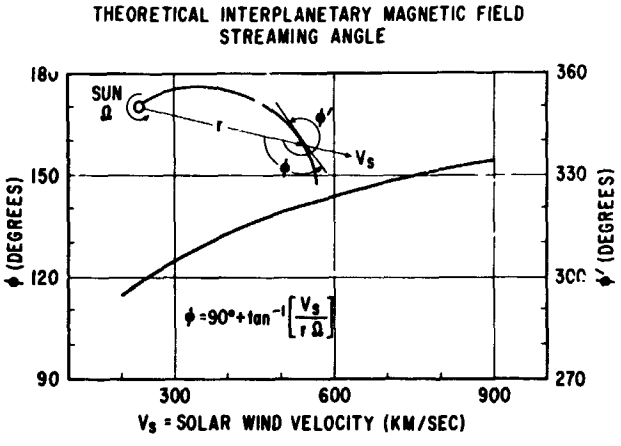


FIGURE 18.—Theoretical direction of interplanetary magnetic field ϕ as a function of solar wind velocity V_s.

or 310 to 330° for the streaming or anti-streaming angles. These directions are consistent with velocities of the solar wind of 300–700 km/sec, values found by the Mariner II probe, (Snyder and Neugebauer, (1963)). The detailed characteristics of the angular deviations of the interplanetary magnetic field are currently being studied and in the future detailed correlations with the plasma probes on board the satellite as well as additional solar and terrestrial indices will be made.

At the present time it appears that the results of the IMP magnetic field experiment have indicated an impressive consistency with the theoretical models proposed mainly by Parker, but also by Gold (1959, 1963). Those of Parker are substantiated on the basis of the interplanetary field direction dependence upon plasma velocity which compares very favorably with observation. The general directions of the field lines empasize the principal fact that a rotating Sun continuously emitting plasma will lead to a gross superposed spiral field structure under most circumstances. The filamentary structures appear to indicate specific sources as proposed by Gold (1963) as

must be considered in the modeling of realistic solar coronal phenomena. It is also possible that the data are consistent with the magnetic "bottle" ideas of Gold (1962) in which cutting of lines of force near the Sun generates field topologies which do not include the Sun.

Additional data from the magnetic field experiment is shown in Figure 19. from orbit No. 15 on

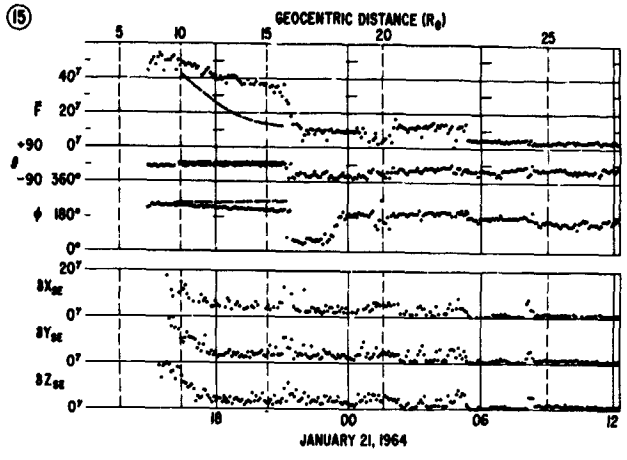


FIGURE 19.—Magnetic field data from orbit No. 15 January 21, 1964, illustrating outbound traversal of magnetopause boundary at 15.7 R_e, shock wave bound. at 22.7 R_e and precursor at 24.5 R_e.

January 21 1964. Again the presentation is in the same form as previous results and clearly illustrates the crossing of the magnetopause at 15.7 R_e while the shock wave traversal is not observed to occur until 22.7 R_e

A particularly significant feature of the outbound traversal of the shock wave on this orbit is the appearance at 24.5 R_e of an increased variance for 10–30 minutes. Not in evidence on every orbit although clearly indicated here, this feature of the IMP–I magnetic field data is very important. This limited increase in variance is also spatially associated with the position of the shock wave and when observed is usually found to be several Earth radii beyond the shock wave. This is identified as a precursor to the shock and is suggested to be indicative of characteristic plasma instabilities or particle resonances whereby the presence of the shock wave can be communicated "upstream" to the plasma by a mechanism that depends upon the individual particle dynamics of the plasma rather than the continuum dynamics of the collective particles. Continuing

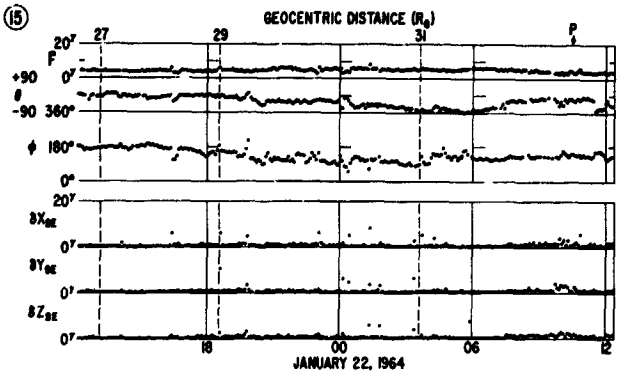


FIGURE 20.—Magnetic field data from orbit No. 15 January 22, 1964, illustrating interplanetary magnetic field.

in orbit No. 15, the interplanetary field is shown in Figure 20. on January 22, 1964. The characteristics of the data are quite similar to those shown for orbit No. 11 although the positions of the boundary are considerably beyond those observed in orbit No. 11 and correspondingly the positions of the boundary detected during orbit No. 11 occur at larger distances than originally detected during the early orbits of the spacecraft. An important feature of the turbulent field structure now is that fluctuations occur about a specific orientation so that there is an appreciable sense of directionality to the magnetic field, even though it is not stable in magnitude.

The data from orbit No. 6 inbound is being investigated in detail since it followed an interval of time during which the interplanetary field was disturbed because of the interaction of the Moon's magnetosphere with the solar wind (see Section 13.). It is clear from Figure 21. that the characteristic boundaries observed on all 37 other traversals are not evidenced here in the same manner. Principally the variance remains at the instrument noise level until a distance of 10.8 R_e is reached. Secondly, the abrupt directional changes in magnetic field are not coincident with abrupt changes in the magnetic field strength. The data strongly suggest the absence at this time of the collisionless shock wave probably due to the increased field strengths and possibly low plasma densities leading to Alfvén velocities sufficiently high that a shock situation does not exist. These data are also consistent with the study by Alfvén, Danielsson, Falthammer and Lindberg (1963) on the direct entry of solar

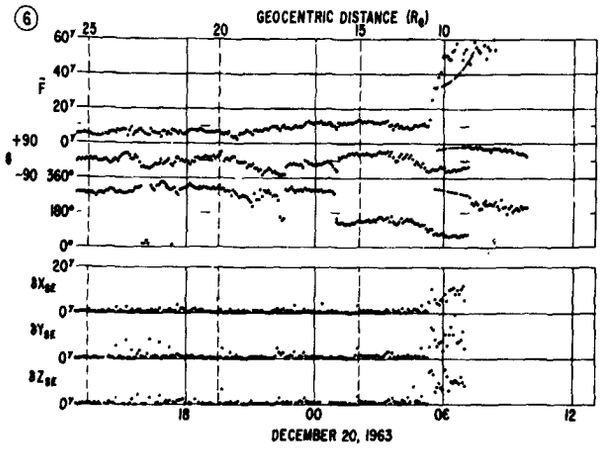


FIGURE 21.—Magnetic field data from orbit No. 6 December 20, 1963, illustrating inbound traversal of magnetopause boundary at 10.8 R_e and questionable shock wave boundary at 16.3 R_e.

plasma into the geomagnetic field when fields external to the cavity are parallel to internal fields.

9.0 SUMMARY CHARACTERISTICS OF MAGNETIC FIELD PHENOMENA

It is appropriate to summarize the many details of the interplanetary magnetic field data and the characteristics of the magnetic field at the shock wave and the magnetopause boundaries as well as the characteristics of the data for the transition region. Certain features are clear and have been already stated. The omission of certain characteristics which may become obvious to the reader should not be taken to indicate their dismissal by the authors as insignificant aspects of the data. It is quite impossible to discuss in this first publication all of the results which have been obtained from this particular experiment.

The orbit of the IMP–I satellite and the data obtained is such that it suggests logically separating the region of space surrounding the Earth into three physical regions for discussions:

1. The *Magnetosphere*, that region of space including the Earth and its radiation belts and enclosed by the magnetopause boundary from region 2. This geomagnetic cavity corresponds to a persistent containment of the geomagnetic field along the concepts initially developed for magnetic storm theory by Chapman and Ferraro (1931, 1932, 1933); Ferraro (1952, 1960a,b) and reviewed by Chapman (1960, 1963).
2. The *transition region* which extends from the magnetosphere to the interplanetary medium or more appropriately the interaction region between the streaming solar plasma and the geomagnetic field which is separated by the collisionless magnetohydrodynamic shock wave from region 3.
3. The *interplanetary medium* in which the plasma flux and magnetic fields are undisturbed and undistorted by the presence of the Earth in interplanetary space. This region of space and its physical characteristics represent the principal region of interest for the IMP–I measurements.

The physical characteristics of the magnetic fields in these three regions and the description and discussion of the boundary properties is given in the following paragraphs.

The dynamics of charged particle motion in the magnetosphere region is dominated by the magnetic field of the Earth. Measurements on the IMP–I show that deviations of direction and magnitude of the observed magnetic field when compared with theory do not become significant until an elevation of 7–10 R_e is reached. The deviations beyond this distance are consistent with the characteristics of a contained geomagnetic field with the strength increasing to at least a factor of two greater than normally expected without the formation of a magnetopause.

The directional distortion is consistent with the fields being oriented approximately tangential to the magnetopause surface. Although it requires further investigation in the future it is possible to state that the magnetic fields near the boundary on the Sunlit hemisphere interior to the magnetopause are not stable as argued by Dessler (1961, 1962) and that appreciable energy must be transmitted into the magnetosphere in the form of magnetohydrodynamic waves. The fluctuations observed are generally smaller than could have been detected by Explorer XII, because of its limited sensitivity of ± 12 gammas (Cahill and Amazeen, 1963) and show characteristic values of 10 gammas RMS or less and corresponding values of $\delta\vec{F}/\vec{F}$ of 0.05 to 0.3.

The magnetopause, reviewed by Hines (1963) is identified by the abrupt discontinuity in the character and quantitative description of the magnetic field as the satellite passes from Region

1 to Region 2. Near the subsolar point this is found to occur at approximately 10 R_e (see Section 10). The magnetic field rapidly de creases from values of 30–60 gammas to 5–20 gammas and corresponding variances from small $\delta\bar{F}/\bar{F}$ to large $\delta\bar{F}/\bar{F}$, being approximately 0.5 to 2.0. The thickness of the magnetopause is difficult to determine unambiguously but upper limits can be placed assuming that the boundary is stationary on a time scale such that the total time for motion of the satellite across the magnetopause is small. At distances of $10R_e$ the satellite radial velocity from the Earth is approximately 2.7 km/sec.

In general the analysis procedures and definitions currently employed to determine the location of the magnetopause are such that ± 2.5 to ± 15 minutes is a good measure of the uncertainty in precisely specifying where the boundary is traversed. Folding the spacecraft velocity and this time uncertainty together yields thicknesses of 810 km to 4860 km. This is quite reasonable since the proton Larmor radius for a 1 kev proton in a magnetic field of 10 gammas is 450 km and should be a reasonable scale length to consider in estimating the magnetopause thickness since all plasma flux is presupposed to be reflected and excluded by the geomagnetic field from entering the magnetosphere (Grad, 1961).

The direction of the fields at the magnetopause boundary strongly suggest that both internally and externally to the magnetopause, the fields are tangential to the magnetopause surface at least locally, i.e., not far from the boundary. Furthermore the directional changes suggest that although the magnetic fields are oppositely directed on either side of the magnetopause for a number of the crossings studied thus far, there are cases in which the fields appear to be crossed perpendicular with respect to each other on opposite sides of the boundary.

The *transition region* is one of great instability of the magnetic field suggesting that turbulent phenomena occur in this region and that statistical descriptions of the field will be most appropriate in their study. The vector field fluctuates rapidly in magnitude and direction with $\delta\bar{F}/\bar{F}$ ranging from 0.5 to 2.0. At times however the magnetic field direction and magnitude stabilize for time scales of 20 to 60 minutes with values corresponding to those of Regions 1 or 3. Although suggested by the interpretation of Explorer XII results, this is not considered to be representative of the general characteristics of Region 2. It is not interpreted to be indicative of rapid spatial movements of the mangetopause as it expands and contracts. Actually in a region of turbulence there is a finite probability that the field orientations will equal those of other Regions. The region of high turbulence encompassing and enclosing the magnetosphere is well established by the IMP data and quantitative statistical descriptions of the magnetic field and its topology in the vicinity of the magnetopause will be presented in future publications.

The collisionless shock wave boundary separates the interplanetary region 3 from the turbulent region 2, and is observed experimentally at 13.4 R_e at the subsolar point. The identification of its position as the satellite traversed it in space is based upon the striking physical differences between Regions 2 and 3. The abruptness of the changes in character of the fields at the outer boundary of the turbulent region is not always as sharply defined as at the inner boundary. In general, the interplanetary fields are very stable and characteristically show very small variances, less than 0.4 gammas for extended times yielding $\delta\bar{F}/\bar{F} < 0.1$. Interior to the shock wave boundary the field is observed to be randomized in direction and the total energy density in the magnetic field increases: both the DC value, as measured by $B^2/8\pi$ and the fluctuating field as measured by the variances δX_{se}, δY_{se}, and δZ_{se}. This means that the total energy of the magnetic field increases by factors of 5–10 and can play an important role in the flow of plasma around the cavity. It should be noted that although the identification of the shock wave boundary is generally unique based upon the above physical criteria, certain samplings of the shock region show a precursor in the Region 3 which stands off from the shock wave by several R_e. This *precursor* does not show the characteristic of larger fields but does display larger variances, up to 5 gammas, in each component for time scales of ≤ 40 minutes. At these distances, the satellite is moving radially outward at approximately 1.8 km/sec. Thus the spatial width of the precursor is ≤ 4000 km. Since the satellite is moving more

slowly when it traverses the shock wave, the longer times required do not indicate a more diffuse boundary than is the case of the magnetopause.

The existence of the collisionless magnetohydrodynamic shock wave was anticipated on the basis of theoretical analyses relying principally on the analogy with high speed aerodynamic flow. Various aspects of this general approach have been reviewed by Levy, Petschek and Siscoe (1963). Their specific interest and results were mainly to deduce certain aspects of the overall plasma flow around the geomagnetic cavity. Analytical studies on the structure of collisionless shocks in magnethydrodynamics have proceeded in some work by modification of the appropriate form of the Rankine-Hugoniot conditions derived by De Hoffman and Teller (1950). Theoretical studies by Adlam and Allen (1958); Auer, Hurwitz and Kilb (1961, 1962); Fishman, Kantrowitz and Petschek (1960); Morawetz (1961, 1962) and Morton (1962) have attempted to treat in varying degree, the detailed description of the shock characteristics and sufficient conditions for its existence. The present situation appears to be one in which the characteristics of the plasma flow are grossly described, at least in the location of the shock and magnetopause, on a basis of the analogy with fluid dynamics. However, the physical structure of the fields and plasmas in the observed three dimensional case of the Earth are not immediately derived by such simplifying considerations. In particular the discovery of the shock precursor is obviously indicative of the necessity to treat the flow of plasma more specifically as a two stream flow of non-cold electrons and protons (Bernstein, Fredricks and Scarf (1964) and Morton (1964)). A more detailed study of the observed shock wave with regard to the theoretical descriptions is presently in progress. However, it is reasonably certain that the IMP-I satellite and its experiment complement has provided the first physical measurements of the physical structure of a collisionless magnetohydrodynamic shock which heretofore has never been observed because of the "scaling problems" in laboratory research.

The interplanetary region 3 is defined most readily by the stability of the magnetic field as measured by the variances δX_{se}, δY_{se}, and δZ_{se}. These are observed to be less than 0.4 gammas with corresponding magnitudes of 4 to 7 gammas although excursions to values as low as 1 and to as high as 10 gammas for extended time intervals have been observed. The extreme stability of the magnitude of the field is in contrast to the more variable directional characteristics of the field. In general the fields lie on average slightly below the plane of the ecliptic ($\theta = 10°$ or $20°$) and at the streaming angle of 130°–150° or anti-streaming angle 310°–330°. The gross structure of the interplanetary field is in remarkably good agreement with the general concepts of spiral field configurations originally proposed by Parker (1958a). Associated with the two preferred orientations of the field, along or opposing the traces of the plasma flux, are regions of abrupt change to low values on time scales of less than 5 minutes. These are identified as neutral surfaces of zero field strength separating magnetic field filaments oppositely directed.

10.0 LOCATION AND GEOMETRY OF MAGNETOPAUSE AND SHOCK WAVE

A summary of the positions of shock wave and magnetopause traversals as a function of time associated with preliminary K indices from the Fredericksburg, Virginia magnetic observatory are shown in Figures 22–25. In addition, the geomagnetic latitude of the subsolar point, χ_{ss},

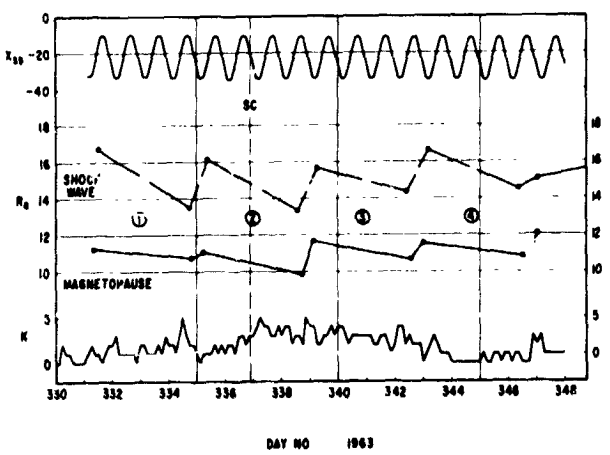


FIGURE 22.—Time characteristics of geomagnetic latitude of subsolar point χ_{ss}, magnetopause and shock wave boundary locations and K indices for days 330 to 348, 1963.

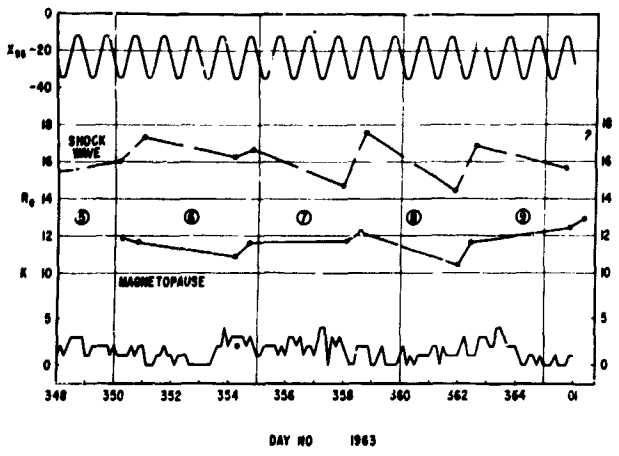


FIGURE 23.—Time characteristics of geomagnetic latitude of subsolar point x_{ss}, magnetopause and shock wave boundary locations and K indices for days 348, 1963, to day 01, 1964.

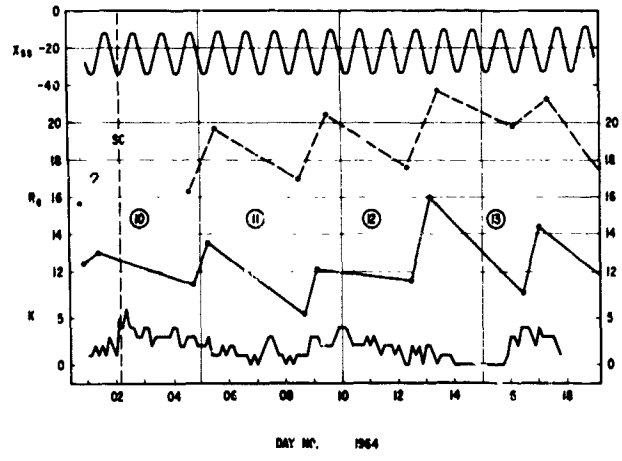


FIGURE 24.—Time characteristics of geomagnetic latitude of subsolar point x_{ss}, magnetopause and shock wave boundary locations and K indices for days 01 to 19, 1964.

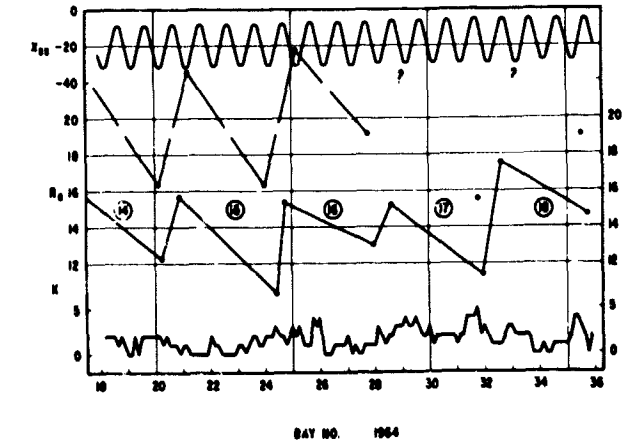


FIGURE 25.—Time characteristics of geomagnetic latitude of subsolar point x_{ss}, magnetopause and shock wave boundary locations and K indices for days 18 to 36, 1964.

has been included since it is an important factor in the interpretation of the position of the boundary crossings and the manner in which they relate to theoretical models of the magnetopause and the shock wave surfaces. In these Figures the orbit number is enclosed in a large circle while the boundary crossings are indicated by small circles. These are connected from inbound to outbound passes by straight lines if successive boundary crossings are uniquely identifiable. In some cases the position of the boundary is not sufficiently clear, as for example on the outbound portion of orbit 5, to uniquely specify its position and a question mark is displayed. This is related to the difficulty associated with the definition of the characteristics of the boundaries as deduced from the magnetic field data alone.

Certain aspects of the position of the boundary crossings are immediately evident upon inspection of these data. Principally seen is the periodic difference in position of the boundary crossings associated with outbound and inbound passes of the satellites. Characteristically the outbound pass shows the position of both boundaries to be greater than that associated with the next successive inbound pass. On a more detailed investigation it is found that following the sudden commencement storm in December 1963, indicated in Figure 22., the magnetopause and shock wave boundaries were located at smaller distances than normally to be expected on the basis of previous passes. Also over the 70 day interval covered by this data presentation, it is clear that the position of the magnetopause boundary and shock wave boundary increase as the satellite moves away from the Earth-Sun line. The two boundaries do not increase by the same amount. The shock wave increases more rapidly implying a wider transition region away from the subsolar point. It is noted that the satellite motion is roughly normal to these surfaces for early orbits, but becomes more oblique for later orbits.

Superimposed upon this spatial variation of the boundaries is the fact that the position of the boundary must be sensitive during successive crossings to the varying geomagnetic latitude of the subsolar point. That this is so is seen by inspecting the condition for balance of magnetic and particle pressure at the boundary. Originally

developed by Chapman and Ferraro (loc. cit.) the recent work by Beard (1960), Davis and Beard (1962), Spreiter and Briggs (1962) has led to agreement of the proper form. The location of the boundary at the stagnation point is given by

$$\frac{B_t^2}{8\pi} = 2mn\, V_s^2 \qquad (10.1)$$

where m is proton mass, n is proton density, V_s is solar wind velocity, and B_t is the net magnetic field tangential to the magnetopause boundary. This is commonly taken to be given by

$$B_t = 2fB_e \qquad (10.2)$$

where f is a factor ≤ 1 (usually taken as unity) and B_e is the theoretical and undisturbed geomagnetic field. However,

$$B_e = B_0 \frac{R_e^3}{R_c^3}\sqrt{1+3\sin^2\chi_{ss}} \qquad (10.3)$$

where χ_{ss} = geomagnetic latitude at the stagnation point where the flow velocity is normal to the surface and B_0 is the terrestrial equatorial field strength.

Combining 10.1, 2 and 3 yields

$$R_c = R_e\left[\frac{B_0^2\,(1+3\sin^2\chi_{ss})}{4\pi\, mn\, V_s^2}\right]^{1/6} \qquad (10.4)$$

where χ_{ss} is the geomagnetic latitude of the subsolar point. Clearly the distance at which the magnetopause is predicted depends upon the angle of the Earth's dipole to the solar plasma.

This concept has been included in the final data summary and has been employed to rectify the position of the boundary crossings to a geometry which allows comparison with the simplified theoretical treatments. In these analyses the solar stream is assumed normally incident upon the axis of the magnetic dipole, implying coincidence of the geomagnetic equatorial plane and the ecliptic plane.

A summary of the location of the magnetopause and shock wave traversals rotated into the ecliptic plane as a function of position of the satellite is presented in Figure 26. The effect of motion of the line of apsides of the satellite relative to the Sun-Earth line is evident. Presented

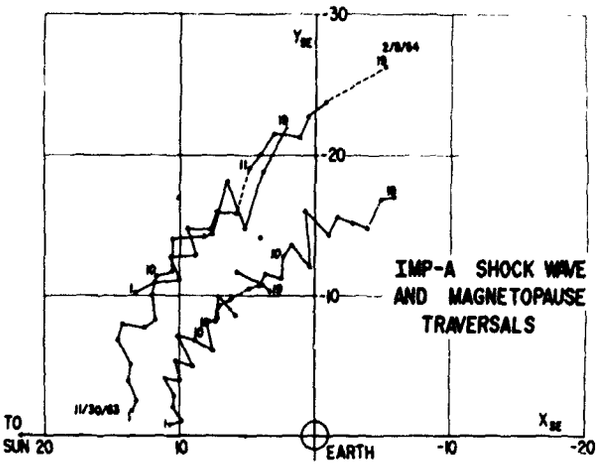


FIGURE 26.—Location of magnetopause and shock wave boundary traversals rotated onto ecliptic XY plane.

in this figure are all boundary crossings which have been uniquely identified on the basis of magnetic field information alone. These data do not include any corrections for time variations of the solar wind, the planetary indices of the geomagnetic field, nor the geomagnetic latitude of the subsolar point, χ_{ss}. These boundary crossing positions have been rotated from their actual position to the ecliptic plane in a meridian plane through the Z_{se} axis. This essentially assumes a constant radius of curvature in each meridian plane. An alternative procedure, not employed, is to rotate the positions into the ecliptic plane about the X_{se} axis which presumes cylindrical symmetry about the X_{se} axis.

Clearly evident in this figure is the fact that the magnetopause is not spherical, but actually on the nightside hemisphere of the Earth, the boundary flares out in consonance with the ideas originally resulting from measurements on the Explorer X satellite (reviewed by Bridge (1963) and Ness (1963)). In addition, it is seen that the shape of the shock wave boundary when projected on the ecliptic plane is approximately hyperbolic. The angle which the hyperbola asymptotically approaches is referred to as the Mach angle and if measured could indicate the velocity of solar wind propagation relative to the interplanetary magnetic field strength and plasma density by defining the Mach number of the super Alfvénic flow of the solar wind.

In order to make a first order correction for the variation of the geomagnetic latitude of the

sub-solar point and thereby take into account the variable magnetic field strength which the solar wind meets at the stagnation point, an expansion factor K has been computed. This is dependent upon the geomagnetic latitude of the sub-solar point and the strength of the magnetic field of an assumed dipole at the center of the Earth. From Equation 10.4, K is defined as:

$$K=[1+3\sin^2\chi_{ss}]^{1/6} \qquad (10.5)$$

The maximum inclination of the Earth's magnetic dipole axis to the solar wind direction is 35.1° being the sum of 11.7°, the tilt of the magnetic axis to the rotation axis, and 23.4°, the obliquity of the rotation axis to the ecliptic. The maximum

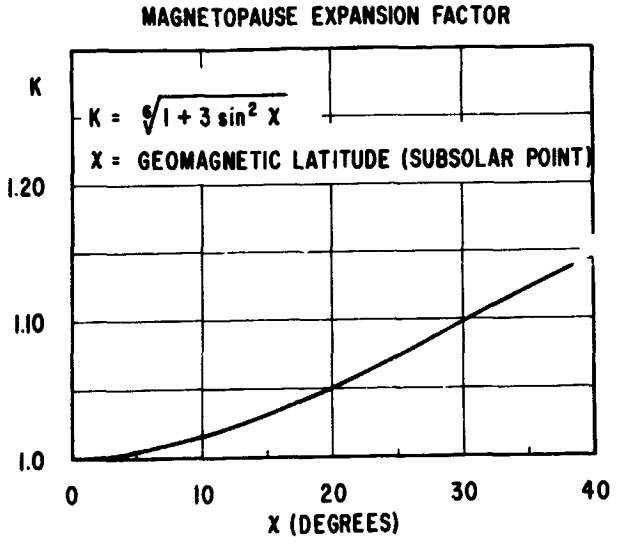


FIGURE 27.—Magnetopause expansion factor K to compensate for variable tilt of Earth's magnetic field axis to "angle-of attack" of solar wind.

expansion is seen in Figure 27. to be 12 percent assuming a constant pressure of the solar wind. The time of launch of IMP–I is such that the maximum inclination was achieved during the first month of operation of the satellite, being winter solstice and since that time the Earth's rotational axis has approached perpendicularity to the Earth-Sun line. This is seen in Figures 22 thru 25 in which the diurnal variation of χ_{ss} is seen to become more negative, reaching a maximum in December 1963 and then becoming less negative as time progresses. The average values of χ_{ss} over one day indicates the angle between the normal to the Earth-Sun line (X_{se} axis) and the direction of the Earth's rotation axis.

In comparing theoretical and observed boundary geometries, it has been assumed that the only model currently available is that provided by the provided by the computations of Spreiter and Jones (1963) using a blunt body profile representing the Earth's magnetosphere to determine the shape of the shock wave surface based upon a strict analogy with high speed aerodynamic flow.

The blunt body profile used is that which has been obtained by rotating about the X_{se} axis the trace of the magnetopause in the geomagnetic equatorial plane. The shape has been determined by a number of workers Beard (1960); Hurley (1961); Midgley and Davis (1962); Slutz (1963); Spreiter and Briggs (1962); Mead and Beard (1964), and all results are in generally good agreement as long as the region near the null points in the polar regions is not considered. None of these analyses have included an interplanetary magnetic field in determining the shape nor physical characteristics of the magnetopause boundary, particularly in view of the importance of the collisionless shock wave. This omission has been partially rectified in the work of Lees (1964) in inclusion of the interplanetary magnetic field. The ramification of the shock on the field configuration external to the magnetopause has been considered by Beard (1964). The results for the particular model which Spreiter and Jones (1963) utilized have been slightly adjusted to match the actual observation of stand-off distances of the geomagnetic cavity boundary and the shock wave boundary at the subsolar point. In addition all boundary crossings have been rectified by dividing the observed distance by the appropriate value of the expansion factor K so as to reduce all distances to a geometry in which the solar wind is normally incident upon the axis of the Earth's magnetic dipole. The results of these rectifications and the adjustment of the theoretical models is compared in Figure 28. with the theoretical results indicated as the solid curves through the experimental data. The theoretical magnetopause geocentric distance at the subsolar point is taken to be 10.25 R_e while that of the shock wave is taken as 13.4 R_e. It has been necessary to rotate the theoretical model by 5° about the Z_{se} axis to obtain the demonstrated

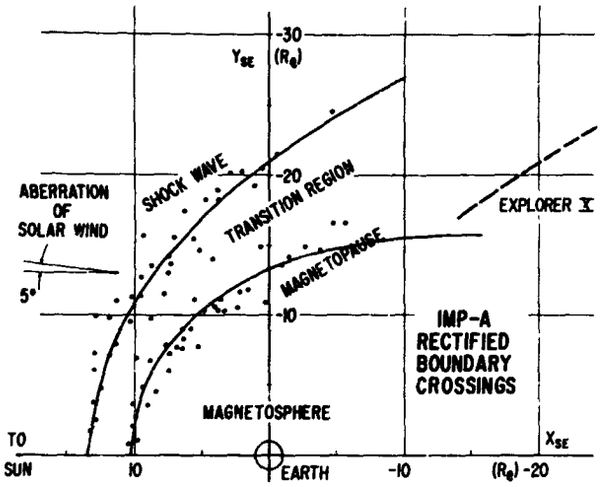


FIGURE 28.—Summary of location of rectified shock wave and magnetopause boundary crossings and comparison with theoretical predictions using analogy with supersonic fluid flow.

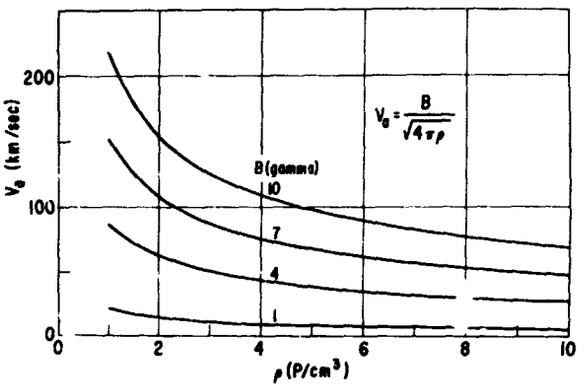


FIGURE 29.—Alfvén magnetohydrodynamic phase velocity of wave propagation as function of magnetic field strength and plasma density.

agreement. This is comparable with the aberration effect due to the Earth's orbital motion of 30 Km/sec when compared to the solar wind velocities. For values of V_s between 300 and 400 Km/sec the aberration effect changes from 5.7° to 4.6° and the value of 5° is taken to be a representative average.

An important parameter describing the flow of the solar plasma and its interaction with any planetary object is the velocity at which disturbances can be propagated. In the highly ionized magnetized plasma being studied, the appropriate modes of propagation are the magneto-acoustic modes. Considering the important mode to be the simple Alfvén transverse magnetohydrodynamic mode, the relevant velocity is given by

$$V_A = \frac{B}{\sqrt{4\pi\rho}} \tag{10.6}$$

This is shown in Figure 29. covering the range of parameters observed in space. Since in general representative values of these parameters are

$$4 \leq B \leq 7 \text{ gammas}, \tag{10.7}$$

$$3 \leq \rho \leq 10 \text{ protons/cm}^3$$

$$\text{then } 25 \leq V_A \leq 90 \text{ Km/sec}$$

Now the directed stream velocity of the solar wind is observed to be

$$300 \leq V_s \leq 700 \text{ km/sec}$$

Thus the ratio of directed velocity to Alfvén velocity measures the super Alfvénic ratio, or Mach number corresponding to the gas dynamic analogy, is

$$3.3 \leq M_A \leq 28 \tag{10.8}$$

and a mean value of

$$\overline{M_A} = 5\text{–}10 \tag{10.9}$$

A general review of the applicability of the gas-dynamic analogy used in treatments of the flow of plasma around the geomagnetic field has been given by Levy, Petschek and Siscoe (1963).

The comparison between theory and observation of the position of the shock wave surface is good. It is clear however that the comparison is possible at the present time only near the stagnation point. Here the shape of the surface is not critically dependent upon Mach number or equivalent gas specific heat ratio used. The experimental data however does suggest an outward flaring of the magnetopause, first observed on Explorer X and is seen to be consistent with the present observations on IMP. The portion of the Explorer X trajectory over which the magnetopause boundary was observed is indicated in Figure 28., having been rotated about the X_{se} axis into the ecliptic plane ahead of the Earth. The extrapolation of the IMP observations coincides reasonably well with these previous measurements. It indicates a fundamental discrepancy between theory and observation in that all theoretical models using a real, thermalized

plasma lead to a closure of the magnetopause at distances of 30–50 R_e on the dark side of the Earth. However, the observations indicate that an outward flaring of the magnetopause surface at a half-cone angle of 10–15° occurs. The magnetic fields interior to the magnetopause for orbits numbered 11 and beyond are directed at $\phi = 205$–215° which is consistent with an outward flaring if the fields are tangential to the bounding geomagnetic cavity surface.

The relative position of the subsolar geocentric distance to the shock wave, R_s, and the magnetopause boundary, R_c, is presented as a function of the Alfvén Mach number in Figure 30. The

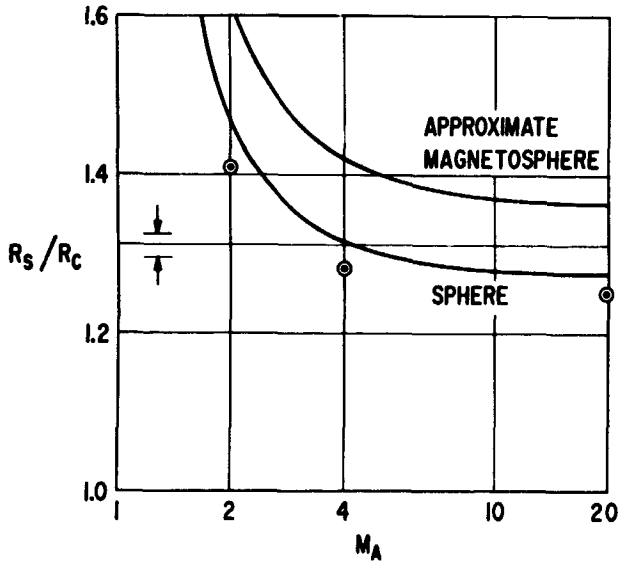


FIGURE 30.—Theoretical stand-off ratio R_s/R_c of shock wave surface to magnetopause boundary and comparison with observations.

solid lines indicate the theoretical models for a sphere and blunt body as predicted by Spreiter and Jones (1963). The circled data represent the values deduced by Kellogg (1962) in his analysis using the analogy of supersonic flow in gasdynamics and the theoretical results of Hida (1953) for a spherical object. The observed value of the ratio of R_s/R_c is shown as a horizontal solid line, experimentally determined to be 1.31 ± 1 percent. The correspondence between theory and observation is very good and although the use of the blunt body profile does not yield the proper ratio, it is increased in the proper sense. The present discrepancy between theory and observation can be readily resolved through use of $\gamma < 2$ and slightly altering the magnetopause shape.

The totality of IMP observations on the location and geometry of the magnetopause and collisionless shock wave boundary are in generally good agreement with theoretical models using the analogies between supersonic fluid flow and the super Alfvénic flow of the solar wind with the interplanetary magnetic field acting to couple the particles together as described by Axford (1962) and Kellogg (1962) rather than direct particle collisions. Obayashi (1964) has reviewed all previous satellite measurements with respect to their significance in the detection and identification of the magnetopause and shock wave boundary positions.

Since the satellite has yet to completely map the nighttime magnetospheric tail, it is not possible to definitely state that an extended flaring of the cavity far behind the Earth is the only possible interpretation of the data. It appears a reasonable assumption to make, however, and it is anticipated that during orbits 20–40 the IMP satellite will provide extremely important data on the actual geometry and physical characteristics of the magnetopause boundary and the magnetic field topology within the geomagnetic tail. Studies on the magnetopause tail magnetic field topology have been performed by Axford and Hines (1961); Harrison (1962); Hones (1963) and Johnson (1960). Quantitative estimates of the field characteristics in the magnetopshere near the stagnation point have been performed by Mead (1964). In the traversals of the magnetopause boundary the direction of the magnetic fields appear to be tangent to the apparent magnetopause surface within $\pm 10°$. At present one of the major theoretical uncertainties is the reason for the outward flaring shape of the cavity boundary on the nightside of the Earth.

11.0 SOLAR TERRESTRIAL TRANSIENTS

The general level of terrestrial magnetic activity throughout the lifetime of the IMP satellite has been extremely low as seen by the K values in Figures 22–25. This is associated with the decrease of solar activity which will meet its absolute minimum within the next eighteen months. Fortunately, however, there have been a few specific isolated magnetic disturbances

associated with the limited solar activity and these have been observed definitively on the IMP satellite in interplanetary space. In particular, the December 2, 1963 sudden commencement storm beginning at 2117 UT occurred at a very favorable time in the second orbit of the satellite. At this time, as shown in Figure 11., the satellite was approaching apogee and essentially was stationary in space except for its motion through the interplanetary medium at the heliocentric orbital velocity of the Earth.

The regular run magnetograms from a number of standard observatories in the U.S. Coast and Geodetic Survey net are presented in Figure 31.

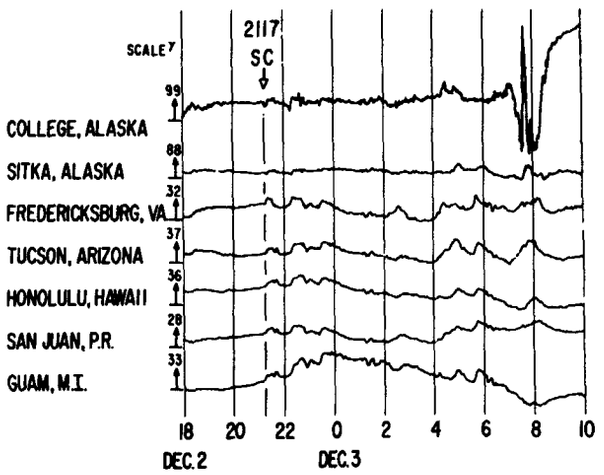


FIGURE 31.—Regular-run observatory H component magnetograms for December 2, 1963 sudden commencement storm period.

for the time interval preceding and following the sudden commencement storm. It is clear from inspection of the records that this sudden commencement showed similar characteristics at a number of stations for several hours after the initial onset. In order to accurately establish the time of occurrence of the sudden commencement at terrestrial stations, the rapid-run magnetograms from certain of these same stations are illustrated in Figure 32. in which the time scale is considerably expanded. It is seen that the onset time of 2117 is easily identifiable as there is no discrepancy in its unique characteristic.

The magnetic field data obtained on the outbound portion of the orbit No. 2 on December 1, 1963 is shown in Figure 33. At this time the boundary of the magnetopause was observed at

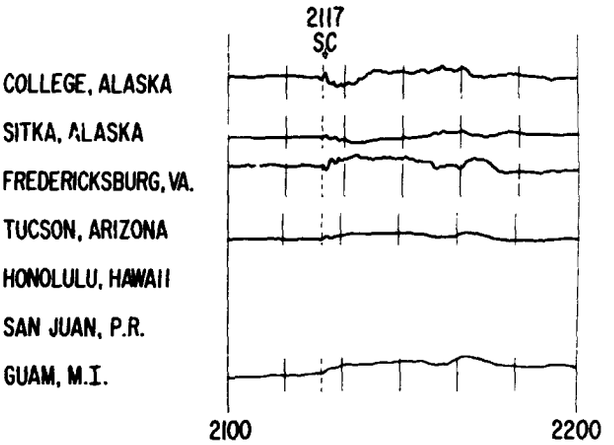


FIGURE 32.—Rapid-run observatory H component magnetograms for December 2, 1963, sudden commencement storm period.

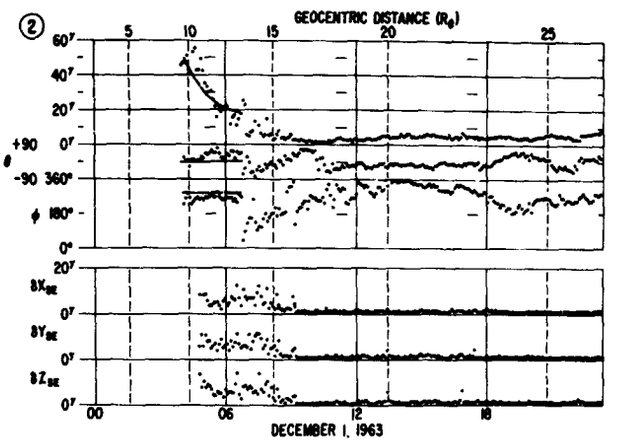


FIGURE 33.—Magnetic field data from orbit No. 2 December 1, 1963 indicating traversal of magnetopause boundary at 11.1 R_e and shock wave boundary at 16.2 R_e.

11.1 R_e while the shock wave was observed at 16.2 R_e. In general the interplanetary magnetic field conditions were quiet although the field was somewhat larger than previously observed, approximately 6 gammas, varying somewhat in direction and illustrating the filamentary structure previously referred to in the summary of the overall characteristics of the interplanetary magnetic field. Data from the portion of the second orbit during which the storm occurred is shown in Figure 34. The interplanetary magnetic field was undisturbed prior to the occurrence of the sudden commencement, although the magnitude increased gradually. Shortly before the terrestrial observations of the sudden commencement

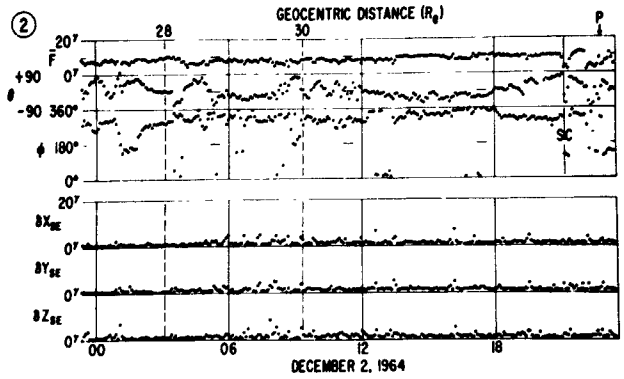


FIGURE 34.—Magnetic field data from orbit No. 2 December 2, 1963, illustrating interplanetary magnetic field at time of sudden commencement storm.

The field decreased very rapidly, and varied somewhat for several hours, eventually returning to a configuration similar to that prior to the storm. An expanded time scale presentation of the interplanetary magnetic field is presented in Figure 35.

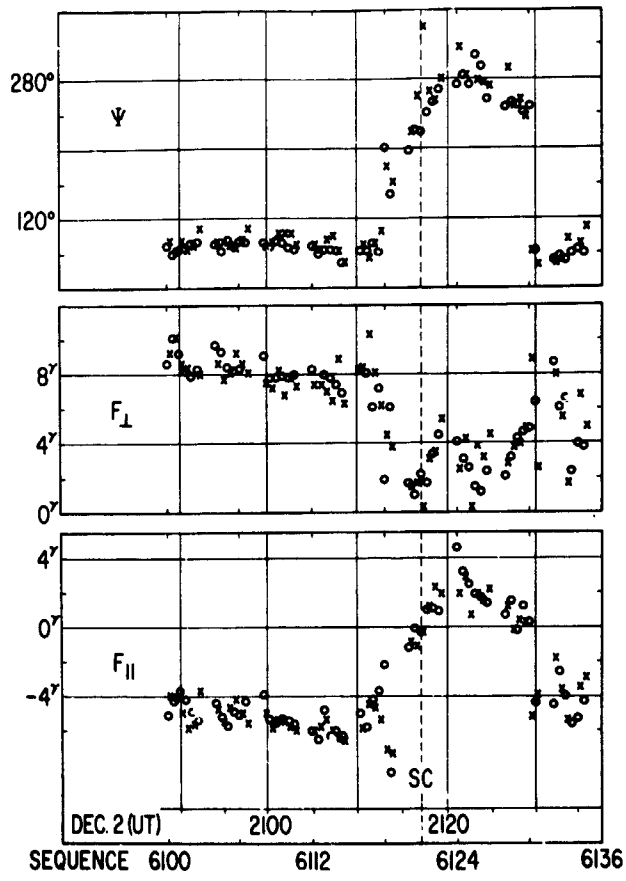


FIGURE 35.—IMP-I magnetic field data on expanded time scale in payload coordinates at time of December 2, 1963, sudden commencement storm. Circles refer to fluxgate *A* data while crosses reference fluxgate *B*.

in a payload coordinate system. Instead of 5.46 minute averages, the 12 sample values in that interval are presented separately with a time spacing of 20 seconds. It is clear that the occurrence of the sudden commencement indicated by the initial onset of the magnetic field disturbance occurs at 2114 on December 2 at the satellite. Assuming radial propagation of the plasma front and the associated magnetic field distortions to the surface of the Earth yields a propagation velocity of 700 kilometers per second. This velocity is an apparent value of velocity of propagation since it does not include the physical propagation of the storm transient itself to the surface of the Earth, assumes the direction of propagation of the storm front and that the magnetic field disturbance and the plasma disturbance occur at the same place in the solar plasma stream. This value of 700 kilometers, however, is consistent with the transit time associated with the projected solar region responsible for the storm, and is equivalent to particle energies of 2.5 kev. On the assumption of a propagation velocity of 700 kilometers per second and observing the time interval during which the magnetic field is disturbed and presumed to be associated with the propagation of the shock front surface, the thickness is 0.005 AU. This is consistent with particle measurements yielding scale lengths of scattering centers in interplanetary space at the time of magnetic disturbances.

Another important feature of the storm is that there is a null surface observed during the transit of the storm front surface across the spacecraft. The structure may well be consistent with Parker's "blast wave" model of sudden commencement storms if one includes the fact that if the compressed interplanetary fields are filamentary in their topology, then null surfaces must occur. This suggests that magnetic field annihilation as proposed by Petschek (1963) may be the mechanism whereby acceleration of particles in the shock front occurs. The present sample of magnetic fields associated with SC's in interplanetary space does not show the large fields, up to 50 gammas, previously reported by Coleman et al., (1961) for Pioneer V results. Indeed the experimental results appear to be in the main contradictory since Pioneer V did not measure a

null surface either. At the present time, it is suggested that these differences only reflect the great variability in storm characteristics and intensities.

12.0 INTERPRETATIONS OF SOLAR STREAM PROPERTIES

The location of the magnetopause and shock wave boundaries as observed by the magnetic field experiment permit an extrapolation to the characteristics of the streaming solar plasma. The streaming angles indicated by the magnetic fields are used in computing the characteristics of the solar stream. The interpretation assumes that the streaming angle provides a reasonable estimate of the particle velocity. Figure 36. gives

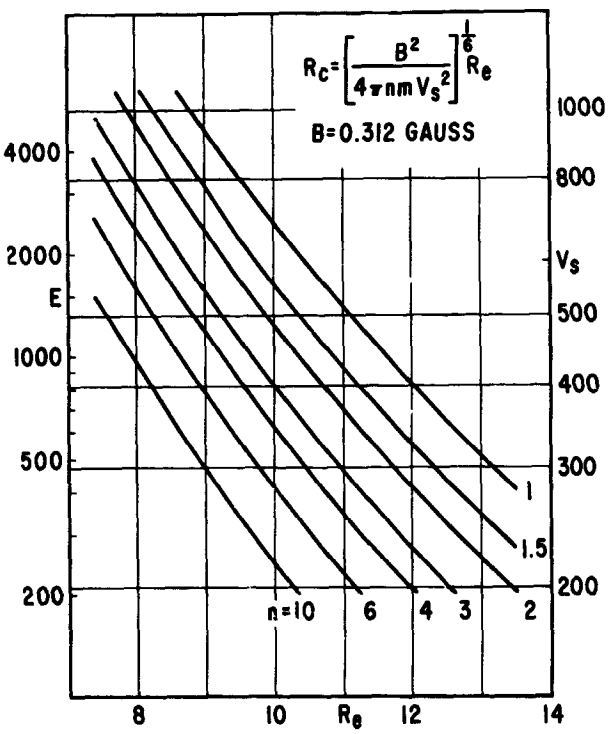


FIGURE 36.—Theoretical stagnation point distance R_c of magnetopause boundary as function of particle velocity (or energy) and plasma density.

the particle density and velocity that would be consistent with the observed location of the magnetopause if the plasma were directly incident upon the magnetopause. This density and velocity represent an effective description of the plasma flux pressure, being balanced by the magnetic field pressure. The actual solar wind velocity and density must be adjusted in terms of the boundary conditions which exist at the shock wave. On the assumption of particle velocities between 300–700 kilometers per second and the observed location of the magnetopause boundary of 10.25 R_e, it is seen that the corresponding proton density is between 1 and 5 protons/cc. This yields a flux value of 0.7 to 1.5×10^8 protons/cm^2/sec, which is in good agreement with previous plasma results extrapolated from Explorer X, (Bonetti et al., 1963) and the direct measurements on the Mariner II spacecraft (Snyder and Neugebauer, 1962).

13.0 MAGNETOHYDRODYNAMIC WAKE OF THE MOON

A unique phenomena occurred in interplanetary space during the fifth orbit of the satellite. At that time the interplanetary magnetic field was observed to fluctuate much more so than had been previously observed or was subsequently observed and also to reach a maximum value of 14.6 gammas for a period of 4 hours on December 14, 1963. This high value of both the magnitude and the variance had never been observed throughout the previous 4 orbits of the spacecraft when in the interplanetary medium and some concern was expressed initially as to whether or not the operation of the instrument and/or the telemetry digitization system was proper at the time. The integrity of the operation of the telemetry system and the sensor has been verified and the conclusion is irrevocably that the basic data is valid. At this time the satellite was located between 28 to 31 R_e from the Earth at approximately an angle of 40° to the Sun-Earth line and far beyond the shock wave location.

Investigation of solar conditions prior to the anomalous field behavior, extending from 0100 UT, December 14 to 1000 UT Decmber 15, has not revealed any significant features that might have led to disturbed conditions in interplanetary space. The characteristics of the fluctuations are quite unlike those of the magnetic fields associated with the sudden commencement storm on December 2, 1963. Indeed the characteristics are similar to those observed when sampling the transition region associated with the solar wind impacting the geomagnetic field.

The explanation offered is that the relative location of the Sun, Moon and satellite was unique and that the interaction of the solar wind with the Moon and its magnetosphere formed a magnetospheric tail which IMP-I traversed. This extended far behind the Moon much in the manner of cometary tails (Biermann 1951) and as may be the case of the Earth's tail. An inspection of the ephemerides of the Moon, Sun and IMP-I satellite has shown that this is a reasonable interpretation. Thus it provides the first direct measurements of magnetic field phenomena associated with the Moon and indeed may provide quantitative estimates of the lunar field strength upon subsequent detailed analysis.

The projected locations of the relevant celestial objects on the ecliptic plane is shown in Figure 37. Included is the location of the magnetopause

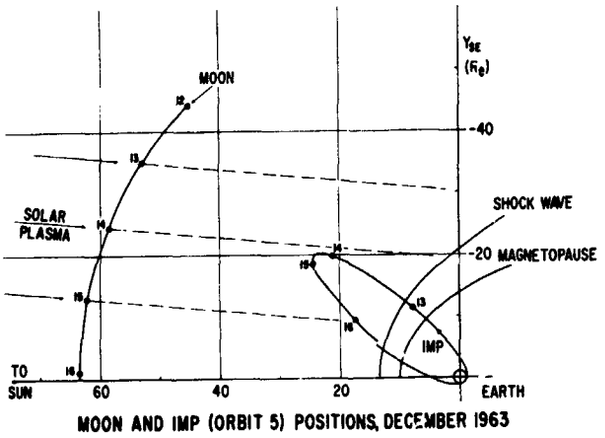


FIGURE 37.—Relative location of IMP-I Satellite, earth and moon during period December 12-16, 1963, when lunar magnetospheric wake was detected by the magnetic field experiment.

and the shock wave boundaries to illustrate the fact that the satellite was not sampling the region of interplanetary space influenced by the presence of the Earth. The numbers on both the IMP-I and Moon orbits indicate the location of the respective object at 00 hours UT on the particular day in December 1963. The projected view of the same days on the solar ecliptic YZ plane is shown in Figure 38. Certain physical parameters important in specifying the relative location of the Sun, Moon and IMP-I are

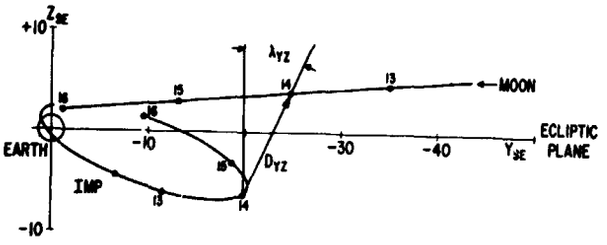


FIGURE 38.—Relative location of IMP-I Satellite, earth and moon projected on plane perpendicular to earth-sun direction, the solar-ecliptic, YZ plane during period December 12-16, 1963.

defined as follows:

$\vec{D}$ = vector distance from satellite to Moon

$$\vec{D} = D_x\vec{i}_x + D_y\vec{i}_y + D_z\vec{i}_z$$

$$D_{sm} = |\vec{D}| = [D_x^2 + D_y^2 + D_z^2]^{\frac{1}{2}}$$

$$D_{xy} = [D_x^2 + D_y^2]^{\frac{1}{2}}$$

= projected distance in XY plane (11.1)

$$D_{yz} = \left[D_y^2 + D_z^2\right]^{1/2} = \text{projected distance in } YZ \text{ plane}$$

cosine $\lambda_{xy} = D_y/D_{xy}$ = cosine of angle between $\vec{D}$ and $\vec{i}_x$ projected into XY plane

cosine $\lambda_{yz} = D_z/D_{yz}$ = cosine of angle between $\vec{D}$ and $\vec{i}_z$ projected into YZ plane

The distance D_{yz} and the angle λ_{yz} are illustrated in Figure 38. For an optical occultation $D_{yz} = 0$, while when $\lambda_{yz} = 0$ it implies that the Sun-Moon-IMP-I define a plane that is perpendicular to the ecliptic plane.

A summary of the above critical parameters as a function of time in December is shown in Figure 39. as well as the indication of where traversals of the shock wave boundary occurred. The parameter D_{yz} refers to the projected distance on a plane rotated by 5° from the YZ plane to take into account the aberration of the solar wind. The "closest" distance from the satellite to the Moon is seen to be 7.9 R_e in the YZ plane while the satellite is 38 R_e from the Moon. Converted to lunar radii, R_m while the apparent diameter of the tail at that distance is at least 30 R_m. These values indicate a magnetospheric wake with a half-cone angle of approximately 11°, a value consistent with the Mach angle associated with the solar wind interaction with a Moon possessing a weak magnetic field.

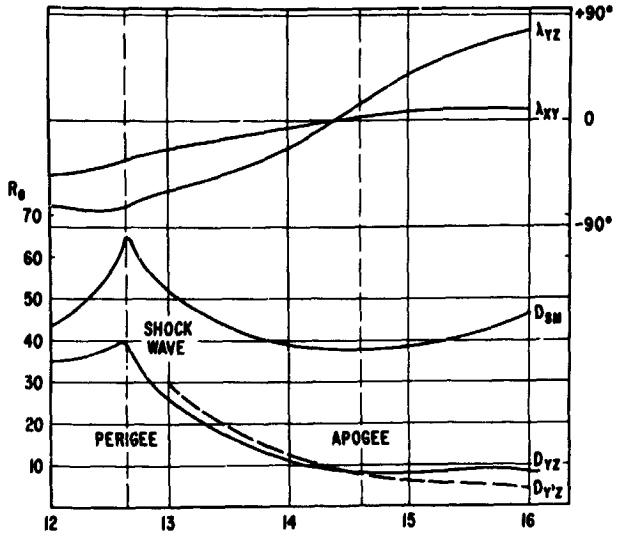


FIGURE 39.—Critical parameters related to "shadowing" by moon's magnetosphere on IMP–I magnetic field measurements during December 12–16, 1963.

The direct measurement of the lunar magnetic field and a precise determination of its geometrical properties is of vital importance in the study of the origin of the Earth-Moon system. The interpretation of terrestrial data suggests the nonexistence of a lunar magnetic field similar in origin to the Earth's in which a dynamo system of currents circulates in a fluid core. The presence of a lunar magnetic field may be indicative of a permanent state of magnetization which reflects an ancient field at the time of origin of the Moon. It has been suggested by Gold (1962) that the streaming magnetized solar plasma may provide a mechanism whereby the interplanetary magnetic field is captured by the finite electrical conductivity of the Moon. This would then lead to a magnetic field configuration which would be interpreted as a lunar magnetic field.

Russian measurements of the lunar magnetic field on the second Cosmic rocket (Dolginov, Eroshenko, Zhugov, Pushkov and Tynrmina, 1960) indicate a surface field of less than 50 to 100 gammas corresponding to a magnetization intensity less than 0.25 percent of the Earth's. The streaming solar plasma is sufficiently strong that it will greatly distort a lunar magnetic field regardless of its origin (Neugebauer, 1960). Thus it can reasonably be expected that a cavity and geomagnetic tail similar to that observed on Explorer X (in the anti-solar direction behind the Earth) and thus far suggested by IMP–I, will also develop around the Moon.

The detailed measurement and accurate vector mapping of such a lunar magnetic field and associated magnetospheric tail can only be accomplished with a spacecraft orbiting the Moon. The present IMP–I measurements may provide information sufficient to indicate the nature of the lunar magnetic field. A detailed study of the magnetic field results in currently being conducted and a future publication will treat the lunar wake separately.

14.0 CONCLUSIONS

The IMP–I satellite has provided the first direct measurements of interplanetary magnetic fields with high accuracy. The spacecraft fields are found by analysis of in-flight data to be less than ± 0.25 gamma. The observed field strengths in the interplanetary medium range generally from 4 to 7 gammas with temporary excursions to values as low as 1 and as high as 10 gammas. The magnitude is impressively steady although directional changes occur fairly frequently on time scales of several hours. The steady magnitudes are a feature of the fields which has not been anticipated.

A significant feature of the vector field and the inferred interplanetary field configuration is the gross feature of spiral structure associated with the rotation of the Sun, in directions consistent with observed solar plasma velocities. At times the field reverses directions on a time scale of 5–20 minutes, going through zero rapidly, indicating neutral surfaces separating filamentary elements of magnetic fields. The overall Archemedian spiral structure imbedded in which are the distinctive filaments strongly suggests a merging of theoretical field topologies principally discussed by Parker (1958) and Gold (1959). However, the fields have only been recently sampled and further analysis will be required to determine whether the "magnetic bottles" of Gold (1959) are also consistent with the data. Indeed with only one sample point in space in space it may not be possible to delineate between such features.

Our observations of the solar wind interactions have experimentally confirmed the existence of the collisionless shock wave predicted by theoretical analogies with supersonic gas dynamics.

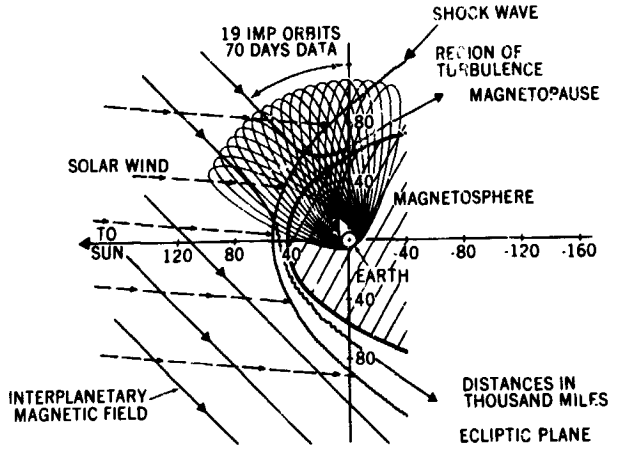


FIGURE 40.—Illustration of IMP—I magnetic field data interpretation showing solar wind, interplanetary fields collisionless shock wave and magnetopause locations on ecliptic plane.

Indeed the position of the shock wave is impressively consistent with the theoretical models showing a departure of only a few percent. The magnetosphere boundary, the magnetopause, is clearly observed and found to be separated from the shock wave by a region of high turbulence in the magnetic field. An interpretive graphical summary of the initial results of interaction of the solar wind with the Earth is shown in Figure 40. The geocentric distance to the magnetopause is found to be 10.25 R_e at the stagnation point. The distance to the shock wave is 13.4 R_e yielding a stand-off ratio of 1.31 ± 1 percent.

The detection of the extended lunar magnetospheric wake into interplanetary space at least

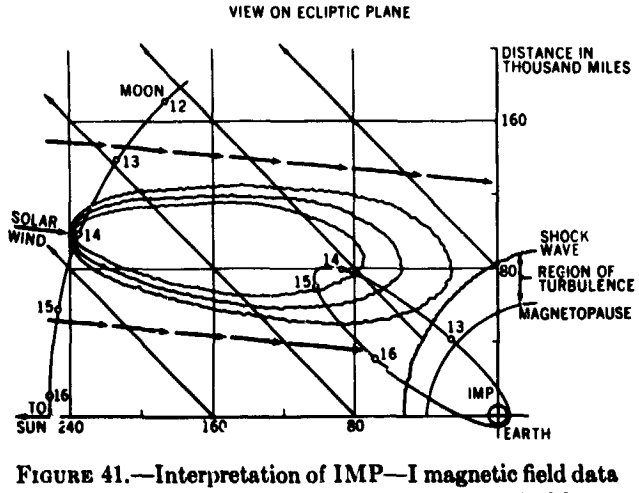


FIGURE 41.—Interpretation of IMP—I magnetic field data indicating characteristics of interplanetary and cislunar space and *MHD* wake of moon at time of lunar magnetosphere detection.

150 lunar radii behind the Moon is important evidence in confirming the general aspects of our understanding of the solar wind interactions with planetary objects. The field strengths and directions are indicative of the magnetohydrodynamic interaction of the solar wind with the Moon's magnetosphere and may provide evidence on the lunar magnetic field strength. The lunar wake geometry is illustrated in Figure 41.

ACKNOWLEDGEMENTS

The success of the magnetic field experiment is due in large part to the outstanding effort by the technical staff at Goddard Space Flight Center, the other experimenters and the IMP project staff in restricting the spacecraft fields to an absolute minimum. We are particularly indebted to those individuals without whose cooperation and contributions the final cleanliness of the spacecraft could not have been achieved. In addition the contributions of additional individuals too numerous to separately mention in the overall IMP spacecraft mission is hereby gratefully acknowledged.

REFERENCES

ADLAM, J. H. and J. E. ALLEN, The structure of strong collision-free hydromagnetic waves, *Phil. Mag.*, **3**, 448–455, 1958.

ALFVEN, H. L. DANIELSSON, C-G FALTHAMMAR and L. LINDBERG, On the penetration of interplanetary plasma into the magnetosphere, Preprint 1963.

AUER, P. L., H. HURWITZ and R. W. KILB, Low Mach number magnetic compression waves in a collision-free plasma, *Phys. Fluids*, **4**, 1105–1121, 1961.

AUER, P. L., H. HURWITZ, JR. and R. W. KILB, Large amplitude magnetic compression of a collision-free plasma, 2; Development of a thermalized plasma, *Phys. Fluids*, **5**, 298–316, 1962.

AXFORD, W. I., The interaction bewteen the solar wind and the earth's magnetosphere, *J. Geophys. Res.*, **67**, 3791–3796, 1962.

AXFORD, W. I. and C. O. HINES, A unifying theory of high latitude geophysical phenomena and geomagnetic storms, *Can. Journ. Phys.*, **39**, 1433–1464, 1961.

BEARD, D. B., The interaction of the terrestrial magnetic field with the solar corpuscular radiation, *J. Geophys, Res.*, **65**, 3559–3568, 1960.

BEARD, D. B., The effect of an interplanetary magnetic field on the solar wind, *J. Geophys. Res.*, **69**, 1159–1168, 1964.

BERNSTEIN, W., R. W. FREDERICKS and F. L. SCARF, A model for a broad disordered transition between the solar wind and the magnetosphere, *J. Geophys. Res.*, **69**, 1201–1210, 1964.

BIERMANN, L. Kometenschweife und solare Korpuscular strahlung, *Z. Astrophys.* **29**, 274– 1951.

BRIDGE, H. S., Plasmas in space, *Physics Today*, **16**, 31–37, 1963.

BLOOM, A. L., Principles of operation of the Rubidium Vapor magnetometer, *Applied Optics*, **1**, 61–68, 1962.

BONETTI, A., H. S. BRIDGE, A. J. LAZARUS, B. ROSSI and F. SCHERB, Explorer X plasma measurements, *J. Geophys. Res.*, **68**, 4017–4063, 1963.

CAHILL, L. J. and P. G. AMAZEEN, The boundary of the geomagnetic field, *J. Geophys. Res.*, **68**, 1835–1844, 1963.

CHAPMAN, S. and FERRARO, V. C. A., A new theory on magnetic storms, *Terr. Mag. and Atmos. Elec.*, **36**, 77-97, 171–186, 1931; **37**, 147–156, 421–429, 1932; **38**, 79–96, 1933.

CHAPMAN, S., Idealized problems of plasma dynamics relating to geomagnetic storms, *Revs. Mod. Phys.*, **32**, 919–933, 1960.

CHAPMAN S., Solar plasma geomagnetism and aurora in geophysics: *The Earth's Environment*, Les Houches 1962 Lectures, edited by C. DeWitt et al., 371–502, Gordon and Breach, New York, 1963.

COLEMAN, P. J., L. DAVIS and C. P. SONETT, Steady component of the interplanetary magnetic field, Pioneer V, *Phys. Rev. Lettrs.* **5**, 43–46, 1960.

COLEMAN, P. J., C. P. SONETT and L. DAVIS, On the interplanetary magnetic storm, Pioneer V, *J. Geophys. Res.*, **66**, 2043–2046, 1961.

COLEMAN, P. J., L. DAVIS, E. J. SMITH and C. P. SONETT, The mission of Mariner II: Preliminary observations, *Science*, **138**, 1095–1100, 1962.

DAVIES, R. D., G. L. VERSCHUOR and P. A. T. WILD, Measurement of magnetic fields in interstellar clouds, *Nature*, 4854, 563, 1962.

DAVIS, L. and D. B. BEARD, A correction to the approximate condition for locating the boundary between a magnetic field and a plasma, *J. Geophys. Res.*, **67**, 4505–4507, 1962.

DE HOFFMAN, F. and E. TELLER, Magnetohydrodynamic shocks, *Phys. Rev.* **80**, 692–703, 1950.

DESSLER, A. J., The stability of the interface between the solar wind and the geomagnetic field, *J. Geophys. Res.*, **66**, 3587–3590, 1961.

DESSLER, A. J., Further comments on the stability of the interface between the solar wind and the geomagnetic field, *J. Geophys. Res.*, **67**, 4892–4894, 1962.

DESSLER, A. J. and J. A. FEJER, Interpretation of K_p index and M-region geomagnetic storms, *Planet. Space Sci.*, **11**, 505–511, 1964.

DESSLER, A. J. and G. K. WALTERS, Hydromagnetic coupling between the solar wind and the magnetosphere, *Planet. Space Sci.*, **12**, 1964.

DOLGINOV, S. H., E. G. EROSHENKO, L. N. ZHUGOV, N. V. PUSHKOV and L. O. TYNRMINA, Magnetic measurements on the second cosmic rocket, *Artificial Earth Satellites-5*, 16–23, 1960. (ARS Journal Suppl., 1640–1643, 1961).

FERRARO, V. C. A., On the theory of the first phase of a geomagnetic storm, *J. Geophys. Res.*, **57**, 15-49, 1952.

FERRARO, V. C. A., An approximate method of estimating the size and shape of the stationary hollow carved out in a neutral ionized stream of corpuscles impinging on the geomagnetic field, *J. Geophys. Res.*, **65**, 3951–3953, 1960a.

FERRARO, V. C. A., Theory of sudden commencement and of the first phase of a magnetic storm, *Revs Modern Phys.* **32**, 934–938, 1960b.

FINCH, H. F. and B. R. LEATON, The earth's main magnetic field-epoch 1955, *MNRAS, GS*, **4**, 314–317, 1957.

FISHMAN, F. J., A. R. KANTROWITZ and H. E. PETSCHEK, Magnetohydrodynamic shock wave in a collision-free plasma, *Rev. Mod. Phys.*, **32**, 959–966, 1960.

FREEMAN, J. W., J. A. VAN ALLEN and L. J. CAHILL, Explorer XII observations of the magnetosphere boundary and the associated solar plasma on September, 13, 1961, *J. Geophys. Res.*, **68**, 2121–2130, 1963.

GOLD, T., Magnetic fields in the solar system, Nuovo Cimento, Suppl., (10) 3, 318–323, 1959.

GOLD, T., The hydrogen atmosphere of the moon, 3rd Int. Space Sci., Symp., Wash., D.C., 1962.

GOLD, T., Magnetic storms, *Space Sci. Rev.*, **1**, 100–114. 1962.

GOLD, T. Magnetic field configurations near the sun and in interplanetary space, *Proc. Pontifical Academy*, 431–457, 1963.

GRAD, H., Boundary layer between a plasma and a magnetic field, *Physics Fluids*, **4**, 1366–1375, 1961.

GREENSTADT, E. W., Private communication, 1963.

GREENSTADT, E. W., A comparison of preliminary Mariner 2 plasma data with solar wind velocities deduced from Pioneer V magnetic field measurements, *J. Geophys. Res.*, **69**, to be published, 1964.

HARRISON, E. R., The Earth's distant geomagnetic field, *Geophys. J.*, **6**, 479-491, 1962.

HEPPNER, J. P., N. F. NESS, T. L. SKILLMAN and C. S. SCEARCE, Explorer X magnetic field measurements, *J. Geophys. Res.*, **68**, 1–46, 1963.

HIDA, K., An approximate study on the detached shock wave in front of a circular cylinder and a sphere, *J. Phys. Soc., Japan*, **8**, 740–745, 1953.

HINES, C. O., The magnetopause: a new frontier in space, *Science*, **141**, 130–136, 1963.

HONES, E. W., Motions of charged particles trapped in the earth's magnetosphere, *J. Geophys. Res.*, **68**, 1209–1219, 1963.

HURLEY, J. D., Interaction between the solar wind and the geomagnetic field, *Phys. Fluids*, **4**, 854–859, 1961.

JOHNSON, F. S., The gross character of the geomagnetic field in the solar wind, *J. Geophys. Res.*, **65**, 3049–3052, 1960.

KELLOGG, P. J., Flow of plasma around the earth, *J. Geophys. Res.*, **67**, 3805–3811, 1962.

LEES, L., Interaction between the solar plasma wind and the geomagnetic cavity, N. Y. meeting of AIAA, January, 1964. Preprint 64-88.

LEVY, R. H., H. E. PETSCHEK and G. L. SISCOE, Aerodynamic aspects of the magnetosphere flow, AVCO Corp. Res. Rep. 170, 1963.

MAER, K and A. J. DRESSLER, Comments on paper by Conway W. Snyder, Marcia Neugebauer, and U. R. Rao, "The Solar Wind velocity and its correlation with cosmic ray variations and with solar and geomagnetic activity," *J. Geophys. Res.* **69**, to be published, 1964.

MCCRACKEN, K. G., The cosmic-ray flare effect, 1., 2. and 3., *J. Geophys. Res.*, **67**, 423–458, 1962.

MCKEEHAN, L. W., Combinations of circular currents for producing uniform magnetic fields, *Rev. Sci. Instrum.* **7**, 150–153, 1936.

MEAD, G. D., Deformation of the geomagnetic field by the solar wind, *J. Geophys. Res.*, **69**, 1181–1195, 1964.

MEAD, G. D. and D. B. BEARD, Shape of the geomagnetic field solar wind boundary, *J. Geophys. Res.*, **69**, 1169–1180, 1964.

MIDGLEY, J. E. and L. DAVIS, Computation of the bounding surface of a dipole field in a plasma by a moment technique, *J. Geophys. Res.*, **67**, 499–504, 1962.

MORAWETZ, C. S., Magnetohydrodynamic shock structure without collisions, *Phys. Fluids*, **4**, 998–1006, 1961.

MORAWETZ, C. S. Modification for magnetohydrodynamic shock structure without collisions, *Phys. Fluids*, **5**, 1447–1450, 1962.

MORTON, K. W., Large-amplitude compression waves in an adiabatic two-fluid model of a collision-free plasma *J. Fluid Mech.*, **14**, 369–384, 1962.

MORTON, K. W., Finite Amplitude compression waves in a collision-free plasma, NYO–10434, Courant Institute of Math. Sciences, New York Univ., 1964.

NESS, N. F., Observations of the geomagnetic cavity boundaries, Catholic U. Symposium on Plasmas in Space, to be published, 1963.

NEUGEBAUER, M., Question of the existence of a lunar magnetic field, *Phys. Rev. Letters*, 4- ' 1960.

NEUGEBAUER, M., and C. W. SNYDER, The mission of Mariner II: Preliminary Observations, Solar Plasma Experiment, *Science*, **138**, 1095–1097, 1962.

OBAYSHI, T., Interaction of solar plasma streams with the outer geomagnetic field, *J. Geophys. Res.*, **69**, 861–868, 1964.

PARKER, E. N., Dynamics of the interplanetary gas and magnetic fields, *Astrophys. J.*, **128**, 667–676, 1958.

PETSCHEK, H. E. Magnetic field annihilation, AVCO Corp. Report AMP 123, 1963.

ROSSI, B., Interplanetary plasma, Space Research III, 529–539, 1963.

RUDDOCK, K. A., Optically pumped rubidium vapor magnetometer for space experiments, Proc. Intern Space Sci. Symp., 2, 692–700, 1961.

SLUTZ, R. J., Numerical method for calculating the equilibrium configuration of free-surface plasma, *Bull. Am. Phys. Soc.*, **8**, 165–167, 1963.

SNYDER, C. W., M. NEUGEBAUER, and U. R. RAO, The solar wind velocity and its correlation with cosmic ray variations and with solar and geomagnetic activity, *J. Geophys. Res.*, **68**, 6361–6370, 1963.

SNYDER, C. W. and M. NEUGEBAUER, Interplanetary solar wind measurements by Mariner II, Space Res., IV, 1964 (to be published).

SPREITER, J. R. and B. R. BRIGGS, Theoretical determination of the form of the boundary of the solar corpuscular stream produced by interaction with the magnetic dipole of the earth, *J. Geophys. Res.*, **67**, 37–51, 1962. Also NASA Tech. Report R-120, 1961.

SPREITER, J. R. and W. P. JONES, On the effect of a weak interplanetary magnetic field on the interaction between the solar wind and the geomagnetic field, *J. Geophys. Res.*, **68**, 3555–3565, 1963.

WALTERS, G. K., Effect of oblique interplanetary magnetic field on magnetosphere shape and behavior, *J. Geophys. Res.*, **69**, to appear, 1964.

ZHIGULEV, V. N. and E. A. ROMISHEVSKII, Concerning the interaction of currents flowing in a conducting medium with the earth's magnetic field, *Soviet Phys. Doklady*, **4**, 879–882, 1960.

ZHIGULEV, V. N., On the phenomenon of magnetic detachment of the flow of a conducting medium, *Soviet Physics, Doklady*, **4**, 344–346, 1959.

THE MAGNETOHYDRODYNAMIC WAKE OF THE MOON*

NORMAN F. NESS

The possible detection of the lee wake of the magnetohydrodynamic interaction of the solar wind with the Moon as observed by the IMP-1 satellite is discussed. The interplanetary magnetic field was found to fluctuate very rapidly and reach anomalously large values when the satellite was approximately eclipsed by the Moon in December 1963. Later data on the interplanetary field in February 1964 suggest that a detached lunar shock wave analogous to that observed by IMP-1 associated with the Earth may not be a permanent feature of the lunar environment. The approximate length of the wake region behind the Moon is 150 lunar radii at which distance the diameter of the region is about 70 lunar radii. A review of related studies on 29.5 day periodicities of K_p is presented. The solar wind interaction with the geomagnetic field extends the magnetosphere far behind the Earth. Hence, lunar synodic periodicities in K_p may reflect the interaction of the Moon with the Earth's magnetic tail rather than the moon's wake with the Earth.

1.0 INTRODUCTION

Lunar associated terrestrial phenomena have been studied extensively for many years in a wide range of scientific disciplines including geomagnetism (Chapman and Bartels, 1940). The principal geophysical effects which have been observed are due primarily to the direct or indirect influence of the Moon's gravitational field. The rotation of the Earth coupled with the elliptic orbit of the Moon introduces characteristic time modulations and fluctuations of geophysical phenomena with frequencies which can be accurately ascribed to a lunar gravitational influence. The various "tidal" phenomena with semi-diurnal to monthly periods occurring in both the solid and fluid Earth contain important lunar constituents (Melchior, 1957; Doodson, 1958; Siebert, 1960). Recently attention has been directed to a study of the expansion of the solar corona into interplanetary space. This can lead to a somewhat different and more complex lunar as well as solar influence on the immediate terrestrial environment. Fejer (1964) has suggested that at least a portion of the classical "tidal" variations in geomagnetism may be directly related to this phenomenon.

Subsequent to the early suggestion by Biermann (1951) that an adequate explanation of cometary tail structures required a substantial and continuous solar corpuscular flux, Parker (1958, 1960) developed the hydrodynamic theory of the coronal expansion, referring to this plasma flow as the "solar wind." Direct measurements by satellite and space probes since 1959 have confirmed experimentally the existence of a solar plasma flowing approximately radially from the Sun with a positive ion flux between 10^7 to 10^8 particles/cm^2/sec and energies between 200ev to 10 Kev (Gringauz, 1961; Bonetti et al, 1963; Snyder and Neugebauer, 1964; Bridge et al, 1964). A direct consequence of the radial solar wind velocity, the highly ionized coronal gases, and the rotation of the Sun is that the photospheric magnetic fields are extended in the plane of the ecliptic and twisted into the form of an Archimedes spiral (Parker, 1963). Accurate measurements of the interplanetary magnetic field have recently been made on the IMP-1 satellite (Ness et al, 1964) confirming certain of the theoretical predictions on the direction of the field and determining the average magnitude to be close to 5 gammas (Ness and Wilcox, 1964).

The interaction of the solar wind with planetary objects is an important factor in determining

*Published as *Goddard Space Flight Center Document X-612-64-333*, October 1964; also published in *Journal of Geophysical Research*, 70(3):517-534, February 1, 1965.

their space environment. The principal result of the solar wind impacting a planetary magnetic field is to confine the field to a region of space surrounding the planet. The existence of a quasi-static Chapman-Ferraro cavity (reviewed by Chapman (1963)) which excludes the direct penetration of solar plasma into the geomagnetic field has been demonstrated by the recent satellite experiments of Bonetti et al (1963), Heppner et al (1963), Cahill and Amazeen (1963), Freeman et al (1963), Ness et al (1964), Anderson et al (1964), Bridge et al (1964), and Freeman (1964). An important feature of the interplanetary medium, with respect to its interaction with planetary objects, is that it is a magnetized plasma and thus supports magnetohydrodynamic wave propagation. In this sense the plasma flow is super-Alfvénic since the velocity of magnetohydrodynamic wave propagation is less than 90 km/sec while the solar plasma velocity is between 300–900 km/sec. This leads to Alfvén Mach numbers usually greater than 4.

This paper is concerned with an expanded discussion of possible direct experimental evidence of the magnetohydrodynamic interaction of the solar wind with the Moon as obtained by the IMP–1 magnetic field experiment (Ness et al, 1964). During the month of December 1963 the IMP–1 satellite appears to have been located within the downstream wake region of the Moon as it interacts with the solar wind. This interpretation of the data as well as the data itself will be discussed in the following sections.

2.0 MAGNETIC FIELD OF THE MOON

The existence or absence of an intrinsic lunar magnetic field is important in the development of adequate theories to explain planetary magnetism and the origin of the solar system. The presently accepted theory for the origin of the geomagnetic field is based upon a fluid core sustaining a regenerative dynamo system of electrical currents (Elsasser, 1950; Inglis, 1955). The possibility of a lunar magnetic field maintained by such fluid motion has been theoretically investigated by Zharkov and Ulinich (1960) concluding that surface fields of 0.1 to 1.0 gauss are possible. The internal constitution of the Moon is not well known (MacDonald, 1961) but the possibility of a dense fluid core analogous to the Earth's seems highly unlikely. Another possibility is that sometime in the ancient past the Moon cooled down through its Curie point while in the presence of an external magnetic field thus acquiring thermoremanent magnetization. This latter suggestion requires that the Moon was at one time nearly entirely molten whether it accreted from a large number of smaller masses or was formed from one large mass. At the present time this hypothesis cannot be conclusively argued (Levin (1960), MacDonald (1963)).

The only direct evidence of the state of a lunar magnetic field has been provided by the Soviet space probe Lunik II (Dolginov et al, 1961). Measurements with a sensitivity of about 30 gammas were performed along an impact trajectory to within a distance of 55 km from the lunar surface. No lunar magnetic field larger than the noise level was detected. The conclusion reached by these authors on the basis of a simple dipole model of the lunar magnetic field was that the effective lunar magnetic moment must be less than 10^{-4} that of the Earth's. As pointed out by Vestine (1957) and Neugebauer (1960) however, the solar wind is sufficiently strong to confine the magnetic field of the Moon to a region very close to or even below its surface. In the latter case the solar wind directly impacts the lunar surface. On the assumption that

(1) the Moon possesses a dipolar magnetic field with equatorial field strength $= B_0$,
(2) the dipole axis is perpendicular to the solar wind velocity and
(3) the solar plasma flows on around to the back side of the Moon rather than to be specularly reflected

then the radius of the lunar magnetic cavity boundary at the subsolar or stagnation point is given by:

$$R_c = R_m \left[\frac{B_0^2}{2\pi m n V_s^2} \right]^{1/6} \qquad (2.1)$$

where V_s = velocity of solar wind (km/sec)
n = proton density (p/cm^3)
m = proton mass (gms)

This is illustrated in Figure 1 and clearly shows that for the observed values of the solar wind any such lunar magnetic field will be compressed rather close to the surface of the Moon.

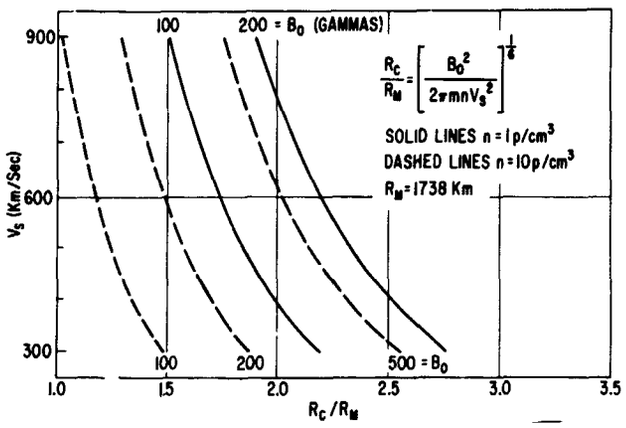


FIGURE 1.—Lunar magnetic cavity radius or curvature (R_c) in units of lunar radius (R_m) as a function of the solar wind flux and equatorial field stenrgth (B_0) of assumed lunar dipole.

Gold (1964) has suggested an origin of a lunar field which incorporates the interaction of the magnetized solar plasma with the lunar body. The finite electrical conductivity of the Moon provides a mechanism for the trapping of the interplanetary magnetic field. The model proposed has not been quantitatively developed. It depends critically upon the magnetic Reynolds number of the Moon, requiring that the interplanetary field be obstructed in its motion through the lunar body so that lines of force are accumulated and compressed on the sunlight hemisphere. This must occur for a time interval sufficiently long that a quasi-static field configuration can develop in spite of the rotation of the Moon and the observed variability of the direction of the interplanetary magnetic field.

3.0 LUNAR SHOCK WAVE

The presence or absence of a lunar magnetic field is important in determining the detailed characteristics of the solar wind interaction with the Moon. Far downstream it is not as clear how important any lunar field may be in determining the characteristics of the resultant interactions in the solar wind as it resumes a free stream flow. The overall situation may well resemble more closely the interaction of the solar wind with comets (Marochnik, 1963) rather than its interaction with the geomagnetic field. For the present study, the most pertinent feature of the solar wind interaction is the distance behind the Moon to which observable effects of such magnetohydrodynamic phenomena would persist. Various authors (Beard, 1960; Johnson, 1960; Piddington, 1960; Harrison, 1963; Lees, 1964; Axford et al, 1964 and Dessler, 1964) have presented qualitative arguments on the length of the geomagnetic "tail" behind the Earth. This region of the magnetosphere has yet to be experimentally probed and existing theoretical treatments are limited in their scope. The best efforts thus far have utilized the analogy with supersonic gas dynamics representing the solar wind as a viscous fluid and translating the particle characteristics of the plasma to a continuum property of the fluid.

The suggestion that a detached collisionless magnetohydrodynamic shock wave encloses the Earth's magnetosphere has been presented by Axford (1962) and Kellogg (1962) among others. Spreiter and Jones (1963) have applied classical supersonic flow theory to determine the location and geometry of the shock wave surface. The agreement between theory and observation is very good, requiring only a minor modification in the choice of equivalent specific heat ratio used in the fluid dynamic analogy to yield improved comparisons with the observations. The most critical point in these studies being the ratio at the stagnation point of the shock wave radius to the geomagnetic cavity radius. The success of this similitude argument has been verified experimentally only near the stagnation point of the solar wind interaction with the Earth's magnetic field. However this suggests its consideration for conditions far from the stagnation point, i.e., the tail region of the solar wind interaction with the geomagnetic field. This has been done by Lees (1964) who showed that a secondary shock or tail shock would develop in the wake of the plasma flow around the earth. Data obtained by the IMP–1 satellite which have yet to be analyzed will partially answer these important questions about the magnetic field topology within the geomagnetic tail.

In the absence of more definite knowledge about a lunar magnetic field the argument for a lunar shock wave is not on as firm a basis as in the case of the Earth. In the relatively weak interplanetary magnetic field the radius of gyration of a solar wind proton is comparable to the lunar radius. Table 1 presents representative values for the proton Larmor radius as a function of

TABLE 1.—*Proton Larmor radius (in Kilometers) as function of particle Energy (or velocity) and magnetic field strength (gammas).*

Magnetic Field Strength (γ)		10	5	2
Proton Energy	Velocity			
4 Kev	876 Km/sec	914	1828	4572
2 Kev	619 Km/sec	646	1292	3232
1 Kev	438 Km/sec	457	914	2286
500 ev	310 Km/sec	323	646	1616
250 ev	219 Km/sec	229	457	1143

field strength and plasma energy (or velocity).

The Larmor radius is equal to or greater than the stand off distance of 550 km. of the shock wave at the stagnation point that is obtained using the value of $R_s/R_c=1.31$ as measured in the case of the Earth (Ness et al, 1964). For the case of a sphere Kellogg (1962) using Hida's (1953) analysis of a shock surrounding a spherical object determined the stand off ratio to be approximately 1.25 for Mach values greater than 4. It may well be that in the case of the moon a collisionless shock wave analogous to the Earth's may not develop at all times, if indeed at all. Whether or not such a shock develops depends very critically upon the physical parameters in the interplanetary medium and the state of the lunar magnetic field. This point has also been discussed by Gold (1964) with regard to the direct penetration of the solar plasma into the lunar body.

The detailed characteristic of the lunar surface and atmosphere are strongly dependent upon whether or not the surface is shielded from direct impact of the solar wind by either a lunar field or a shock wave. Studies on these and related lunar surface characteristics have been performed by Bernstein et al (1963), Wehner et al (1963), Hapke (1964), Hinton and Taeusch (1964) and Kopal and Rackham (1964).

4.0 OBSERVATIONS OF A LUNAR MAGNETOHYDRODYNAMIC WAKE

The IMP–1 magnetic field experiment has been described in an earlier publication (Ness et al, 1964) which should be consulted for a detailed description of the satellite, its orbit, the instrumentation and telemetry system and the general characteristics of the interplanetary magnetic field. In addition the interaction of the solar wind with the geomagnetic field is discussed. In the following presentation the positions of the satellite, earth and moon are initially presented in the solar ecliptic coordinate system defined as: X_{se} is the axis from the Earth to the Sun, Z_{se} is the axis normal to the ecliptic plane and in the same sense as the Earth's angular momentum and Y_{se} which forms a right handed coordinate system. All distances are given in units of an Earth radius ($R_e=6378.4$ Km) although in the final interpretation of the data a translation will be made to a lunar associated coordinate system using a scale in Moon radii ($R_m=1738$ Km).

The positions of the Moon and IMP–1 during December 1963 are shown in Figure 2 as projected

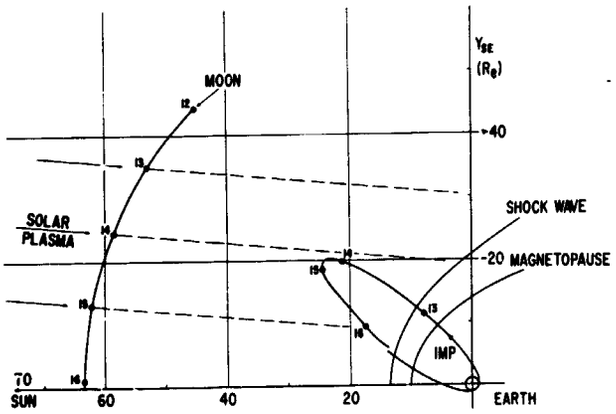


FIGURE 2.—Relative positions of IMP–1 satellite, earth and moon projected on ecliptic plane during period December 12–16, 1963, when the wake of the moon was detected in interplanetary space.

on the plane of the ecliptic. The solar wind direction has been adjusted to appear to come from 5° west of the Sun in order to represent the aberration effect due to the heliocentric motion of the Earth. The orbital velocity of the Earth is 30 Km/sec and for a solar wind velocity of approximately 330 Km/sec the angular displacement is 5°. Ness and Wilcox (1964) have determined from a correlation of the solar photospheric magnetic field and the interplanetary field as measured on IMP–1 that the average solar wind velocity was approximately 385 Km/sec during the time interval November 27, 1963 to February 19, 1964.

Thus the correction for a 5° aberration is slightly larger than would be justified on the basis of experimental data. The projected traces of the average position of the magnetopause and shock wave boundaries are also included in the figure in order to show when the satellite was outside of these boundaries and within the relatively undisturbed interplanetary medium. It is seen that on December 14 the satellite was close to apogee and approximately in line with the Moon with respect to the propagation of the solar wind. This is a measure only in the XY plane however, and does not include the distance of the satellite either above or below the Sun-Moon line.

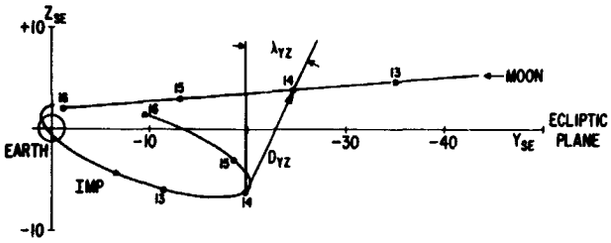


FIGURE 3.—Relative positions of IMP–1 satellite, earth and moon projected on solar ecliptic Y-Z plane during period December 12–16, 1963, when the wake of the moon was detected in interplanetary space.

In Figure 3 the positions of the satellite and the Moon during the same time interval are shown projected onto the plane perpendicular to the X_{se} axis. The satellite was several Earth radii below the Moon during the entire period of interest, December 14–15. The location of the satellite with respect to the Moon as a function of time is summarized in Figure 4. The two angles λ_{xy} and λ_{yz} are introduced as a measure of the apparent angular displacement of the Moon with respect to the satellite as projected onto the XY and YZ planes respectively. The quantity D_{yz} measures the projected distance in the YZ plane, D_{sm} measures the distance from the satellite to the Moon and $D_y'_z$ measures the distance projected onto a YZ plane which has been rotated by 5° West so as to be perpendicular to the aberrated solar wind. These parameters show that on December 14 at approximately 0915 the satellite was close to being immediately behind the Moon. At this time the satellite was 37.8 R_e from the Moon and 8.7 R_e below the Sun-Moon line. The aberration of the solar wind by 5° leads

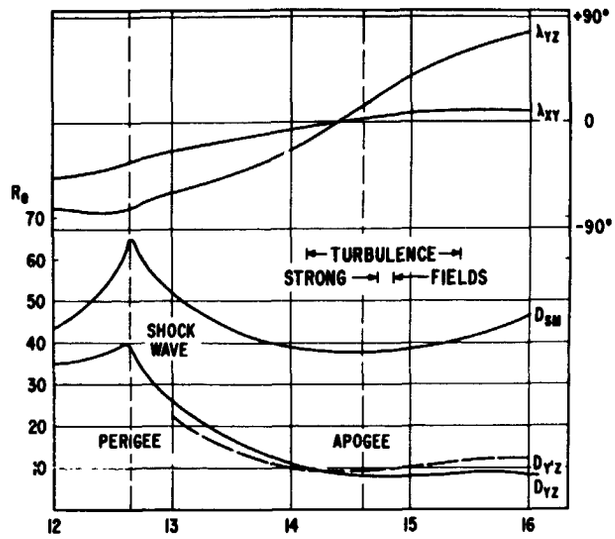


FIGURE 4.—Variation of relevant earth, moon and IMP–1 satellite position parameters during period December 12–16, 1963. The wake of the moon was detected between 0130 December 14 to 0900 December 15. The time intervals of turbulence and of unusually strong fields are indicated.

to a change in the alignment of the satellite with regard to the Moon and solar wind as shown by the dashed curve in Figure 4. The time at which the satellite is closest to being aligned with the Moon-Sun direction occurs several hours in advance of the non-aberrated case although the distance is approximately the same. On the basis of these data the most probable time for the observation of solar wind interactions with the lunar body would be centered around 0900 December 14, on the assumption that the lunar wake extended in the general flow direction behind the Moon.

The 5.46 minute averages of the magnetic field obtained by this experiment for the time interval 1600 December 13 to 1600 December 15 are shown in Figures 5 and 6. The format for the presentation is identical to that used by Ness et al (1964). The two angles θ and ϕ measured the latitude and longitude respectively of the magnetic field vector in the solar ecliptic coordinate system. In figure 5 the first 9 hours of data show the characteristic behavior of the interplanetary magnetic field with a stable orientation and magnitude. At this time the magnetic field was directed approximately in the plane of the ecliptic ($\theta = 0 \pm 30°$) and pointed back toward the Sun

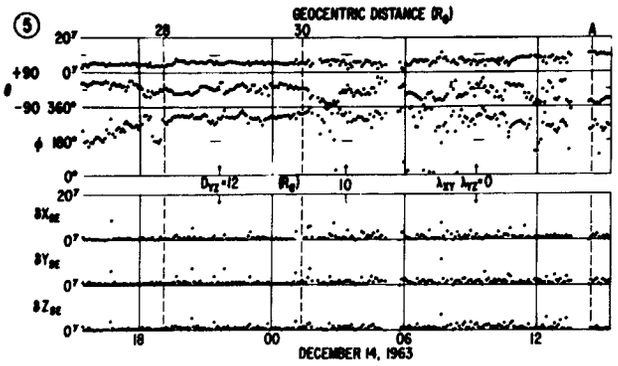


FIGURE 5.—Magnetic field data from orbit No. 5 December 14, 1963, illustrating IMP–1 entry into lunar magnetohydrodynamic wake from relatively undisturbed interplanetary space at 0130.

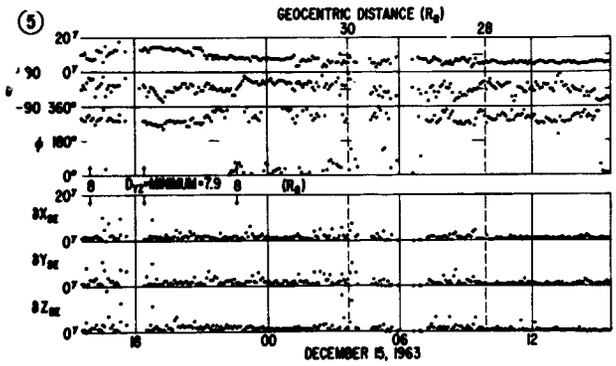


FIGURE 6.—Magnetic field data from orbit No. 5 December 15, 1963, illustrating IMP–1 exit from lunar magnetohydrodynamic wake into relatively undisturbed interplanetary space at 0900 with large field magnitudes observed between 1800 to 2100 December 14.

along the Archimedian spiral angle predicted for the interplanetary field ($\phi = 315° \pm 20°$). During this time the satellite is close to apogee (geocentric distance = 31.7 R_e), and far removed from any direct effects of the solar wind interaction with the Earth. As time progresses the field becomes variable in both magnitude and direction. The onset of this turbulence occurs at approximately 0130 December 14, and persists until 0900 December 15. Midway in this time interval the field magnitude reaches the anomalously large value of about 15 gammas for more than 3 hours, from 1800–2100 December 14, while the turbulent characteristics of the field as measured by the angle ϕ noticeably decreases. Following 0900 December 15 the magnetic field again resumes its characteristic stable configuration although it is not as stable as it was prior to these events.

Superimposed on the time plots are specific values of parameters related to the position of the satellite with respect to the Moon. It should be noted that when the satellite is closest to being in line ($\lambda_{xy} = 0$) the distance D_{yz} is not necessarily at a minimum. Indeed the minimum value of D_{yz} occurs at 1830 December 14 and is approximately the same time as the midpoint of the time interval during which the turbulent magnetic field observations were made. The closest aberrated distance, $D_{y'z} = 9.3\ R_e$, occurs at 0840. The characteristics of the magnetic field during this time interval are unlike those during other occasions on which the satellite was in the interplanetary medium. The turbulence and large field values, as well as the orientation during the time of large field magnitudes suggest that the IMP–1 satellite passed through the turbulent lee wake of the Moon in the solar wind. The strong fields observed correspond to the magnetosphere of the Moon and the turbulent fields to the transition region between the lunar magnetosphere and the undisturbed interplanetary medium.

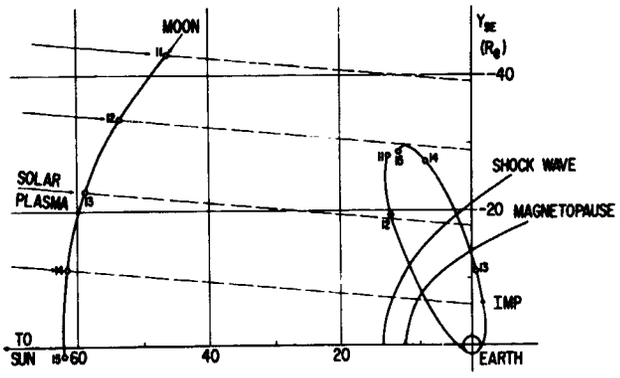


FIGURE 7.—Relative positions of IMP–1 satellite, earth and moon projected on ecliptic plane during period January 11–15, 1964.

In Figure 7 the positions of IMP–1 and the Moon during January 1964 are shown as projected onto the ecliptic plane. This presentation is identical to that for December as shown in Figure 2. Here, however, the relative positions of the satellite and the Moon are seen to be quite different. The time at which IMP–1 is roughly aligned with the Moon occurs at 0730 January 13 but at this time the satellite is enclosed within the Earth's shock wave. Figure 8 presents the relative positions of IMP–1 and the Moon as projected onto

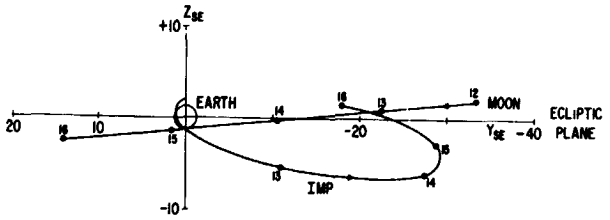


FIGURE 8.—Relative positions of IMP–1 satellite, earth and moon projected on solar ecliptic *Y-Z* plane during period January 12–16, 1964.

the *YZ* plane. The satellite is about in the same position relative to the plane of the ecliptic as in December but the Moon is considerably lower. A summary of the relevant position parameters related to the relative positions of IMP–1 and the Moon is given in Figure 9. The format is identical to that of Figure 4 and illustrates that during the 13th orbit when the satellite was in the lee of the Moon it was also within the region of space surrounding the Earth enclosed by the shock wave surface. The time at which the Earth's shock wave was traversed by the satellite is indicated by the solid vertical line. It occurred several hours after the satellite was aligned with the Moon and the Sun. The situation with respect to the aberrated solar wind direction does not favor the observation of a lunar wake since the satellite is seen to spend a longer time within the region of space associated with the solar wind interaction with the geomagnetic field. Thus the

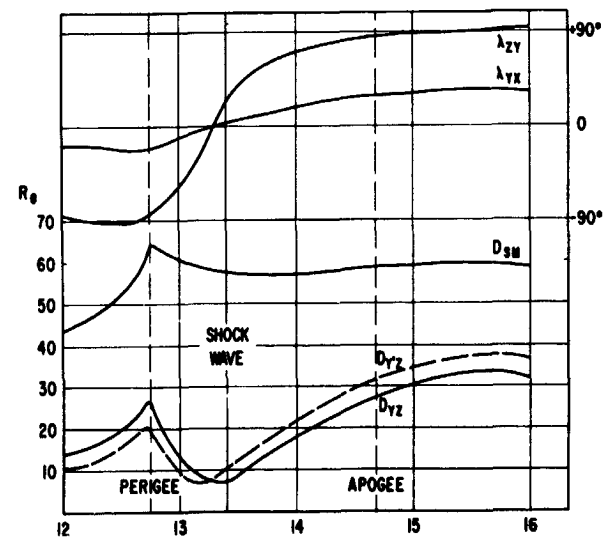


FIGURE 9.—Variation of relevant earth, moon and IMP–1 satellite position parameters during period January 12–16, 1964.

satellite motion in the month of January 1964, was such as to preclude being in a favorable position with respect to observation of the lunar wake. In spite of this the interplanetary magnetic field data for January 13 and 14 has been inspected in detail since the December observation extended over a time period of more than one day. It has not disclosed any characteristics which are suggestive of a lunar wake, either in turbulence or field magnitude.

The last opportunity for observation of the lunar wake with IMP–1 occurred in February 1964 during orbit. 20. On the succeeding opportunities for wake observations during the next six months, the IMP–1 satellite was enclosed completely within the interaction region of the solar wind with the geomagnetic field. The steady precessional motion of the IMP–1 orbit in the solar ecliptic coordinate system is evident by comparing Figure 10 for February 1964 with that of Figure 2 for December 1963. The relative positions

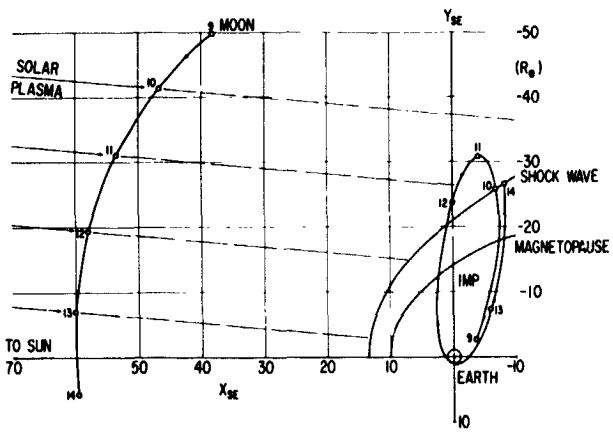


FIGURE 10.—Relative positions of IMP–1 satellite, earth and moon projected on ecliptic plane during period February 9–14, 1964, when a weak wake of the moon was possibly observed.

of IMP–1 and the Moon in February are such that they are approximately aligned at 0020 February 11. The position of these two objects in February projected onto the *YZ* plane is shown in Figure 11. Both the elongation of the IMP–1 orbit relative to Figures 3 and 8 and the location of the Moon below the ecliptic plane are a result of the heliocentric motion of the Earth about the Sun. It can be seen from this figure, however, that the two objects are very close to being in

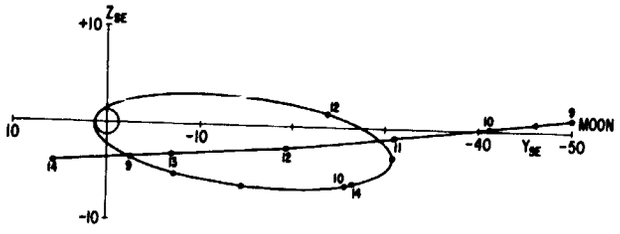


FIGURE 11.—Relative positions of IMP-1 satellite, earth and moon projected on solar ecliptic *Y-Z* plane during period February 9–14, 1964, when a weak wake of the moon was possibly observed.

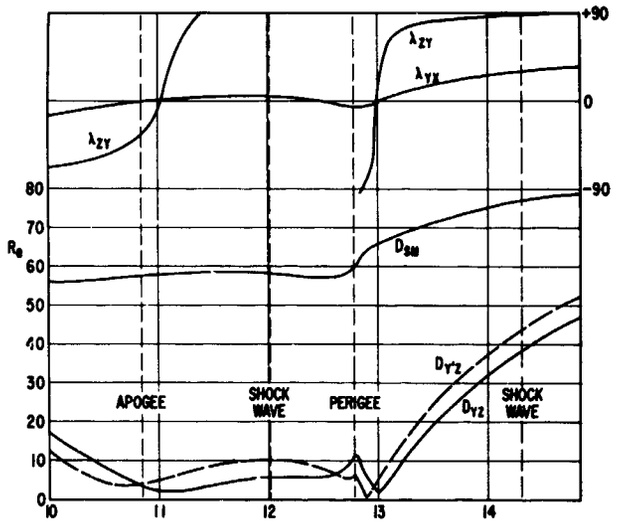


FIGURE 12.—Variation of relevant earth, moon and IMP-1 satellite position parameters during period February 10–15 when a weak wake of the moon was possibly observed between 0400 to 1200 February 10.

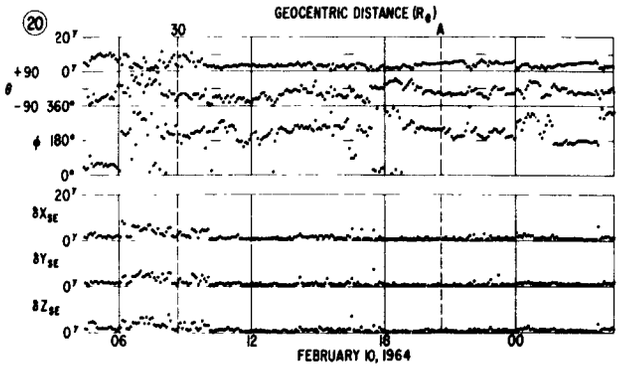


FIGURE 13.—Magnetic field data from orbit No. 20, February 10, 1964, illustrating IMP-1 sampling of the interplanetary magnetic field and possible lunar magnetohydrodynamic wake after exit from transition region of the solar wind interaction with the geomagnetic field. The Earth's shock wave was traversed at a geocentric distance of 30.3 R_e.

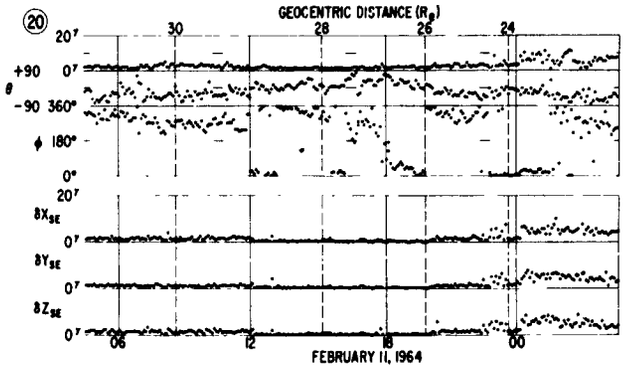


FIGURE 14.—Magnetic field data from orbit No. 20, February 11, 1964, illustrating IMP-1 sampling of the interplanetary magnetic field and possible lunar magnetohydrodynamic wake before entry into transition region of solar wind interaction with geomagnetic field. The earth's shock was traversed at a geocentric distance of 23.6 R_e with a possible "precursor" detected at 24.2 R_e.

precise alignment. Figure 12 summarizes the relative positions of IMP-1 and the Moon as a function of time in a format identical to that used for Figures 4 and 9. The distance from the satellite to the Moon is seen to have increased now to 57.9 R_e while the minimum value for the distance in the *YZ* plane is 1.7 R_e. For the aberrated solar wind the closest distance $D_{y'z}$ of 3.7 R_e occurs at 1640 February 10.

The magnetic field data for the time interval 0430 February 10 to 0430 February 12 is shown in Figure 13 and 14. The data in Figure 13 illustrates the outbound traversal of the shock wave associated with the Earth's magnetosphere at a distance of 30.3 R_e at 0958 February 10. The characteristic of large turbulent fields in the transition region is clearly evidenced. Also shown and to be discussed in a future separate publication on the magnetic field characteristics in the transition region is the compressed and reasonably well aligned magnetic field such as were first observed by Explorer X (Heppner et al, 1963). The abruptness of the shock boundary, as shown by the variance plot, is a persistent feature of the shock phenomena in which the thickness is small and comparable with the ion gyro-radius. Subsequent to the outbound shock wave traversal the magnetic field in interplanetary space is somewhat variable in direction as measured by θ and ϕ but is not turbulent in the same sense as in the transition region as measured by the variance characteristics. However, during the time interval 0400 to 1200 February 11 there is a small

increase above the noise level in the turbulence as measured by the variance. At the same time the variability between successive values of θ and ϕ is increased. However, during the time interval while the variance is higher than the noise level, the field is at its lowest value, less than 3 gammas, during this sample of the interplanetary medium on orbit No. 20.

Figure 14 contains the remainder of the interplanetary magnetic field data observed on orbit 20 and also illustrates the inbound traversal of the Earth's shock wave by the IMP–1 satellite. Again the characteristics of the transition region on the inbound pass are quite similar to those observed on the outbound traversal although they occur at a different geocentric distance, 23.6 R_e, with a suggestion of a "precursor" at 24.3 R_e. The nonspherical symmetry of the shock wave surface enclosing the Earth coupled with the fact that the satellite traverses these boundaries at considerably different satellite-earth-sun angles on successive transits leads to these differences. Throughout the time interval during which the interplanetary medium was sampled, features similar to the December 1963 observation can not be seen. At this time it is suggested that the February 1964 data show characteristics which are similar in some respects but are not as conclusive an evidence of a lunar wake as observed in December 1963.

5.0 FLUID DYNAMIC CONSIDERATIONS

The classical magnetohydrodynamic problem of the interaction of a spherical object with a conducting fluid in the presence of a magnetic field has been treated by Stewartson (1956), Imai (1960) and Ludford and Murray (1960) among others. For the case of either compressible or incompressible fluid flow a downstream wake always develops for large or small magnetic Reynolds numbers as long as the magnetic field strength is small (Sears (1960)). This is measured by the ratio of free stream velocity (V_s) to the pure Alfvén velocity (V_a) or the geometric mean (V_a^*) of the Alfvén and sound velocities. This ratio, Vs/Va or Vs/V_a^*, must be greater than unity in order that a lee wake develop and effectively determines whether or not the presence of the body in the flow field can be communicated upstream by magnetohydrodynamic wave propagation. A major limitation in the quantitative use of the presently available analytical treatments is that they employ restrictive assumptions which are not justified in the case of the interaction of the solar wind with the Moon. In general these include one or all of the following: infinitely conducting fluid, magnetic fields aligned with the flow velocity, nonconducting or infinitely conducting body, and collision mean free paths much smaller than other lengths characteristic of the flow. The case of the Moon in the solar wind satisfies few of these assumptions and thus direct comparisons are not possible. Further development of analytical studies on collisionless magnetohydrodynamic shocks (see review by Kantrowitz and Petschek (1964)) is required for application to the present problem.

In the case of the interaction of the solar wind with the geomagnetic field rather successful results, at least near the stagnation point, have been obtained through use of a pure fluid dynamic analogy. A general review of the use of the aerodynamic similitude and its application to the solar wind interaction with the Earth's magnetic field has been given by Levy et al (1964) and Lees (1964). On this basis an important aspect of the interpretation of the magnetic field data on IMP–1 is that the free stream Mach number can provide a measure of the angular width of the shock surface associated with the interaction. (Michel (1964) has treated this problem assuming complete magnetohydrodynamic flow of the solar wind.)

However, the detailed characteristics within the interaction region and more importantly the wake region may not be completely specified. The Mach number in the case of the solar wind is measured by the ratio of the plasma bulk velocity (V_s) to the magnetohydrodynamic wave phase velocity (V_a). In the interplanetary medium the velocity V_a is approximately given by the Alfvén velocity so that the Mach number is given by

$$M_a = \frac{V_s}{V_a} = \frac{V_s\sqrt{4\pi mn}}{B} \qquad (5.1)$$

where B is the strength of the interplanetary field.

A plot of the Alfvén velocity as a function of the plasma density and magnetic field strength is shown in Figure 15. From classical supersonic

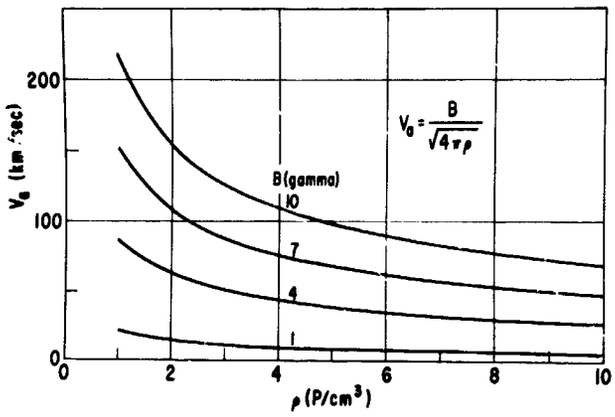


FIGURE 15.—The Alfvén velocity as a function of the interplanetary magnetic field strength and solar plasma flux parameters. The average magnetic field in interplanetary space was 5 gammas with limits of 1 to 10γ.

flow (Hayes and Probstein (1959)) the Mach angle β associated with a specific Mach number is given by

$$\beta = \sin^{-1}\left(\frac{1}{M_a}\right) \tag{5.2}$$

and represents the asymptotic direction of the shock surfaces far downstream from the obstruction. The Mach angle is plotted in Figure 16 as a function of the Mach number.

The average magnetic field in the interplanetary medium has been determined by the IMP–1 measurements to be 5.0 gammas. The average solar wind velocity has been estimated to be approximately 385 Km. Using these average values for the magnetic field and solar wind velocity and

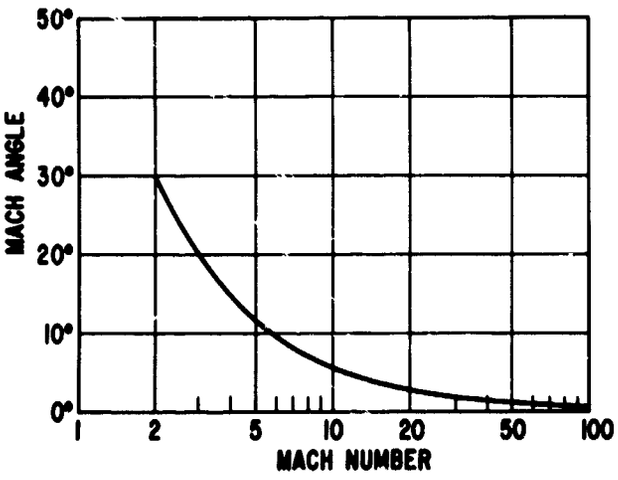


FIGURE 16.—Mach angle as function of the corresponding Alfvén mach number.

assuming a plasma density between 1.5 to 35 p/cm^3 yields a Mach number between 4 and 20. For these two values of Mach number the limits on the Mach angle are 14.5° and 2.8° respectively.

Obayashi (1964), in a review of previous satellite measurements related to the solar wind interaction, has adapted classical hydrodynamic equations for approximating the geometry of the shock surface surrounding a spherical object. Defining:

R_s = the radial distance to the shock from the center of the sphere at the stagnation point

R = the radial distance to the shock surface

ϕ = the angle between R and R_s, then he deduces that:

$$R = R_s\left[\frac{1+\sec\beta}{1+\sec\beta\cos\phi}\right] \tag{5.3}$$

A translation of the position of the IMP–1 satellite to a lunar set of coordinates has been made for the pertinent time periods in December 1963 and February 1964. The position of the satellite in these lunar associated coordinates is

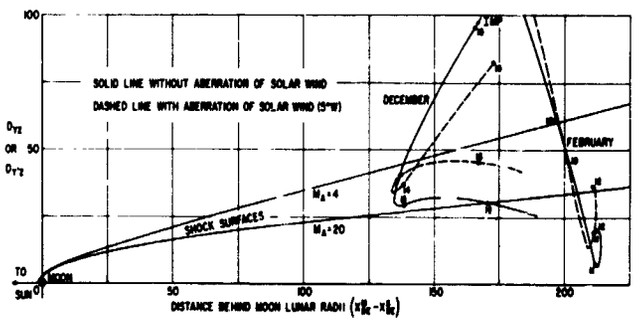


FIGURE 17.—Relative position of IMP–1 satellite and moon in lunar associated corrdinates illustrating envelopment within lunar shock wave assuming aerodynamic similitude during December 1963 and February 1964.

shown in Figure 17 with the abscissa being given by D_{xy} or D_{xy}' where

$$D_{xy} = \sqrt{(D_{sm})^2 - (D_{ys})^2}$$
$$D_{xy}' = \sqrt{(D_{sm})^2 - (D_{y's})^2} \tag{5.4}$$

and the ordinate by D_{yz} or $D_{y'z}$. This presentation essentially assumes that far down stream from the Moon the lunar wake region is cylindrically symmetrical about the solar wind flow direction. Superimposed on these trajectory positions

are the two shock surfaces corresponding to Mach numbers 4 and 20. These surfaces assume that the lunar magnetic field strength is inherently less than a few hundred gammas and that the shock stand off ratio $R_s/R_m=1.25$. It is seen that if the flow is such that the Mach number is less than about 10 then the position of the satellite in December 1963 is well within the shock region of the lunar wake. If the Mach number of the flow is less than about 40 then the satellite is within the shock for February 1964.

The detection of a lunar wake is definitely consistent with supersonic flow theory for the orbit No. 5 observation and with the orbit 20 observation. However, even though the wake was possibly observed in December 1963, the interaction of the solar wind with the Moon is probably so variable that appropriate interplanetary conditions in the vicinity of the Moon are required in order to initiate the formation of a collisionless magnetohydrodynamic shock. Indeed using the average velocity of 385 km/sec and the average field strength of 5 gammas leads to a proton Larmor radius of 830 Km. This implies that on average the shock wave radius assumed for the Moon, 2175 Km, is only 2.6 times this critical value. The stand off distance is about 340 Km so that as previously noted it is somewhat questionable whether a lunar shock can develop that is similar to the case of the Earth. In the case of the Earth the values of the parameters are such that the radius of the shock surface at the stagnation point is more than 100 times the proton Larmor radius. If one considers the field internal to the magnetosphere the radius of the magnetosphere is some 500 times the proton gyro radius since the magnetic field is 50 gammas or greater at the magnetopause boundary.

6.0 *Kp* AND RELATED STUDIES

During the three month interval, December 1963 to February 1964, solar conditions were reasonably undisturbed. A geomagnetic storm observed on December 2, 1963, at 2117 UT was also measured in interplanetary space by the magnetic field experiment (Ness et al, 1964). During the remainder of the interval however, terrestrial magnetic activity as measured by the planetary index K_p was generally low. Because the solar-terrestrial conditions were relatively quiescent those disturbances which were observed and identified as the lunar wake have not been associated with any solar origin that might cause these variable properties of the interplanetary medium.

Recent studies (Bell and Defouw, 1964; Davidson and Martyn, 1964; Michel et al, 1964; Stolov and Cameron, 1964) on K_p have been directed to a resolution of the important question raised by the work of Bigg (1963a, b; 1964) on lunar influences on geomagnetic activity with a synodic period of 29.53 days. The original work was based on the physical supposition that any lunar magnetic field would shield the Earth from propagation of solar particles and thereby affect the level of terrestrial magnetic activity. The specific result of this paper is that a turbulent lunar wake associated with a magnetohydrodynamic interaction of the solar wind with the Moon does exist. If such turbulence in the magnetic field implies equivalent turbulence in the interplanetary plasma it suggests further that fluctuations of magnetic activity could possibly be enhanced at about the time of new Moon.

The statistical significance of the early work has come under severe criticism, although the use of K_p as a quantity obeying the normal laws of algebra is inherently presumed by all participants in these discussions. In spite of this, separate but very similar correlative studies by two groups using identical data have been interpreted quite differently. Stolov and Cameron (1964) have concluded that a lunar effect exists but with the interesting result that the phase corresponds to a full Moon disturbance. Michel et al (1964) on the contrary state that their results are not statistically significant. Bell and Defouw (1964) using a superposed epoch method however concluded a lunar effect at full moon does exist statistically.

Davidson and Martyn (1964) using a superposed epoch and harmonic dial analyses on A_p conclude that no statistically significant effect exists. These authors conducted "a simple statistical experiment" to illustrate their case. A major problem in the analysis of a dimensionless index such as K_p or A_p is whether or not the statistical signifiance tests are based on valid statistical properties of the variable being investigated. In the case of these indices it is by no means clear that definitive answers can be obtained. Davidson

and Martyn (1964) also state that smoothing the data such as done by Bigg can introduce spurious periodicities. This is not true for the weighting function used by Bigg (running average by three's) and is in general not true for most classical numerical filtering procedures. The general situation with regard to statistical studies on K_p or A_p is definitely not yet resolved concerning lunar periodicities. Finally a very recent study by Dodson and Hedeman (1964) indicated an anomalous periodicity of 29.5 days in solar cosmic ray data. This latest result is certain to contribute to our apparent lack of understanding of possible lunar related phenomena.

If the geomagnetic tail extends far behind the Earth to the distance of or beyond the Moon, that is at least 60 R_e, then the Moon will interact with the tail and this may lead to temporal variations in the characteristics of the magnetospheric fields and particle populations independent of solar activity. Indeed such an extended magnetospheric tail has recently been proposed by Dessler (1964). A more pertinent discussion with respect to the magnetic field topology interior to the tail has been given by Axford et al (1964). Detailed studies on the IMP-1 data obtained while in the geomagnetic tail region are presently being conducted to determine the possible lunar effects during March, April and May of 1964, as well as the detailed topological characteristics of the field. On the basis of these IMP-1 results and others obtained on the Explorer X satellite (Heppner et al (1963) and Explorer XIV (Cahill (1964) it appears certain that the geomagnetic field and its tail extend far away from the Earth in an anti-solar direction. Thus, interaction of the Moon with the Earth's magnetic tail can be reasonably anticipated. The studies of Stolov and Cameron (1964) and Bell and Defouw (1964) lend support to such a conclusion.

Howard et al (1964) have suggested that variations in the electron density in the cislunar medium during a solar eclipse are related to a shadowing effect in the lee of the Moon as it interacts with the solar wind. Although the maximum change in solar wind density required to explain the observations is rather high, 100 p/cm^3, there is considerable difficulty in accurately removing changes caused by the ionsphere and magnetosphere. This propagation experiment on lunar associated changes in the cislunar medium may indicate a separate technique for the investigation of the magnetohydrodynamic wake of the moon during the time of new moon.

SUMMARY AND CONCLUSIONS

During the period in which the IMP-1 satellite was located in interplanetary space beyond the direct effects of the solar wind, interaction with the geomagnetic field a unique disturbance was observed. This was characteristic of the transition region and magnetosphere tail associated with the solar wind interaction with the geomagnetic field. This temporal disturbance showed turbulent magnetic fields and enhanced magnetic flux values well above the normal interplanetary condition. This event is identified as the magnetohydrodynamic wake of the Moon's interaction with the solar wind. The use of an aerodynamic analogy, which has proven so successful in studying the Earth's interaction with the solar wind, shows that the satellite was within the primary shock region associated with a super Alfvénic solar wind flow. A second opportunity to investigate the lunar wake in January 1964 was obscured by the location of the satellite within the shock region of the solar wind interaction with the geomagnetic field. A third opportunity in February 1964, possibly detected a wake but it did not demonstrate the same characteristic features. At this time the satellite was some 200 lunar radii from the Moon.

The great distance at which the lunar wake is observed and the lack of quantitative descriptions regarding the Earth's magnetic tail preclude an estimate of the strength of any lunar magnetic field. However, the fact that a lunar wake was observed during December 1963 is suggestive of a lunar magnetic field at the equatorial surface of several hundred gammas or more.

A major question raised by this paper is whether or not a shock wave due to the moon is expected to develop at all times in the interplanetary medium unless an intrinsic lunar magnetic field is present. It would appear that the absence of a strong field, and the small size of the Moon preclude a condition analogous to that in the case of the Earth in which a shock wave is usually expected to occur. The Larmor radius of the average solar wind proton is of the same order of

magnitude as the lunar radius and thus the analogy with fluid dynamics is probably not justified.

The Russian measurements on the Lunik II spacecraft impacting the lunar surface indicate that the magnetic field of the Moon is not sufficiently strong to produce field strengths larger than 100 gammas at a distance of 55 Km. This can be understood on the basis of a compression of a lunar magnetic field by the solar wind. However, at the same time that this data was obtained scintillation counter data (Vernon et al, 1960) obtained at a distance of less than 1000 Km from the lunar surface indicated energetic particle flux values 100 times those found typically in interplanetary space. This value of 1000 Km is tantalizingly close to the lunar shock wave stand off distance. Indeed the magnetic field data for distances less than 1000 Km are also larger and more variable also. This may indicate that shock region characteristics similar to that observed surrounding the Earth may have been present surrounding the Moon at that time.

The fact that the Earth's magnetospheric tail does extend far behind the Earth may well lead to lunar associated variations of K_p or A_p at full moon as the Moon sweeps through the Earth's tail.

ACKNOWLEDGEMENTS

I am indebted to C. S. Scearce and J. B. Seek for their important participation in the IMP–1 magnetic field experiment. Stimulating discussions of the lunar wake with Drs. E. K. Bigg, A. G. W. Cameron and T. Obayashi are hereby acknowledged. The comments of Dr. J. M. Wilcox are particularly appreciated.

REFERENCES

Anderson, K. A., H. K. Harris and R. J. Paoli, Energetic Electron fluxes in and beyond the Earth's Outer Magnetosphere, preprint, 1964.

Axford, W. I., The interaction between the solar wind and the Earth's magnetosphere, J. Geophys. Res., 67, 3791–3796, 1962.

Axford, W. I., H. E. Petschek and G. L. Siscoe, The Tail of the Magnetosphere, AVCO-Everett Research Report R190, August, 1964.

Beard, D. B., The interaction of the terrestrial magnetic field with the solar corpuscular radiation, J. Geophys. Res., 65, 3559–3658, 1960.

Bell, B. and R. J. Defouw, Concerning a Lunar Modulation of Geomagnetic Activity, J. Geophys. Res., 69, 3169–3174, 1964.

Bernstein, W., R. W. Fredricks, J. L. Vogl and William A. Fowler, The Lunar Atmosphere and the Solar Wind, Icarus 2, 233–248, 1963.

Biermann, L., Kometenschweife and Solare Korpuscular Strahlung, Z. Astrophys. 29, 274–286, 1951.

Bigg, E. K., The influence of the Moon on geomagnetic disturbances, J. Geophys. Res. 68, 1409–1413, 1963a.

Bigg, E. K., Lunar and Planetary Influences on Geomagnetic Disturbances, J. Geophys. Res., 68, 4099–4104, 1963b.

Bigg, E. K., Lunar Influences on the Frequency of Magnetic Storms, preprint, 1964.

Bonetti, A., H. S. Bridge, A. J. Lazarus, B. Rossi and F. Scherb, Explorer 10 plasma measurements, J. Geophys. Res., 68, 4017–4063, 1963.

Bridge, H., A. Egidi, A. Lazarus, E. Lyon and L. Jacobson, Preliminary results of plasma measurements on IMP-A, COSPAR presentation, Florence, Italy, 1964.

Cahill, L. J., and R. G. Amazeen, The Boundary of the Geomagnetic Field, J. Geophys. Res., 68, 1835–1844, 1963.

Cahill, L. J., Preliminary Results of Magnetic Field Measurements in the tail of the Geomagnetic cavity, IGY Bulletin 7b, 1964

Chapman, S., Solar plasma, geomagnetism and aurora in geophysics, in *The Earth's Environment*, Les Houches 1962 Lectures, edited by C. Dewitt et al, pp. 371–502, Gordon and Breach, New York, 1963.

Chapman, S. and J. Bartels, *Geomagnetism*, Clarendon Press, Oxford, 1940.

Davidson, T. W. and D. F. Martyn, A Supposed Dependence of Geomagnetic Storminess of Lunar Phase, J. Geophys. Res., 69, 3973–3979, 1964.

Dessler, A. J., Length of the Magnetospheric Tail, J. Geophys. Res., 69, 3913–3918, 1964.

Dodson, Helen W. and E. Ruth Hedeman, An unexpected effect in Solar Cosmic Ray Data related to 29.5 Days, J. Geophys. Res., 69, 3965–3971, 1964.

Dolginov, S. H., E. G. Eroshenko, L. N. Zhugov, N. V. Pushkov and L. O. Tynrmina, Investigation of the Magnetic Field of the Moon, Geomagnetism and Aeronomy, 1, 21–29, 1961.

Doodson, A. T., Oceanic tides, *Advan. in Geophys.*, 5, 118–152, Academic Press, New York, 1958.

Elsasser, W. M., The Earth's Interior and Geomagnetism, Revs. Mod. Phys., 22, 1–35, 1950.

Fejer, J. A., Atmospheric tides and associated magnetic effects, Rev. of Geophysics, 2, 275–310. 1964.

Freeman, J. W., J. A., Van Allen and L. J., Cahill, Explorer 12 observations of the magnetosphere boundary and the associated solar plasma on Sept. 13, 1961, J. Geophys. Res., 68, 2121–2130, 1963.

Freeman, J. W., Electron distribution in the outer radiation zone, J. Geophys. Res., 69, 1691–1724, 1964.

Gold, T., The magnetosphere of the Moon, Cornell Univ., Report CTSR–177, 1964.

Gringauz, K. I., Some results of experiments in interplanetary space by means of charged particle traps on Soviet space probes, *Space Research II*, 539–553, North-Holland, Amsterdam, 1961.

HAPKE, B. W., Effect of the Solar Wind on the photometric properties of the lunar surface, Trans. A. G. U., 45, 347, 1964.

HARRISON, E. R., Solar Corpuscular radiation and the nature of the Earth's Immediate Interplanetary environment, Proc. Int. School of Physics "Enrico Fermi", Course XIX, 65–84, edited by B. Peters, Academic Press, New York, 1963.

HAYES, W. D. and R. F. Probstein, Hypersonic Flow theory, Academic Press, New York, 1959.

HEPPNER, J. P., N. F. NESS, T. L. SKILLMAN and C. S. SCEARCE, Explorer X Magnetic Field Measurements, J. Geophys. Res., 68, 1–46, 1963.

HIDA, K, An approximate study on the detached shock wave in front of a circular cylinder and a sphere, J. Phys. Soc., Japan, 8, 740–745, 1953.

HINTON, F. L. and D. R. TAEUSCH, Variation of the lunar atmosphere with the strength of the solar wind, J. Geophys. Res., 69, 1341–1347, 1964.

HOWARD, H. T., B. B. LUSIGNAN, P. YOH and V. R. ESHLEMAN, Radar Doppler and Faraday Polarization Measurements of the Cislunar Medium during the July 20, 1963 Solar Eclipse, J. Geophys. Res, 69, 540–544, 1964.

IMAI, I., On Flows of Conducting Fluids past Bodies, Revs. Mod. Phys., 32, 992–999, 1960.

INGLIS, D. R., Theories of the Earth's Magnetism, Revs. Mod. Phys., 27, 212–248, 1955.

JOHNSON, F. S., The gross character of the geomagnetic field in the solar wind, J. Geophys. Res., 65, 3049–3051, 1960.

KANTROWITZ, A. and H. E. PETSCHEK, MHD Characteristics and Shock Waves, AVCO-Everett Research Report R185, July, 1964.

KELLOGG, P. J., Flow of Plasma around the earth, J. Geophys. Res., 67, 3805–3811, 1962.

KOPAL, Z. and T. W. RACKHAM, Excitation of Lunar luminescence by Solar flares, Nature, 201 (4916), 239–241, 1964.

LEES, L., Interaction between the solar plasma wind and the geomagnetic cavity, AIAA preprint 64–88, New York, 1964.

LEES, L. and L. HROMAS, Turbulent diffusion in the wake of a blunt nosed body at hypersonic speeds, J. Aero. Sci., 29, 976–993, 1962.

LEVIN, B., Thermal History of the Moon, in *The Moon* edited by Kopal and Mikhailov, pp. 157–167, Academic Press, New York, 1960.

LEVY, R. H., H. E. PETSCHEK and G. L. SISCOE, Aerodynamic Aspects of the Magnetosphere Flow, AVCO-Everett-R170, 1964.

LUDFORD, G S. S. and J. D. MURRAY, On the flow of a conducting fluid past a magnetized sphere, J. Fl. Mech., 7, 516–528, 1960.

MACDONALD, G. J. F., Interior of the Moon, Science, 133, 1045–1050, 1961.

MACDONALD, G. J. F., Internal constitution of the inner planets and the Moon, Space Science Review, 2, 473-557, 1963.

MAROCHNIK, L. S., The Interaction of Solar Corpuscular Streams with the Atmospheres of Comets, Geomagnetism and Aeronomy, 3, 608–615, 1963.

MELCHIOR, PAUL, Earth Tides, *Advan in Geophys. 4*, 392–445, Academic Press, New York, 1957.

MICHEL, F. G., A. J. DESSLER and G. K. WALTERS, A Search for Correlation Between K_p and the Lunar Phase, J Geophys. Res., 69, 4177–4181, 1964.

MICHEL, F. G., Detectability of disturbances in the Solar Wind, J. Geophys. Res., to appear, 1964.

NESS, N. F., C. S. SCEARCE, J. B. SEEK, Initial results of the IMP-1 Magnetic Field Experiment, J. Geophys. Res., 69, 3531–3570, 1964.

NESS, N. F. and J. M. WILCOX, The solar origin of the Interplanetary Magnetic Field, Phys. Rev. Letters, 13 (15), 461–4, 1964.

NEUGEBAUER, M., Question of the existence of a lunar magnetic field, Phys. Rev. Letters, 4 (1), 6–8, 1960.

OBAYASHI, T., Interaction of Solar Plasma Streams with the outer Geomagnetic field, J. Geophys. Res., 69, 861–868, 1964.

PARKER, E. N., Dynamics of the interplanetary gas and magnetic field, Astrophys. J., 128, 664–676, 1958.

PARKER, E. N., The hydrodynamic theory of solar corpuscular radiation and stellar winds, Astrophys. J., 132, 821–866, 1960.

PARKER, E. N., Interplanetary Dynamical Processes, pp. 131, Interscience, New York 1963.

PIDDINGTON, J. H., Geomagnetic storm theory, J. Geophys, Res., 65, 93–105, 1960.

SEARS, W. R., Some remarks about flow past bodies, Revs. Mod. Phys., 32, 701–705, 1960.

SIEBERT, M., Atmosphere tides, Advan. in Geophys. 7, 105–187, Academic Press, New York, 1960.

SNYDER, C. and NEUGEBAUER, Interplanetary solar wind measurements by Mariner 2, Space Research IV, edited by P. Muller, North-Holland Publ. Co., Amsterdam, 1960.

SPREITER, J. R. and W. P. JONES, On the effect of a weak interplanetary magnetic field on the interaction between the solar wind and the geomagnetic field, J. Geophys. Res., 68, 3555–3565, 1963.

STEWARTSON, K., Motion of a sphere through a conducting fluid in the presence of a strong magnetic field, Proc. Camb. Phil. Soc., 52, 301–316, 1956.

STOLOV, H. L. and A. G. W. CAMERON, Variations of Geomagnetic Activity with Lunar Phase, J. Geophys. Res., to appear, 1964.

VERNOV, S. N., A. E. CHUDAKOV, P. V. VALUKOV, YU. I. LOGACHEV and A. G. NIKOLAYEV, Radiation measurements during the flight of the second Moon Rocket, Dokl. Acad. Nauk SSSR, 130, 517–520, 1960.

VESTINE, E. H., Utilization of a Moon-Rocket System for Measurement of the Lunar Magnetic Field, RAND report RM–1933, 1957.

WEHNER, G. K., C. E. KENKNIGHT and D. ROSENBERG, Modification of the lunar surface by the solar-wind bombardment, Planet. Space Sci., 11, 1257–1261, 1963.

ZHARKOV, V. N. and F. R. ULINICH, Can the Moon sustain a magnetic field by hydromagnetic amplification? Proceedings of Geophysical Institute of O. Yn. Shmidt No. 11 (178), 50–55, 1960. (Translated for NSF 1963).

THE MAGNETOSPHERE AND ITS BOUNDARY LAYER*

NORMAN F. NESS

The sun is an ubiquitous source of the Earth's energy input. More than 30 years ago in an attempt to explain the characteristic storm-time fluctuations of the geomagnetic field, a solar origin was postulated by Chapman and Ferraro (1933). It was suggested that the sun, at times of solar disturbances such as flares, emitted a neutral but ionized gas referred to as a plasma. When this plasma reached the earth it compressed the earth's magnetic field and contained it in a region of space surrounding the earth. The cavity in the solar plasma thus formed has been termed Chapman-Ferraro cavity and the mechanism of its formation has been reasonably successful in explaining the temporal characteristics of various geomagnetic sudden commencement and other storm phenomenon. A naive representation of the interaction of the solar plasma with the earth's magnetic field is shown in Figure 1. Here the individual particles are assumed to be specularly reflected at the boundary of the earth's magnetic field. The region within the boundary the geomagnetic cavity has been referred to as the *magnetosphere* since the dominant factor influencing charged particle motion within this region of space is the earth's magnetic field. In the remainder of this paper the terminology will utilize magnetosphere rather than the Chapman-Ferraro cavity.

In an attempt to explain the fluctuations and characteristics of type I cometary tails, Biermann (1951) early in the 1950's suggested that a continual flux of solar plasma was required. This was similar to that postulated by Chapman and Ferraro (1933) in their theoretical studies. Subsequent to this, Parker (1958) developed his hydrodynamic theory of the expansion of the solar corona referring to the phenomena as the "solar wind." This was predicted to consist of ionized

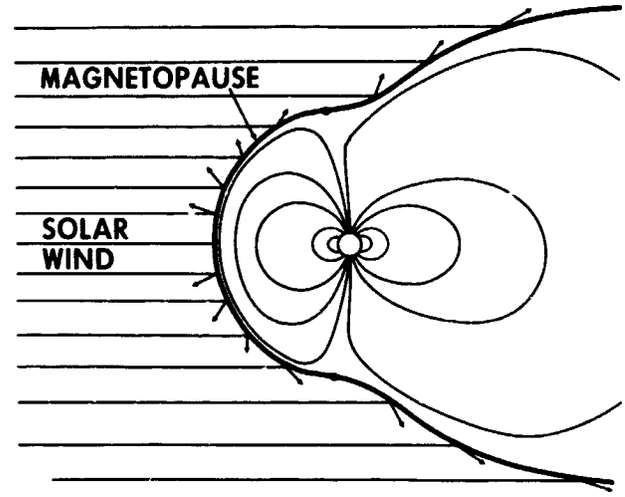


FIGURE 1.—Naive representation of the interaction of the solar plasma with the geomagnetic field. Direct impact of the plasma with the magnetic field is shown as being specularly reflected from the geomagnetic or Chapman-Ferraro boundary. The distance to the boundary at the subsolar point on this basis is given by $R_c = \left[\frac{B_0^2}{4\pi m n V_s^2}\right]^{1/6} R_e$ where R_e is the radius of the Earth, B_0 the equatorial magnetic field strength, V_s the velocity of the solar plasma and n, the plasma density (m being proton mass). See Figure 8.

gas with the principal constituent being hydrogen and flowing radially from the sun with flux values of 10^7 to 10^{10} particles/cm^2/sec. The energy of the particles was assumed to be approximately 1 Kev. Direct measurements of this solar wind or plasma have recently been performed by means of satellite measurements conducted both by this country (Bonetti et al, 1963; Snyder and Neugebauer, 1963) and by the USSR (Gringauz, 1962). As a part of the overall NASA program investigating the characteristics of the interplanetary medium on a continuing basis, a series of interplanetary Explorer satellites has been developed. Figure 2 presents a photograph of the IMP-1 satellite, the first Interplanetary Monitoring Platform in this series which was successfully launched

*Published as *Goddard Space Flight Center Document X-612-64-356*, November 1964.

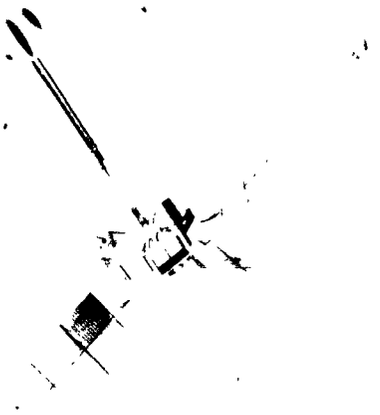


FIGURE 2.—Photograph of the first Interplanetary Monitoring Platform, IMP–1, launched Nobemver 27, 1963. The unique appendages extending from the spacecraft octagonal body support magnetometers at remote distances so that the magnetic fields of the electronic components do not contaminate the low field measurements. The satellite weighs 140 lbs and measures 14 ft from tip to tip of the fluxgate magnetometer booms.

IMP EXPERIMENT REPERTOIRE

#	Study	Experiment	Characteristics
1	Cosmic Rays	Range / Energy Loss	100 kev < P < 200 mev
2	Cosmic Rays	Total Energy / Energy Loss	Energy, Charge Spectra
3.	Cosmic Rays	Neher Ionization Chamber	Total Ionization
4	Cosmic Rays	Orthogonal Geiger–Counter Telescope	Spatial Isotropy CR Events
5	Magnetic Fields	Rubidium Vapor Scaler Mag.	$\|\vec{B}\|$ < 2000 Gammas
6.	Magnetic Fields	Fluxgate Vector Sensor Mag.	$\vec{B}$ < 40 Gammas
7.	Solar Wind	Proton Flux — Electrostatic	200 ev < P < 20 kev
8.	Solar Wind	Proton Flux — Faraday Cup	10 k/s < V_p < 1000 k/s
9.	Solar Wind	Thermal Ion–Electrons (Charged–Particle Trap)	few ev < electrons, ions

FIGURE 3.—Experiment repetoire on the IMP–1 satellite. The separate experiments were provided by both NASA and University laboratories including: Goddard Space Flight Center, Ames Research Center and the Universities of Chicago, California and MIT. Details of the various energy ranges and sensor techniques are included.

November 27, 1963. It transmitted information on the characteristics of magnetic fields, plasmas and energetic particles in the region surrounding the earth for a period of more than six months. The apogee of the satellite was 31.7 R_e (earth radii) or 197,616 km with an orbital period of 93 hours. The interaction of the solar wind with the earth's magnetic field leads to a distortion of the earth's magnetic field as well as creating a disturbance in the flow field of the solar wind. This paper is concerned principally with the distortion of the earth's magnetic field and the resultant boundary layer region between the magnetosphere and the undisturbed interplanetary medium as measured by the IMP–1 satellite.

A broad compliment of experiments in the measurement of energetic particles, low energy plasmas and magnetic fields was instrumented for flight on the IMP–1 satellite. Figure 3 presents a summary of the various instruments, their measurement range and energy characteristics. Figure 4 presents the solar-ecliptic coordinate system appropriate for studying the interaction of the solar wind with the earth's magnetic field. In this coordinate system the X-axis is directed at all times from the earth's center to the sun, the Z-axis is chosen to be perpendicular to the ecliptic plane and the Y-axis forms a right handed co-

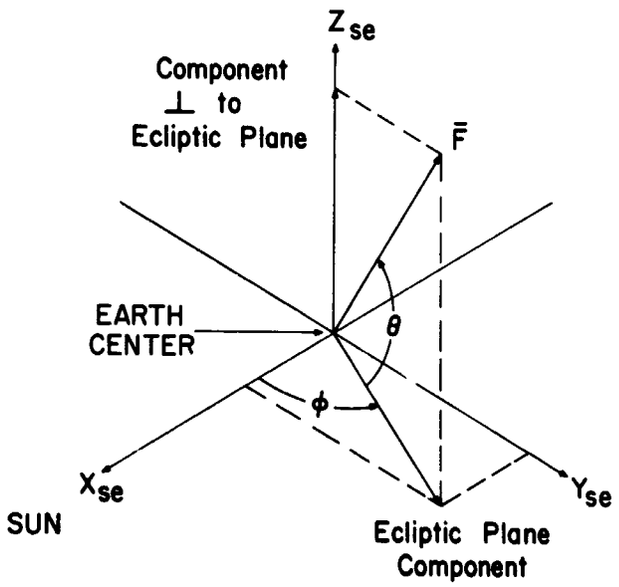


FIGURE 4.—Illustration of the solar-ecliptic coordinate system employed to study the characteristics of the interaction of the solar wind with the geomagnetic field. See text for an explanation of the specific parameters X_{se}, Y_{se}, Z_{se}, θ and ϕ.

ordinate system. In addition two angles are defined to represent a vector field: θ being the latitude, positive about the plane of the ecliptic and negative below and Φ the longitude, being 0° directed to the sun and 180° when pointed away from the sun. The characteristics of the highly eccentric IMP–1 orbit are shown in Figure 5 as projected on the ecliptic plane. The first four

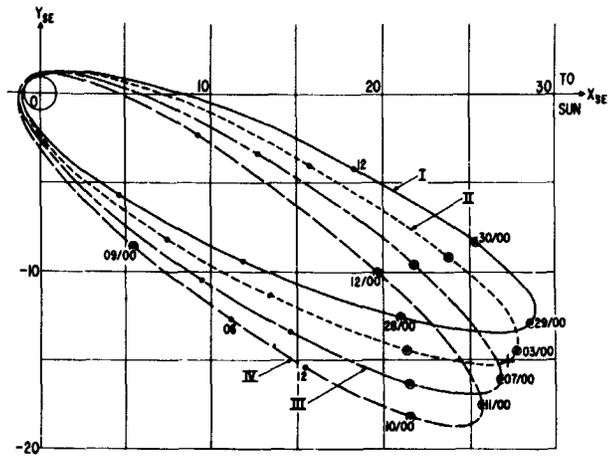


FIGURE 5.—Projection of the first four orbits of IMP–1 satellite as viewed on the $X_{ss}-Y_{ss}$ plane, the ecliptic plane. The values adjacent each trajectory orbit refer to the date and/or time in hours when the satellite was located at the indicated points.

orbits are shown with the figures adjacent to the trajectory indicating the time at which the satellite was at a particular position in space. Upon inspection of this figure it is noted that for approximately 60 percent of each orbit the satellite is well beyond 20 R_e distance from the center of the earth. A corresponding view of the orbit projected on a plane perpendicular to the ecliptic plane is shown in Figure 6. It is seen from these two figures that the orbit of IMP–1 is a very elongated ellipse. This paper shall utilize the experimental results obtained from the magnetometers (Ness et al, 1964) and the plasma probe

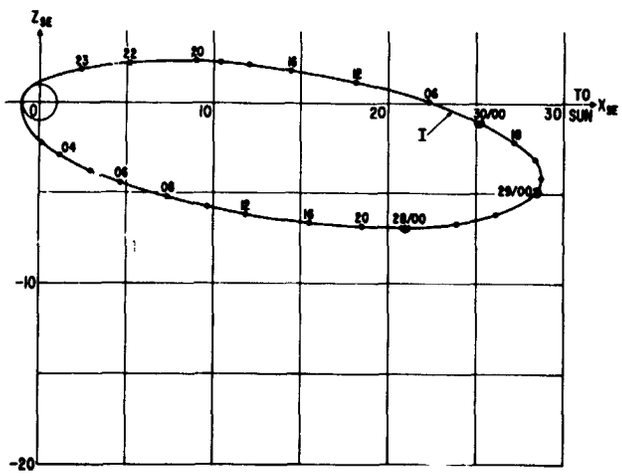


FIGURE 6.—Projection of the first orbit of the IMP–1 satellite as viewed on the $X_{ss}-Z_{ss}$ plane. The positive Z_{ss} axis points towards the north ecliptic pole.

(Bridge et al, 1964) to illustrate the characteristics of the magnetosphere and its boundary region.

The results of the magnetic field measurements on the inbound portion of orbit 1 are shown in Figure 7. The experimental data is presented as

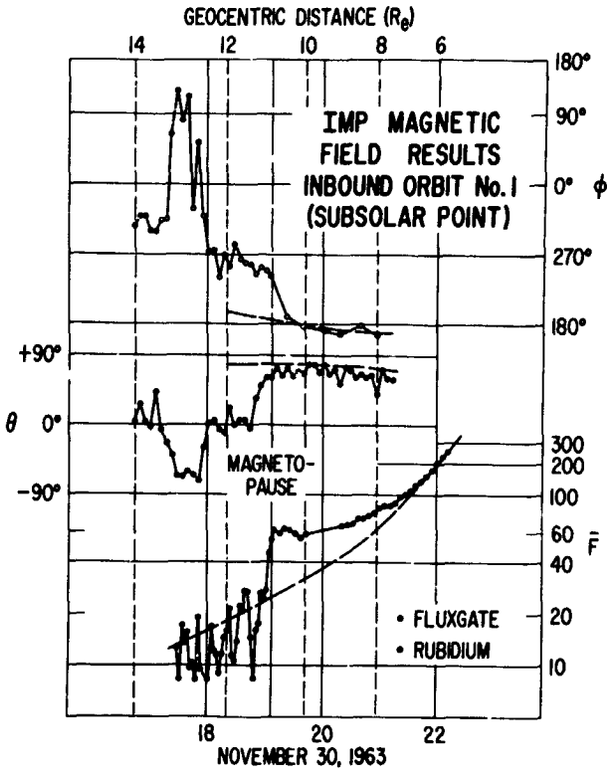


FIGURE 7.—Magnetic field measurements of the boundary of the magnetosphere from IMP–1 inbound orbit No. 1. The abrupt discontinuity in magnitude and direction of the field at 10.7 R_e is identified as the boundary. Theoretical values for F, θ and ϕ are shown as dashed curves and do not include any compression of the earth's field by the solar plasma.

a magnitude F and two angles θ and Φ. Each data point represents the average of the vector magnetic field over a time interval of 5.46 minutes. The satellite is moving approximately 2 km/sec in this region of space so that over this time scale the satellite traverses a radial distance of approximately 660 kilometers. The measurements at geocentric distances beyond 10.7 R_e are seen to be highly variable in both magnitude and direction of the magnetic field. However, at a distance of 10.7 R_e the magnetic field abruptly increases in magnitude to a value of 60 gammas and assumes a stable configuration. The theoretical magnetic field to be measured in space, extrapolated by spherical harmonic analysis from

surface measurements, is shown as dashed lines in this figure. The abruptness in both magnitude and direction as well as the temporal characteristics of the magnetic field at and beyond 10.7 R_e are identified as the boundary of the magnetosphere. It is seen that the observed magnitude is approximately twice that which would be theoretically predicted by considering the magnetic field in space to be only that due to the earth's magnetic field.

The containment of the earth's field by the solar plasma essentially doubles the magnetic field strength at the boundary surface. This can be understood simply by viewing the plasma impacting the geomagnetic field as being represented by a plane boundary across which the normal component of magnetic field must be zero. This is related to the phenomenon that in a highly conducting plasma, such as the solar wind, the magnetic field is "frozen into" the plasma motion. Hence as a plasma stream interacts with a magnetic field it does so by compressing the lines of force ahead of it. Mathematically this can be represented by placing an image dipole on the sunward side of the earth at an equal distance from the boundary and thus the normal component of the field is zero. This can explain both the magnitude increase of the observed magnetic field as well as the preservation of its direction. This simplified viewpoint is not completely correct but is substantially valid as long as the discussion refers only to the boundary near the subsolar or stagnation point.

Utilizing this simplified theoretical model of the solar plasma directly impacting the earth's magnetic field permits an interpretation of the solar stream properties on the basis of the size of the earth's magnetosphere. On the assumption that the subsolar radial distance to the magnetosphere boundary is 10.7 R_e it is seen in Figure 8 the plasma density ranges from 1 to 10 protons per cubic centimeter for velocities between 200 and 600 km/sec. These plasma values are representative of those which have been measured on previous satellites and space probes. Thus, it would appear that the general characteristics of the bounding of the earth's magnetic field by the solar plasma and the distance at which it occurs can be reasonably well understood on the elementary individual particle basis.

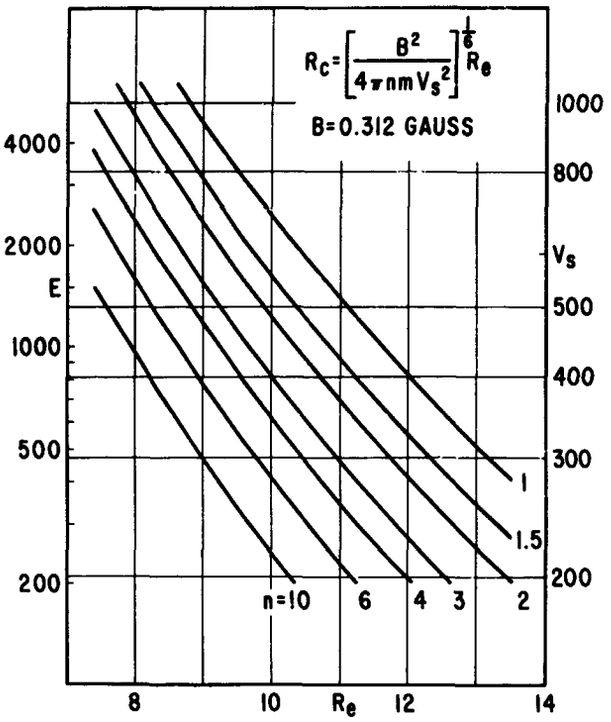


FIGURE 8.—Theoretical size of the magnetosphere at the subsolar point assuming normal impact of the solar plasma on the geomagnetic field. For a magnetosphere of 10.7 R_e and assumed velocity of 400 km/sec the deduced plasma density is found to be 2 p/cm^3.

However, this is not the complete story on the characteristics of the magnetosphere and its boundary region. A correlated set of data from the MIT plasma probe and the GSFC magnetic field experiment is shown in Figure 9. This included the same interval shown in greater detail in Figure 7. The important feature of the MIT plasma detector, a Faraday cup, is that it is directionally sensitive to the flow of plasma. As the satellite rotates, the acceptance aperture of the detector scans the celestral sphere and includes orientations directly toward and away from the sun. Shown on this figure are the plasma flux values when the detector is pointed almost directly toward the sun and directly away from the sun. The difference between these two measurements is a measure of the anisotropy of the plasma flow. It is seen that at very large distances from the earth, greater than 16 R_e, the plasma flow is principally from the sun. However, at a distance of 13.6 R_e the flow of plasma suddenly comes apparently from all directions, i.e. the flux is isotropic. It is also at this point that the fluctuations in the magnetic

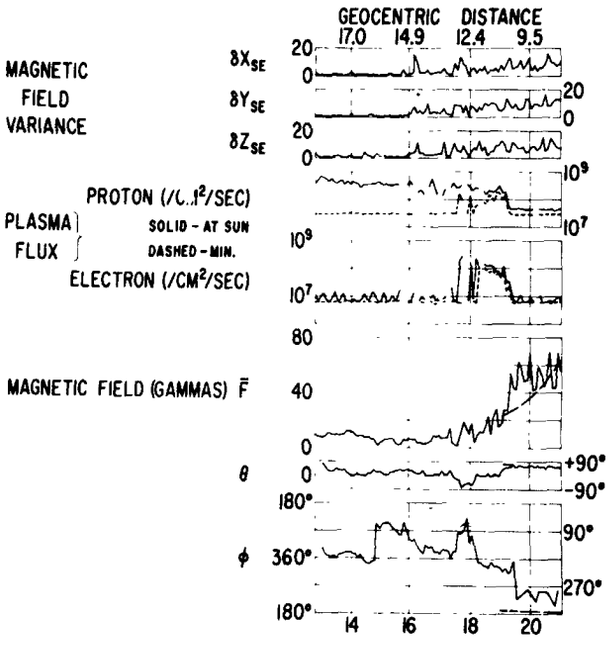


FIGURE 9.—Time correlated MIT plasma and GSFC magnetic field measurements on IMP–1 inbound orbit 1. The root-mean-squared deviation or variance estimate is of the magnetic field is shown as SX_{se}, SY_{se} and SZ_{se}. The plasma flux values in both a solar and antisolar direction are shown for the plasma probe. Isotropic fluxes immediately outside the magnetosphere boundary terminate at 13.7 R_e. Directional fluxes are observed beyond this boundary, which is identified as a collisionless magnetohydodynamic shock wave.

field increase appreciably. This is measured by the root-mean-squared deviations shown in the top most three curves illustrating the Z, Y, Z, components of the deviation of the magnetic field over the 5.46 minute time intervals. The fluctuating magnetic field and the isotropic plasma are observed until the distance at which the magnetic field abruptly increases to a very large value. This region of space surrounding the earth's magnetosphere in which a thermalized or isotropic plasma flux is observed to be correlated with fluctuating magnetic fields is termed the transition region of the magnetosphere boundary layer. It is this boundary layer which this paper studies, discusses its characteristics and attempts to present the current concepts related to its formation.

The first measurements clearly suggesting a continual containment of the earth's magnetic field were provided on the leeward side of the solar wind plasma flow by the Explorer X satellite in March, 1961 (Heppner et al, 1963). Over an interval of 48 hours the magnetic field and plasma was observed in a characteristic pattern in which strong fields directed radially from the earth were exchanged with periods during which radial plasma flow from the sun and fluctuating magnetic fields were observed. Conclusive experimental evidence for the bounding of the geomagnetic field by the solar wind was provided by the Explorer XII satellite measurements of the magnetic field and trapped particle fluxes as reported by Cahill and Amazeen (1963) and Freeman, Van Allen and Cahill (1963). Subsequent to the Explorer XII, the Explorer XIV satellite provided additional information on these characteristics. Thus far only limited summaries of the magnetic field in these regions have appeared although detailed discussions of the particle flux measurements have been presented in numerous articles. The plasma probes on board the Explorer XII and XIV did not reveal the isotropic fluxes observed on IMP–1.

Although the purpose of the IMP–1 satellite was primarily to investigate the characteristics of the interplanetary medium, the fact that the satellite is gravitationally anchored to the earth implies a traversal of the magnetosphere boundary region twice each orbit. The results of the IMP–1 satellite obtained in these traversals have substantially confirmed and extended our knowledge of the magnetosphere boundary layer. Our overall interpretation of the results is based upon an analogy with high speed aerodynamic flow. It is assumed that the magnetosphere acts as a blunt body which deflects the flow of the solar plasma. An important aspect of the rarefied solar plasma flow is that it contains a magnetic field. The average interplanetary magnetic field value has been accurately established by the IMP–1 satellite (Ness and Wilcox, 1964). For the average solar proton of 1 Kev energy this leads to a Larmor radius of approximately 500 kilometers. This small characteristic length permits the use of a fluid continuum approximation. This is also about the spatial resolution with which the boundaries of both the magnetosphere and the transition region are sampled. In this magnetized plasma the propagation of disturbances is by magnetohydrodynamic waves as contrasted to the supersonic gas dynamic case when propagation is by acoustic waves. The appropriate propagation velocity, the Alfvén mode is presented

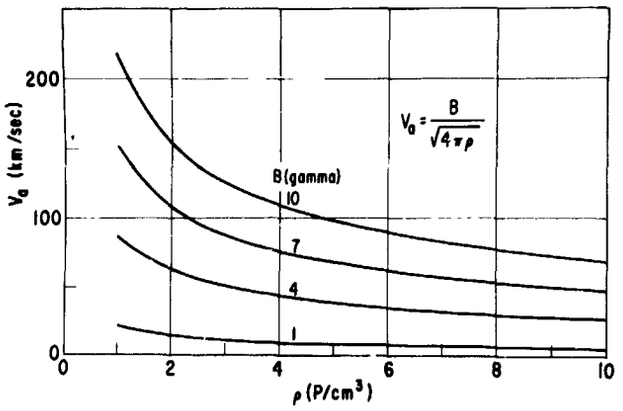


FIGURE 10.—The characteristic velocity of propagation of disturbances in the interplanetary medium is the Alfven velocity. This magnetohydrodynamic mode is shown as a function of magnetic field strength B and plasma density, ρ.

in figure 10 as a function of plasma density and magnetic field strength. The important feature of this diagram is that it shows that for the interplanetary medium the Alfvén velocity is characteristically less than 100 Km/sec. The estimated velocity of the solar plasma is 385 Km/sec from the interpretations of solar magnetic fields and the interplanetary magnetic field (Ness and Wilcox, 1964). Thus, the flow of the solar wind is supersonic in the magnetohydrodynamic sense. Actually the flow is hypersonic since the equivalent Mach number or more appropriately the Alfvén number is greater than 4. Under such conditions the well known phenomenon of a detached shock wave develops in the gas dynamic case which encloses the disturbing body in a region of space with a boundary across which discontinuous changes in parameter values occur. At the present time the detailed quantitative study of the physical properties of the boundaries as observed by the IMP-1 satellite have yet to be completed. A particular limitation to their detailed study will be the spatial and time resolution limitations inherent in the spacecraft orbit and telemetry system. The detached shock wave which is observed in gas dynamics has characteristics which closely resemble that of the earth's magnetic field interacting with the flow of the solar wind. The termination of the turbulent transition region observed as the satellite moves radially away from the earth is interpreted to be the collisionless magnetohydrodynamic shock wave associated with the interaction of the solar wind with the geomagnetic field. The IMP-1 data has provided the first accurate measurements of this phenomenon and indeed have mapped in detail its position relative to the earth sun line. This is a most important feature of the boundary of the magnetosphere since it may provide mechanisms for acceleration of charged particles.

Within and adjacent to the transition region, satellite detectors have shown transient fluxes of energetic electrons having energies greater than 45 Kev and total fluxes of $10^6/cm^2/sec$ an order of magnitude greater than background (Anderson et al, 1964; Fan et al, 1964). The IMP results suggest these observations are related to the formation of the magnetosphere and the shock wave boundary. The experimental evidence is very recent and the full theoretical significance of these data has yet to be completely evaluated. The particle fluxes which are observed are substantially less than those observed within the trapped particle belts within the earth's magnetic field. Hence it is not possible to consider these as hazards to manned space flight travel or to satellite hardware systems when one considers the more important contributions due to the Van Allen Radiation Belts.

A summary of the observed positions of the shock wave boundary and the magnetosphere boundary is shown in Figure 11. In this presen-

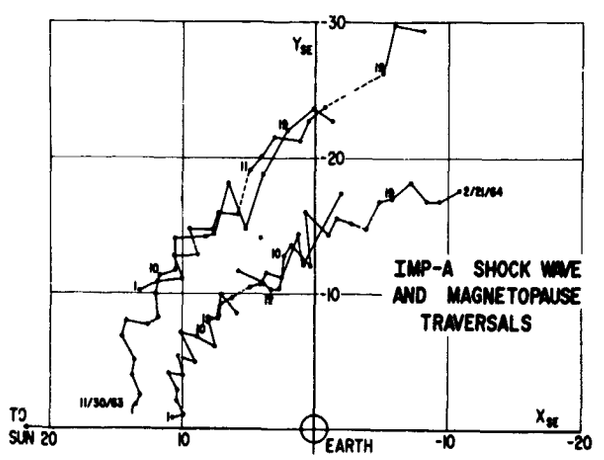


FIGURE 11.—Summary of the observed positions of the magnetosphere boundary and shock wave as determined by the magnetic field experiment on IMP-1. Successive traversals of the boundaries are connected by straight line segment. Orbits number 1 through 23 are shown and are seen to be approximately parabolic in shape.

tation the boundaries of the magnetosphere and transition region as detected by the magnetic field experiment are illustrated. It is seen that the geocentric distance to the shock wave at the stagnation point is approximately 13.4 R_e but this distance increases away from the subsolar or stagnation point. This indicates an increase in the thickness of the transition region. The data also indicates that the magnetosphere is not closed, at least to the distance of 10 to 20 R_e behind the earth. The data are suggestive that the magnetosphere trails out far behind the earth in the fashion analogous to cometary tails. On this basis it is reasonable to expect the moon to intersect the earth's magnetosphere once each month (Ness, 1964).

A comparison of the theoretical shape and position of the shock wave boundary and magnetosphere boundary with observations is shown in Figure 12. Using a gas dynamic ratio of less than 2 but more than 5/3 permits exact comparison of the data. The small scatter in the position of the boundary crossings is related to the variability of the solar plasma flow. The comparison with theory (Spreiter and Jones, 1963) is very good and indicates a fundamental characteristic of the interplanetary plasma near the stagnation point and on the scale on which the observations are made. The standoff ratio between the shock

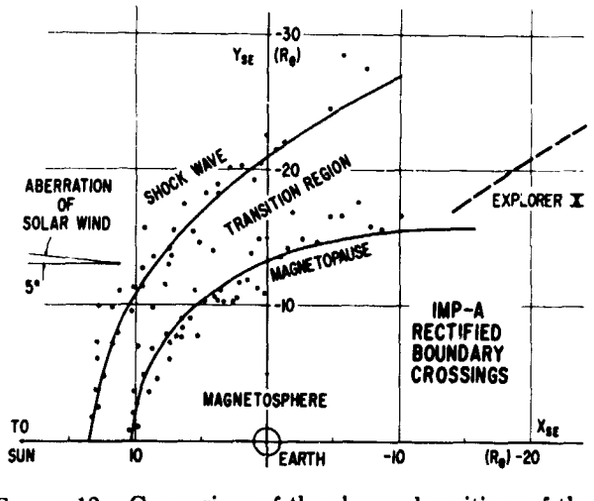


FIGURE 12.—Comparison of the observed positions of the boundary of the magnetosphere and shock wave with the theoretical positions according to Spreiter and Jones (1963). Very good agreement is obtained by slightly modifying their treatment to accommodate the values of magnetosphere radius (10.25 R_e) and shock wave radius (13.4 R_e) actually observed.

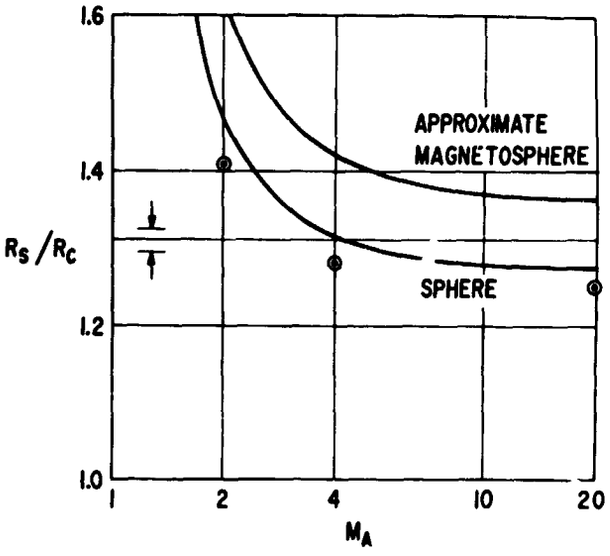


FIGURE 13.—Theoretical standoff ratios (R_s/R_c) for the magnetosphere assuming it to be a sphere or an extended blunt object as a function of Mach number. The observed values 1.31 shown intermediate to these two cases.

wave distance and the magnetosphere boundary is shown in Figure 13 as a function of Mach number for two models of the shape of this magnetosphere. One is that of a sphere utilizing the theoretical results by Hida (1953) and the other utilizes the various models in generally good agreement as represented by Beard (1960) and Spreiter and Jones (1963). The observed value of 1.31 ± 1 percent is seen to be between the two limits. For the observed Mach numbers the standoff ratio is reasonably insensitive to the exact value of Mach number. Hence time variations in the characteristics of the solar plasma do not affect to first order the standoff ratio, as do the magnetosphere shape and specific heat ratio used in the gas dynamic analogy.

A summary of the description of the magnetosphere and its boundary layer as projected on the plane of the ecliptic is shown in Figure 14. In this figure the interplanetary magnetic field is shown at an angle of 135 degrees to the Earth Sun line and in a sense which is positive with respect to flux lines extending from the sun into interplanetary space. The projected positions of the IMP satellite are shown for the first 19 orbits and the positions of the magnetosphere boundary and shock wave are shown as average positions. Within the transition region is a turbulent plasma

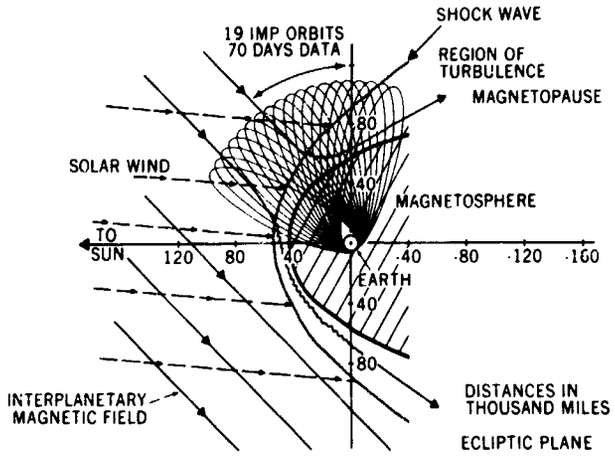


FIGURE 14.—Summary schematic illustration of the magnetosphere shape and boundary layer thickness as deduced from magnetic field measurements on the IMP-1 satellite. The flow of solar plasma, the solar wind, is taken to be aberated by 5° west of the sun due to the heliocentric orbital motion of the earth.

flow of very high temperatures with fluctuating magnetic fields. Within the magnetosphere a distorted geomagnetic field is observed, dependent upon the strength of the earth's magnetic field and the strength of the solar wind containing it. Present measurements do not indicate a termination of the magnetosphere on the leeward side of the solar wind flow. It is very possible that the earth's magnetic field trails out 100 R_e or more behind it, intersecting the orbit of the Moon (Ness, 1964).

ACKNOWLEDGMENTS

I acknowledge the contribution of the other experimenters on the IMP-1 satellite, in addition to my own co-experimenters, Mr. C. S. Scearce and Mr. J. B. Seek.

REFERENCES

ANDERSON, K. A., H. K. HARRIS and R. J. PAOLI, Energetic Election Fluxes in and beyond the Earth's Outer Magnetosphere, preprint, 1964.

BEARD, D. B., The interaction of the terrestrial magnetic field with the solar corpuscular radiation, *J. Geophys. Res.*, **65**, 3559–3568, 1960.

BIERMANN, L. Kometenschweife und solare Korpuscular strahlung, *Z. Astrophys.* **29**, 274–286, 1951.

BONETTI, A., H. S. BRIDGE, A. J. LAZARUS, B. ROSSI and F. SCHERB, Explorer X plasma measurements, *J. Geophys. Res.*, **68**, 4017–4063, 1963.

BRIDGE, H., A. EGIDI, A. LAZARUS, E. LYON and L. JACOBSON Preliminary results of plasma measurements on IMP-A, COSPAR presentation, Florence, Italy, 1964.

CAHILL, L. J. and P. G. AMAZEEN, The boundary of the geomagnetic field, *J. Geophys. Res.*, **68**, 1835–1844, 1963.

CHAPMAN, S. and FERRARO, V. C. A., A new theory on magnetic storms, *Terr. Mag. and Atmos. Elec.*, **36**, 77–97, 171–186, 1931; **37**, 147–156, 421–429, 1932; **38**, 79–96, 1933.

FAN, C. Y., G. GLOECKLER, and J. A. Simpson, Evidence for > 30 Kev electrons accelerated in the shock transition region beyond the earth's magnetospheric boundary, Phys. Rev. Letters, to be published, 1964.

FREEMAN, J. W., J. A. VAN ALLEN and L. J. CAHILL, Explorer XII observations of the magnetosphere boundary and the associated solar plasma on Spetember, 13, 1961, *J. Geophys. Res.*, **68**, 2121–2130, 1963.

GRINGAUZ, K. I., Some results of experiments in interplanetary space by means of charged particle traps on Soviet space probes, *Space Research II*, 539–553, North Holland, Amsterdam, 1961.

HEPPNER, J. P., N. F. NESS, T. L. SKILLMAN and C. S. SCEARCE, Explorer X magnetic field measurements, *J. Geophys. Res.*, **68**, 1–46, 1963.

HIDA, K., An approximate study on the detached shock wave in front of a circular cylinder and a sphere, *J. Phys. Soc., Japan*, **8**, 740–745, 1963.

NESS, N. F., SCEARCE, C. S. and SEEK, J. B., Initial Results of the IMP-1 Magnetic Field Experiment, *J. Geophys. Res.* **69**, 3531–3570, 1964.

NESS, N. F. and J. M. WILCOX, The solar origin of the Interplanetary Magnetic Field, Phys. Rev. Letters, 13 (15), 461–4, 1964.

NESS, N. F., The Magnetohydrodynamic Wake of the Moon, Goddard Report, X-612-64-323, October 1964.

PARKER, E. N., Dynamics of the interplanetary gas and magnetic fields, *Astrophys. J.*, **128**, 667–676, 1958.

SNYDER, C. W. and M. NEUGEBAUER, Interplanetary solar wind measurements by Mariner II, Space Res., IV, 1964 (to be published).

SPREITER, J. R. and W. P. Jones, On the effect of a weak interplanetary magnetic field on the interaction between the solar wind and the geomagnetic field, *J. Geophys. Res.*, **68**, 3555–3565, 1963.

ON THE SOLAR ORIGIN OF THE INTERPLANETARY MAGNETIC FIELD*

NORMAN F. NESS AND JOHN M. WILCOX†

Direct measurements of the interplanetary magnetic field at 1 A.U. have recently been made by an experiment[1] conducted on the Interplanetary Monitoring Platform (IMP-1). The 63 kg spin stabilized satellite was launched on Nov. 27, 1963 into a highly eccentric orbit with a geocentric apogee of 197,616 km. and an orbital period of 93.5 hours. The Sun-Earth-apogee angle was initially 25° west of the Sun and increased approximately 1°/day due to the Earth's helicoentric motion. Successful transmission of scientific data continued for 180 days until May 30, 1964 when the Sun-Earth-apogee angle had increased to 201°. The interplanetary data is based on the average magnetic field computed every 5.46 minutes using 12 successive data transmissions at 20.5 second intervals. The absolute uncertainty of the data is less than 0.25 γ ($1\gamma = 10^{-5}$ gauss). This letter presents the initial results of a detailed comparison of these interplanetary magnetic field measurements with solar magnetograph observations of the photospheric magnetic fields as measured at the Mt. Wilson Observatory.

Previous space probe measurements[2] of the interplanetary magnetic field by Pioneer V in 1960 were interpreted as evidence of a 2.5 γ field normal to the ecliptic plane. A recent reanalysis[3] of this data has yielded the result that the direction of the interplanetary field was nearly parallel to the plane of the ecliptic and that the magnitude was between 5–10 γ. The measurements[4] of the solar plasma efflux in 1962 by the Mariner II Venus probe have confirmed the initial restricted results[5] of the Explorer X satellite obtained in 1961 and the Lunik II[6] moon probe in 1960. The Mariner II results showed that a continual flow of plasma, the solar wind[7], existed with an average streaming velocity of 504 km/sec.

The interaction of the solar wind with the geomagnetic field disturbs the interplanetary medium in the immediate vicinity of the Earth and spatially confines the Earth's field. Thus measurements of the relatively undisturbed interplanetary medium by IMP/1 are restricted to times centered about the satellite apogee and outside the interaction region. This region is limited to about 85,000 km at the subsolar point, but is not spherically symmetrical[1] and instead broadens out away from the stagnation point and eventually trails out behind the Earth. Hence as the orbit of IMP-1 precessed in the solar-ecliptic coordinate system[1] the maximum length of time possible for measurements of the interplanetary medium decreased from about 3 days during Orbit 1 to 1 day on Orbit 22. The interplanetary magnetic field data obtained during this portion of the IMP-1 lifetime were distributed over the time interval Nov. 27, 1963 to Feb. 17, 1964 covering the three solar rotations 1784 to 1786.

The initial results of these measurements are consistent with the general model[8] of an interplanetary field of solar origin in which the effects of a radial solar wind velocity and the solar rotation combine to produce magnetic lines of force which are twisted in the plane of the ecliptic in the form of an Archimedes spiral. Figure 1 summarizes the orientation histograms for the interplanetary magnetic field data. The angular distribution of the field component in the ecliptic plane is shown on the left-hand side with an angular increment of 20° selected on the basis of an uncertainty of direction of less than 5° for each 5.46 minute average. A total of 12,510 data points were obtained during the 83 day interval

*Published as *Goddard Space Flight Center Document X-612-64-250* also published in *Physical Review Letters*, 13(15):461-464, October 12, 1964.

†Space Sciences Laboratory, University of California, Berkeley, California.

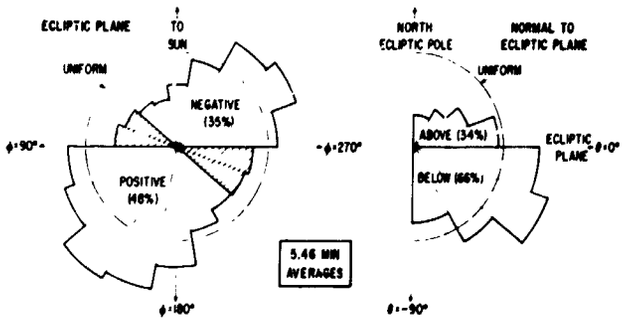


FIGURE 1.—Distribution of the interplanetary magnetic field direction in the plane of the ecliptic and normal to the ecliptic. Both histograms show the field angular distribution per unit solid angle; the dashed circles would correspond to an isotropic distribution of the same number of vectors.

covered by this data. The decided tendency of the field to be oriented parallel or anti-parallel to the theoretical angle (~135°) proposed by Parker[8] is clearly indicated. The asymmetry between the oppositely directed fields is presently being investigated. The figure also identifies those sectors in which the direction of the field is predominantly away from the Sun ($\phi = 90°$ to 230°, "positive" direction) and those in which the field is predominantly towards the Sun ($\phi = 270°$ to 50°, "negative" direction). Also shown in Figure 1 is the angular distribution of the field component perpendicular to the ecliptic plane. This illustrates the tendency of the neaby interplanetary field to be directed nearly parallel to the ecliptic plane, although a slight but distinct southward directed component is apparent. These facts suggest that magnetic field filaments exist in the interplanetary medium. The average field magnitude is close to 5γ and ranges principally between 3–6γ.

For direct comparison with the solar magnetograph data the IMP–1 data has been averaged over a 12 hour interval. Each such interval was divided into 4 subintervals which were assigned a polarity on the basis of the direction of the most frequent occurrence of the initial 5.46 minute averages. If 3 or 4 of the subintervals were of the same polarity the 12 hour interval was assigned that polarity. Other possible results including incomplete data coverage were considered too ambiguous and were not assigned any polarity. The purpose of this procedure was to obtain the large scale polarity distribution of the interplanetary magnetic field. These IMP–1 data were first examined for evidence of a 27 day recurrence period, on the assumption that this would be a measure of the solar influence on the measured interplanetary field. A prominent peak with an amplitude of 0.55 ± 0.05 in the auto-correlation[9] curve was observed at a lag of about 27 days. It was observed that if the first 8 days of IMP observations (including the magnetic storm of Dec. 2, 1963) were omitted, the remaining data yielded a 27 day auto-correlation peak with an amplitude of about 0.85 ± 0.10. Therefore this 75 day interval of data was selected for further analysis on the assumption that this represented a period of time in which the large scale features of the interplanetary magnetic field were quasi-static.

Dr. Robert Howard of the Mt. Wilson Observatory kindly made available the daily solar magnetograms that were obtained with the solar magnetograph.[10] With this instrument the photospheric magnetic field is measured utilizing the longitudinal Zeeman effect. The minimum detectable field was 2 gauss (averaged over an area on the Sun of 4 x 10^8 km^2) in the observations used in this note. Each magnetic region occupied an appreciable portion of the visible disk (i.e. this note is not directly concerned with the much stronger but much more localized fields associated with sunspots; discrete events such as flares also do not enter into this analysis). The direction (out of or into the Sun) of the photospheric field at the center of the visible disk was assigned a polarity (plus or minus) during the three solar rotations of interest. An increment of 6.7° in longitude (corresponding to the solar rotation in 12 hours) and of 10° in latitude was chosen. Occasionally the field direction in an increment of area was ambiguous and such an area was omitted in the analysis. There are also gaps in the photospheric data for about 25 days caused by weather conditions at Mt. Wilson. A cross-correlation between the IMP–1 field direction and the solar field direction was then computed. The computation was repeated at 5° intervals for latitudes from 25°N to 25°S of the center of the visible disk.

Three such cross-correlations are shown in Figure 2 as a function of the time lag from central meridian passage of the photospheric field area to the time the field was measured by IMP–1. The prominent positive peak at a lag of approximately

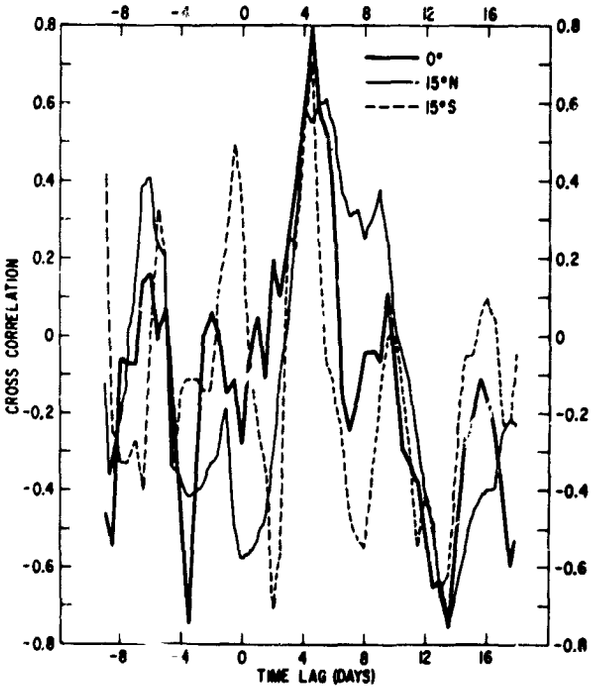


FIGURE 2.—Cross-correlation between IMP magnetic field direction (toward or away from the sun) and the photospheric field direction (into or out of the sun) for three latitudes on the sun. The positive peak at about 1½ days would correspond to a uniform radial solar wind velocity of about 385 km/sec. The negative peaks at −3½ days and +13 days can be understood in terms of a quasiperiodicity in the data. The latitudes are measured from the center of the visible disk.

4½ days is common to all of the latitudes examined. Because several of the photospheric field regions have a considerable extent in latitude the limited data sample available does not permit a statistically significant determination of the solar latitude at which the best correlation with IMP is obtained. The peak in the correlation at a lag of $4\frac{1}{2}\pm\frac{1}{2}$ days would correspond to an average radial solar wind velocity of 385 ± 45 km/sec. The average radial solar wind velocity over the first seven IMP orbits measured by the MIT plasma detector experiment[11] was 398 km/sec.

The latitude of the photospheric source of the interplanetary magnetic field has been investigated by means of the differential rotation of the photospheric magnetic field. Figure 3 shows the rotation period as a function of latitude as determined from auto-correlations of the photospheric field data previously described[12]. Also shown is the recurrence period determined from auto-correlation of the IMP data. This comparison suggests that the latitude of the photospheric source of the interplanetary field was within 10 or 15 degrees of the center of the visible disk (or of the equator; the present analysis cannot make a distinction between the two).

These results indicate that during three solar rotations near the minimum of the solar cycle, some of the magnetic field lines passing through the photosphere near the center of the visible disk tended to be dragged out by the solar wind plasma to become part of the near-by interplanetary magnetic field. The best correlation was obtained

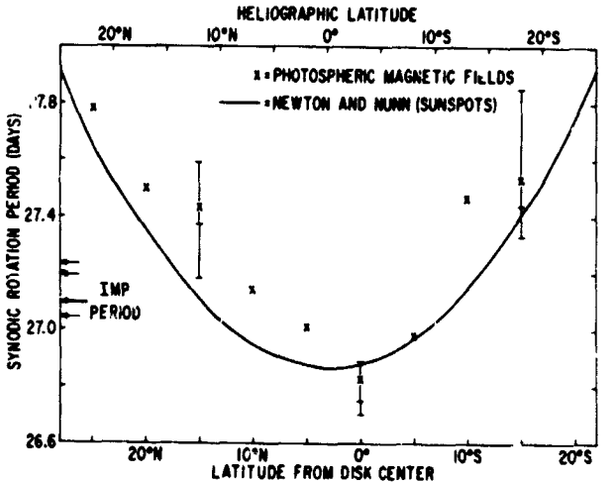


FIGURE 3.—Differential rotation periods of the photospheric magnetic field. The latitude of the photospheric observations is measured from the center of the visible disk, whose average heliographic latitude was 3°S during the time considered here. The error bars were estimated by dividing the data into three parts and analyzing each separately. The solid line is from the sunspot observations of Newton and Nunn [13]. The arrows at the left indicate the IMP recurrence period; the position of the heavy arrow was computed using all the data and the light arrows represent a division of the data into three parts.

when the photospheric field was in the same direction throughout an area corresponding to at least two or three days rotation. These conclusions are consistent with the model suggested by Ahluwalia and Dessler[13] in which the sense of the interplanetary magnetic field filaments is related to the sense of photospheric magnetic field regions.

We thank Dr. Robert Howard for making the solar magnetograms available, and Dr. Howard and Dr. V. Rumba for several valuable discussions.

One of us (JMW) thanks the director of the Mt. Wilson Observatory for guest investigator privileges at the observatory.

REFERENCES

1. N. F. Ness, C. S. Scearce and J. B. Seek, J. Geophys. Res. *69*, 3531 (1964).
2. P. J. Coleman, L. Davis and C. P. Sonett, Phys. Rev. Letters *5*, 43 (1960).
3. E. W. Greenstadt, J. Geophys. Res. (to be published).
4. C. W. Snyder and M. Neugebauer, Space Research IV (to appear).
5. A. Bonetti, H. S. Bridge, A. J. Lazarus, B. Rossi and F. Scherb, J. Geophys. Res. *68*, 4017 (1963).
6. K. E. Gringauz, V. V. Bezruvkikh, V. D. Ozerov, and R. E. Rybchinskii, Sov. Phys. Dokl. *5*, 361 (1960).
7. E. N. Parker, Ap. J. *132*, 821 (1960).
8. E. N. Parker, Ap. J. *128*, 664 (1958).
9. In computing the "correlations" described in this note a weight of +1 was assigned to a field directed out of or away from the Sun, and vice versa for −1. Each individual set of data did not add up exactly to zero. If the data sample were enough that each set of data would sum to zero then this computation would yield a formal correlation. The error is estimated by dividing the data sample into three parts and repeating the computation on each.
10. H. W. Babcock, Ap. J. *118*, 387 (1953).
11. H. Bridge, A. Egidi, A. Lazarus, E. Lyon and L. Jacobson, Private Communication, 1964.
12. The differential rotation of the photospheric magnetic field observed over several years will be reported in a future publication.
13. H. S. Ahluwalia and A. J. Dessler, Planet. Space Sci. *9*, 195 (1962).
14. H. W. Newton and M. L. Nunn, Roy. Astron. Soc. M. N. *111*, 413 (1951).

SPINOR FORMULATION OF MAGNETOGAS DYNAMICS*

LAWRENCE A. SCHMID

It is shown that all the variables needed for a classical description of the dynamical behavior of a fluid consisting of electrically charged particles having spin can be incorporated into two spinors having a total of four complex elements. The particle spin is included, not because it plays any significant role in magnetogas-dynamical problems, but because it is needed to account for all the degrees of freedom of the spinors.

The link between the spinors and the familiar quantities that describe the fluid is provided by the "particle tetrapod" consisting of one time-like and three space-like 4-vectors. These four vectors constitute an orthonormal system and are normalized to the particle density of the fluid in its local rest-frame. The time-like vector is identified with the flux density of the fluid; one space-like vector is used to specify the orientation of the particle spin axis; and the remaining degree of freedom, an angle, is postulated to be proportional to the de Broglie phase, the proportionality constant being Planck's constant.

Incorporating the de Broglie phase into the tetrapod has two consequences: First, Planck's constant is introduced into the formalism. Second, because the canonical momentum is the gradient of the de Broglie phase, the tetrapod contains all the information needed for a dynamical, as well as a kinematical, description of the fluid.

It is shown that the vectors of the tetrapod can be generated by four different bilinear forms involving the elements of two spinors and their complex conjugates. In this sense, the spinors may be regarded as the "square roots" of the tetrapod. The phase angle common to the two spinors is the only one of the eight degrees of freedom of the spinors that does not make itself felt in generating the tetrapod. This phase angle is used to specify the sign of the particle charge.

If the spinors that describe the fluid satisfy a first-order linear partial differential equation that involves the 4-vector electromagnetic potential and the particle mass, regarded as a linear function of the gravitational potential and the specific enthalpy, then it turns out that, for the case of adiabatic flow, the quantities involved in the tetrapod automatically satisfy a 4-vector equation that has the form of the relativistic Euler equation in the presence of electromagnetic, gravitational, and pressure fields. Spin-dependent forces proportional to Planck's constant also appear in this equation but, for problems on an macroscopic scale, these forces are completely negligible. It is shown that it is possible to drop the condition of adiabatic flow, and admit fluid viscosity, or energy injection or loss (e.g. through nuclear reactions or radiation, respectively), without changing the form of the spinor equation to be solved. Only the form of the equation that determines the specific enthalpy in terms of other fluid quantities is altered.

Because the spinor equation is linear for the case of gravitational, electromagnetic, and thermal fields regarded as fixed functions of the space-time coordinates, wheras the corresponding Euler equation is nonlinear, it is suggested that use of the spinor alternative to the Euler equation would facilitate the solution of the complete megnatogas-dynamical problem, involving both fluid-dynamical and field equations, by means of a straight-forward iteration procedure. This approach should be especially fruitful in the case of the self-excited dynamo problem, and it is in terms of this problem that the physical interpretation of the formalism is made.

The spinor alternative to Euler's equation has the form of the Dirac equation, except that the particle mass is regarded as a scalar function of the space-time coordinates, rather than a constant. The theory, however, except for the incorporation of the de Broglie phase into the tetrapod, is completely classical in spirit. In particular, no quantization process, and nothing corresponding to the Exclusion Principle have been introduced into the formalism. Thus the theory developed

*Published as *Goddard Space Flight Center Document X-640-64-55*, March 1964.

here could not be expected to yield valid solutions for problems in which the charged fluid should become degenerate as, for instance, at the centers of certain stars.

As a preliminary to the introduction of the particle tetrapod and the spinor equation, consideration is given to the relativistic problem of electron and proton gases interacting with the gravitational, electromagnetic, and pressure fields produced by these same charged gases. The analysis is first carried out without particle spin, and is then modified to take spin effects into account. Finally, the particle tetrapod and the spinor equation are introduced. The detailed mathematical work involving spinors is relegated to an appendix, only the results being summarized and interpreted in the body of the paper, an understanding of which requires no prior knowledge of spinor analysis.

I INTRODUCTION

Self-Excited Dynamo

The basic equations of magnetogas dynamics are composed of two categories of equations: the fluid-dynamical equations, and the field equations. It is the purpose of this paper to develop a new formulation of the fluid-dynamical equations that will simplify the problem of the simultaneous solution of the fluid-dynamical and field equations. Although this new formulation will be applicable to a wide range of problems, the problem we shall use as a guide in discussing physical questions will be that of an isolated spheroid of fully ionized plasma whose fluid and electric currents are such that it constitutes a self-excited dynamo. A discussion of this type problem and review of the literature has been given by Elsasser (1), and in a more condensed form by Cowling (2).

The only serious attempt at a quantitative solution of this problem has been the one carried out by Bullard and his collaborators (3). In this calculation the fluid dynamical half of the problem was neglected, except for imposing the continuity condition. The calculation assumes an incompressible fluid (core of the earth), and so would not be valid, excpet in a qualitative way, for a star. Because of the immense difficulty of the calculation, it was necessary to truncate the harmonic expansion of the solution after the first few terms, and for this reason the solution is only approximate and some doubt remains concerning its convergence. Finally, the calculation was carried out for only one possible flow pattern. Other patterns, possibly of great physical interest, have not as yet been attempted. Calculations [(4) and (5)] have been made, however, for two other flow patterns that do not pretend to resemble the actual flow within the earth's core. The primary purpose of these calculations was merely to demonstrate the possibility of the existence of a self-excited dynamo in a homogeneous, dissipative, conducting sphere. Accordingly, the flow patterns were chosen to facilitate the calculation rather than (as in Bullard's calculation) to resemble a possible flow pattern in the core of the earth. These calculations, like Bullard's, neglected the dynamical half of the problem except for the condition that the fluid flow be solenoidal.

The extreme difficulty of these calculations clearly illustrates the desirability of finding an alternative formulation of magnetogas dynamics that will ease the calculational burden.

Two-Fluid Formulation of the Problem

Rather than formulate the problem in terms of a single conducting fluid through which an electric current flows, we shall work with two electrically charged fluids—the electron and proton gases. (Simply by changing the particle mass, we would have a gas of positive ions, instead of a proton gas.) This two-fluid approach is more rigorous than the single-fluid approach in that it obviates the need to introduce the conductivity which under certain conditions becomes an untenable concept.

The electron and proton gases are postulated to be independent perfect gases that interact only through common gravitational and electromagnetic fields. The partial pressures of the two charged gases are taken into account, but viscosity is neglected. Each of the charged fluids must independently satisfy Euler's equation, which is just the expression of Newton's second law for the fluid.

Sources of Nonlinearity

The magnetogas-dynamical problem is intrinsically nonlinear inasmuch as the interaction of

the electron and proton gases with each other, and the self-interaction of each, means that superposability of solutions is not possible. However, even when we eliminate the interaction by requiring that the gravitational, electromagnetic, and pressure fields be fixed functions of the space-time coordinates (i.e., independent of the fluid flow solutions), we find that Euler's equation is still nonlinear because it contains terms that are quadratic in the fluid velocity. Because we know from physical reasoning that in the case of fixed fields superposability of solutions is possible, it would appear that the nonlinearity in Euler's equation caused by the terms that are quadratic in velocity must be spurious in the sense that there must exist an alternative way to formulate the fluid-dynamical problem that would be linear for the case of fixed fields.

We shall see that it is possible to replace Euler's equation by an equivalent spinor equation that is just the desired linear alternative (in the case of fixed fields). Having this linear alternative to Euler's equation, a straight-forward iteration solution of the total magnetogas-dynamical problem is now possible: we start with a zero-order solution of the linear spinor equations (one for each of the two charged fluids), for a certain choice of zero-order fields; we use this solution to determine the zero-order source terms in the field equations; we solve the field equations for the first-order corrections to the fields, which we then use in the spinor equations to calculate the first-order corrections to the fluid flow of the electron and proton gases; then the whole process is repeated as many times as needed to give the desired accuracy. Such a straight-forward iteration procedure would be impossible using Euler's equation instead of the spinor equation because the solution of the fluid flow for the case of given fixed fields would be blocked by the nonlinearity of Euler's equation arising from the terms that are quadratic in velocity.

Spinor Equations of Motion

As a prerequisite for deriving the spinor equations, we must introduce the electron and proton spin as additional degrees of freedom. Since particle spin plays no significant role in magnetogas-dynamical problems, it would seem that this is an unnecessary complication of an already too complicated problem. In actual fact, however, in the solution of a problem like the self-excited dynamo, the introduction of spin coordinates does not increase the calculational difficulty because it merely means that we work with an expansion in spinor harmonics instead of an expansion of vector harmonics.

We shall see that it is possible to regard the spinors that describe the electron gas, for example, as the "square roots" of the electron flux density 4-vector. In the same way that we find that taking the square root of a real number gives us an extra degree of freedom (the sign of the root) that must be specified by some physical condition, taking the "square root" of the flux 4-vector gives rise to extra degrees of freedom, among which are those corresponding to particle spin. Thus, the introduction of particle spin is the necessary price we must pay for the convenience of working with the linear spinor equation, rather than with the nonlinear Euler (4-vector) equation.

We shall find that the spinor equations of motion have the form of the Dirac equation in which the particle mass, instead of being a constant, is a scalar function of the space-time coordinates that includes the mass per particle associated with the gravitational and thermal energy of the fluid, as well as the constant rest-mass of the particle. In spite of the fact that the spinor equation of motion has the form of the Dirac equation, the theory developed in this paper is purely classical (except for the fact that it incorporates the de Broglie Hypothesis, which introduces Planck's constant into the theory). In particular, no quantization process is introduced. Thus the Exclusion Principle, and consequently degeneracy of the electron and proton gases at high densities, do not follow from the theory.

In recent years there has been an increased interest in the problem of finding a classical interpretation of the Dirac equation. Noteworthy in this respect is the long paper by Takabayasi (6), and the book by Halbwachs (7) in which most of the work up to 1960 is reviewed. More recent work has been done by Schiller (8) and by Grossmann and Peres (9). (The latter reference gives a bibliography for the most recent work.) All this work differs from the theory developed in this paper in that the authors cited were ultimately concerned with finding a deeper or more intuitively

appealing interpretation of quantum mechanics, rather than in the simplification of the equations of classical magnetogas dynamics. Thus, although some formal similarities exist between the present work and certain of the papers cited, vital differences also exist, with the result that any comparison must be made with great care. In any case, the present paper is intended to be self-contained and no use is made of any previous work.

Assumptions and Approximations

A realistic solution of the self-excited dynamo problems corresponding to the sun or the core of the earth would involve taking into account many detailed and complicated physical effects, some of which are not well understood. Such a detailed program would be premature. Rather, we shall aim at solving a well-defined, relatively simple idealization that, with only minor modifications can be brought into close enough correspondence with the real problems occurring in nature to afford some physical insight. We shall now discuss the assumptions and approximations that denfine this idealization.

First, we treat the electron and proton gases as fluids, rather than as distributions of particles having different velocities, spins, etc. Thus, for example, at a given point of space-time, there is only a single electron velocity, and this velocity is a continuous function of the space-time coordinates.

Further, we postulate that the electron and proton gases are perfect, classical (no degeneracy) gases.

We postulate that we have at every point of space-time a fully ionized plasma in which the electrons or protons (or positive ions) are individually conserved.

We shall neglect fluid viscosity (but not pressure). This is a valid approximation in situations such as we find in the interior of the sun and earth's core, since in such cases the magnetic forces are very large compared with the viscous forces. In the case of very tenuous plasmas, such as we find in the corona of the sun, or in interstellar space, this is no longer true, and it would be necessary to take viscosity into account. This can be done in fact, by means of a relatively minor modification of the theory, which is discussed in Appendix A. For the sake of definiteness, however, we assume throughout the body of the paper that fluid viscosity may be neglected.

We impose two different adiabatic conditions: first, we postulate that a given bubble of electron or proton gas loses or gains no heat energy. Second, we postulate that if we were to make a small adiabatic virtual displacement of a given bubble of gas into a neighboring position, then the bubble will have the same thermodynamic properties as the fluid surrounding its new position. We shall see that for quasistatic (i.e., reversible) flow, these two adiabatic conditions can be combined into the single condition that the 4-gradient of the specific entropy vanish. This means that the specific entropy is constant throughout the gas for all time.

The adiabatic condition with respect to virtual displacement of a bubble of gas is a natural one, since it corresponds to the fact that even though the macroscopic flow is well-ordered, there is always small-scale turbulence that tends to maintain the assumed adiabatic condition.

The assumption that a given bubble of gas gains or loses no heat energy as time goes by is less realistic inasmuch as it neglects energy lost through radiation and the possibility of energy injection through nuclear reactions or some other mechanism. Moreover, we are neglecting heat gained or lost by conduction. It is, in fact, not difficult to modify the theory so as to allow for energy injection or loss. The way in which this can be done is discussed in Appendix A. For definiteness, however, through the body of the paper we shall neglect the possibility of energy loss or injection. We justify this by taking the point of view that we divide the total problem into two steps: (1) Solution of the dynamical problem subject to the adiabatic condition; (2) Solution of the energy balance problem subject to the assumption that dropping the adiabatic condition has no effect on the dynamical solution other than to cause a slow change in its parameters (e.g., the slow expansion or contraction of a star). This two-step point of view is illustrated in Figure 1: On the left side is shown a possible flow pattern in the sun or the earth's core. Fluid rises from the center C in the equatorial plane, floes along the surface through a typical surface point S to the poles, where it returns to the center. On

the right side of Figure 1 is shown the pressure (P) vs. specific volume (V_1) plot for a typical bubble of fluid that goes through a complete cycle in the indicated flow pattern. The points C and S on the P-V_1 plot indicate the thermodynamic state of the bubble at the corresponding points in the pattern. Thus, assuming the adiabatic condition, we note that as the bubble moves from the center C to the surface S, on the P-V_1 plot it moves

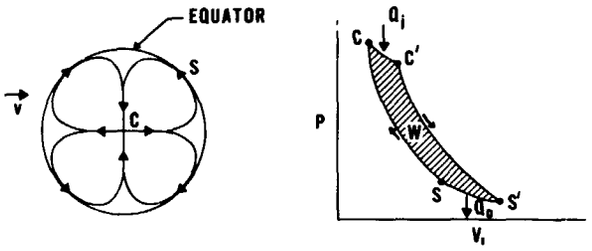


FIGURE 1.—Heat transfer associated with convection.

downward from C to S, corresponding to an expansion. The return trip to the center of the core carries it along the same path in the P-V_1 plot back to the point C. Thus, on the outward trip, the bubble performs an amount of work equal to the area under the curve C-S; that is, heat energy in this amount is converted into some other form of energy, notably electromagnetic energy. But on the return trip the bubble retrieves this energy, so there is no net transfer to electromagnetic field.

Now let us imagine that, when the bubble is at C, an amount of energy Q_i is injected isothermally by means of a nuclear reaction so that while the bubble is still at the center of the core C, it traverses the isothermal C-C'. Then it expands adiabatically reaching the surface in the thermodynamic state S'. At the surface it isothermally radiates the energy Q_o, and then traverses the path S-C back to the center. Thus we have a Carnot cycle in which the net amount of energy W is transferred to the electromganetic field to compensate losses due to Joule heating.

The two-step procedure mentioned above for solving the total problem is valid only if the Carnot cycle is very narrow. In such a case, the energy injection and loss will have no appreciable effect on the dynamics of the problem. This amounts to assuming that the energy injected and radiated during each cycle is small compared with the adiabatic work performed during the trip from the center to the surface. Obviously, this will not be the case if energy is injected too quickly. In such cases the energy injection does affect the dynamical problem, and it is necessary to modify the theory in the manner discussed in Appendix A.

As a final observation concerning the adiabatic condition, we note that, since this condition is applied to the electron and proton gases independently, it would seem that we are assuming that there is no energy transfer between the two gases; that is, we are assuming that the thermodynamic properties, in particular the temperatures, of the two gases are completely independent. We shall find, however, that it follows as a consequence of the approximate equality of the two gas densities that, if the temperatures are approximately equal at one point in space-time, then they must be approximately equal to everywhere. The approximate equality of the two charged gas densities is, of course, automatically maintained by the fact that strong electric fields cannot exist in a plasma. The approximate equality of the two temperatures at a single point, such as the center of the sun or earth's core, must be imposed as a boundary condition.

Completing the list of approximations that define our idealized model, we note that we employ a scalar theory of gravitation which, although it is a covariant theory, will not give the correct higher-order velocity dependence (such as we encounter in computing the advance of the perihelion of Mercury). This requires a theory based on a symmetric tensor, as in General Relativity, rather than one based on a scalar potential.

A scalar theory, however, is perfectly adequate for the needs of magnetogas dynamics. We shall also violate the Equivalence Principle as far as the mass associated with electromagnetic, thermal, and gravitational energy is concerned. That is, although all these contributions will be included in the inertial mass, they will not be included in either the active or passive gravitational mass, for which we use just the rest-mass. This approximation should also be adequate for the needs of almost all of the existing problems in magnetogas dynamics.

Plan of Presentation

As a preliminary to writing down the fundamental equation of motion for each of the two charged fluids (Euler's equation), it is shown in Section II that the necessary thermal properties of each fluid can be completely described in terms of the specific enthalpy of the fluid and the particle density. The great advantage of doing this is that the pressure term in Euler's equation can then be replaced by the gradient of the specific enthalpy, which then plays the role of a "thermal potential function" that is completely analogous to the scalar gravitational potential. This fact allows us to absorb both the gravitational and the thermal energy into the particle rest-mass, which is regarded as a scalar function of the space-time coordinates. In this way the particle rest-mass is made to play a dual role: It describes the inertial property of the particle, and at the same time serves as a potential function for the scalar force fields acting on the particle. In Appendix A, it is shown that this procedure has the added advantage that, when we replace the adiabatic condition by one that allows for fluid viscosity or energy injection, the form of Euler's equation (or the spinor equation that replaces it) remains unchanged. Only the form of the "thermal field equation," i.e. the equation that determines the specific enthalpy as a function of the fluid variables, is altered. This same remark is also valid if we drop the assumption that the two charged fluids are perfect gases.

After writing down Euler's equation in terms of specific enthalpy instead of pressure, the rest of Section II is devoted to showing that, if we limit ourselves to a certain subset of solutions that correspond to imposing certain physically reasonable conditions, Euler's equation can be replaced by a tensor equation that is linear except for the fact that the quadratic normalization condition for the 4-velocity must be satisfied.

In Section III, it is shown that it is this tensor equation, rather than Euler's equation, that results from the Hamilton-Jacobi formulation of the problem. In order to make the discussion more physical (as well as to simplify it), the Hamilton-Jacobi formulation is presented simply as the implementation of the deBroglie Hypothesis, reguarded as an experimental fact.

The conclusion drawn from Sections II and III is that the tensor equation is more fundamental than Euler's equation. In Section VIII it is shown that the spinor equations correspond to the tensor equation, rather than directly to Euler's equation. (Any solution of the tensor equation, of course, is a solution of Euler's equation, although the converse in general is not true.)

In Section IV the electromagnetic and gravitational field equations are presented.

In Section V the changes in the fluid-dynamical equations that result when we introduce particle spins are discussed.

Section VI is a summary of the results of the preceding sections, all of which have been derived within the framework of Special Relativity, but without any mention of spinors.

In Section VII spinors are introduced as the "square roots" of 4-vectors, and all the necessary dynamical quantities are defined in terms of spinors. The discussion is intended to be self-contained, and no prior knowledge of spinor analysis is assumed.

In Section VIII the postulated spinor equation of motion is presented, and the corresponding scalar, vector, and tensor equations that follow as direct consequences of the spinor equation and the definitions of Section VII are presented and discussed. The identification of these equations with the equations of magnetogas dynamics that were summarized in Section VI is made, and the advantages of the spinor formulation are discussed. No prior knowledge of spinor analysis is necessary to follow the discussion of this section. The detailed derivation of the results presented in this section is given in Appendix C, and here some familiarity with spinor analysis is assumed.

Choice of Metric

All the work of this paper is done within the framework of Special Relativity. The metric tensor g^{jk} that is used throughout is chosen to have the following form:

$$g^{00}=g_{00}=1;\ g^{jj}=g_{jj}=-1;\ g^{jk}=g_{jk}=0 \qquad (j,k=1,2,3)\ j\neq k \qquad (1\text{-}1)$$

This choice of metric, which is the most convenient for the transition to spinors, means that the

spatial components of the contravariant and covariant forms of a 4-vector differ by a sign. We shall always identify the *contravariant* form as being the four-dimensional generalization of the corresponding 3-vector. (The only exception is the gradient operator, discussed below.) Thus, if $\vec{v}$ is the particle 3-velocity, the 4-velocity u^j has the following form:

$$u^0=\frac{1}{\sqrt{1-\beta^2}}=u_0;\ \beta=\frac{v}{c}$$
$$u^j=\frac{u^0}{c}v^j=-u_j(j=1,\ 2,\ 3) \tag{1-2}$$

Note that this specification of u^j corresponds to a normalization to unity, rather than to c^2:

$$u^j u_j=1 \tag{1-3}$$

Because $(x^1,\ x^2,\ x^3)$ (rather than the covariant components) are identified with the coordinates $(x,\ y,\ z)$, it follows that the *covariant* form of the gradient operator ∂_j (rather than the contravariant form ∂^j) is the generalization of the 3-gradient $\vec{\nabla}=(\partial_x,\partial_y,\partial_z)$:

$$\partial_0=\frac{\partial}{c\partial t}=\partial^0$$
$$\partial_j=\frac{\partial}{\partial x^j}=-\partial^j\ (j=1,\ 2,\ 3) \tag{1-4}$$

and

$$\vec{\nabla}=(\partial_x,\ \partial_y,\ \partial_z)=(\partial_1,\ \partial_2,\ \partial_3) \tag{1-5}$$

The d'Alembertian operator $\Box$ is defined in terms of the 4-gradient ∂_j as follows:

$$\Box=\partial_j\partial^j=\frac{1}{c^2}\frac{\partial^2}{\partial t^2}-\vec{\nabla}^2 \tag{1-6}$$

We shall have occasion to use the completely antisymmetric unit tensor ϵ^{ijkl}, which is zero if any two of its indices are equal, and $+1$ or -1 if the indices are all different and their order differs from the order (0 1 2 3) by an even or odd number of transpositions respectively. The covariant form of the anti-symmetric unit tensor ϵ_{ijkl} is derived from the contravariant form ϵ^{mnpq} according to the usual rule:

$$\epsilon_{ijkl}=g_{im}\,g_{jn}\,g_{kp}\,g_{lq}\,\epsilon^{mnpq} \tag{1-7}$$

This relation has the consequence that the definition of the covariant form of the tensor differs bay sign from that of the contravariant form. Thus ϵ_{ijkl} is -1 (instead of $+1$) when the order of the indices $(ijkl)$ differs from the order (0 1 2 3)by an even number of transpositions.

Inasmuch as the definition of the contravariant form ϵ^{ijkl} is just the extension of the familiar three-dimensional antisymmetric unit tensor, whereas the definition of the covariant form differs by a sign, we see that these definitions are consistent with the general rule that it is the contravariant form of a quantity, rather than the covariant form, that we regard as the generalization to four dimensions of the corresponding three-dimensional quantity.

II EULER FORMULATION

Thermodynamic Properties

As a preliminary to writing down the Euler equation for each of the charged fluids, we shall first consider the thermodynamic properties of the fluids. All these properties will appear in duplicate, one for each fluid. This fact will be indicated by the symbol $(\pm)$ used usually as a subscript, but sometimes as a superscript when the symbol in question already has a subscript. When we have a product of quantities all referring to one of the two fluids, the symbol $(\pm)$ will be appended to the entire product instead of to each symbol in the product individually. Our notation for the necessary thermodynamic quantities is defined as follows:

$P_{(\pm)}\equiv$ pressure
$T_{(\pm)}\equiv$ absolute temperature
$m_{(\pm)}\equiv$ particle rest-mass
$\rho_{(\pm)}\equiv$ particle density in fluid rest-frame
$V_1{}^{(\pm)}\equiv$ volume per unit mass (specific volume)
$k\equiv$ Boltzmann's constant $=1.3708\times10^{-16}$ *erg/°k*
$f_{(\pm)}\equiv$ number of degrees of freedom per particle
$u_{(\pm)}\equiv$ internal thermal energy per unit mass
$h_{(\pm)}\equiv$ enthalpy per unit mass (specific enthalpy)
$s_{(\pm)}\equiv$ entropy per unit mass (specific entropy)
$c_v{}^{(\pm)}\equiv$ constant volume specific heat referred to unit mass
$c_p{}^{(\pm)}\equiv$ constant pressure specific heat referred to unit mass
$\gamma_{(\pm)}\equiv$ ratio of specific heat $=(c_p/c_v)_{(\pm)}$

We note that because $u_{(\pm)}$ and $h_{(\pm)}$ are referred to unit mass, $(mu)_{(\pm)}$ and $(mh)_{(\pm)}$ represent respectively the internal thermal energy per particle, and the enthalpy per particle.

By definition

$$h_{(\pm)} \equiv u_{(\pm)} + (PV_1)_{(\pm)} \tag{2-1}$$

PV_1 is the amount of work that could be done by the gas surrounding a bubble of unit mass of the gas if this bubble were removed (or its molecules cooled to absolute zero) and the surrounding gas allowed to move in and occupy the space that had been occupied by the bubble. (We assume, of course, that the bubble is so small compared with the surrounding atmosphere of gas that, when the surrounding gas moves in to occupy the space previously occupied by the bubble, the pressure P of the gas does not change significantly. Because the energy PV_1, as well as the energy u, is theoretically available for conversion into work, h, rather than u, is to be regarded as the potential thermal energy associated with a bubble of unit mass of the gas. (Because the energy PV_1 is really provided by the gas surrounding the bubble, rather than the gas within the bubble, it would be more accurate to say that h is the potential thermal energy to be associated with the *region of space occupied by the bubble.*)

The above argument indicates why h is often referred to as "the heat function" and is regarded as the total thermal energy content of the gas. It is not surprising that, as we shall see below, mh, rather than mu, is the thermal potential energy of the particle, analogous to the gravitational potential energy mg, where g is the gravitational scalar potential.

Further intuitive support for the idea that it is h, rather than u, that plays the role of a thermal potential comes from the fact that, for adiabatic flow of a gas through an insulated pipe in which the gas performs no work, the governing equation has the following form: kinetic energy + potential energy + enthalpy = constant. Thus the enthalpy enters the equation in the same way as potential energy. In the case of large-scale adiabatic convection flow within a fluid, we may think of the fluid as flowing in pipes whose walls are defined by the flow lines.

The above arguments are intended only to provide an intuitive understanding of the results ensuing from the formalism below, but no use will be made of them in the formalism.

We now postulate that both charged fluids are perfect gases. Thus the following relations are valid:

$$c_v^{(\pm)} = \frac{k}{2}\left(\frac{f}{m}\right)_{(\pm)} \tag{2-2a}$$

$$c_p^{(\pm)} = c_v^{(\pm)} + \frac{k}{m_{(\pm)}} = \frac{k}{m_{(\pm)}}\left(1 + \frac{1}{2}f_{(\pm)}\right) \tag{2-2b}$$

$$\gamma_{(\pm)} = (c_p/c_v)_{(\pm)} = 1 + \frac{2}{f_{(\pm)}} \tag{2-3}$$

$$u_{(\pm)} = (c_v T)_{(\pm)} \tag{2-4}$$

$$h_{(\pm)} = (c_p T)_{(\pm)} \tag{2-5}$$

$$P_{(\pm)} = k(\rho T)_{(\pm)} \tag{2-6}$$

$$s_{(\pm)} = [c_v \ln T - \frac{k}{m} \ln \rho + \text{constant}]_{(\pm)} \tag{2-7}$$

We choose $h_{(\pm)}$ and $\rho_{(\pm)}$ as our thermodynamic variables, and note that the other quantities can be expressed in terms of these variables as follows:

$$P_{(\pm)} = k\left(\frac{\rho h}{c_p}\right)_{(\pm)} = \left[\rho m h\left(1 - \frac{1}{\gamma}\right)\right]_{(\pm)} \tag{2-8}$$

$$T_{(\pm)} = (h/c_p)_{(\pm)} = \left[\frac{mh}{k(1+\frac{1}{2}f)}\right]_{(\pm)} \tag{2-9}$$

$$u_{(\pm)} = (h/\gamma)_{(\pm)} \tag{2-10}$$

$$s_{(\pm)} = \left[\frac{k}{m}(\tfrac{1}{2}f \ln h - \ln \rho) + \text{constant}\right]_{(\pm)} \tag{2-11}$$

From (2-1) we have

$$\partial_j h_{(\pm)} = \left[(\partial_j u + P\partial_j V_1) + \frac{1}{\rho m}\partial_j P\right]_{(\pm)} \tag{2-12}$$

where we have used the fact that, by definition

$$V_1^{(\pm)} = \frac{1}{(\rho m)_{(\pm)}} \tag{2-13}$$

Since $\Delta u - P\Delta V_1$ is the heat injected into unit mass of the gas during the change indicated by the

increment symbol Δ, we can interpret $(\Delta x^j)(\partial_j u - P\partial_j V_1)$ as the heat injected into unit mass of the gas during the virtual displacement (in time or space) described by Δx^j. Thus $\partial^j u - P\partial^j V_1$ can be regarded as the 4-force per unit mass caused by heat injection. Let Q^j be this force. Thus

$$Q^j{}_{(\pm)} = [\partial^j u + P\partial^j V_1]_{(\pm)} \tag{2-14}$$

We note that, using (2-8) and (2-10), we can rewrite (2-14) as follows:

$$[\rho\partial^j h - (\gamma-1)h\partial^j\rho - \rho\gamma Q^j]_{(\pm)} = 0 \tag{2-15}$$

Substituting (2-14) into (2-12), we have

$$[\rho m\partial^j h]_{(\pm)} = [\rho m Q^j - \partial^j P]_{(\pm)} \tag{2-16}$$

In order to exploit this relation, we shall introduce a variable particle mass $\breve{m}_{(\pm)}$, defined as follows:

$$\breve{m}_{(\pm)} = m_{(\pm)}(1 + g/c^2 + h_{(\pm)}/c^2) \tag{2-17}$$

where g is the gravitational potential. Thus $\breve{m}_{(\pm)}$ is the total particle mass, taking the mass associated with the gravitational and thermal energy as well as the rest-mass, into account. Because, in the situations in which we shall be primarily interested (e.g., the interior of the sun), the positive thermal energy is insufficient to overcome the negative gravitational energy, we shall refer to $\breve{m}_{(\pm)}$ as the "bound particle mass" and $\breve{m}_{(\pm)}$ as the "free particle mass."

From (2-16) and 2-17), we have

$$[\rho\partial^j(\breve{m}c^2)]_{(\pm)} = [\rho m\partial^j g + \partial^j P \cdot + \rho m Q^j]_{(\pm)} \tag{2-18}$$

The significance of this relation becomes apparent when we note that the three terms on the right are the force-per-unit-volume terms that appear in Euler's equation. (Equation (2-18) has been written in contravariant form to emphasize that this, rather than the covariant form, is the generalization of 3-force.)

Euler's equation for the charged fluids has the following form:

$$\left[\rho\frac{d(\breve{m}cu^j)}{d\tau}\right]_{(\pm)} = [\rho m\partial^j g + \partial^j P + \rho m Q^j]_{(\pm)} \pm qF^{jk}(\rho u_k)_{(\pm)} \tag{2-19}$$

where

$\tau_{(\pm)} \equiv$ proper time

$q \equiv$ magnitude of particle charge (always positive)

$F^{jk} \equiv$ Maxwell field tensor

In terms of the electric field intensity $\vec{E}$ and the magnetic flux density $\vec{B}$, the tensor F^{jk} is defined or follows:

$$(F^{10}, F^{20}, F^{30}) \equiv (E_x, E_y, E_z) = \vec{E} \text{ (statvolts/cm)}$$

$$(F^{23}, F^{31}, F^{12}) \equiv -(B_x, B_y, B_z) = -\vec{B} \text{ (Gauss)}$$

$$F^{jk} = -F^{kj} \tag{2-20}$$

where the units have been given to indicate that we shall use Gaussian units throughout. (All dynamical quantities will be expressed in *cgs* absolute units.)

Referring to (1-2), we note the following:

$$\frac{d}{d\tau} = cu^j\partial_j = u^0\left[\frac{\partial}{\partial t} + \vec{v}\cdot\vec{\nabla}\right] \tag{2-21}$$

Using (2-20) and (2-21), we can write (2-19) in the more familiar three-dimensional form:

$$\left\{(\rho u^0)\left[\frac{\partial(\breve{m}u^0c^2)}{\partial t} + \vec{v}\cdot\vec{\nabla}(\breve{m}u^0c^2)\right]\right\}_{(\pm)} = \left[\rho m\frac{\partial g}{\partial t} + \frac{\partial p}{\partial t} + \rho mcQ^0\right]_{(\pm)} \pm q(\rho u^0\vec{v})_{(\pm)}\cdot\vec{E} \tag{2-22a}$$

$$\left\{(\rho u^0)\left[\frac{\partial(\breve{m}u^0\vec{v})}{\partial t} + \vec{v}\cdot\vec{\nabla}(\breve{m}u^0\vec{v})\right]\right\}_{(\pm)} = [-\rho m\vec{\nabla}g - \vec{\nabla}p + \rho m\vec{Q}]_{(\pm)} \pm q(\rho u^0)_{(\pm)}\left(\vec{E} + \frac{1}{c}\vec{v}_{(\pm)}\times\vec{B}\right) \tag{2-22b}$$

where

$$\vec{Q}_{(\pm)} = (Q^1, Q^2, Q^3)_{(\pm)} = -(Q_1, Q_2, Q_3)_{(\pm)} \tag{2-23}$$

The physical interpretation of (2-22) is immediately evident when we recall that ρu^0 and $\breve{m}u^0$ are respectively particle density and particle mass *in the observer's reference frame.* $(\rho mcQ^0)_{(\pm)}$ obviously represents the rate at which energy is

injected per unit volume by nuclear or other reactions, and/or the heat energy generated by fluid viscosity, and $(\rho m\vec{Q})_{(\pm)}$ represents the viscous force, and/or the effective force resulting from heat injection caused by nuclear or other reactions.

The advantage of working with specific enthalpy becomes apparent when we substitute (2–18) into (2–19):

$$\left[\frac{d(\breve{m}cu^j)}{d\tau}\right]_{(\pm)} = \partial^j(\breve{m}_{(\pm)}c^2) \pm qF^{jk}u_k^{(\pm)} \quad (2\text{–}24)$$

We have assumed in writing (2–24) that $\rho_{(\pm)}$ is everywhere non-zero, so that it is permissible to divide an equation through by it.

The advantage of (2-24) over (2–19) is twofold: First, we have eliminated $Q^j{}_{(\pm)}$ from Euler's equation, and transferred it instead to (2–15) which may be regarded as a "thermal field equation" on a par with the gravitational and electromagnetic field equations. Thus, regardless of whether or not we introduce fluid viscosity or heat injection by nuclear reactions, we can always regard (2–24), with $\breve{m}_{(\pm)}$ defined by (2–17), as the governing fluid dynamical equation for each charged fluid. In fact, this is true even if we drop the postulate that the charged fluids are perfect gases. Second, $\rho_{(\pm)}$ does not appear in (2–24), which has the form of a single-particle equation. This would appear to allow us to solve first for $u^j{}_{(\pm)}$ without concerning ourselves about $\rho_{(\pm)}$ and then, knowing $u^j{}_{(\pm)}$, we would find $\rho_{(\pm)}$ from the continuity equation:

$$\partial_j(\rho u^j)_{(\pm)} = 0 \quad (2\text{–}25)$$

whose three-dimensional form is

$$\frac{\partial(\rho u^0)_{(\pm)}}{\partial t} + \vec{\nabla}\cdot[(\rho u^0)\vec{v}]_{(\pm)} = 0 \quad (2\text{–}26)$$

In actual fact, of course, $\rho_{(\pm)}$ enters into (2–24) implicitly through $h_{(\pm)}$ which, as we see from (2–15), is a function of $\rho_{(\pm)}$. In an iteration solution, however, we may regard $h_{(\pm)}$ as a specified function of the space-time coordinates. In such a case, use of (2–24) allows a separation of the solution of $\rho_{(\pm)}$ and $u^j{}_{(\pm)}$ whereas (2–19) does not.

Thus far we have not introduced the assumption that our fluid flow is reversible. If we now introduce this assumption, we have the following relation:

$$Q^j_{(\pm)} = [T\partial^j s]_{(\pm)} = \left[\frac{h}{c_p}\partial^j s\right]_{(\pm)} \quad (2\text{–}27)$$

Using this, we can integrate (2–15) to obtain the following expression for $h_{(\pm)}$:

$$h_{(\pm)} = [H(\rho/\rho_0)^{(\gamma-1)}e^{(s-s_0)/c_p}]_{(\pm)}$$

(reversible flow) (2–28)

where $\rho_0{}^{(\pm)}$ and $s_0{}^{(\pm)}$ are the density and specific entropy respectively at a certain fixed point $x^j_{(0)}$ in space-time (and hence constants). $H_{(+)}$ and $H_{(-)}$ are two constants that must be specified as part of the boundary conditions of the problem. Obviously,

$$H_{(\pm)} = [h(x^j_{(0)})]_{(\pm)} = [c_p T(x^j_{(0)})]_{(\pm)} \quad (2\text{–}29)$$

Thus, the constants $H_{(+)}$ and $H_{(-)}$ can be determined from a knowledge of the fluid temperatures at the point $x^j_{(0)}$.

The equation (2–28) may be used as the "thermal field equation" in preference to (2–15). Equation (2–15) has the advantage, however, that it is valid even when condition (2–27) is not satisfied. If, for example, we wished to introduce viscosity into the problem by requiring that Q^j be a suitable function of the derivatives of u^j, then in general (2–27) will not be satisfied. In such a case (2–28) would not be valid, whereas (2–15) would.

We have indicated the way in which viscosity or heat injection through nuclear reactions would be introduced into the problem by specifying an appropriate functional form for $Q^j{}_{(\pm)}$. We have shown that such a procedure would in no way alter our basic dynamical equation (2–24). Only the "thermal field equation" (2–15) would be altered. The ultimate purpose of this paper is to find a spinor alternative to the vector equation (2–24), and the way we choose to handle $Q^j{}_{(\pm)}$ makes no difference in the form of this equation. Both for the sake of definiteness, and because we noted in Section I that the adiabatic case was a sensible first step of a two-step process for solving the total self-excited dynamo problem, we shall now introduce the *adiabatic condition:*

$$Q^j{}_{(\pm)} = 0 \quad \text{(adiabatic condition)} \quad (2\text{–}30)$$

In Appendix A several alternatives to (2-30) are discussed, but for the rest of the body of the paper we shall assume that (2-30) holds. Thus the "thermal field equation" (2-15) becomes

$$[\rho\partial_j h-(\gamma-1)h\partial_j\rho]_{(\pm)}=0 \quad \text{(adiabatic case)} \quad (2\text{-}31)$$

and its integral is

$$h_{(\pm)}=[H(\rho/\rho_0)^{(\gamma-1)}]_{(\pm)} \quad \text{(adiabatic case)} \quad (2\text{-}32)$$

We may use either (2-31) or (2-32) as our adiabatic "thermal field equation."

Now we note that if the following relations are valid

$$\gamma_{(+)}=\gamma_{(-)} \quad (2\text{-}33a)$$

$$\rho_{(+)}\approx\rho_{(-)} \quad (2\text{-}33b)$$

$$(mH)_{(+)}\approx(mH)_{(-)} \quad (2\text{-}33c)$$

then it follows from (2-32) that

$$(mh)_{(+)}\approx(mh)_{(-)} \quad (2\text{-}34)$$

and from (2-8) and (2-9) that

$$P_{(+)}\approx P_{(-)} \quad (2\text{-}35)$$

and

$$T_{(+)}\approx T_{(-)} \quad (2\text{-}36)$$

(2-33a) is valid because the electron and proton gases are both monatomic gases. The approximate equality (2-33b) will be maintained automatically by the electrostatic screening effect. (2-33c) must be imposed as a boundary condition. From (2-2b) and (2-29) it is evident that this amounts to requiring that the electron and proton temperatures be approximately equal at a particular point in space-time (e.g., the center of the sun at $t=0$). (2-35) and (2-36) then tell us that the two pressures and temperatures are approximately equal throughout space for all time. The significance of this statement is that, even through our way of formulating the problem in terms of two independent specific enthalpies $h_{(+)}$ and $h_{(-)}$ would at first glance appear to let the electron and proton pressures and temperatures differ wildly from each other, in actual fact (if we impose the boundary condition (2-34)) the two pressures and the two temperatures are automatically constrained to remain very nearly equal to each other. This is a satisfying state of affairs since we know that because of particle collisions the two temperatures cannot differ greatly.

Tensor Alternative to Euler's Equation

Equation (2-24), which is the form of Euler's equation that we shall use throughout, is a 4-vector equation. We shall now derive an alternative equation in which all the terms are antisymmetric tensors. This is most easily done by starting with the three-dimensional form given in (2-22b) (except that the bracket on the right is replaced by $-]\rho\vec{\nabla}(\breve{m}c^2)]_{(\pm)}$ and the term $[\vec{v}\cdot\vec{\nabla}(\breve{m}u^0\vec{v})]_{(\pm)}$ on the left is transformed by means of a well-known vector identity):

$$\frac{\partial(\breve{m}u^0\vec{v})}{\partial t}_{(\pm)}+\vec{\nabla}(\breve{m}u^0c^2)_{(\pm)}-\vec{v}_{(\pm)}\times[\vec{\nabla}\times(\breve{m}u^0\vec{v})_{(\pm)}]$$

$$\pm q\left(\vec{E}+\frac{1}{c}\vec{v}_{(\pm)}\times\vec{B}\right) \quad (2\text{-}37)$$

It can be shown that (2-22a) results from the dot product of both sides of (2-37) with $\vec{v}_{(\pm)}$.

Now we note that if we impose the condition

$$\vec{\nabla}\times(\breve{m}u^0\vec{v})_{(\pm)}=\mp\frac{q}{c}\vec{B} \quad (2\text{-}38)$$

(2-37) reduces to

$$\frac{\partial(\breve{m}u^0\vec{v})_{(\pm)}}{\partial t}+\vec{\nabla}(\breve{m}u^0c^2)_{(\pm)}=\pm q\vec{E} \quad (2\text{-}39)$$

These two 3-vector equations can be combined into the following single tensor equation:

$$\partial^j(\breve{m}cu^k)_{(\pm)}-\partial^k(\breve{m}cu^j)_{(\pm)}=\mp\frac{q}{c}F^{jk} \quad (2\text{-}40)$$

This is the tensor alternative to Euler's equation (2-24). Regarding $\breve{m}$ and F^{jk} as fixed functions of the space-time coordinates, (2-40) is a linear equation, whereas Euler's equation is nonlinear. The linearity of (2-40) is spoiled, however, by the fact that we must maintain the normalization condition (1-3) for the 4-velocity:

$$(u^ju_j)_{(\pm)}=1 \quad (2\text{-}41)$$

Thus we must still contend with the algebraic nonlinearity of (2-41). In the case of Euler's equation, the condition (2-41) must also be satisfied. In addition, we have what we might term

a "differential nonlinearity" in Euler's equations itself because of the presence of the term $[u^l\partial_l(\mathring{m}cu^j)]_{(\pm)}$. Thus, by replacing (2–24) with (2–40), we have rid ourselves of the differential nonlinearity, but are still left with the algebraic nonlinearity (2–41).

It is evident that the tensor equation (2–40) is not exactly equivalent to Euler's equation, because it corresponds only to the subset of solutions of (2–37) that satisfy the condition (2–38). Either we regard (2–37) as more fundamental than (2–40), with the result that by working with (2–40) we are excluding physically meaningful, and perhaps important, solutions from consideration; or, we take the point of view that (2–38) is a physically necessary condition, and the solutions of (2–37) that do not satisfy (2–38) have no physical validity. In order to help convince ourselves that the latter situation is actually the case, we shall first investigate the physical meaning of condition (2–38).

To this end, we let $\mathring{\vec{B}} = [\vec{B}]_{v(\pm)=0}$ be the magnetic field in the local fluid rest-frame.

Then from (2–38) we have

$$[(\vec{\nabla}\times\vec{v})_{v=0}]_{(\pm)} = -\frac{q}{\mathring{m}_{(\pm)}c}\mathring{\vec{B}} \qquad (2\text{–}42)$$

The left-hand side is the fluid vorticity The presence of vorticity means that at the microscopic level the fluid can be regarded as rotating like a rigid body with the angular velocity $\mathring{\vec{\Omega}}_{(\pm)}$ where

$$\mathring{\Omega}_{(\pm)} = \tfrac{1}{2}[(\vec{\nabla}\times\vec{v})_{v=0}]_{(\pm)} \qquad (2\text{–}43)$$

Thus we have

$$\mathring{\vec{\Omega}}_{(\pm)} = \mp\frac{q}{2\mathring{m}_{(\pm)}c}\mathring{\vec{B}} \qquad (2\text{–}44)$$

which is just the Larmor Condition. The physical interpretation of this condition is that there can be no local vorticity, or rotation, in the charged fluid without a magnetic field to provide, via the Lorentz force, the necessary Coriolis force.

From (2–38) we note that large-scale vorticity ($\vec{v}\neq 0$) is possible even in the absence of a magnetic field, because then the gravitational force and/or the pressure gradient can provide the necessary Coriolis and centripetal forces.

Thus we see that for a charge *fluid*, (2–38) is a physically necessary condition. It must be emphasized that the Larmor Condition is necessary for a fluid, but not for a single particle. In the case of a single particle moving in a circular orbit, for example, the condition relating angular velocity and magnetic field is that the Lorentz force must provide the necessary centripetal force. This condition, which defines the so-called cyclotron frequency, is identical to (2–44) except that the factor 2 is absent in the denominator on the right. The Euler equation guarantees that this condition is fulfilled, since it is just the generalization of the condition on the Lorentz force for arbitrary particle orbits.

In the case of a fluid, we must allow for the fact that, in addition to the rotational velocity which requires a centripetal force, there is a random thermal velocity which, together with the rotation of the fluid, gives rise to a Coriolis accelcration which must also be taken into account. For given angular velocity, the required centripetal force is proportional to the size of the fluid vortex being considered, and so vanishes in the limit of zero vortex radius, which is the limit involved in the evaluation of $\vec{\nabla}\times\vec{v}$ in (2–43). The Coriolis force, however, remains finite in this limit since the thermal velocity on which it depends is independent of vortex size. This is the reason why, for a fluid, the Lorentz force must be equated to the Coriolis force alone.

We noted that, in the case of a single particle, the Euler equation itself guaranteed that the Lorentz force would have the proper value in relation to the characteristics of the particle orbit. In the case of a fluid, however, an extra side condition, namely (2–38), is needed to guarantee that the magnetic field strength will have the proper relation to the vorticity. The reason for the need for an extra side condition is that this condition arises from the presence of thermal velocity, which is not represented in the Euler equation. (Only the *energy* of thermal motion is represented via the enthalpy, but the detailed *momentum* and *force* considerations associated with thermal motion—namely the need for a Coriolis force—cannot be represented in Euler's equation.)

Thus, we conclude that the solutions of (2–37) which do *not* satisfy the condition (2–38), and hence are lost when we use (2–40) instead of

(2–37), are those corresponding to an absolutely cold fluid with no thermal motion. In such a case there is no need to impose the Larmor Condition, and the problem of calculating fluid flow is no different from that of calculating particle orbits. If, however, thermal motions are present in the fluid (i.e., $h \neq 0$), the Larmor Condition must be imposed. Except for this extra side condition, the fluid problem is still the same as the problem of calculating particle orbits (except that we must introduce the thermal potential h), since solutions of the tensor equation (2–40) are always solutions of (2–24), which has the form of a single particle equation.

We have used above a physical argument to convince ourselves that the tensor equation is more fundamental (for real fluids) than Euler's equation. In the next section we shall use a more formal argument to arrive at the same conslusion, in that we shall show that the Hamilton-Jacobian formulation of the magnetogas-dynamical problem leads to the tensor equation, rather than to Euler's equation. To simplify the discussion, as well as to lend it greater physical signifiance, we shall present the Hamilton-Jacobian formulation as the incorporation of the de Broglie Hypothesis, regarded as an experimental fact, into the formalism.

III HAMILTON—JACOBI FORMULATION

We take as our starting point the following well-known expression for the canonical particle momentum p^j in the presence of an electromagnetic field represented by the 4-vector potential A^j:

$$p^j_{(\pm)} = (\check{m}cu^j)_{(\pm)} \pm \frac{q}{c}A^j \tag{3–1}$$

The feature of the Hamilton-Jacobi formalism that is vital to our present considerations is the fact that p^j can be represented as the gradient of a scalar function Φ, which is Hamilton's Characteristic Function. Thus

$$p^j_{(\pm)} = -\partial^j \Phi_{(\pm)} \tag{3–2}$$

or, in three-dimensional form

$$p^0_{(\pm)} = -\frac{1}{c}\frac{\partial \Phi_{(\pm)}}{\partial t} \tag{3–3a}$$

$$\vec{p}_{(\pm)} = (p^1, p^2, p^3) = \vec{\nabla}\Phi_{(\pm)} \tag{3–3b}$$

From the point of view of the Hamilton-Jacobi formalism, (3–2) is part of a contact transformation that carries the Hamiltonian function over into a form that allows an easy solution. It is physically more significant, however, if we take the point of view that (3–2) is just the expression of the de Broglie Hypothesis, which we regard as an experimental fact that must be incorporated into our theory. In its usual form, the de Broglie Hypothesis states that for every *particle* there exists a scalar function whose gradient is the particle momentum. In order to adapt this statement to a *fluid*, it is necessary to impose the requirement that Φ be a *continuous* function of the space-time coordinates. Thus, in the same way that the fluid picture of a gas (as opposed to the kinetic theory picture) requires that the fluid velocity must not vary in a rapid and random way, but rather must be a smoothly varying function, we require that the phase function of the fluid must also be a smoothly varying function. This continuity postulate is implicit in (3–2).

Substituting (3–2) into (3–1), we have

$$-\partial^j \Phi_{(\pm)} = (\check{m}cu^j)_{(\pm)} \pm \frac{q}{c}A^j \tag{3–4}$$

Taking the curl of (3–4), we have

$$\partial^j(\check{m}cu^k)_{(\pm)} - \partial^k(\check{m}cu^j)_{(\pm)} = \mp\frac{q}{c}(\partial^j A^k - \partial^k A^j) = \mp\frac{q}{c}F^{jk} \tag{3–5}$$

where we have used the fact that the curl of A^j is the Maxwell field tensor F^{jk}. This equation is just the tensor alternative (2–40) to Euler's equation that was introduced in Section II. As was noted there, this equation (like Euler's) must be supplemented by the continuity and normalization conditions:

$$\partial_j(\rho u^j)_{(\pm)} = 0 \tag{3–6}$$

$$(u^j u_j)_{(\pm)} = 1 \tag{3–7}$$

If we solve (3.4) for $u^j_{(\pm)}$ and substitute into (3–7), we obtain

$$\left(\partial_j \Phi_{(\pm)} \pm \frac{q}{c}A_j\right)\left(\partial^j \Phi_{(\pm)} \pm \frac{q}{c}A^j\right) = (\check{m}_{(\pm)}c)^2 \tag{3–8}$$

which is just the well-known Hamilton-Jacobi equation.

If we regard A^j and $\check{m}_{(\pm)}$ as given space-time functions, then in principle we can solve (3–8) for $\Phi_{(\pm)}$ and use this in (3–4) to find $u^j_{(\pm)}$, which would then automatically satisfy (3–7). (It would still be necessary to impose (3–6).)

Alternatively, we could ignore (3–8), and regard (3–5) – (3–7) as our basic system of equations.

To return to the question of weighing the tensor equation (3–5) against Euler's equation (2–24), we note that if we contract (3–5) with u_j, and use the relations

$$u_j\partial^j=\frac{1}{c}\frac{d}{d\tau} \tag{3–9}$$

and

$$u_j\partial^k u^j=\tfrac{1}{2}\partial^k(u_ju^j)=0 \tag{3–10}$$

the result is

$$\left[\frac{d(\check{m}cu^k)}{d\tau}\right]_{(\pm)}=\partial^k(\check{m}c^2)_{(\pm)}\pm qF^{kj}u_j^{(\pm)} \tag{3–11}$$

which is just Euler's equation.

Thus we have seen that it is the tensor equation, rather than Euler's equation, that results from the Hamilton-Jacobi formulation of the problem, and that Euler's equation (which is nonlinear) results from the tensor equation (which is linear) by contracting the latter with u_j. It is this contraction process that introduces the term quadratic in u_j, and hence the nonlinearity, into Euler's equation. Thus the nonlinearity of Euler's equation can be regarded as physically spurious, resulting from the unnecessary contraction of the tensor equation with u_j.

IV FIELD EQUATIONS

It was noted in Section I that the purpose of this paper is to transform the dynamical half of the magnetogas-dynamical problem, replacing Euler's equation by a spinor equation. The source equations for the electromagnetic and gravitational fields will be just Maxwell's equation and the four-dimensional Poisson equation respectively. In order to define the notation and to provide a convenient reference, the relevant field equations will be briefly discussed in this section.

Electromagnetic Field

The definition for the elements of the Maxwell field tensor F^{jk} in terms of the electric field intensity $\vec{E}$ and the magnetic flux density $\vec{B}$ has been given in (2–20). The four-dimensional form of Maxwell's equations (in unrationalized Gaussian units) is

$$\partial_jF^{jk}=4\pi q[(\rho u^k)_{(+)}-(\rho u^k)_{(-)}] \tag{4–1}$$

where $q=4.802\times10^{-10}$ statcoulomb is the magnitude (always positive) of the electron and proton charge, and $(\rho u^k)_{(+)}$ and $(\rho u^k)_{(-)}$ are the flux densities of the proton and electron gases respectively. The three-dimensional form of (4–1) is

$$\vec{\nabla}\cdot\vec{E}=4\pi q[(\rho u^0)_{(+)}-(\rho u^0)_{(-)}] \tag{4–2a}$$

$$\vec{\nabla}\times\vec{B}=\frac{4\pi}{c}q[(\rho u^0\vec{v})_{(\pm)}-(\rho u^0\vec{v})_{(-)}]+\frac{1}{c}\frac{\partial\vec{E}}{\partial t} \tag{4–2b}$$

The field tensor F^{jk} is related to the 4-vector potential in the usual way:

$$F^{jk}=\partial^jA^k-\partial^kA^j \tag{4–3}$$

The vector equivalent of (4–1) is

$$\Box A^k-\partial^k(\partial_jA^j)=4\pi q[(\rho u^k)_{(+)}-(\rho u^k)_{(-)}] \tag{4–4}$$

where $\Box$ is the d'Alembertian operator defined in (1–6). The choice of gauge in (4–4) has been left arbitrary.

Let $\hat{F}^{jk}$ be the dual to F^{jk}. It is defined as follows:

$$\hat{F}^{jk}\equiv\tfrac{1}{2}\epsilon^{jkln}F_{ln}=\epsilon^{jkln}\partial_lA_n \tag{4–5}$$

where ϵ^{jkln} is the antisymmetric unit tensor defined at the end of Section I. Using (2–20) we have

$$(\hat{F}^{10},\hat{F}^{20},\hat{F}^{30})=(B_x,B_y,B_z)=\vec{B}$$

$$(\hat{F}^{23},\hat{F}^{31},\hat{F}^{12})=E_x,E_y,E_z)=\vec{E}$$

$$\hat{F}^{jk}=-\hat{F}^{kj} \tag{4–6}$$

From (4–5) we have

$$\partial_j\hat{F}^{jk}=\epsilon^{jkln}\partial_j\partial_lA_n=0 \tag{4–7}$$

Using (4–6), we have for the three-dimensional form of (4–7)

$$\vec{\nabla}\cdot\vec{B}=0 \tag{4–8a}$$

$$\vec{\nabla}\times\vec{E}+\frac{1}{c}\frac{\partial\vec{B}}{\partial t}=0 \tag{4–8b}$$

which, together with (4–2), constitute Maxwell's equations in three-dimensional form. We note that the equations (4–8) follow directly from the introduction of the 4-vector potential; that is, these equations are the justification for introducing the 4-vector potential.

We remark, incidentally, that using (4–5), it is easy to show that (4–7) can be written in the following alternative form:

$$\partial^j F^{kl}+\partial^k F^{lj}+\partial^l F^{jk}=0 \tag{4–9}$$

We note that, because of the antisymmetry of F^{jk}, the left side of this equation vanishes identically when any two indices have the same value.

Let E^{jk} be the electromagnetic stress-energy tensor. As is well-known, it has the following form:

$$E^{jk}=\frac{1}{4\pi}\left[F^{jl}F^{k}_{\ l}+\frac{1}{4}g^{jk}(F^{ln}F_{ln})\right] \tag{4–10}$$

where g^{jk} is the metric tensor specified in (1–1). Thus $4\pi E^{jk}$ may be regarded as the square (in the sense of matrix multiplication) of the field tensor F^{jk}, plus the diagonal tensor $\frac{1}{4}g^{jk}(F^{ln}F_{ln})$ where, using (2–20), we find

$$\tfrac{1}{4}(F^{ln}F_{ln})=\tfrac{1}{2}(\vec{B}^2-\vec{E}^2) \tag{4–11}$$

Using (4–10) and (2–20), we find the following well-known expressions for the elements of E^{jk}:

$$E^{00}=\frac{1}{8\pi}(\vec{E}^2+\vec{B}^2)$$
$$\equiv\text{electromagnetic energy density} \tag{4–12a}$$

$$c(E^{01}, E^{02}, E^{03})=c(E^{10}, E^{20}, E^{30})=\frac{c}{4\pi}(\vec{E}\times\vec{B})$$
$$\equiv\vec{U}_{em}\equiv\text{electromagnetic energy flux density (Poynting's vector)} \tag{4–12b}$$

$$E^{jk}=-\left[E^{00}g^{jk}+\frac{1}{4\pi}(E^jE^k+B^jB^k)\right]$$
$$=\text{electromagnetic stress tensor}$$
$$(j, k=1, 2, 3) \tag{4–12c}$$

where

$$(E^1, E^2, E^3)\equiv(E_x, E_y, E_z)$$
$$(B^1, B^2, B^3)\equiv(B_x, B_y, B_z) \tag{4–12d}$$

Using the fact that

$$g^j_{\ j}=4 \tag{4–13}$$

it is evident from (4–10) that

$$E^j_{\ j}=0 \tag{4–14}$$

Using (4–1), (4–9), and (4–10), we find the following well-known relation:

$$\partial_l E^{kl}=-gF^{kl}[(\rho u_l)_{(+)}-(\rho u_l)_{(-)}] \tag{4–15}$$

The time-like part of this, which is the statement of the conservation of energy, can be written as follows:

$$\frac{\partial E^{00}}{\partial t}+\vec{\nabla}\cdot\vec{U}_{em}=-q\vec{E}\cdot[(\rho u^0\vec{v})_{(+)}-(\rho u^0\vec{v})_{(-)}] \tag{4–16}$$

The expression on the right side represents the energy transferred per unit time per unit volume *from* the charged fluids *to* the electromagnetic field.

Gravitational Field

We postulate the following source equation for the scalar gravitational potential g:

$$\Box g=-4\pi\Gamma[(\rho m)_{(+)}+(\rho m)_{(-)}] \tag{4–17}$$

where $\Gamma=6.668\times10^{-8}$ dyne cm^2/gm^2 is the gravitational constant.

For the time-independent case in which the mass on the right is distributed with spherical symmetry, we find for $(-\vec{\nabla}g)$, which is the gravitational force per unit mass, the following:

$$-\vec{\nabla}g=-\frac{\Gamma m_s}{r^2}\vec{r}_1 \tag{4–18}$$

where $\vec{r}_1$ is the unit radial vector, and m_s is the total mass contained within the sphere of radius r. This, of course, is just the familiar Newtonian gravitational force. Thus (4–17) corresponds to a suitably covariant extension of Newtonian gravitation.

Because we are employing scalar theory, however, rather than a tensor theory as in General Relativity (in which the symmetric metric tensor effectively plays the role of a gravitational potential), we must expect to find that the velocity dependence of the gravitational force on a moving mass is not correct to higher orders of v/c. Moreover, since only the rest-mass appears on the right side of (4–17), our scalar theory neglects the gravitational effect of the mass arising from the thermal, electromagnetic, and gravitational energy of the electron and proton gases. However, these errors are negligible in most magnetogasdynamical problems of physical interest.

Let G^{jk} be the stress-energy tensor of the scalar gravitational field. It has the following form:

$$G^{jk}=\frac{1}{4\pi\Gamma}\left[(\partial^j g)(\partial^k g)-\frac{1}{2}g^{jk}(\partial^l g)(\partial_l g)\right] \tag{4–19}$$

That this is indeed the correct expression for G^{jk} can be verified by calculating $\partial_j G^{jk}$:

$$\partial_l G^{kl}=\frac{1}{4\pi\Gamma}[(\partial_j\partial^j g)\partial^k g+(\partial_j g)(\partial^j\partial^k g)-(\partial_l g)(\partial^k\partial^l g)]=\frac{1}{4\pi\Gamma}[(\Box g)(\partial^k g)]$$
$$=-[(\rho m)_{(+)}+(\rho m)_{(-)}]\partial^k g \tag{4–20}$$

The right side of (4–20) is just the negative of the gravitational force density acting on the two fluid fields. This is completely analogous with (4–15) in which the right side is the negative of the electromagnetic force density acting on the two charged fluid fields.

From (4–19) we find the following expressions for the elements of G^{jk}:

$$G^{00}=\frac{1}{8\pi\Gamma}\left[\frac{1}{c^2}\left(\frac{\partial g}{\partial t}\right)^2+(\vec{\nabla}g)^2\right] \tag{4–21a}$$
$\equiv$ gravitational energy density

$$c(G^{01}, G^{02}, G^{03})=c(G^{10}, G^{20}, G^{30})=-\frac{1}{4\pi\Gamma}\left(\frac{\partial g}{\partial t}\right)(\vec{\nabla}g)$$
$\equiv \vec{U}_{gr} \equiv$ gravitational energy flux density 4(–21b)

$$G^{jk}=\frac{1}{4\pi\Gamma}\left\{\frac{1}{2}\left[(\vec{\nabla}g)^2-\frac{1}{c}\left(\frac{\partial g}{\partial t}\right)^2\right]g^{jk}+(\partial^j g)(\partial^k g)\right\}$$
$(j, k=1, 2, 3) \equiv$ gravitational stress tensor (4–21c)

From (4–13) and (4–19) we find

$$G^j=-\frac{1}{4\pi\Gamma}(\partial^l g)(\partial_l g)=\frac{1}{4\pi\Gamma}\left[(\vec{\nabla}g)^2-\frac{1}{c^2}\left(\frac{\partial g}{\partial t}\right)^2\right] \tag{4–22}$$

Finally, we note that the time-like component of (4–20), which is the statement of conservation of energy, can be written as follows:

$$\frac{\partial G^{00}}{\partial t}+\vec{\nabla}\cdot\vec{U}_{gr}=-[(\rho m)_{(+)}+(\rho m)_{(-)}]\frac{\partial g}{\partial t} \tag{4–23}$$

where the right-hand side is just the energy transferred per unit time per unit volume *from* the two fluids *to* the gravitational field.

V SPIN EFFECTS

As was noted in Section I, we must take particle spin into account, not because it is physically significant in magnetogasdynamical problems, but rather because it is the price we must pay for the benefits of using the spinor formulation of the problem.

Magnetic Moment

Spin makes itself felt in two ways: through its magnetic moment, and through its angular momentum. We shall consider first the effects of the magnetic moment.

Let the direction of the magnetic moment of a particle in its rest-frame be specified by the 3-vector $\vec{\mu}_{(\pm)}$, which can be generalized to the 4-vector $\mu^j{}_{(\pm)}$ by specifying that the time-like component be zero in the particle rest-frame. We specify the magnitude of the magnetic moment of a particle by the scalar $\mu_{(\pm)}$. (We allow the possibility that $\mu_{(\pm)}$ is not a constant, but rather a function of the fluid and field variables. In Sector VIII we shall see that the functional form of $\mu_{(\pm)}$ is automatically specified by the spinor equations.)

The magnetic moment of a particle may also be described in terms of the antisymmetric tensor $(\mu s^{jk})_{(\pm)}$ where $s^{jk}{}_{(\pm)}$ is derived from $\mu^j{}_{(\pm)}$ and the particle 4-velocity $u^j{}_{(\pm)}$ as follows:

$$s^{jk}{}_{(\pm)} \equiv \epsilon^{jklm}(\mu_l u^n)_{(\pm)} \tag{5–1}$$

Thus

$$[\rho\mu(s^{23}, s^{31}, s^{12})]_{(\pm)}=[(\rho u^0)(\mu\dot{\vec{\mu}})]_{(\pm)} \tag{5–2a}$$

$$[\rho\mu(s^{10}, s^{20}, s^{30})]_{(\pm)}=\left\{(\rho u^0)\left[\frac{1}{c}\vec{v}\times(\mu\dot{\vec{\mu}})\right]\right\}_{(\pm)} \tag{5–2b}$$

where in (5–2b) we have made use of the fact that $\vec{v}x\vec{\mu}=\vec{v}x\dot{\vec{\mu}}$, which follows directly from the form of the Lorentz transformation.

Since ρu° is just the particle density in the observer's frame, (5–2a) tells us that the space-space part of $\rho\mu s^{jk}$ is just the magnetization of the fluid. We note that the cross-product inside the brackets on the right side of (5–2b) is just the well-known electric dipole associated with a moving magnetic dipole. Thus, even though electrons and protons have no electric dipole moment at rest, a spin-dependent electric polarization of the fluid is possible when there is motion.

Now we know that in a magnetized body there is an effective current density given by the curl of the magnetization. By analogy, we would expect a 4-momentum density to be associated with the angular momentum density arising from particle spin. If $\delta p^j{}_{(\pm)}$ is this extra 4-momentum density divided by ρ, i.e. the extra momentum per particle, than (3–4) must be replaced by

$$-\partial^j\Phi_{(\pm)}=(\tilde{m}cu^j)_{(\pm)}\pm\frac{q}{c}A^j+\delta p^j_{(\pm)} \qquad (5\text{–}3)$$

The form of this equation indicates that it should be possible to absorb $\delta p^j{}_{(\pm)}$ into the 4-potential, i.e., to express the spin effects as a small correction to the electromagnetic field. In order to be thorough about this, we must keep in mind that every particle has a contribution to its rest-mass $\Delta m_{(\pm)}$ arising from the interaction of the magnetic field and the magnetic dipole moment. If $\dot{\vec{B}}$ and $(\dot{\vec{\mu\mu}})_{(\pm)}$ are respectively the magnetic field and the magnetic dipole moment in the particle rest-frame, then the magnetic contribution to the particle rest-mass is

$$\Delta m_{(\pm)}=-\frac{1}{c^2}(\dot{\vec{\mu\mu}})_{(\pm)}\cdot\dot{\vec{B}}=\frac{1}{2c^2}(\mu s^{jk})_{(\pm)}F_{jk} \qquad (5\text{–}4)$$

where, in the second step, we have made use of (2–20) and (5–2).

If we wish to express the spins effects in terms of a small correction $a^j{}_{(\pm)}$ to the 4-vector potential A^j, then we must introduce a corresponding correction $f^{jk}{}_{(\pm)}$ to the field tensor F^{jk} where, in analogy, to (4–3), we have

$$f^{jk}{}_{(\pm)}=\partial^j a^k{}_{(\pm)}-\partial^k a^j{}_{(\pm)} \qquad (5\text{–}5)$$

Analogous to (2–20), we can define corrections $\vec{e}_{(\pm)}$ and $\vec{b}_{(\pm)}$ to the electric and magnetic field intensities $\vec{E}$ and $\vec{B}$ as follows:

$$\begin{aligned}&(f^{10}, f^{20}, f^{30})_{(\pm)}\equiv(e_x, e_y, e_z)_{(\pm)}=\vec{e}_{(\pm)}\\&(f^{23}, f^{31}, f^{12})_{(\pm)}\equiv-(b_x, b_y, b_z)_{(\pm)}=-\vec{b}_{(\pm)}\\&f^{jk}{}_{(\pm)}=-f^{jk}{}_{(\pm)}\end{aligned} \qquad (5\text{–}6)$$

If $f^{jk}{}_{(\pm)}$ is to be regarded as completely analogous to F^{jk}, then it should produce a contribution $\delta m_{(\pm)}$ to the particle rest-mass having the form of (5–4) with F_{jk} replaced by $f_{jk}{}^{(\pm)}$:

$$\delta m_{(\pm)}=-\frac{1}{c^2}[(\dot{\vec{\mu\mu}})\cdot\dot{\vec{b}}]_{(\pm)}=\frac{1}{2c^2}(\mu s^{jk}f_{jk})_{(\pm)} \qquad (5\text{–}7)$$

Let $\tilde{m}_{(\pm)}$ be the total magnetic contribution to the particle rest-mass. Then

$$\begin{aligned}\tilde{m}_{(\pm)}=\Delta m_{(\pm)}+\delta m_{(\pm)}&=-\frac{1}{c^2}[(\dot{\vec{\mu\mu}})\cdot(\dot{\vec{B}}+\dot{\vec{b}})]_{(\pm)}\\&=\frac{1}{2c^2}(\mu s_{jk})_{(\pm)}(F^{jk}+f^{jk}{}_{(\pm)})\end{aligned} \qquad (5\text{–}8)$$

If we regard $u^j{}_{(\pm)}$, $(\mu s^{jk})_{(\pm)}$, and $\delta p^j{}_{(\pm)}$ as known functions, then it is evident from (5–3) that $a^j{}_{(\pm)}$, is determined by the condition

$$\delta p^j_{(\pm)}=(\tilde{m}cu^j)_{(\pm)}\pm\frac{q}{c}a^j{}_{(\pm)} \qquad (5\text{–}9)$$

where from (5–5) and (5–8) we see that $\tilde{m}_{(\pm)}$ is a function of the derivatives of $a^j{}_{(\pm)}$. Thus (5–9) is a first-order differential equation for $a^j{}_{(\pm)}$. In actual fact, it will not be necessary to solve such an equation since, as we shall see in Section VIII, the spinor equations give us an expression for $a^j{}_{(\pm)}$ that automatically satisfies (5–9).

If we let $M_{(\pm)}$ be the total particle rest-mass, we have from (2–17) and (5–8)

$$\begin{aligned}M_{(\pm)}=\tilde{m}_{(\pm)}+\tilde{m}_{(\mp)}&=m_{(\pm)}(1+g/c^2+h_{(\pm)}/c^2)\\&+\frac{1}{2c^2}(\mu s_{jk})_{(\pm)}[F^{jk}+f^{jk}_{(\pm)}]\end{aligned} \qquad (5\text{–}10)$$

Substituting (5–9) and (5–10) into (5–3), we obtain

$$-\partial^j\Phi=(Mcu^j)_{(\pm)}\pm\frac{q}{c}[A^j+a^j_{(\pm)}] \qquad (5\text{–}11)$$

This way of formulating the dynamical effects of spin gives us an immediate intuitive feeling for these effects. Moreover it provides a neat way of determining when these effects can be neglected: If we can show that $|a^{j}_{(\pm)}| \ll |A^{j}|$, then we know that the dynamical effects of spin are negligible. We shall see in Section VIII that it is very easy to estimate the magnitude of the $a^{j}_{(\pm)}$ that results from the spinor equations, and that, in all magnetogas-dynamical problems on a laboratory or astronomical scale, $a^{j}_{(\pm)}$ is completely negligible compared with the magnitudes of A^{j} encountered in such problems. If, however, we were to attempt to apply our spinor formulation of magnetogas dynamics on an atomic or nuclear scale, then it turns out that the spin effects represented by the effective potential $a^{j}_{(\pm)}$ would be far from negligible.

We note that in general $a^{j}_{(+)} \neq a^{j}_{(-)}$. This is not surprising since the correction potentials $a^{j}_{(+)}$ and $a^{j}_{(-)}$ have their origins in dynamical effects of the spin fields $(\rho\mu^{j})_{(+)}$ and $(\rho\mu^{j})_{(-)}$ which are in general different. From a purely formal point of view, the equations that follow treat $a_{(+)}{}^{j}$, for example, as an electro-magnetic field which produces a self-interaction within the proton gas, but which produces no interaction between the electron and proton gases. Similarly for $a^{j}_{(-)}$. In this respect, the effects of $a^{j}_{(\pm)}$ are analogous to the behavior of the exchange forces in quantum mechanics which act only between identical particles, and not between unlike particles.

To complete the derivation of the dynamical equations including the effects of spin, we take the curl of (5–11), and arrive at the following:

$$\partial^{j}(Mcu^{k})_{(\pm)} - \partial^{k}(Mcu^{j})_{(\pm)} = \mp\frac{q}{c}[F^{jk} + f^{jk}_{(\pm)}] \qquad (5\text{–}12)$$

This is just the generalization of the tensor equation (2–40). Contracting (5–12) with $\mu_{j}{}^{(\pm)}$, we arrive at the following generalization of the Euler equation in the form (2–24):

$$\left[\frac{d(Mcu^{k})}{d\tau}\right]_{(\pm)} = \partial^{k}(Mc^{2})_{(\pm)} \pm q[F^{kl} + f^{kl}_{(\pm)}]u_{l}{}^{(\pm)} \qquad (5\text{–}13)$$

Fluid Stress-Energy Tensor

As a preliminary to introducing the stress-energy tensor for each of the two fluid fields, we multiply (5–13) by $\rho_{(\pm)}$ and write it in the following form:

$$\left[\rho\frac{d(\tilde{m}cu^{j})}{d\tau}\right]_{(\pm)} + \left[\rho\frac{d(\tilde{m}cu^{j})}{d\tau} - \rho\partial^{j}(\tilde{m}c^{2}) \mp q\rho f^{jl}u_{l}\right]_{(\pm)} = k^{j}_{(\pm)} \qquad (5\text{–}14)$$

where $k^{j}_{(\pm)}$ is the force density acting on the proton or electron gas:

$$k^{j}_{(\pm)} = (\rho m)_{(\pm)}\partial^{j}g + \rho^{j}p_{(\pm)} \pm qF^{jl}(\rho u_{l})_{(\pm)} = \rho_{(\pm)}\partial^{j}(\tilde{m}c^{2})_{(\pm)} \pm qF^{jl}(\rho\mu_{l})_{(\pm)} \qquad (5\text{–}15)$$

We define the fluid stress-energy tensor $T_{(\pm)}{}^{jk}$ for either the proton or electron gas as follows:

$$T^{jk}{}_{(\pm)} = (\rho\tilde{m}c^{2}u^{j}\mu^{k})_{(\pm)} + t^{jk}{}_{(\pm)} \qquad (5\text{–}16)$$

where the tensor $t^{jk}{}_{(\pm)}$ contains all the spin dependences. We impose the following condition on $t^{jk}{}_{(\pm)}$:

$$\partial_{k}t^{jk}_{(\pm)} = \left[\rho\frac{d(\tilde{m}cu^{j})}{d\tau} - \rho\partial^{j}(\tilde{m}c^{2}) \mp q\rho f^{jl}u_{l}\right]_{(\pm)} \qquad (5\text{–}17)$$

(In Section VIII we shall see that the spinor equations specify the functional dependence of $t^{jk}{}_{(\pm)}$ in such a way that condition (5–17) is automatically satisfied.)

We augment (5–17) with the continuity condition:

$$\partial_{j}(\rho\mu^{j})_{(\pm)} = 0 \qquad (5\text{–}18)$$

These two conditions, together with the definitions (5–16), suffice to guarantee the validity of the following relation:

$$\partial_{l}T^{jl}{}_{(\pm)} = k^{j}{}_{(\pm)} \qquad (5\text{–}19)$$

This equation, which is identical with (5–13) when (5–17) and (5–18) are taken into account, is the justification for regarding the tensor defined in (5–16) as the fluid stress-energy tensor.

We now make the following identifications:

$$\text{fluid energy density} \equiv T_{(\pm)}{}^{00} \qquad (5\text{–}20a)$$

$$\text{fluid energy flux density} \equiv \vec{U}_{(\pm)} \equiv c(T^{01}, T^{02}, T^{03})_{(\pm)} \qquad (5\text{–}20b)$$

$$\text{fluid momentum density} \equiv \vec{G}_{\pm} = (G^{1}, G^{2}, G^{3})_{\pm} \equiv \left(\frac{1}{c}T^{10}T^{20}T^{30}\right)_{\pm} \qquad (5\text{–}20c)$$

In terms of these quantities, (5–19) can be written as follows:

$$\frac{\partial T^{00}_{(\pm)}}{\partial t}+\vec{\nabla}\cdot\vec{U}_{(\pm)}=ck^{0}_{(\pm)} \tag{5–21a}$$

$$\frac{\partial G^{j}_{(\pm)}}{\partial t}=-\partial_l T^{jl}_{(\pm)}+k^{j}_{(\pm)}(j,\ l=1,\ 2,\ 3) \tag{5–21b}$$

These equations are the justification for the identifications made in (5–20).

Fluid Angular Momentum Tensor

If the spin angular momentum density of either the proton or electron gas is designated by $S^{jkl}{}_{(\pm)}$, and the total angular momentum density of either charged gas by $M^{jkl}{}_{(\pm)}$, then

$$M^{kjl}_{(\pm)}\equiv S^{jkl}_{(\pm)}+\frac{1}{c}[x^j T^{kl}_{(\pm)}-x^k T^{jl}_{(\pm)}] \tag{5–22}$$

The intuitive significance of this definition is more apparent if it is written in the following form:

$$\vec{M}_{(\pm)}=\vec{S}_{(\pm)}+\vec{r}\times\vec{G}_{(\pm)} \tag{5–23}$$

where

$$\vec{M}_{(\pm)}=(M_x,\ M_y,\ M_z)_{(\pm)}\equiv(M^{230},\ M^{310},\ M^{120})_{(\pm)} \tag{5–24}$$

$$\vec{S}_{(\pm)}=(S_x,\ S_y,\ S_z)_{(\pm)}\equiv(S^{230},\ S^{310},\ S^{120})_{(\pm)} \tag{5–25}$$

$$\vec{r}=(x,\ y,\ z)\equiv(x^1,\ x^2,\ x^3) \tag{5–26}$$

and $\vec{G}_{(\pm)}$ is defined in (5–20c). The 3-vectors $\vec{S}_{(\pm)}$ and $\vec{M}_{(\pm)}$ are the quantities to be regarded respectively as the spin and total angular momentum densities, but the complete tensors $S^{jkl}{}_{(\pm)}$ and $M^{jkl}{}_{(\pm)}$ are necessary for a covariant formulation of the theory.

The conservation of angular momentum requires that the following condition be satisfied:

$$\partial_l M^{jkl}_{(\pm)}=\frac{1}{c}[x^j k^{k}_{(\pm)}-x^k k^{j}_{(\pm)}] \tag{5–27}$$

The integral form of this equation makes its significance more apparent. To derive this we first write the space-space part of (5–27) as follows:

$$\frac{\partial\vec{M}_{(\pm)}}{\partial t}+c\partial_l\vec{M}^{l}_{(\pm)}=\vec{r}\times\vec{k}_{(\pm)}\ (l=1,\ 2,\ 3) \tag{5–28a}$$

where

$$\vec{M}^{l}_{(\pm)}=(M^{l}_{x},\ M^{l}_{y},\ M^{l}_{z})_{(\pm)}$$
$$\equiv(M^{23l},\ M^{31l},\ M^{12l})_{(\pm)}(l=1,\ 2,\ 3) \tag{5–28b}$$

$$\vec{k}_{(\pm)}=(k_x,\ k_y,\ k_{z\,(\pm)}\equiv k^1,\ k^2,\ k^3)_{(\pm)} \tag{5–28c}$$

We integrate (5–28a) over the 3-volume V_3 contained within a large sphere of radius R. Since the term involving $\vec{M}l_{(\pm)}$ is a 3-divergence, it can be transformed by means of Gauss' theorem into an integral over the surface of the sphere. If the two charged fluids are contained completely within the sphere, or if their distribution is such that as $R\rightarrow\infty$, the magnitude of $\vec{M}^l$ on the surface of the sphere falls off faster than R^{-2} then the integral over all 3-space of the terms in (5–28a) involving $\vec{M}^l$ will vanish and we are led to the desired result:

$$\frac{d}{dt}\int_{V3}\vec{M}_{(\pm)}dV_3=\int_{V3}(\vec{r}\times\vec{k}_{(\pm)})dV_3 \tag{5–29}$$

where the integration extends over all 3-space. This equation states that the time rate of change of the total angular momentum of one of the charged gases equals the total torque acting on the gas.

Substituting (5–22) into (5–27) and using (5–19), we arrive at the following condition on $S^{jkl}{}_{(\pm)}$:

$$\partial_l S^{jkl}_{(\pm)}=\frac{1}{c}(T^{jk}-T^{kj})_{(\pm)}=\frac{1}{c}(t^{jk}-t^{kj})_{(\pm)} \tag{5–30}$$

This shows that when a fluid has spin, its stress-energy tensor cannot, in general, be completely symmetric. We shall see in Section VIII that the spinor equations specify the functional forms of $S^{jkl}{}_{(\pm)}$ and $t^{jk}{}_{(\pm)}$ in such a way that the condition (5–30) is automatically satisfied.

Total Stress-Energy Tensor

From (5–15) and (5–19) we have

$$\partial_l[T^{jl}{}_{(+)}+T^{jl}{}_{(-)}]=\partial^j[P_{(+)}+P_{(-)}]$$
$$+[(\rho m)_{(+)}+(\rho m)_{(-)}]\partial^j{}_g+qF^{jl}[(\rho\mu_l)_{(+)}$$
$$-(\rho\mu_l)_{(-)}]=\partial^j[P_{(+)}+P_{(-)}]-\partial_l G^{jl}-\partial E^{jl} \tag{5–31}$$

where we have made use of (4–15) and (4–20)

Thus

$$\partial_i T^{jl}=0 \tag{5-32}$$

where

$$\begin{aligned} T^{jl} &= T^{jl}{}_{(+)}+T^{jl}{}_{(-)}-[P_{(+)}+P_{(-)}]g^{jl}+G^{jl}+E^{jl} \\ &= (\rho\breve{m}c^2\mu^j\mu l-Pg^{jl})_{(+)}+(\rho\breve{m}c^2\mu^j\mu^l-Pg^{jl})_{(-)} \\ &\quad +t^{jl}_{(+)}+t^{jl}_{(-)}+G^{jl}+E^{jl} \end{aligned} \tag{5-33}$$

Obviously, T^{jl} is the total stress-energy tensor of the entire magnetogas-dynamical system including charged fluids and fields.

We note, incidentally, that using (2-1), (2-13$_j$, and (2-17) we have

$$(\rho\breve{m}c^2\mu^j\mu^l-Pg^{jl})_{(\pm)}=[(\epsilon+P)\mu^j\mu^l-Pg^{jl}]_{(\pm)} \tag{5-34}$$

where

$$\epsilon_{(\pm)}=[\rho(mc^2+mg+m\mu)]_{(\pm)} \tag{5-35}$$

is the energy density of the fluid in its rest-frame. The right side of (5-34) is the usual relativistic form for the stress-energy tensor of a perfect fluid (viscosity neglected).*

We now make the following identifications:

$$\text{total energy density}\equiv T^{00} \tag{5-36a}$$

$$\begin{aligned} \text{total energy flux density}\equiv \vec{U} &\equiv c(T^{01},\ T^{02},\ T^{03}) \\ &= \vec{U}_{(\pm)}+\vec{U}_{(-)}+\vec{U}_{em}+\vec{U}_{gr} \end{aligned} \tag{5-36b}$$

$$\begin{aligned} \text{total momentum density}\equiv \vec{G} &= (G^1,\ G^2,\ G^3) \\ &\equiv \frac{1}{c}(T^{10},\ T^{20},\ T^{30}) \\ &= \vec{G}_{(+)}+\vec{G}_{(-)}+\vec{G}_{em}+\vec{G}_{gr} \end{aligned} \tag{5-36c}$$

In terms of these quantities (5-32) can be written as follows:

$$\frac{\partial T^{\infty}}{\partial t}+\vec{\nabla}\cdot\vec{U}=0 \tag{5-37a}$$

$$\frac{\partial G^j}{\partial t}=-\partial_i T^{jl}\ (j,l=1,\ 2,\ 3) \tag{5-37b}$$

*See, for example, Møller (ref. 10) p. 182, eq. 104.

These equations, of course, are the justification for the identifications made in (5-36)

We can integrate the equations (5-37) in just the same way we integrated (5-28a). Doing this, we find

$$\epsilon=\int_{V_3} T^{00}dV_3=\text{constant} \tag{5-38}$$

$$\vec{p}=\int_{V_3}\vec{G}dV_3=\text{constant} \tag{5-39}$$

where ϵ is the total energy of the entire magnetogas-dynamical system, and $\vec{p}$ is the total momentum.

Total Angular Momentum Tensor

If M^{jkl} is the total angular momentum density tensor for the entire magnetogas-dynamical system, we have

$$M^{jkl}\equiv S^{jkl}_{(+)}+S^{jkl}_{(-)}+\frac{1}{c}(x^jT^{kl}-x^kT^{jl}) \tag{5-40}$$

The condition for conservation of angular momentum is

$$\begin{aligned} O=\partial_i M^{jkl} &= \partial_i[S^{jkl}_{(+)}+S^{jkl}_{(-)}]+\frac{1}{c}(T^{kj}-T^{jk}) \\ &= \left[\partial_i S^{jkl}-\frac{1}{c}(t^{jk}-t^{kj})\right]_{(+)}+\left[\partial_i S^{jkl}-\frac{1}{c}(t^{jk}-t^{kj})\right]_{(-)} \end{aligned} \tag{5-41}$$

where we have used the fact that $t^{jk}{}_{(+)}$ and $t^{jk}{}_{(-)}$ are the only parts of T^{jk} that contain any antisymmetry. Obviously, if the condition (5-30) is satisfied, conservation of total angular momentum follows automatically from the definition (5-40).

Integrating (5-41) over all 3-space in the same way we integrated (5-27), we have the following result:

$$M^{jk}=\int_{V_3} M^{jk0}dV_3=\text{constant} \tag{5-42}$$

where M^{jk} is the total angular momentum of the entire system. If we define the 3-vector momentum $\vec{M}$ as

$$\vec{M}=(M_x,\ M_y,\ M_z\equiv(M^{23},\ M^{31},\ M^{12}) \tag{5-43}$$

then from (5–42), (5–41), (5–36c) and (5–25) we have

$$\vec{M}=\int_{V_3}\vec{\mathrm{M}}dV_3=\int_{V_3}[\vec{S}_{(+)}+\vec{S}_{(-)}+\vec{r}\times\vec{G}]dV_3 \\ =\text{constant} \qquad (5\text{–}44)$$

It is possible to express the total angular momentum in terms of a 4-vector Σ^j instead of the antisymmetric tensor M^{jk} once we have defined the 4-velocity V^j of the entire system.

If $\vec{V}$ is the 3-velocity corresponding to V^j then by definition

$$V^0\equiv[1-\vec{V}^2/c^2]^{-1/2} \qquad (5\text{–}45a)$$

$$V^j\equiv\frac{V^0}{c}V^j \ (j=1,2,3) \qquad (5\text{–}45b)$$

$$V^jV_j=1 \qquad (5\text{–}45c)$$

V^j is defined in terms of ϵ and $\vec{p}=(p^1, p^2, p^3)$ as follows:

$$V^0\equiv\epsilon[\epsilon^2-c^2\vec{p}^2]^{-\frac{1}{2}} \qquad (5\text{–}46a)$$

$$V^j\equiv cp^j[\epsilon^2-c^2\vec{p}^2]^{-\frac{1}{2}} \\ (j=1,2,3) \qquad (5\text{–}46b)$$

From (5–45) and (5–46), it follows that

$$\vec{V}=\frac{\vec{p}}{(\epsilon/c^2)} \qquad (5\text{–}47)$$

where the quantity in the denominator on the right is just the total mass of the system in the observer's reference frame. The square root in the denominator of (5–46) corresponds to the rest-energy of the entire system.

Using the 4-velocity defined in (5–46), we define the 4-vector angular momentum Σ^j as follows:

$$\Sigma^j\equiv-\tfrac{1}{2}\epsilon^{jkln}V_kM_{ln} \qquad (5\text{–}48)$$

In the rest-frame of the total system ($\vec{V}=0$,) we have

$$\left.\begin{aligned}\overset{0}{\Sigma}{}^0&=0\\(\overset{0}{\Sigma}{}^1,\overset{0}{\Sigma}{}^2,\overset{0}{\Sigma}{}^3)&=\vec{M}\end{aligned}\right\}\ \text{for } \vec{V}=0 \qquad (5\text{–}49)$$

VI SUMMARY OF FOREGOING RESULTS

We have seen that the dynamical side of the magnetogas-dynamical problem can be specified by the following system of equations:

$$\left[\frac{d(Mcu^j)}{d\tau}\right]_{(\pm)}=\partial^j(Mc^2)_{(\pm)}\pm q[F^{jk}+f^{jk}_{(\pm)}]u_k^{(\pm)} \qquad (6\text{–}1a)$$

$$\partial_j(\rho u^j)_{(\pm)}=0 \qquad (6\text{–}1b)$$

$$(u_ju^j)_{(\pm)}=1 \qquad (6\text{–}1c)$$

where

$$M_{(\pm)}=\mathring{m}_{(\pm)}+\tilde{m}_{(\pm)} \qquad (6\text{–}1d)$$

$$\mathring{m}_{(\pm)}=m_{(\pm)}[1+g/c^2+h_{(\pm)}/c^2) \qquad (6\text{–}1e)$$

$$\tilde{m}_{(\pm)}=\frac{1}{2c^2}(\mu s_{jk})_{(\pm)}[F^{jk}+f^{jk}_{(\pm)}] \qquad (6\text{–}1f)$$

We impose the adiabatic condition

$$[\rho\partial j\ h-(\gamma-1)h\partial j\ \rho]_{(\pm)}=0 \qquad (6\text{–}2a)$$

or

$$h_{(\pm)}=[H(\rho/\rho_0)^{(\gamma-1)}]_{(\pm)} \qquad (6\text{–}2b)$$

where

$$\overset{0}{\rho}_{(\pm)}=[\rho(x^j{}_{(0)})]_{(\pm)};\quad \overset{0}{H}_{(\pm)}=[h(x^j{}_{(0)})]_{(\pm)} \qquad (6\text{–}2c)$$

where $x_{(0)}{}^j$ are the coordinates of a particular point in space-time. If the condition (6–2) is satisfied, then

$$[\rho\partial^j(Mc^2)]_{(\pm)}=(\rho m)_{(\pm)}\partial^j g+\partial^j p_{(\pm)} \\ +[\rho\partial^j(\tilde{m}c^2)]_{(\pm)} \qquad (6\text{–}3)$$

It is this relation which allows us to replace the force densities in Euler's equation that are associated with gravitation, pressure, and spin-field interaction by the particle density times the gradient of the total particle rest-energy $M_{(\pm)}C^2$.

The specific enthalpy $h_{(\pm)}$ is the only thermodynamic function we need know, since the other two thermodynamic functions of principal interest, $T_{(\pm)}$ and $P_{(\pm)}$, can be found from the hollowing relations:

$$T_{(\pm)}=(h/c_p)_{(\pm)} \qquad (6\text{–}4)$$

and

$$P_{(\pm)}=\left[(\rho mh)\left(1-\frac{1}{\gamma}\right)\right]_{(\pm)} \qquad (6\text{–}5)$$

We have seen that there exists a simplified alternative to the system (6–1):

$$-\partial^j \Phi_{(\pm)} = (Mcu^j)_{(\pm)} \pm \frac{q}{c}\left[A^j + a^j_{(\pm)}\right] \quad (6\text{–}6a)$$

$$\partial j\,(\rho u^j)_{(\pm)} = 0 \quad (6\text{–}6b)$$

$$(u_j\, u^j)_{(\pm)} = 0 \quad (6\text{–}6c)$$

Taking the curl of (6–9a) yields the tensor alternative to Euler's equation:

$$\partial^j (Mcu^k)_{(\pm)} - \partial^k (Mcu^j)_{(\pm)} = \mp\frac{q}{c}\left[F^{jk} + f^{jk}_{(\pm)}\right] \quad (6\text{–}7)$$

Contracting (6–7) with $u_j{}^{(\pm)}$ yields Euler's equation in the form (6–1a). Thus any solution of (6–7) or (6–6a) is also a solution of (6–1a), but the converse is not in general true. Thus (6–6a) and (6–7) correspond to a more restricted class of solutions than (6–1a). The source of this restriction can be stated in either of two ways: (1) The restriction is just the requirement that there can be no local vorticity in either charged gas without a magnetic field to provide the necessary Coriolis force arising from the random thermal motion; or (2) The restriction results from the requirement (3–2) that the generalized particle momentum must be the gradient of a scalar, which is just the incorporation of the de Broglie Hypothesis into the formalism.

Obviously a choice must be made between the systems (6–1) and (6–6), and we have chosen in favor of (6–6) which is the more fundamental formulation of magnetogas dynamics, because the added restriction it imposes is physically necessary for a charged fluid with random thermal motion.

Thus the fluid dynamical half of the problem is specified by the system (6–6). The field half of the problem is specified by the following group of equations:

$$[\rho\partial_j\, h - (\gamma-1)h\partial_j\,\rho]_{(\pm)} = 0 \quad (6\text{–}8a)$$

or

$$h_{(\pm)} = [H(\rho/\rho_0{}^{(\gamma-1)})]_{(\pm)} \quad (6\text{–}8b)$$

$$\Box\, A^j - \partial^j(\partial_k A^k) = 4\pi q[(\rho u^j)_{(+)} - (\rho u^j)_{(-)}] \quad (6\text{–}8c)$$

$$\Box\, g = -4\pi\Gamma[(\rho m)_{(+)} + (\rho m)_{(-)}] \quad (6\text{–}8d)$$

Equations (6–8a) and (6–8b) are just the equations (6–2), and have been repeated here to emphasize that we are treating $h_{(\pm)}$ as a "thermal field" on a par with the gravitational and electromagnetic fields represented by g and A^j. We have a choice between (6–8a) and (6–8b), and can use the one that is simpler for calculational purposes. We noted in (2–33) – (2–36) that if the constants $H_{(+)}$ and $H_{(-)}$ satisfy the condition

$$(mH)_{(+)} \approx (mH)_{(-)} \quad (6\text{–}9)$$

then the electron and proton temperatures are everywhere approximately equal.

The stress-energy tensor T^{jk} of the entire system, including both charged fluids and the fields, has the following form:

$$T^{jk} \equiv T^{jk}{}_{(+)} + T^{jk}{}_{(-)} - [P_{(+)} + P_{(-)}]g^{jk} + E^{jk} + G^{jk} \quad (6\text{–}10)$$

where

$$T^{jk}{}_{(\pm)} \equiv (\rho\tilde{m}c^2 u^j u^k)_{(\pm)} + t^{jk}{}_{(\pm)} \quad (6\text{–}11)$$

$$E^{jk} = \frac{1}{4\pi}\left[F^{jl}F_l{}^k + \frac{1}{4}g^{jk}(F^{lm}F_{lm})\right] \quad (6\text{–}12)$$

$$G^{jk} = \frac{1}{4\pi\Gamma}\left[(\partial^j g)(\partial^k g) - \frac{1}{2}g^{jk}(\partial^l g)(\partial_l g)\right] \quad (6\text{–}13)$$

These tensors satisfy the following divergence relations:

$$\partial_k T^{jk} = 0 \quad (6\text{–}14)$$

$$\partial_k T^{jk}{}_{(\pm)} = k^j{}_{(\pm)} \quad (6\text{–}15)$$

where

$$k^j{}_{(\pm)} = (\rho m)_{(\pm)}\partial^j g + \partial^j p_{(\pm)} \pm qF^{jk}(\rho u_k)_{(\pm)} \quad (6\text{–}16)$$

$$= [\rho\partial^j(\tilde{m}c^2)]_{(\pm)} \pm qF^{jk}(\rho u_k)_{(\pm)}$$

$$\partial_k E^{jk} = -qF^{jl}[(\rho u_l)_{(\pm)} - (\rho u_l)_{(-)}] \quad (6\text{–}17)$$

$$\partial_k G^{jk} = -[(\rho m)_{(+)} + (\rho m)_{(-)}]\partial^j g \quad (6\text{–}18)$$

Equation (6–14) justifies the following identifications:

$$\text{total energy density} \equiv T^{00} \quad (6\text{–}19)$$

$$\text{total energy flux density} \equiv \vec{U} \equiv c(T^{01}, T^{02}, T^{03}) \quad (6\text{–}20)$$

$$\text{total momentum density} \equiv \vec{G} \equiv \frac{1}{c}(T^{10}, T^{20}, T^{30}) \quad (6\text{–}21)$$

The invariant trace of the total stress-energy tensor is

$$T^{j}_{j}=(\rho\breve{m}c^2+t^{j}_{j}-4P)_{(+)}+(\rho\breve{m}c^2+t^{j}_{j}-4P)_{(-)}$$
$$+\frac{1}{4\pi\Gamma}\left[(\vec{\nabla}g)^2-\frac{1}{c^2}\left(\frac{\partial g}{\partial t}\right)^2\right]=[(\epsilon-3P)+t^{j}_{j}]_{(+)}$$
$$+[(\epsilon-3P)+t^{j}_{j}]_{(-)}+\frac{1}{4\pi\Gamma}\left[(\vec{\nabla}g)^2-\frac{1}{c^2}\left(\frac{\partial g}{\partial t}\right)^2\right] \quad (6\text{–}22)$$

where $\epsilon_{(\pm)}$ is the fluid energy density defined in (5–35).

If ϵ and $\vec{p}$ are respectively the total energy and momentum of a finite system, it follows from (6–14) that

$$\epsilon=\int_{V_3} T^{00}dV_3=\text{constant} \quad (6\text{–}23)$$

$$p=\int_{V_3}\vec{G}dV_3=\text{constant} \quad (6\text{–}24)$$

where the integration extends over all 3-space. The 4-velocity V^j of the entire system is defined as

$$U^0\equiv\epsilon[\epsilon^2-c^2\vec{p}^2]^{-1/2} \quad (6\text{–}25a)$$

$$V^j\equiv cp^j[\epsilon^2-c^2\vec{p}^2]^{-1/2}\ (j=1,2,3) \quad (6\text{–}25b)$$

which corresponds to a 3-velocity $\vec{V}$ given by

$$\vec{V}=\frac{\vec{p}}{(\epsilon/c^2)} \quad (6\text{–}26)$$

The square root in the denominator of (6–25) is the total rest-energy of the system, and the factor (ϵ/c^2) in the denominator of (6–26) is the total mass of the system in the observer's reference frame.

If $S^{jkl}{}_{(\pm)}$ is spin angular momentum density of either charged fluid, then the total fluid angular momentum $M^{jkl}{}_{(\pm)}$ is defined as follows:

$$M^{jkl}_{(\pm)}\equiv S^{jkl}_{(\pm)}+\frac{1}{c}[x^jT^{kl}_{(\pm)}-x^kT^{jl}_{(\pm)}] \quad (6\text{–}27)$$

This definition can be written in terms of 3-vectors as follows:

$$\vec{M}_{(\pm)}\equiv\vec{S}_{(\pm)}+\vec{r}\times\vec{G}_{(\pm)} \quad (6\text{–}28)$$

where

$$\vec{M}_{(\pm)}\equiv(M_x, M_y, M_z)_{(\pm)}\equiv(M^{230}, M^{310}, M^{120})_{(\pm)} \quad (6\text{–}29)$$

$$\vec{S}_{(\pm)}\equiv(S_x, S_y, S_z)_{(\pm)}\equiv(S^{230}, S^{310}, S^{120})_{(\pm)} \quad (6\text{–}30)$$

$$\vec{r}\equiv(x, y, z)\equiv(x^1, x^2, x^3) \quad (6\text{–}31)$$

Conservation of fluid angular momentum requires that

$$\partial_l M^{jkl}_{(\pm)}=\frac{1}{c}[x^jk^k_{(\pm)}-x^kk^j_{(\pm)}] \quad (6\text{–}32)$$

where $k^j{}_{(\pm)}$ is given by (6–16). This equation is equivalent to the following condition on $S_{(\pm)}{}^{jkl}$:

$$\partial_l S^{jkl}_{(\pm)}=\frac{1}{c}(T^{jk}-T^{kj})_{(\pm)}=\frac{1}{c}(t^{jk}-t^{kj})_{(\pm)} \quad (6\text{–}33)$$

The definitions for the total angular momentum density M^{jkl} for the entire magnetogas-dynamical system including fluids and fields is analogous to (6–27):

$$M^{jkl}\equiv S^{jkl}_{(+)}+S^{jkl}_{(-)}+\frac{1}{c}(x^jT^{kl}-x^kT^{jl}) \quad (6\text{–}34)$$

The condition for conservation of total angular momentum is

$$\partial_l M^{jkl}=0 \quad (6\text{–}35)$$

which is automatically satisfied if condition (6–33) is satisfied.

The integral form of (6–35) is

$$M^{jk}\equiv\int_{V_3} M^{jk0}dV_3=\text{constant} \quad (6\text{–}36)$$

or

$$\vec{M}=\int_{V_3}\vec{\mathrm{M}}dV_3=\int_{V_3}[\vec{S}_{(+)}+\vec{S}_{(-)}+\vec{r}\times\vec{G}]dV_3$$
$$=\text{constant} \quad (6\text{–}37)$$

where

$$\vec{M}=(M_x, M_y, M_z)\equiv(M^{23}, M^{31}, M^{12}) \quad (6\text{–}38)$$

It is possible to define a 4-vector total angular momentum Σ^j in terms of the antisymmetric tensor M^{jk} and the 4-velocity V^j of the total system as follows:

$$\Sigma^j\equiv-\tfrac{1}{2}\epsilon^{jkln}V_kM_{ln} \quad (6\text{–}39)$$

In the rest-frame of the total system ($\vec{V}=0$), we have

$$\left.\begin{aligned}&\dot{\Sigma}^0=0\\&(\dot{\Sigma}^1,\ \dot{\Sigma}^2,\ \dot{\Sigma}^3)=\dot{\vec{M}}\end{aligned}\right\}\ \text{for}\ \vec{V}\equiv 0 \qquad (6\text{–}40)$$

VII SPINOR DESCRIPTION OF A PARTICLE

Tetrapod Description

The particles with which we are dealing, electrons and protons, have an internal structure characterized by a symmetry axis, as evidenced by the existence of particle spin and magnetic dipole moment. They also possess a de Broglie phase, which must be regarded as an intrinsic property of the particle. Thus, for a complete description of a particle, in addition to its velocity, we must specify the orientation of its symmetry axis and the magnitude of its de Broglie phase. As illustrated in Figure 2, this can be accomplished by means of a set of four orthonormal 4-vectors—one time-like and three space-like vectors. This set of 4-vectors is called the particle tetrapod.

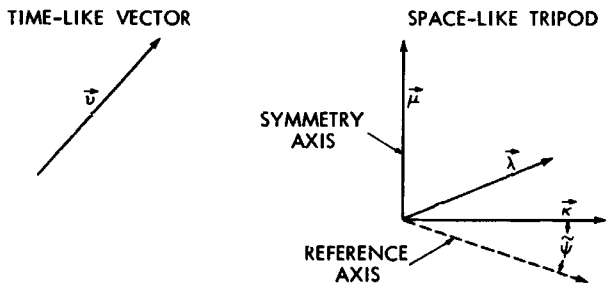


FIGURE 2.—Tetrapod representation of a particle.

The time-like vector of the tetrapod is the 4-velocity u^j (designated in the figure by the 3-velocity $\vec{v}$). One of these three space-like vectors $\vec{\mu}$ is used to specify the symmetry axis of the particle. The remaining degree of freedom, the angle $\vec{\Psi}$ needed to specify the orientation of the other two space-like vectors $\vec{k}$ and $\vec{\lambda}$ about the axis $\vec{\mu}$, is measured between $\vec{k}$ and a fixed reference axis in the observer's frame of reference. The 3-vectors $\vec{k}$, $\vec{\lambda}$, and $\vec{\mu}$ are, of course, the space-like parts of three 4-vectors, k^j, λ^j, and μ^j. In the particle rest-frame, the time-like components of these vectors are all zero.

We postulate that the angle $\vec{\Psi}$ is, to within a proportionality constant, the de Broglie phase Φ, and we postulate the proportionality constant to be $\hbar/2$ where $\hbar=1.0542\times10^{-27}$ erg-sec in Planck's constant. (The factor $\frac{1}{2}$ is necessary in order for the theory to yield a particle spin angular momentum of $\hbar/2$, which is regarded as an experimental fact with which the theory must agree.) Thus we have

$$\Phi=\frac{\hbar}{2}\tilde{\Psi} \qquad (7\text{–}1)$$

(We shall later introduce the sign of the particle charge into this relation, so that (7–1), like all the relations until (7–8), is valid only for positively charged particles. For the moment, however, we shall ignore the question of the sign of the particle charge.)

It is with (7–1) that Planck's constant is introduced into the theory. It should be emphasized, however, that in spite of the incorporation of the de Broglie Hypothesis into the theory which results from (7–1), the theory is classical, rather than quantum-mechanical, because no quantization process is introduced at any point.

From (3–2) and (7 – 1) we have

$$p^j=-\partial^j\Phi=-\frac{\hbar}{2}\partial^j\tilde{\Psi} \qquad (7\text{–}2)$$

In the particle rest-frame this becomes

$$\mathring{p}^0=\frac{\overset{0}{E}}{c}=mc+\frac{q}{c}\mathring{A}^0=-\frac{\hbar}{2c}\frac{d\tilde{\Psi}}{d\tau} \qquad (7\text{–}3a)$$

$$\mathring{\vec{p}}=\frac{q}{c}\mathring{\vec{A}}=\frac{\hbar}{2}\mathring{\vec{\nabla}}\tilde{\Psi} \qquad (7\text{–}3b)$$

where $\dot{E}$ is the particle rest-energy, $\dot{\vec{A}}$ is the space-like part of the 4-vector potential, and $\mathring{A}^\circ$ is the time-like part which, by means of a proper choice of gauge, could be made to vanish. Thus we see that the angular velocity of rotation of $\mathring{\vec{k}}$ and $\mathring{\vec{\lambda}}$ about $\mathring{\vec{\mu}}$ in the particle rest-frame determines the particle rest-energy, and the spatial rate of change in the rest-frame, when we go from one particle to another, determines the particle momentum in the rest-frame which is just $\frac{q}{c}\mathring{\vec{A}}$. Since a knowledge of the particle rest-energy at all points of space amounts to a knowledge of the function $\mathring{m}$ which determines the inertial properties of the particle as well as its scalar potentials, we see that a knowledge of $\tilde{\Psi}$ at every point of space-time

contains all the information necessary for a complete dynamical description of the particle.

Unfortunately, the function $\tilde{\Psi}$ is rather difficult to work with. In the case of a distribution of particles all at rest with respect to the observer, and having their $\mathring{\vec{\mu}}$ axes all aligned, $\tilde{\Psi}$ is easily determined. In the general case, however, $\tilde{\Psi}$ is one of three Euler angles that describe the orientation of the particle tripod in its own rest-frame, which in general is not the observer's frame. This matter is discussed at greater length in Appendix B. For present purposes, however, it is sufficient merely to appreciate the desirability of introducing a function having the essential properties of $\tilde{\Psi}$, but simpler to define and work with. From (7-2) we see that it is really the gradient of $\tilde{\Psi}$, rather than $\tilde{\Psi}$ itself, that interests us, so it will suffice to find a simply defined 4-vector ω^j that approximates the gradient of $\tilde{\Psi}$. (The choice we make for ω^j will actually approximate $-\frac{1}{2}\partial^j\tilde{\Psi}\cdot$) We define ω^j as follows:

$$\omega^j \equiv \tfrac{1}{2}\lambda_\epsilon \partial^j k^l \tag{7-4}$$

In order to see that

$$\omega^j \approx -\tfrac{1}{2}\partial^j\tilde{\Psi} \tag{7-5}$$

we note that in the particular rest-frame, ω^j can be written

$$\mathring{\omega}^j = -\tfrac{1}{2}\mathring{\vec{\lambda}}\cdot(\mathring{\partial}^j\mathring{\vec{k}}) \tag{7-6}$$

where $\partial^{\circ j}$ is the gradient operator in the particle rest-frames. Now if one pictures two neighboring tripods whose orientations are nearly the same (or a single tripod whose orientation is changing with time), it is clear that the contribution to (7-6) resulting from a difference (or a time change) in the orientation of $\mathring{\vec{\mu}}$ is very small compared with the contribution resulting from the angular displacement of $\vec{k}$ and $\mathring{\vec{\lambda}}$ about $\mathring{\vec{\mu}}$. The fact that one of the two tripods we are comparing may not be exactly at rest, but rather have a small velocity, makes only a very small difference in the right side of (7-6). Thus, the value of $\dot{\omega}^j$ is very close to what we would find in the case for which all the tripods are at rest and have their $\mathring{\vec{\mu}}$ axes aligned. This then is the basis for the approximate equality given in (7-5).

It should be emphasized, however, that this approximate equality has been introduced only to give an intuitive feeling for the physical significance of ω^j, but no use of (7-5) will be made in the development of the formalism to follow.

We conclude this discussion of the physical significance of ω^j by noting that its time-like component is essentially the de Broglie frequency of the particle, and its space-like part the de Broglie wave-vector.

Rewriting (7-3a), we have

$$\frac{d\tilde{\Psi}}{d\tau} = -\frac{2\mathring{E}}{\hbar} \tag{7-7}$$

Since the particle rest-energy $\dot{E}$ must be positive, we see from Figure 2 that the direction of rotation of $\mathring{\vec{k}}$ about $\mathring{\vec{\mu}}$ is given by the *left*-hand screw rule. ($\mathring{\vec{k}}$, $\mathring{\vec{\lambda}}$, and $\mathring{\vec{\mu}}$, in that order, constitute a right-hand system of axes.)

Now let us identify the direction of rotation of $\mathring{\vec{k}}$ about $\mathring{\vec{\mu}}$ with the sign σ ($\sigma = \pm 1$) of the electric charge of the particle. Thus our discussion has so far concerned itself only with a positively charged particle. In the case of a negatively charged particle, we assert that the direction of rotation of $\mathring{\vec{k}}$ about $\mathring{\vec{\mu}}$ is given by the *right*-hand screw rule. Thus, if the (+) and (−) subscripts or superscripts refer respectively to positively or negatively charged particles, the foregoing relations are generalized as follows:

$$\Phi_{(\pm)} = \pm\frac{\hbar}{2}\tilde{\Psi}_{(\pm)} \tag{7-8}$$

$$p^j_{(\pm)} = -\partial^j\Phi_{(\pm)} = \mp\frac{\hbar}{2}\partial^j\tilde{\Psi}_{(\pm)} \tag{7-9}$$

$$\omega^j_{(\pm)} = \tfrac{1}{2}\lambda_l^{(\pm)}\partial^j k_{l(\pm)} \tag{7-10}$$

$$\omega^j_{(\pm)} \approx -\tfrac{1}{2}\partial^j\tilde{\Psi}_{(\pm)} \tag{7-11}$$

$$\left(\frac{d\tilde{\Psi}}{d\tau}\right)_{(\pm)} = \mp\frac{2\mathring{E}_{(\pm)}}{\hbar} \tag{7-12}$$

From (7-9) and (7-11) we have

$$p^j{}_{(\pm)} \approx \pm\hbar\omega^j{}_{(\pm)} \tag{7-13}$$

and

$$\mathring{\omega}^\circ{}_{(\pm)} \approx \pm\frac{\mathring{p}^\circ{}_{(\pm)}}{\hbar} = \pm\frac{\mathring{E}_{(\pm)}}{\hbar c} \tag{7-14}$$

Thus, in contradistinction to $\dot{p}^{\circ}{}_{(\pm)}$ which is always positive, $\overset{\circ}{\omega}{}^{\circ}{}_{(\pm)}$ has the same sign as the electric charge of the particle.

We now complete our tetrapod description by normalizing the four 4-vectors to $\rho_{(\pm)}$, the invariant particle density, rather than to unity. In this way we have packed all the information we need for a complete description of a charged fluid into the particle tetrapod. Altogether this description involves seven degrees of freedom: three for $\vec{v}$, two for $\vec{\vec{\mu}}$, and one apiece for $\vec{\vec{\Psi}}$ and ρ. Since each of the 4-vectors has four elements, however, our tetrapod involves a total of sixteen elements. But it has only seven degrees of freedom. Thus the elements of the tetrapod do not represent the "normal coordinates" of the problem. This fact makes the tetrapod difficult to use in any formalism.

We shall now show that it is possible to express the four 4-vectors of the tetrapod as bilinear [illegible] of the complex elements of two spinors, each having two components. Thus the two spinors together have eight degrees of freedom. One of these degrees of freedom, the phase common to the two spinors, will be used to specify the sign of the particle charge. (It is shown in Appendix C that this identification, together with the spinor equations of motion, causes the direction or rotation of $\vec{k}$ and $\vec{\lambda}$ about $\vec{v}$ to be dependent on the sign of the particle charge, as indicated by (7–14).) If we regard the sign of the particle charge as a degree of freedom, then we can assert that the two spinors together are completely specified by eight real functions and, since the charged fluid has eight degrees of freedom, the two spinors constitute the "normal mode description" of the problem.

Spinor Description

As a preliminary to introducing the spinors that will be used to describe the charged fluids, we must introduce the irreducible form of a 4-vector. For reasons that will shortly become clear, the superscripts and subscripts used to label the elements of a 4-vector in irreducible form will not be the usual ones ($j=0, 1, 2, 3$). Rather, two indices α and $\bar{\beta}$, where $\alpha=1, 2$ and $\bar{\beta}=\bar{1}, \bar{2}$, will be used. The overhead bar in $\bar{\beta}$ indicates complex conjugation, and we shall see shortly how the need for such a notation enters the formalism.

First, we must define the irreducible vector elements in terms of the ordinary (reducible) elements. Using the 4-vector $(\rho u^j)_{(\pm)}$ as an example, and omitting for the time-being the $_{(\pm)}$ subscript, we have

$$\rho u^{1\bar{1}}=\frac{\rho}{\sqrt{2}}(u^{\circ}+u^{3})=\frac{\rho}{\sqrt{2}}e^{\nu}$$

$$\rho u^{2\bar{2}}=\frac{\rho}{\sqrt{2}}(u^{\circ}-u^{3})=\frac{\rho}{\sqrt{2}}e^{-\nu} \qquad (7\text{–}15)$$

$$\rho u^{1\bar{2}}=\frac{\rho}{\sqrt{2}}(u^{1}-iu^{2})=\frac{\rho}{\sqrt{2}}\sinh\nu\cos\hat{\theta}\,e^{-i\hat{\phi}}$$

$$\rho u^{2\bar{1}}=\frac{\rho}{\sqrt{2}}(u^{1}+iu^{2})=\frac{\rho}{\sqrt{2}}\sinh\nu\cos\hat{\theta}\,e^{i\hat{\phi}}$$

where

$$\nu=\tanh^{-1}(v/c) \qquad (7\text{–}16)$$

and $\hat{\theta}$ and $\hat{\phi}$ are respectively the polar and azimuthal angles of a system of spherical coordinates referred to the z axis as pole.

It is obvious that, for a rotation of the coordinate system through an angle $\Delta\hat{\phi}$ around the z axis, we have

$$(\rho u^{1\bar{2}})'=e^{-i\Delta\hat{\phi}}(\rho u^{1\bar{2}})$$
$$(\rho u^{2\bar{1}})'=e^{i\Delta\hat{\phi}}(\rho u^{2\bar{1}}) \qquad (7\text{–}17)$$

where the primes indicate the values of the elements after the rotation. Under this same symmetry operation, the components ρu^1 and ρu^2 are not, however, simply multiplied by a constant as in (7–17). Rather they become shuffled, i.e., the expression for $(\rho u^1)'$ involves *both* ρu^1 and ρu^2, and similarly for $(\rho u^2)'$. Thus $\rho u^{1\bar{2}}$ and $\rho u^{2\bar{1}}$ are irreducible representations of the symmetry group of rotations about the z axis, whereas ρu^1 and ρu^2 are not. Similarly, $\rho u^{1\bar{1}}$ and $\rho u^{2\bar{2}}$ are irreducible representations of the symmetry group consisting of all Lorentz transformations in the z direction, whereas ρu° and ρu^3 are not.

It is well-known that the equations of physics assume their simplest form when they are expressed in terms of irreducible quantities. For this reason, the components $\rho u^{\alpha\bar{\beta}}$ on the left side of (7–15) are to be regarded as more fundamental than the usual components ρu^j.

We now introduce the two fundamental spinors ζ^α and χ^α by means of the following relation:

$$\rho\begin{pmatrix} u^{\bar{1}1} u^{\bar{1}2} \\ u^{21} u^{22} \end{pmatrix} = \begin{pmatrix} \zeta^1 \\ \zeta^2 \end{pmatrix}(\zeta^{\bar{1}} \zeta^{\bar{2}}) + \begin{pmatrix} \chi^1 \\ \chi^2 \end{pmatrix}(\chi^{\bar{1}} \chi^{\bar{2}}) \quad (7\text{–}18)$$

where

$$\overline{\rho^\alpha} = \overline{\rho^\alpha}\ ;\ \overline{\chi^\alpha} = \overline{\chi^\alpha} (\alpha = 1, 2) \quad (7\text{–}19)$$

On the left side of (7–18) we have arranged the elements of $\rho u^{\alpha\bar{\beta}}$ in the form of a 2×2 matrix, and on the right side we have the sum of two outer products in which the pre-factor of each product is a column matrix and the postfactor is the complex conjugate of this column matrix written in row form, i.e., its Hermitean adjoint.

It is convenient, as indicated in (7–19), to designate complex conjugation of a spinor element by means of a bar over only the index(*), rather than over the entire symbol for the element. Thus, the indices $\alpha = 1, 2$ and $\bar{\beta} = \bar{1}, \bar{2}$ are regarded as entirely independent, i.e., it is not necessary in the case of $\rho u^{\alpha\bar{\beta}}$ that, for example, $\beta = \bar{1}$ when $\alpha = 1$. A 4-vector in irreducible form is characterized by one of each type of index. It can be shown that an antisymmetric world tensor of rank is characterized by either two barred or two unbarred indices. A symmetic world tensor of second rank is characterized by four indices, two of each kind. In keeping with the customary notation, vector and tensor indices ranging over the values 0, 1, 2, 3 will be designated by lower case Roman letters, whereas spinor indices will always be designated by lower case Greek letters.

In order to explain why two outer products, rather than just one, are needed on the right side of (7–18), we note that, although the spinor ζ^α (or χ^α) itself has four degrees of freedom, one of these, namely the phase factor that is common to both ζ^1 and ζ^2, is suppressed when we carry out the outer multiplication between ζ^α and its conjugate $\zeta^{\bar{\alpha}}$. Thus ζ^α by itself can be used only to represent a 4-vector with no more than three degrees of freedom, i.e., a null vector. To represent an arbitrary time-like or space-like 4-vector, two spinors are necessary.

It is in the sense of the relation (7–18) that we may think of spinors as the "square roots" of 4-vectors. In the same way that, when we take the square root of a real number, we get an extra degree of freedom, namely the sign of the square root, we find that the "square root" of a 4-vector hase xtra degrees of freedom. This is obviously the case because the 4-vector ρu^j (or $\rho u^{\alpha\beta}$) has four degrees of freedom, whereas the two spinors ζ^α and χ^α together have a total of eight degrees of freedom.

The physical significance of these extra degrees of freedom becomes clear when we note that, having the two spinors ζ^α and χ^α, it is possible to generate the complete particle tetrapod as follows:

$$\begin{aligned} \rho u^{\alpha\bar{\beta}} &= \zeta^\alpha \zeta^{\bar{\beta}} + \chi^\alpha \chi^{\bar{\beta}} \\ \rho \mu^{\alpha\bar{\beta}} &= \zeta^\alpha \zeta^{\bar{\beta}} - \chi^\alpha \chi^{\bar{\beta}} \\ \rho k^{\alpha\bar{\beta}} &= \zeta^\alpha \chi^{\bar{\beta}} + \chi^\alpha \zeta^{\bar{\beta}} \\ \rho \lambda^{\alpha\bar{\beta}} &= -i(\zeta^\alpha \chi^{\bar{\beta}} - \chi^\alpha \zeta^{\bar{\beta}}) \end{aligned} \quad (7\text{–}20)$$

Thus the extra degrees of freedom are to be associated with the particle tripod (k^j, λ^j, μ^j), and hence are the degrees of freedom specifying the orientation of the spin axis, and the de Broglie phase. The phase angle common to ζ^α and χ^α, which does not make itself directly felt in the generation of the tetrapod, will be used to specify the sign of the particle charge. (In Appendix C it is shown that this identification has the consequence that, for solutions of the spinor equations of motion, the direction of rotation of $\vec{k}$ and $\vec{\lambda}$ about $\vec{\mu}$ is dependent on the sign of the particle charge. It is only in this indirect way that the phase common ζ^α and χ^α makes itself felt—via the spinor equations of motion—in the particle tetrapod.) Finally, we note that it can be shown that the tetrapod defined by (7–20 automatically satisfies the necessary orthonormality requirements.

Equations (7–20) provide the bridge between the spinors ζ^α and χ^α, which we regard as constituting the most fundamental description of a fluid of charged particles, and the particle tetrapod in irreducible form. To complete the formalism we need a bridge that connects the spinors

*A commonly used spinor notation employs a dot, rather than a bar, over the index. This notation is appropriate when the spinor so designated, e.g. $\zeta^{\dot{\alpha}}$, transforms like $\zeta^{\bar{\alpha}}$, but is not equal to $\zeta^{\bar{\alpha}}$, i.e., it differs from $\zeta^{\bar{\alpha}}$ by a factor that is a scalar invariant. In order to emphasise that $\zeta^{\bar{\alpha}}$ not only transforms like $\zeta^{\bar{\alpha}}$, but is exactly equal to it, we use a bar over the index rather than a dot.

directly with the 4-vectors of the tetrapod expressed in the usual form, rather than with the irreducible form as in (7–20). To accomplish this, we first rewrite (7–15) in matrix form as follows:

$$\rho\begin{pmatrix} u^{1\bar{1}} u^{1\bar{2}} \\ u^{2\bar{1}} u^{2\bar{2}} \end{pmatrix} = \frac{\rho}{\sqrt{2}}\begin{pmatrix} [u^0+u^3][u^1-iu^2] \\ [u^1+iu^2][u^0-u^3] \end{pmatrix} \tag{7–21}$$

$$= \frac{\rho}{\sqrt{2}}\left[u^0\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + u^1\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + u^2\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} + u^3\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\right]$$

We introduce the constant matrices $\sigma_j^{\alpha\bar{\beta}}$ defined as follows:

$$\sigma_0^{\alpha\bar{\beta}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \sigma_1^{\alpha\bar{\beta}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$
$$\sigma_2^{\alpha\bar{\beta}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \qquad \sigma_3^{\alpha\bar{\beta}} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{7–22}$$

where, following the usual convention, the first index α indicates the row of the matrix, and the second index $\bar{\beta}$ indicates the column. The sematrices, of course, are just the Pauli spin matrices. Using these matrices, (7–15) or (7-21) can be written in the following condensed form:

$$\rho u^{\alpha\bar{\beta}} = \sigma_j^{\alpha\bar{\beta}}(\rho u^j) \tag{7–23}$$

where, as always, a repeated index means contraction.

The matrices $\sigma_j^{\alpha\bar{\beta}}$ satisfy the following important relations:

$$\sigma_j^{\alpha\bar{\gamma}}\sigma_{\gamma\bar{\beta}}^{j} = \delta_\beta^\alpha \delta_{\bar{\gamma}}^{\bar{\gamma}} \tag{7–24a}$$

$$\sigma_j^{\alpha\bar{\beta}}\sigma_{\beta\bar{\alpha}}^{k} = \delta_j^k \tag{7–24b}$$

and

$$\overline{\sigma_j^{\alpha\bar{\beta}}} = \sigma_j^{\bar{\beta}\alpha} \qquad \overline{\sigma_{\alpha\bar{\beta}}^{j}} = \sigma_{\bar{\beta}\alpha}^{j} \tag{7–24c}$$

where $\delta_\beta^\alpha \cdot \delta_\nu^\mu$, and δ_j^k are Kronecker delta functions, and the matrices $\sigma_{\alpha\beta}^j$ have the same form as the matrices $\sigma_j^{\alpha\beta}$, i.e.

$$\sigma_{\alpha\beta}^j = \begin{pmatrix} \sigma_1^j & \sigma_2^j \\ \sigma_{\bar{1}}^j & \sigma_{\bar{2}}^j \end{pmatrix} = \begin{pmatrix} \sigma_j^{1\bar{1}} & \sigma_j^{1\bar{2}} \\ \sigma_j^{2\bar{1}} & \sigma_j^{2\bar{2}} \end{pmatrix} = \sigma_j^{\alpha\beta} \tag{7–25}$$

(Note that in the case of $\sigma_{\alpha\bar{\beta}}^j$ the barred index, which designates the row of the matrix, is written first whereas, in the case of $\sigma_j^{\alpha\bar{\beta}}$, it is written second and designates the column.) Using (7–24b), it is possible to invert (7–23):

$$\rho u^j = \sigma_{\bar{\beta}\alpha}^{j}(\rho u^{\alpha\bar{\beta}}) \tag{7–26}$$

Multiplying the equations (7–20) by $\sigma_{\bar{\beta}\alpha}^{j}$ and contracting the $\bar{\beta}$ and α, we arrive at the desired bridge between the spinors and the entire tetrapod in its usual (reducible) form:

$$\begin{aligned} \rho u^j &= \sigma_{\bar{\beta}\alpha}^{j}(\zeta^\alpha \zeta^{\bar{\beta}} + \chi^\alpha \chi^{\bar{\beta}}) \\ \rho \mu^j &= \sigma_{\bar{\beta}\alpha}^{j}(\zeta^\alpha \zeta^{\bar{\beta}} - \chi^\alpha \chi^{\bar{\beta}}) \\ \rho k^j &= \sigma_{\bar{\beta}\alpha}^{j}(\zeta^\alpha \chi^{\bar{\beta}} + \chi^\alpha \zeta^{\bar{\beta}}) \\ \rho \lambda^j &= -i\sigma_{\bar{\beta}\alpha}^{j}(\zeta^\alpha \chi^{\bar{\beta}} - \chi^\alpha \zeta^{\bar{\beta}}) \end{aligned} \tag{7–27}$$

Thus far we have used only contravariant spinors, whose indices have been indicated by superscripts rather than subscripts. Knowing the contravariant form of a spinor, we can define the covariant form of this same spinor by means of the requirement that the contraction of the two must be an invariant. Furthermore, this invariant must be zero. Otherwise, as it can easily be shown, the invariant norm of the null vector that is the "square" of the spinor would have to be non-zero, and this cannot be. Thus we define the covariant spinors ζ_α and χ_α as follows:

$$\begin{aligned} \zeta_1 &= -\zeta^2;\ \zeta_2 = \zeta^1 \\ \chi_1 &= -\chi^2;\ \chi_2 = \chi^1 \end{aligned} \tag{7–28}$$

Thus

$$\zeta_\alpha \zeta^\alpha = \zeta_1\zeta^1 + \zeta_2\zeta^2 = -\zeta^2\zeta^1 + \zeta^1\zeta^2 = 0 \tag{7–29}$$

Similarly

$$\chi_\alpha \chi^\alpha = 0 \tag{7–30}$$

The contraction of ζ_α and χ^α is, however, non-zero. It follows from (7–28) that

$$\zeta_\alpha \chi^\alpha = -\zeta^\alpha \chi_\alpha \tag{7–31}$$

Thus raising one index in a contraction and lowering the other changes the sign of the result. This is to be contrasted with the fact that when

we do the same thing with the contraction of two 4-vectors, the result is unaltered. For example,

$$p_j u^j = p^j u_j \tag{7-32}$$

For a further discussion of the algebraic aspects of spinor analysis, the reader is referred to the papers of van der Waerden (11), Laporte and Uhlenbeck (12), and Bade and Jehle (13).

By means of the relations (7–27), it is possible to write the components of $\zeta^{\alpha}{}_{(\pm)}$ and $\chi^{\alpha}{}_{(\pm)}$ for each of the two charged fields in terms of the Euler angels that specify the orientation of the corresponding space-like tripod in the rest-frame of the fluid in question, and in terms of the magnitude and direction of the fluid velocity. This is carried out in Appendix B. However, for our present purposes the complete result of Appendix B is not needed. We only need to know that $\zeta^{\alpha}{}_{(\pm)}$ and $\chi^{\alpha}{}_{(\pm)}$ can be expressed in terms of two normalized spinors $\hat{\zeta}^{\alpha}{}_{(\pm)}$ and $\hat{\chi}^{\alpha}{}_{(\pm)}$ as follows:

$$\zeta^{\alpha}_{(\pm)} = \left(\frac{\rho e^{i\epsilon}}{\sqrt{2}}\right)^{1/2}_{(\pm)} e^{-i\Psi/2} \hat{\zeta}^{\alpha}_{(\pm)} \tag{7-33a}$$

$$\chi^{\alpha}_{(\pm)} = \left(\frac{\rho e^{i\epsilon}}{\sqrt{2}}\right)^{1/2}_{(\pm)} e^{i\Psi/2} \hat{\chi}^{\alpha}_{(\pm)} \tag{7-33b}$$

where

$$(\hat{\zeta}_{\alpha}\hat{\chi}^{\alpha})_{(\pm)} = 1 \tag{7-34}$$

and $\hat{\zeta}_{(\pm)}{}^{\alpha}$ and $\hat{\chi}_{(\pm)}{}^{\alpha}$ are not functions of Ψ, which means that they contain none of the high-frequency (of order of the de Broglie frequency) time or space oscillation associated with Ψ, but only the slow variation associated with changes in fluid velocity and spin orientation. From (7–33) and (7–34), it follows that

$$(\zeta_{\alpha}\chi^{\alpha})_{(\pm)} = \left(\frac{\rho e^{i\epsilon}}{\sqrt{2}}\right)_{(\pm)} \tag{7-35}$$

ϵ is the phase that is common to both ζ^{α} and χ^{α}. It is obvious from (7–27) that ϵ has no direct effect on the vectors of the tetrapod. Thus ϵ is a disposable degree of freedom that we are free to use to specify the sign of the particle charge. Thus we impose the following condition on $\epsilon_{(\pm)}$:

$$\cos \epsilon_{(\pm)} = \pm 1 \tag{7-36a}$$

which is equivalent to

$$e^{i\epsilon(\pm)} = \pm 1 \tag{7-36b}$$

Thus (7–35) becomes

$$(\zeta_{\alpha}\chi^{\alpha})_{(\pm)} = \pm \frac{\rho_{(\pm)}}{\sqrt{2}} \tag{7-37}$$

The condition (7–36) is fulfilled if we impose the following constraint on our spinors:

$$I(\zeta_{\alpha}\chi^{\alpha})_{(\pm)} = 0 \tag{7-38}$$

where I designates the imaginary part of the argument. The only constraint on the real part of $(\zeta_{\alpha}\chi^{\alpha})_{(\pm)}$ is that it have the same sign at every point of space-time.

From (7–27) and (7–35), we have the following expression for the 4-velocity u^j as a function of the spinors:

$$u^j{}_{(\pm)} = \{[2(\zeta_{\overline{\alpha}}\chi^{\overline{\alpha}})(\zeta_{\nu}\chi^{\nu})]^{-\frac{1}{2}}[\sigma^j{}_{\alpha\overline{\beta}}(\zeta^{\beta}\zeta^{\overline{\alpha}} + \chi^{\beta}\chi^{\overline{\alpha}})]\}_{(\pm)} \tag{7-39}$$

This identity is valid regardless of whether or not the constraint (7–38) is satisfied. From (7–29)—(7–31), (7–24), and (7–39) it follows automatically that

$$(u^j u_j)_{(\pm)} = 1 \tag{7-40}$$

Thus the $u_{(\pm)}{}^j$ that we calculate from the spinors is automatically normalized, and so when we work with spinors thate is no need for the troublesome nonlinear side condition that we encountered in (6–1c) or (6–6c). This is true even when the constraint (7–38) is dropped which, as we shall later show, is permissible in problems on a laboratory or astronomical scale.

Now let us estimate the magnitude of the error involved in the approximate equalities (7–11) and (7–13). To do this, we substitute the expressions for k^j and λ^j given in (7–27) into the definition of ω^j given in (7–10) which yields

$$\omega^j{}_{(\pm)} = -\tfrac{1}{2}\partial^j \Psi_{(\pm)} + \gamma^j{}_{(\pm)} \tag{7-41}$$

where

$$\gamma^j{}_{(\pm)} = I(\hat{\chi}^{\alpha}\partial^j \hat{\zeta}_{\alpha})_{(\pm)} \tag{7-42}$$

Comparing (7–41) with (7–11), we see that $\gamma^j{}_{(\pm)}$ is the amount by which (7–11) fails of being an

exact equality. It is easy to estimate the magnitude of $\gamma^{j}{}_{(\pm)}$. Let L be a length characterizing the size of the system in which we are interested. Now we have already noted that because $\hat{\zeta}^{\alpha}{}_{(\pm)}$ and $\hat{\chi}^{\alpha}{}_{(\pm)}$ are not functions of $\tilde{\Psi}$, which contains all the fast space-time oscillation associated with the de Broglie phase, they are functions only of particle velocity and spin orientation. We shall limit ourselves to solutions in which no fast precession of the spin axis is present. For such cases it is always true that appreciable changes in $\hat{\zeta}^{\alpha}{}_{(\pm)}$ and $\hat{\chi}^{\alpha}{}_{(\pm)}$ occur only over distances comparable with L and in times greater than L/c. Thus it is true that, in such cases,

$$\gamma^{j}_{(\pm)} \sim \frac{1}{L} \tag{7-43}$$

where the symbol $\sim$ is to be read "is of the order of."

Substituting (7–41) into (7–9), we have

$$p^{j}{}_{(\pm)} = -\partial^{j}\Phi_{(\pm)} = \pm\hbar(\omega^{j} - \gamma^{j})_{(\pm)} \tag{7-44}$$

which is the exact version of (7–13).

We shall now estimate the importance of $\gamma_{(\pm)}{}^{j}$ in (7–41) and (7–44). In making these estimates we shall, to simplify the notation, drop the $(\pm)$ subscript with the understanding that the results apply equally well to either charged fluid. First we note that, since the effect of γ^{j} in (7–44) is (as we shall see) small, we have

$$|\hbar\omega^{\circ}| \sim mc \tag{7-45a}$$

$$|\hbar\vec{\omega}| \sim |\vec{p}| \sim |m\vec{v}| = mv \tag{7-45b}$$

where, for order of magnitude estimates, the contribution $\frac{q}{c}A^{j}$ to p^{j} has been neglected. (The magnitude of A^{j} is, in any case, uncertain because of the arbitrariness in the choice of gauge.) From (7–43) and (7–45) we have

$$\frac{|\gamma^{\circ}|}{|\omega^{\circ}|} \sim \frac{\hbar L^{-1}}{mc} = \frac{\lambda\!\!\!^{-}}{L} \tag{7-46a}$$

$$\frac{|\vec{\gamma}|}{|\vec{\omega}|} \sim \frac{\hbar L^{-1}}{mv} = \left(\frac{\lambda\!\!\!^{-}}{L}\right)\left(\frac{c}{v}\right) \tag{7-46b}$$

where

$$\lambda\!\!\!^{-} = \frac{\hbar}{mc} \tag{7-47}$$

is the Compton wavelength of the particle. Let $O(n)$ designate "a function of the order of magnitude of the number n." Using (7–46), we can write (7–41) and (7–44) as follows:

$$\frac{\partial\tilde{\gamma}}{\partial t} = -2c\omega^{\circ}[1 + O(\lambda\!\!\!^{-}/L)] \tag{7-48a}$$

$$\vec{\nabla}\tilde{\gamma} = 2\vec{\omega}\left[1 + O\left(\frac{\lambda\!\!\!^{-}c}{Lv}\right)\right] \tag{7-48b}$$

$$p^{\circ} = \sigma\hbar\omega^{\circ}[1 + O(\lambda\!\!\!^{-}/L)] \tag{7-49a}$$

$$\vec{p} = \sigma\hbar\vec{\omega}\left[1 + O\left(\frac{\lambda\!\!\!^{-}c}{Lv}\right)\right] \tag{7-49b}$$

where $\sigma = \pm 1$ is the sign of the particle charge.

To estimate $(\lambda\!\!\!^{-}/L$ and $(\lambda\!\!\!^{-}/L)(c/v)$ we use $\lambda\!\!\!^{-} \sim 10^{-11}cm$, which is appropriate for an electron; $L \sim 10^{11}cm$, which is the order of magnitude of the sun's radius; and $v \sim 10^{5}$ cm/sec, which is the order of magnitude of the rotational velocity at the surface of the sun. For these values we have

$$\frac{\lambda\!\!\!^{-}}{L} \sim 10^{-24} \tag{7-50a}$$

$$\frac{\lambda\!\!\!^{-}c}{Lv} \sim 10^{-19} \tag{7-50b}$$

which are very small indeed. We note that, in the neighborhood of a stagnation point where $v = 0$, the contribution $\frac{q}{c}A^{j}$ to p^{j}, which we neglected in (7–45), becomes important. In such a region we should replace v in (7–45)—(7–50) by $|\frac{q}{mc}A|$. This guarantees that the ratio (7–46b) remains small, even when $v = 0$, since in general $\vec{A} \neq 0$ at such points. (In fact, we can guarantee that this is the case by an appropriate choice of gauge).

The sole purpose of these estimates has been to demonstrate that, at least for the type of problem in which we are interested, the vector ω^{j} that we introduced to replace $-\frac{1}{2}\partial^{j}\tilde{\Psi}$, which was difficult to work with in calculations, is indeed very nearly equal to $-\frac{1}{2}\partial^{j}\tilde{\Psi}$. (7–48) also provides the quantitative justification for calling ω° the de Broglie frequency and $\vec{\omega}$ the de Broglie wave-vector. When we recall that, according to (3–2), p^{j} must be the gradient of a scalar, we note from (7–48) and (7–49) that this is also very nearly the case for ω^{j}.

All of the relations presented in this section have followed directly from the definitions that provide the link between the spinors ζ^α, χ^α and the kinetic and dynamical properties described by the particle tetrapod. It is evident that any spinor equation of motion that determines the space-time behavior of ζ^α and χ^α for given electromagnetic, gravitational, and thermal fields will indirectly specify the equations of motion of the fluid quantities described by the tetrapod, in particular the equations of motion of ρ and u^j. If it should turn out that the equations so determined are just the systems (6–1) or (6–6), then the postulated spinor equations of motion for ζ^α and χ^α is the desired spinor alternative to Euler's equation.

In the next section, we shall postulate a certain spinor equation of motion, and then using the definitions of the present section we shall derive the equations of motion satisfied by the quantities described by the tetrapod. We shall find that the resulting equations are identical to the system (6–6) supplemented by an equation of motion for the particle spin, which is interesting but, for the purposes of magnetogas dynamics, not of primary importance. The spin-dependent potential $a^j{}_{(\pm)}$ in (5–1a) will be given as an explicit function of $\zeta^\alpha{}_{(\pm)}$ and $\chi^\alpha{}_{(\pm)}$, and it will be shown that, for macroscopic problems, $a^j{}_{(\pm)}$ is completely negligible compared with the potentials A^j that are encountered in such problems.

Section VIII will merely summarize results. The derivation of these results is carried out in Appendix C. For a detailed understanding of Appendix C, some prior familiarity with spinor analysis would be helpful, and for this the reader is referred to references (11), (12), and (13).

As a necessary preliminary to discussing the spinor equation of motion, we note that the 4-gradient operator ∂_j can also be written with spinor indices. Thus, analogous to (7–21), we have (7-51)

$$\begin{pmatrix} \partial_{\bar{1}1}\partial_{\bar{1}2} \\ \partial_{\bar{2}1}\partial_{\bar{2}2} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \left[\frac{1}{c}\frac{\partial}{\partial t}+\frac{\partial}{\partial z}\right]\left[\frac{\partial}{\partial x}-i\frac{\partial}{\partial y}\right] \\ \left[\frac{\partial}{\partial x}+i\frac{\partial}{\partial y}\right]\left[\frac{1}{c}\frac{\partial}{\partial t}-\frac{\partial}{\partial z}\right] \end{pmatrix}$$

which can also be written

$$\partial_{\bar{\alpha}\beta} = \sigma^j{}_{\bar{\alpha}\beta}\partial_j \qquad (7\text{–}52)$$

whose inverse is

$$\partial_j = \sigma_j{}^{\beta\bar{\alpha}}\partial_{\bar{\alpha}\beta} \qquad (7\text{–}53)$$

The following operator identities are also of interest:

$$\partial^{\alpha\bar{\lambda}}\partial_{\bar{\lambda}\beta} = \partial_{\bar{\lambda}\beta}\partial^{\alpha\bar{\lambda}} = \tfrac{1}{2}\delta^\alpha_\beta\partial_l\partial^l = \tfrac{1}{2}\delta^\alpha_\beta\Box \qquad (7\text{–}54)$$

$$\sigma_j{}^{\beta\bar{\gamma}}\sigma_{\bar{\gamma}\alpha}{}^k\partial^{\alpha\bar{\lambda}}\partial_{\bar{\lambda}\beta} = \tfrac{1}{2}\delta^k_j\partial_l\partial^l = \tfrac{1}{2}\delta^k_j\Box \qquad (7\text{–}55)$$

From (7–54) it follows that

$$\Box = \partial_l\partial^l = \partial_{\bar{\alpha}\beta}\partial^{\beta\bar{\alpha}} \qquad (7\text{–}56)$$

In the next section we shall work with the irreducible covariant form $A_{\bar{\alpha}\beta}$ of the 4-vector potential where

$$\begin{pmatrix} A_{\bar{1}1}A_{\bar{1}2} \\ A_{\bar{2}1}A_{\bar{2}2} \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} [A_0+A_3][A_1-iA_2] \\ [A_1+iA_2][A_0-A_3] \end{pmatrix} = \frac{1}{\sqrt{2}}\begin{pmatrix} [A^0-A_z][-A_x+iA_y] \\ [-A_x-iA_y][A^0+Az] \end{pmatrix} \qquad (7\text{–}57)$$

where we have made use of the fact that it is the contravariant form A^j, rather than the covariant form A_j, that is regarded as the generalization to four dimensions of the 3-vector $\vec{A} = (A_x, A_y, A_z)$.

The following identity is useful for evaluating F^{jk} from $A_{\bar{\alpha}\beta}$:

$$2g^{jl}\sigma_l{}^{\beta\bar{\lambda}}\sigma_{\bar{\lambda}\alpha}{}^k(\partial^{\alpha\bar{\gamma}}A_{\bar{\gamma}\beta}) = g^{jk}(\partial_l A^l) - (\partial^j A^k - \partial^k A^j) - i\epsilon^{jkln}\partial_l A_n \qquad (7\text{–}58)$$
$$= g^{jk}(\partial_l A^l) - F^{jk} - i\hat{F}^{jk}$$

where we have used (4–3) and (4–5).

VIII SPINOR EQUATIONS OF MOTION

Spinor Equations

We postulate that the spinor equations of motion have the following form:

$$i\hbar\partial_{\bar{\alpha}\beta}\zeta^\beta_{(\pm)} = \frac{\mathfrak{m}_{(\pm)}c}{\sqrt{2}}\chi^{\bar{\alpha}}_{(\pm)} - \frac{q}{c}A_{\bar{\alpha}\beta}\zeta^\beta_{(\pm)} \qquad (8\text{–}1a)$$

$$i\hbar\partial_{\alpha\beta}\chi^\beta_{(\pm)} = \frac{\mathfrak{m}_{(\pm)}c}{\sqrt{2}}\zeta^{\bar{\alpha}}_{(\pm)} + \frac{q}{c}A_{\bar{\alpha}\beta}\chi^\beta_{(\pm)} \qquad (8\text{–}1b)$$

where

$$\mathfrak{m}_{(\pm)} = m_{(\pm)}[1+g/c^2+h_{(\pm)}/c^2] \qquad (8\text{–}1c)$$

We supplement these equations with the constraint (7-38):

$$I(\zeta_\alpha \chi^\alpha)_{(\pm)} = 0 \tag{8-1d}$$

which is equivalent to requiring that

$$\cos \epsilon_{(\pm)} = 0; \quad \sigma = \pm 1 \tag{8-1e}$$

where σ is the sign of the particle charge of the fluid under consideration.

We note that the signs preceding q in the above spinor equations are the same for both charged fluids. The reason for this is that the sign of the particle charge is treated as part of the solution of the spinor equations (the value of the phase ϵ), rather than as a parameter in the equations.

The equations (8-1) are just the spinor form of the Dirac equation with the generalization that the mass includes contributions arising from the gravitational and thermal energy.* It should be emphasized, however, that, except for the incorporation of the de Broglie Hypothesis into the theory, which introduced Planck's constant as a proportionality factor relating $\Phi_{(\pm)}$ and $\tilde{\Psi}_{(\pm)}$ (cf. (7-8)), the theory developed in this paper is completely classical in spirit. In particular, no quantization process has been introduced.

It is really not surprising that a Dirac-type equation should be the first candidate for investigation in a theory like the one developed here, since this equation is the simplest first-order linear equation that incorporates a scalar field ($\tilde{m}_{(\pm)}$) and a vector field ($A_{\alpha\beta}$), and also guarantees that the observable quantities of physical interest (ρ, u^j, μ^j, ω^j)$_{(\pm)}$ do not contain oscillatory terms with frequency of the order of the de Broglie frequency.

Derivation of the Euler System of Equations

Using the spinor equations of motion (8-1), and the definitions of Section VII that relate the observable quantities described by the tetrapod for the fluid to the spinors $\zeta^\alpha{}_{(\pm)}$ and $\chi^\alpha{}_{(\pm)}$, it is shown in Appendix C that we can arrive at the fluid equations of motion in terms of these observable quantities. In this section, we shall merely exhibit these equations, and discuss their physical significance. The most important of these equations are the following:

$$\partial_j(\rho u^j)_{(\pm)} = 0 \tag{8-2}$$

$$p^j_{(\pm)} = -\partial^j \Phi_{(\pm)} = (MCu^j)_{(\pm)} + \frac{\sigma q}{c}\left[A^j + a^j_{(\pm)}\right] \tag{8-3}$$

where

$$a^j_{(\pm)} = \frac{\hbar c}{q}(\pi^j - \gamma^j)_{(\pm)} \tag{8-4}$$

where $\gamma^j{}_{(\pm)}$ is defined by (7-42) and

$$\pi^j_{(\pm)} = \frac{1}{2\rho}\left\{\partial_k(\rho s^{jk}) - [u_k \partial_l(\rho s^{kl})]u^j\right\}_{(\pm)} \tag{8-5}$$
$$= \frac{1}{2}\left[\rho^{-1}\epsilon^{jklm}u_k \partial_l(\rho\mu_m) - s^{jk}u^l \partial_l u_k\right]_{(\pm)}$$

The $M_{(\pm)}$ that appears in (8-3) is defined as follows:

$$M_{(\pm)} \equiv \mathring{m}_{(\pm)} + \tilde{m}_{(\pm)} \tag{8-6}$$

where

$$\tilde{m}_{(\pm)} = \frac{1}{2c^2}\left[\frac{q\hbar}{2M_{(\pm)}c} s^{jk}_{(\pm)}(F_{jk} + f^{(\pm)}_{jk})\right]$$
$$= -\frac{1}{c^2}\left[\frac{q\hbar}{2M_{(\pm)}c}\vec{\mu}\cdot(\vec{B} + \vec{b})\right]_{(\pm)} \tag{8-7}$$

where

$$f^{jk}{}_{(\pm)} = \partial^j a^k{}_{(\pm)} - \partial^k a^j{}_{(\pm)} \tag{8-8a}$$

$$(f^{10}, f^{20}, f^{30})_{(\pm)} \equiv (e_x, e_y, e_z)_{(\pm)} = \vec{e}_{(\pm)} \tag{8-8b}$$
$$-(f^{23}, f^{31}, f^{12})_{(\pm)} \equiv (b_x, b_y, b_z)_{(\pm)} = \vec{b}_{(\pm)}$$

and the overhead "o" in (8-7) indicates the value of the 3-vector in question as seen in the particle rest-frame. (Note that, in general, $\vec{B}$ in the case of the proton gas is different from $\vec{B}$ in the case of the electron gas because the two rest-frames are different.)

Note that the definition for $M_{(\pm)}$ in (8-6) is implicit, rather than explicit, because $\tilde{m}_{(\pm)}$ involves $M_{(\pm)}$ itself (rather than $m_{(\pm)}$) in the denominator.

Comparing (8-7) with (6-1f), we make the following identification:

$$\mu_{(\pm)} = \frac{q\hbar}{2M_{(\pm)}c} \tag{8-9}$$

*See, for example, ref. 12, p. 1393, eq. 1. This equation does not include the factor $\sqrt{2}$ appearing in (8-1a) and (8-1b) simply because of a difference in the definition of the matrices $\sigma^j_{\alpha\beta}$.

where $\mu_{(\pm)}$ is the magnitude of the magnetic dipole moment of the particle. (8–9) is just the dipole moment that the Dirac theory yields (g-factor equals 2), with the modification that the total mass $M_{(\pm)}$, rather than the free-particle mass $m_{(\pm)}$, appears in the denominator.

The equations (8–2) and (8–3), together with the automatic normalization of the 4-velocity $\mu^j{}_{(\pm)}$ that was demonstrated in (7–40), constitute the system of equations given in (6–6). From this system we can derive, as was shown in Section VI, the Euler system of equations (6–1). Thus we have shown that a solution of the system of spinor equations (8–1) automatically yields, via the identities of Section VII, a solution of the Euler system of equations (6–1).

Magnitude of Spin-Dependent Electromagnetic Field

Now we shall demonstrate that, for magnetogas-dynamical systems on a laboratory or astronomical scale, the effects of the spin-dependent field $f^{jk}{}_{(\pm)}$ are completely negligible. (Recall that $f^{jk}{}_{(\pm)}$ has its origin in the small contribution $\zeta\rho^j{}_{(\pm)}$ to the fluid momentum density that resulted from the presence of the spin angular momentum density. Thus $f^{jk}{}_{(\pm)}$ is really a spin-dependent mechanical effect, although it has been treated, for reasons of intuitive clarity, as a contribution to the electromagnetic field.)

In Appendix C it is shown that, for solutions of (8–1) that yield functions for ρ, u^j, μ^j, and ω^j that change significantly only over distances of the order L and in time intervals greater than L/c, we have

$$|f^{jk}{}_{(\pm)}| \sim \frac{q}{\alpha L^2} \tag{8–10}$$

where α is the fine-structure constant:

$$\alpha = \frac{q^2}{\hbar c} = \frac{1}{137.04} \tag{8–11}$$

Table I summarizes the comparison between $|f^{jk}{}_{(\pm)}|$ and the magnitudes $|F^{jk}|$ of the electromagnetic fields encountered in the sun and in a system of laboratory dimensions. Only crude estimates for the lower limits of $|F^{jk}|$ have been given, but these suffice to indicate that in macroscopic systems the effects of $f^{jk}{}_{(\pm)}$ are completely negligible. The third line of Table 1 has been included merely to demonstrate that, if one were to attempt to use the spinor formulation of magnetogas dynamics to construct models of nuclei or elementary particles, the effects of the spin-dependent field $f^{jk}{}_{(\pm)}$ would be not at all negligible.

TABLE 1.

	L(cm)	$\lvert f^{jk}{}_{(\pm)}\rvert$ (Gauss)	$\lvert F^{jk}\rvert$ (Gauss)
Sun	10^{11}	10^{-29}	>1
Laboratory	1	10^{-7}	>1
Nucleus	10^{-13}	10^{19}	10^{15}

Effect of Variable ϵ

The side condition (8–1d) is analogous to the 4-velocity normalization condition (6–1c) in the Euler system of equations in that both conditions introduce an algebraic nonlinearity into the problem, since they are quadratic in the unknown functions. We shall now demonstrate, however, that the spinor system of equations has the great advantage that, for problems on an astronomical or laboratory scale, the nonlinear side condition (8–1d) may be ignored.

In this regard, we already noted following (7–40) that dropping the constraint (8–1d) has no effect on the automatic normalization of $\mu^j{}_{(\pm)}$. Moreover, in Appendix C it is shown that, if we ignore this constraint, no change in the continuity equation (8–2) results, and the effect on (8–3) can be represented as a spurious mass contribution $(\Delta_\epsilon m)_{(\pm)}$ and a spurious 4-vector potential $(\Delta_\epsilon A^j)_{(\pm)}$. Under the same conditions used in estimating $|f^{jk}{}_{(\pm)}|$, we find

$$\left|(\Delta_\epsilon A^j)_{(\pm)}\right| \sim \left|a^j{}_{(\pm)}\right| \left(\frac{\lambda\!\!\!^{-}(\pm)}{L}\right) \tag{8–12}$$

where $\lambda\!\!\!^{-}_{(\pm)}$ is the Compton wavelength introduced in (7–47). Thus, for macroscopic systems, the effect of $(\Delta_\epsilon A^j)_{(\pm)}$ is even smaller than that of $a^j{}_{(\pm)}$, whereas, for systems of nuclear dimensions, this is obviously no longer the case.

In order to estimate the effect of the spurious mass $(\Delta_\epsilon m)_{(\pm)}$, we note that this may be regarded as arising from a spurious contribution to the specific enthalpy, which in turn may be regarded

as arising from an error $(\Delta_\epsilon T)_{(\pm)}$ in the fluid temperature. In Appendix C, it is shown that

$$|(\Delta_\epsilon T)_{(\pm)}| \sim \left(\frac{m_{(\pm)}c^2}{k}\right)\left(\frac{\lambda\!\!\!^{-}_{(\pm)}}{L}\right)^2 \qquad (8\text{–}13)$$

where k is the Boltzmann constant. In Table 2 $|\Delta_\epsilon T_{(-)}|$ (i.e., the error in the electron temperature) is compared with the range of temperatures encountered in the sun, laboratory, and in nuclear systems. (The case for protons leads to even smaller estimates.)

TABLE 2.

	L(cm)	$\lvert\Delta_\epsilon T_{(-)}\rvert$ (°K)	T(°K)
Sun	10^{11}	10^{-33}	10^3-10^7
Lab	1	10^{-11}	$0-10^3$
Nucleus	10^{-13}	10^{15}	$0-10^{10}$

Obviously, the effect of $\Delta_\epsilon T_{(\pm)}$ is completely negligible in the first two cases, and not at all negligible in the last case.

Finally, it is shown in Appendix C that

$$|\epsilon_{(+)}| \sim \frac{\lambda\!\!\!^{-}_{(+)}}{L}; \quad |\pi - \epsilon_{(-)}| \sim \frac{\lambda\!\!\!^{-}_{(-)}}{L} \qquad (8\text{–}14)$$

Thus, neglecting the side condition (8–1d) in the case of macroscopic systems will not produce a change in the sign of cos $\epsilon_{(\pm)}$, which we interpret as the sign of the electric charge of the particle. Another way of interpreting (8–14) is to say that, for macroscopic systems, the equations (8–1a) and (8–1b) come very close to satisfying the condition (8–1d) automatically, without anything being done to bring this about.

In summation, we conclude that, in the case of macroscopic magnetogas-dynamical systems, ignoring the side condition (8–1d) will introduce no significant error into the solution.

Stress-Energy Tensor

In Appendix C it is shown that the following equation results from the spinor equations (8–1):

$$\partial_k T^{jk}_{(\pm)} = \partial_k T^{kj}_{(\pm)} = [\rho\partial^j(\breve{m}c^2)]_{(\pm)} + F^{jk}(\sigma_Q \rho\mu_k)_{(\pm)} = k^j{}_{(\pm)} \qquad (8\text{–}15)$$

where

$$T^{jk}_{(\pm)} = \left[\rho\breve{m}c^2u^ju^k + \frac{\sigma\hbar c}{2}u^j\partial_l(\rho s^{kl}) - \frac{\sigma\hbar c}{2}\rho s^{jl}\partial^k u_l\right] \qquad (8\text{–}16)$$

and $k^j{}_{(\pm)}$ is the 4-force density acting on the fluid.

We see in (8–15) that it makes no difference whether we contract ∂_k with $T^{jk}{}_{(\pm)}$ or $T^{kj}{}_{(\pm)}$. This is a non-trivial result, inasmuch as the tensor $T^{jk}{}_{(\pm)}$ is not symmetric.

Equation (8–15) justifies our regarding $T^{jk}{}_{(\pm)}$ as the stress-energy tensor of the fluid. Comparing (8–16) with (6–11), we see that $t^{jk}{}_{(\pm)}$, the spin-dependent part of the stress-energy tensor, has the following form:

$$t^{jk}_{(\pm)} = \frac{\sigma\hbar c}{2}[u^j\partial_l(\rho s^{kl}) - \rho s^{jl}\partial^k u_l]_{(\pm)} \qquad (8\text{–}17)$$

Angular Momentum

In Appendix C it is shown that the following equation results from the system of spinor equations (8–1):

$$\partial_l M^{jkl}_{(\pm)} = \frac{1}{c}[x^j k^k_{(\pm)} - x^k k^j_{(\pm)}] \qquad (8\text{–}18)$$

where

$$M^{jkl}_{(\pm)} = -\frac{\sigma\hbar}{2}\epsilon^{jkln}(\rho\mu_n)_{(\pm)} + \frac{1}{c}[x^jT^{kl}_{(\pm)} - x^kT^{jl}_{(\pm)}] \qquad (8\text{–}19)$$

Comparing (8–19) with (6–27) and (6–30), we make the following identifications:

$$S^{jkl}_{(\pm)} = -\frac{\sigma\hbar}{2}\epsilon^{jkln}(\rho\mu_n)_{(\pm)} \qquad (8\text{–}20)$$

and

$$\vec{S}_{(\pm)} \equiv (S^{230}, S^{310}, S^{120})_{(\pm)} = \frac{\sigma\hbar}{2}(\rho\vec{\mu})_{(\pm)} = \left[(\rho u^0)\left(\frac{\sigma\hbar}{2}\frac{\vec{\mu}}{u^0}\right)\right]_{(\pm)} \qquad (8\text{–}21)$$

Thus $\frac{\sigma\hbar}{2}(\rho\vec{\mu})_{(\pm)}$ is the spin angular momentum density and, since $(\rho\mu^0)_{(\pm)}$ is the particle density in the observer's frame, $\left(\frac{\sigma\hbar}{2}\frac{\vec{\mu}}{u^0}\right)_{(\pm)}$ is the spin angular momentum of a single particle in the observer's frame.

In this regard, it is interesting to note that it follows directly from the Lorentz transformation that

$$\frac{\vec{\mu}}{u^0}=\frac{\overset{0}{\vec{\mu}}_{\parallel}}{}+\frac{\overset{0}{\vec{\mu}}_{\perp}}{u^0} \tag{8-22a}$$

where

$$\overset{0}{\vec{\mu}}_{\parallel}=\left(\frac{\overset{0}{\vec{\mu}}\cdot\vec{v}}{v^2}\right)\vec{v} \tag{8-22b}$$

is the component of $\overset{0}{\vec{\mu}}$ parallel to $\vec{v}$, and

$$\overset{0}{\vec{\mu}}_{\perp}=\frac{(v\times\overset{0}{\vec{\mu}})\times\vec{v}}{v^2} \tag{8-22c}$$

is the component of $\overset{0}{\vec{\mu}}$ perpendicular to $\vec{v}$. Thus

$$\lim_{\vec{v}\to c}\left(\frac{\vec{\mu}}{u^0}\right)\to\overset{0}{\vec{\mu}}_{\parallel} \tag{8-23}$$

which corresponds to the well-known property of particle spin to align itself parallel or antiparallel (depending on the sign of $\overset{0}{\vec{\mu}}\cdot\vec{v}$ in (8–22b)) to the particle velocity at high speeds.

Finally, it is shown in Appendix C that in the fluid rest-frame the following equation is valid:

$$\left[\frac{d\left(\frac{\sigma\hbar}{2}\overset{0}{\vec{\mu}}\right)}{d\tau}\right]_{(\pm)}=\left[\left(\frac{q\hbar\overset{0}{\vec{\mu}}}{2Mc}\right)\times(\overset{0}{\vec{B}}+\overset{0}{\vec{b}})\right]_{(\pm)} \tag{8-24}$$

This equation describes the precession in its own rest-frame of a particle having a spin angular momentum $\left(\frac{\sigma\hbar}{2}\overset{0}{\vec{\mu}}\right)_{(\pm)}$ and a magnetic moment $\left(\frac{q\hbar}{2Mc}\overset{0}{\vec{\mu}}\right)_{(\pm)}$ corresponding to a gyromagnetic ratio $\left(\frac{\sigma q}{Mc}\right)_{(\pm)}$ which in turn corresponds to a g-factor of 2 (referred, of course, to the total particle mass $M_{(\pm)}$ rather than the free particle mass $m_{(\pm)}$). The precession produced by $\overset{0}{\vec{b}}_{(\pm)}$ can be regarded as the result of a self-interaction of the electron or proton fluid with itself. In any case, Table 1 shows us that $|\overset{0}{\vec{b}}_{(\pm)}|\ll|\overset{0}{\vec{B}}|$ for macroscopic systems.

Dimensionless Form of Equations

Let us refer the particle masses to the electron mass m_e where

$$m_e=9.107\times10^{-28}gm \tag{8-25}$$

Thus, if $m_{(-)}$ and $m_{(+)}$ are the electron and proton masses respectively, we have

$$\overset{\prime}{m}_{(-)}=\frac{m_{(-)}}{m_e}=1 \tag{8-26a}$$

$$\overset{\prime}{m}_{(\pm)}=\frac{m_{(+)}}{m_e}=1836.1 \tag{8-26b}$$

where the dimensionless masses have been indicated by an overhead stroke. We shall refer all lengths to the Compton wavelength of the electron $\lambda\!\!\!^{-}_e$ where

$$\lambda\!\!\!^{-}_e=\frac{\hbar}{m_ec}=3.862\times10^{-11}\,\text{cm} \tag{8-27}$$

Thus we have the following relations between the dimensional and dimensionless (indicated by a stroke) quantities of principal interest:

$$t=(\lambda\!\!\!^{-}_e/c)\overset{\prime}{t}{}^{\circ}=1.2882\times10^{-21}\overset{\prime}{t}{}^{\circ}\ \text{sec} \tag{8-28a}$$

$$x^j=\lambda\!\!\!^{-}_e\overset{\prime}{x}{}^j=3.862\times10^{-11}\overset{\prime}{x}{}^j\ \text{cm} \tag{8-28b}$$

$$\partial_j=\lambda\!\!\!^{-}_e{}^{-1}\overset{\prime}{\partial}_j=2.590\times10^{10}\overset{\prime}{\partial}_j\ \text{cm}^{-1} \tag{8-28c}$$

$$\Box=\lambda\!\!\!^{-}_e{}^{-2}\overset{\prime}{\Box}=6.707\times10^{20}\overset{\prime}{\Box}\ \text{cm}^{-2} \tag{8-28d}$$

$$u^j{}_{(\pm)}=\overset{\prime}{u}{}^j{}_{(\pm)} \tag{8-28e}$$

$$\omega_{(\pm)}{}^j=\lambda\!\!\!^{-}_e{}^{-1}\omega_{(\pm)}{}^j=2.590\times10^{10}\overset{\prime}{\omega}_{(\pm)}{}^j\ \text{cm}^{-1} \tag{8-28f}$$

$$\rho_{(\pm)}=\lambda\!\!\!^{-}_e{}^{-3}\rho_{(\pm)}=1.7372\times10^{30}\overset{\prime}{\rho}_{(\pm)}cm^{-3} \tag{8-28g}$$

$$(\zeta^\alpha,\chi^\alpha)_{(\pm)}=\lambda\!\!\!^{-}_e{}^{-3/2}(\overset{\prime}{\zeta}{}^\alpha,\overset{\prime}{\chi}{}^\alpha)=4.168\times10^{15}(\overset{\prime}{\zeta}{}^\alpha,\overset{\prime}{\chi}{}^\alpha)\ \text{cm}^{-3/2} \tag{8-28h}$$

$$A^j=(q/\lambda\!\!\!^{-}_e)\overset{\prime}{A}{}^j=12.434\overset{\prime}{A}{}^j\ \text{statvolt} \tag{8-28i}$$

$$N_{(\pm)}=qN_{(\pm)}=4.802\times10^{-10}\overset{\prime}{N}_{(\pm)}\ \text{statvolt}-\text{cm} \tag{8-28j}$$

$$F^{jk}=(q/\lambda\!\!\!^{-}_e{}^2)\overset{\prime}{F}{}^{jk}=3.220\times10^{11}\overset{\prime}{F}{}^{jk}\ \text{statvolt/cm} \tag{8-28k}$$

$$g=(\alpha c^2)\overset{\prime}{g}=6.55\times10^{18}\overset{\prime}{g}\ \text{erg/gm} \tag{8-28l}$$

$$h_{(\pm)}=(\alpha c^2)(\overset{\prime}{H}\overset{\prime}{h})_{(\pm)}=6.558\times10^{18}(\overset{\prime}{H}\overset{\prime}{h})_{(\pm)}\ \text{erg/gm} \tag{8-28m}$$

Two dimensionless coupling constants will be needed: the fine-structure constant α,

$$\alpha=\frac{q^2}{\hbar c}=7.297\times10^{-3}=\frac{1}{137.04} \tag{8-29}$$

and the dimensionless gravitational constant $\overset{\prime}{\Gamma}$,

$$\overset{\prime}{\Gamma}=\Gamma\left(\frac{m_e}{q}\right)^2=2.398\times10^{-43} \tag{8-30}$$

The dimensionless spinor equations (8–1) have the form

$$i\partial'_{\bar{\alpha}\beta}\zeta'^{\beta}_{(\pm)}=\frac{m'_{(\pm)}}{\sqrt{2}}\chi'^{(\pm)}_{\alpha}+\alpha\{m'_{(\pm)}[g'+(H'h')_{(\pm)}]\chi'_{\bar{\alpha}}-[A'_{\bar{\alpha}\beta}+\partial'_{\alpha\bar{\beta}}N'_{(\pm)}]\zeta'^{\beta}_{(\pm)}\} \quad (8\text{–}31a)$$

$$i\partial'_{\bar{\alpha}\beta}\chi'^{\beta}_{(\pm)}=\frac{m'_{(\pm)}}{\sqrt{2}}\zeta'^{(\pm)}_{\bar{\alpha}}+\alpha\{m'_{(\pm)}[g'+(H'h')_{(\pm)}]\zeta'_{\bar{\alpha}}+[A'_{\bar{\alpha}\beta}+\partial'_{\bar{\alpha}\beta}N'_{(\pm)}]\chi'^{\beta}_{(\pm)}\} \quad (8\text{–}31b)$$

where $N'_{(-)}$ and $N'_{(+)}$ are dimensionless real scalar functions of the space-time coordinates that determine the gauges of the 4-vector potential that one chooses to use in the solution of the spinor equations for the electron and proton gases. There is no need that the same gauge be used in the two cases. These scalar functions, which did not appear in (8–1), have been inserted in (8–31) simply to make the arbitrariness of the gauge more explicit.

The algebraically nonlinear condition (8–1d) has not been included in (8–31) because we saw that ignoring it introduced no significant error into the solutions of macroscopic problems.

The field equations (6–8) in dimensionless form are as follows:

$$\left.\begin{aligned}&\rho'_{(\pm)}\partial'_{\beta}h'_{(\pm)}=[\gamma_{(\pm)}-1]h'_{(\pm)}\partial'_{\beta}\rho'_{(\pm)} \quad (8\text{–}32a)\\ \text{or}\\ &h'_{(\pm)}=[(\rho'/\rho'_{0})^{(\gamma-1)}]_{(\pm)} \quad (8\text{–}32b)\end{aligned}\right\}$$

$$\Box'A'_{\bar{\alpha}\beta}-\partial'_{\bar{\alpha}\beta}(\partial'^{\bar{\nu}\lambda}A'_{\bar{\lambda}\nu})=4\pi[(\rho'\mu'_{\bar{\alpha}\beta})_{(+)}-(\rho'\mu'_{\bar{\alpha}\beta})_{(-)}] \quad (8\text{–}32c)$$

$$\Box'g'=-4\pi\Gamma'[(\rho'm')_{(+)}+(\rho'm')_{(-)}] \quad (8\text{–}32d)$$

where from (7–28) and (7–35)

$$\rho'_{(\pm)}=\sqrt{2}\,|\zeta'_{\alpha}\chi'^{\alpha}|_{(\pm)}-\sqrt{2}\,|\zeta'^{1}\chi'^{2}-\zeta'^{2}\chi'^{1}|_{(\pm)} \quad (8\text{–}32e)$$

and from (7–20)

$$(\rho'u'_{\bar{\alpha}\beta})_{(\pm)}=(\zeta'_{\bar{\alpha}}\zeta'_{\beta}+\chi'_{\bar{\alpha}}\chi'_{\beta})_{(\pm)} \quad (8\text{–}32f)$$

or, using (7–28)

$$(\rho'u'_{\bar{1}1})_{(\pm)}=(\zeta'^{\bar{2}}\zeta'^{2}+\chi'^{\bar{2}}\chi'^{2})_{(\pm)}$$

$$(\rho'u'_{\bar{2}2})_{(\pm)}=(\zeta'^{\bar{1}}\zeta'^{1}+\chi'^{\bar{1}}\chi'^{1})_{(\pm)}$$

$$(\rho'u'_{\bar{1}2})_{(\pm)}=-(\zeta'^{\bar{2}}\zeta'^{1}+\chi'^{\bar{2}}\chi'^{1})_{(\pm)}$$

$$(\rho'u'_{\bar{2}1})_{(\pm)}=-(\zeta'^{\bar{1}}\zeta'^{2}+\chi'^{\bar{1}}\chi'^{2})_{(\pm)}$$

where the vertical bars in (8–32e) indicate the modulus of the invariant $(\zeta'_{\alpha}\chi'^{\alpha})_{(\pm)}$ which, if (8–1d) is not satisfied, will be complex (although the imaginary part will be extremely small for macroscopic problems).

Note that we have written the electromagnetic field equation (8–32c) in terms of the irreducible form $A_{\bar{\alpha}\beta}$ of the 4-potential instead of the usual form A^{j} because it is $A_{\bar{\alpha}\beta}$ that appears in the spinor equations (8–31).

In (8–32a) and (8–32b) both the differential and integrated forms of the "thermal field equation" have been given. For problems involving harmonic expansions, the differential form (8–32a) would probably be more convenient, but for the purposes of the present discussion we shall refer to the integrated form (8–32b).

It will be noted from (8–32) that the source terms of all the field equations are algebraically nonlinear expressions of the spinor components; but, given the functional dependence of the spinor components on the space-time coordinates, it is a simple matter to determine the source terms and then solve the field equations. (Working with (8–32b), one would use the binomial theorem to generate a series involving the ratio of the higher harmonics in the harmonic expansion for ρ to the lowest harmonic, and then terminate the series after enough terms to provide sufficient accuracy.) The fields found by this procedure could then be substituted into the spinor equations (8–31) in order to calculate spinor functions more accurate than those used to calculate the fields. The solution of the equations (8–31) for given g', h', and $A'_{\bar{\alpha}\beta}$ is no problem since the equations are linear. This procedure is the basis for a straight-forward iteration solution of the complete magnetogasdynamical problem. The equations (8–31) have been cast into a form in which all the interaction terms are contained in the square braces preceded by the coupling constant $\alpha=\frac{1}{137}$. The smallness of this constant encourages the hope that the convergence of the solution would be rapid.

It has been noted that the gauge function $N'_{(\pm)}$ is completely arbitrary, and its presence in the spinor equations is a great asset since it can be chosen partially to cancel out unwanted higher harmonics in g' and $h'_{(\pm)}$ which necessarily arise because the source terms of the field equations are algebraically nonlinear functions of the spinor components. To a lesser degree the constants $H'_{(\pm)}$ and $H'_{(-)}$ may also be adjusted to minimize

the effects of higher harmonics, but we must recall that the choice of these two constants is restricted by the constraint (6–9) which in dimensionless form reads

$$\overset{\prime}{m}_{(-)}\overset{\prime}{H}_{(-)} \approx \overset{\prime}{m}_{(+)}\overset{\prime}{H}_{(+)} \qquad (8\text{–}33)$$

How close this relation must come to being an exact equality depends on the magnitude of the difference between $\rho_0{}^{(-)}$ and $\rho_0{}^{(+)}$, and how great a difference between the temperatures of the electron and proton gases one is willing to regard as physically admissible.

It has already been noted that there is no need that $\overset{\prime}{N}_{(+)}$ and $\overset{\prime}{N}_{(-)}$ be equal. Moreover, we are free to choose different $\overset{\prime}{N}_{(\pm)}$ and $\overset{\prime}{H}_{(\pm)}$ at each different stage of the iteration. It may even be possible in certain time-independent problems to choose $\overset{\prime}{N}_{(\pm)}$ and $\overset{\prime}{H}_{(\pm)}$ in such a way as to allow a separation of the angular and radial dependence of the equations. In such a case an iteration procedure would be unnecessary since the system of partial differential equations would reduce to a nonlinear system of ordinary differential equations which could be solved by numerical means.

In order to make the significance of $\overset{\prime}{N}_{(\pm)}$ clearer, it has been shown in Appendix C that introducing a gauge transformation characterized by the real scalar function $\overset{\prime}{N}_{(\pm)}$ is equivalent to multiplying $\overset{\prime}{\zeta}{}^{\beta}{}_{(\pm)}$ by $e^{-i\alpha\overset{\prime}{n}(\pm)}$ and $\overset{\prime}{\chi}{}^{\beta}{}_{(\pm)}$ by $e^{i\alpha\overset{\prime}{n}(\pm)}$.

APPENDIX A: ENERGY INJECTION AND VISCOSITY

Energy Injection

In Section II we saw that the possibility existed of taking energy injection into account by choosing an appropriate functional form for $Q^{j}{}_{(\pm)}$. Let us postulate the following form for $Q^{j}{}_{(\pm)}$:

$$Q^{j}{}_{(\pm)} = (\Lambda c\mu^{j})_{(\pm)} \qquad (A\text{–}1)$$

where $\Lambda_{(\pm)}$ is a scalar function of the fluid coordinates whose physical significance is discussed below. Substituting (A–1) into (2–22), we find that Euler's equation in the fluid rest-frame has the following form:

$$\left[\rho\frac{\partial(\breve{m}c^{2})}{\partial\tau}\right]_{(\pm)} = \left[\rho m\frac{\partial g}{\partial\tau}+\frac{\partial p}{\partial\tau}+\rho(\Lambda mc^{2})\right]_{(\pm)} \qquad (A\text{–}2a)$$

$$\left\{\rho\left[\frac{\partial(\breve{m}\breve{v})}{\partial\tau}\right]_{v=0}\right\}_{(\pm)} = -[\rho m\overset{\circ}{\nabla}g+\overset{\circ}{\nabla}p]_{(\pm)} \pm q\rho_{(\pm)}\overset{\circ}{E}$$

Thus the choice (A–1) produces no new 3-force in the fluid rest-frame, but it does produce a new term $\rho(\Lambda mc^{2})$ in (A–2a). This corresponds to an energy injection into the fluid in which Λ represents the fraction of the particle rest-energy that is injected per particle per unit time. Obviously Λ has the units of inverse time.

In general, $(\Lambda mc^{2})_{(+)} \neq (\Lambda mc^{2})_{(-)}$. The reason for this is that the nuclear reactions give the injected energy to the protons rather than to the electrons, and it is only through collisions that the electrons receive their share of the energy. If, however, we assume that energy equipartition establishes itself instantaneously (i.e., in a time very short compared with the time for the fluid to move a significant distance), then

$$(\Lambda m)_{(+)} = (\Lambda m)_{(-)} \qquad \text{(instantaneous equipartition)} \qquad (A\text{–}3)$$

Although energy injection introduces no new 3-force in the fluid rest-frame, we note from (2–22b) and (A–1) that in the observer's reference frame a 3-force $(\rho\mu^{0})\,(\Lambda m\vec{v})$ appears, where $\Lambda m\vec{v}$ is just the momentum per particle that is injected along with the energy. (If the particle energy, hence its mass, were increased without the supplementary force $\Lambda m\vec{v}$, then, by Newton's Second Law, the particle would necessarily slow down.)

Obviously, (A–1) applies to the case of energy loss (through radiation), as well as to energy injection. It is only necessary to make Λ negative instead of positive. For negative Λ, the force $\Lambda m\vec{v}$ represents a drag on the particle. This is the radiation drag that occurs when a charged particle loses energy by radiation.

As was emphasized in Section II, the use of (A–1) instead of the adiabatic condition (2–30) makes no difference in the form of Euler's equation written in the form (2–24). It does, however, change the "thermal field equation." From (2–15) and (A–1) we see that, instead of (2–31), we have

$$[\rho\partial^{j}h-(\gamma-1)h\partial^{j}\rho-\gamma\Lambda c(\rho\mu^{j};)]_{(\pm)}=0 \qquad (A\text{–}4)$$

Viscosity

For the sake of illustration, a rough and simple way will be given to take into account the viscous interaction between the two charged fluids. Let

$$Q^{j}{}_{(\pm)} = \frac{F}{(\rho m)_{(\pm)}}[(\rho u^{j})_{(\mp)}-(\rho u^{j})_{(\pm)}] \qquad (A\text{–}5)$$

where F is a suitably chosen invariant function of the fluid variables which has the dimensions of a force.

Now we note that for the proton gas, for example, the term $(\rho mcQ^0)_{(+)}$ on the right side of (2–22a), which represents the rate of energy injection per unit volume, has the following form:

$$(\rho mcQ^0)_{(+)} = cF[(\rho\mu^0)_{(-)} - (\rho\mu^0)_{(+)}] \quad \text{(A–6)}$$

The quantity in the brackets is the total change density which, because of the electrostatic screening effect, must be everywhere very close to zero. Thus, the energy injection is also very close to zero. The energy injection for the electron gas is just the negative of (A–6). Thus the energy gained by one field is lost by the other, and the total energy generated is zero.

We note that the force term $(\rho m\vec{Q})_{(+)}$ on the right side of (2–22b), which represents the viscous force on the proton gas, has the form

$$(\rho m\vec{Q})_{(+)} = \frac{F}{c}\left[(\rho u^0\vec{v})_{(-)} - (\rho u^0\vec{v})_{(+)}\right] \quad \text{(A–7)}$$

The second term in the brackets represents a dragging force that acts to slow the proton gas down. The first term is a force which tends to accelerate the proton gas in the direction in which the electron gas is moving. These are just the effects we would expect viscosity to produce. The viscous force acting on the electron gas is just the negative of the viscous force acting on the proton gas. If, for any reason, we wished to free ourselves of this constraint, we could do so very easily by modifying (A–5) so that $Q^j{}_{(+)}$ and $Q^j{}_{(-)}$ involved different invariant functions $F_{(+)}$ and $F_{(-)}$. In this case, the total energy injected into both fields at each point in space by means of the viscous forces is not in general zero.

From (2–15), we see that when $Q^j{}_{(\pm)}$ is given by (A–5) the "thermal field equation" has the following form:

$$\left[\rho\partial^j h - (\gamma-1)h\partial^j\rho + \frac{\gamma F}{m}(\rho u^j)\right]_{(\pm)} = F\gamma_{(\pm)}\left[\frac{(\rho u^j)}{m}\right]_{(\mp)} \quad \text{(A–8)}$$

The approximation based on (A–5) does not take into account the self-interaction of either fluid with itself that results because of viscosity. To do this we would have to introduce derivatives of the velocities into (A–5). In particular, we would introduce a term involving a suitable covariant generalization of the expression $(\vec{\nabla}\cdot\vec{\nabla})\vec{v}_{(\pm)}$.

APPENDIX B: EXPLICIT EXPRESSIONS FOR THE PARTICLE SPINORS

We saw in Section VII that the particle tetrapod has seven degrees of freedom, one of which, the particle density ρ, drops out if we work with a tetrapod normalized to unity. The remaining six degrees of freedom are to be identified with the three components of the particle velocity, and the three Euler angles needed to specify the orientation of the particle tripod in the rest-frame of the particle.

Rather than use the three Cartesian components of the particle velocity, we shall use $\beta = v/c$ and the two spherical angles $\hat{\theta}$ and $\hat{\varphi}$ that specify the direction of the particle velocity in the observer's reference frame. (The caret over $\hat{\theta}$ and $\hat{\varphi}$ merely indicates that these angles refer to the velocity, and distinguishes them from the angles θ, φ which will be used to specify the orientation of the particle tripod.) Thus we write the particle 4-velocity as follows:

$$\begin{aligned}
&\mu^0 = (1-\beta^2)^{-\frac{1}{2}};\ \beta = v/c \\
&\mu^1 = \mu^0\beta \sin\hat{\theta}\cos\hat{\varphi} \\
&\mu^2 = \mu^0\beta \sin\hat{\theta}\sin\hat{\varphi} \\
&\mu^3 = \mu^0\beta \cos\hat{\theta}
\end{aligned} \quad \text{(B–1)}$$

(There will be no need in this appendix for (+) and (−) subscripts since it is understood that everything applies equally well to either of the two charged fluids.)

In order to specify the angular orientation of the particle tripod, we imagine that the observer accelerates himself and his coordinate system in the direction of the particle velocity until his velocity is the same as that of the particle, i.e., until he is in the particle rest-frame. This is accomplished without rotating his coordinate system. The orientation of the tripod $\vec{k}$, $\vec{\lambda}$, and $\vec{\mu}$ is then easily specified in terms of the Euler angles φ, θ, and Ψ that would carry the observer's

own x, y, and z axes over into the positions occupied by the $\vec{\hat{k}}$, $\vec{\hat{\lambda}}$, and $\vec{\hat{\mu}}$ vectors respectively. Thus, we first rotate the observer's tripod of coordinates axes an angle $\tilde{\varphi}$ about his z axis, then an angle $\tilde{\theta}$ about his y *axis*, and finally an angle $\tilde{\Psi}$ about the direction in which the observer's z axis finds itself after the first two rotations have been carried out. The observer's tripod, like the tripod $(\vec{\hat{k}}, \vec{\hat{\lambda}}, \vec{\hat{\mu}})$ is assumed to constitute a right-handed system, and the positive directions for the rotations are given by the right-hand screw rule. Note that $\tilde{\theta}$ specifies the rotation about the y *axis*, rather than the x axis as is frequently the case. The reason for this choice is that it makes the angular dependence of $\vec{\hat{\mu}}$ similar to that of (μ^1, μ^2, μ^3) in (B–1). Thus we have

$$\begin{aligned}\hat{\mu}^1 &= \sin\tilde{\theta}\cos\tilde{\varphi}\\ \hat{\mu}^2 &= \sin\tilde{\theta}\sin\tilde{\varphi}\\ \hat{\mu}^3 &= \cos\tilde{\theta}\end{aligned} \tag{B–2a}$$

$$\begin{aligned}\hat{k}^1 &= \cos\tilde{\theta}\cos\tilde{\varphi}\cos\tilde{\Psi} - \sin\tilde{\varphi}\sin\tilde{\Psi}\\ \hat{k}^2 &= \cos\tilde{\theta}\sin\tilde{\varphi}\cos\tilde{\Psi} + \cos\tilde{\varphi}\sin\tilde{\Psi}\\ \hat{k}^3 &= -\sin\tilde{\theta}\cos\tilde{\Psi}\end{aligned} \tag{B–2b}$$

$$\begin{aligned}\hat{\lambda}^1 &= -(\cos\tilde{\theta}\cos\tilde{\varphi}\sin\tilde{\Psi} + \sin\tilde{\varphi}\cos\tilde{\Psi})\\ \hat{\lambda}^2 &= -(\cos\tilde{\theta}\sin\tilde{\varphi}\sin\tilde{\Psi} - \cos\tilde{\varphi}\cos\tilde{\Psi})\\ \hat{\lambda}^3 &= \sin\tilde{\theta}\sin\tilde{\Psi}\end{aligned} \tag{B–2c}$$

The expressions (B–2) are valid in the particle rest-frame. It is now a straight-forward matter to transform these three vectors to the observer's original frame of reference. Let $\vec{v}_1$ be the velocity unit 3-vector, i.e.

$$\vec{v}_1 = \vec{v}/v = (\sin\hat{\theta}\cos\hat{\varphi},\ \sin\hat{\theta}\sin\hat{\varphi},\ \cos\hat{\theta}) \tag{B–3}$$

Then from the Lorentz transformation we have

$$\mu^0 = \mu^0\beta(\vec{\hat{\mu}}\cdot\vec{v}_1) \tag{B–4a}$$

$$\vec{\mu} = (\mu^1, \mu^2, \mu^3) = \vec{\hat{\mu}} + (u^0 - 1)(\vec{\hat{\mu}}\cdot\vec{v}_1)\vec{v}_1 \tag{B–4b}$$

where

$$\vec{\hat{\mu}}\cdot\vec{v}_1 = \sin\hat{\theta}\sin\tilde{\theta}\cos(\hat{\varphi}-\tilde{\varphi}) + \cos\hat{\theta}\cos\tilde{\theta} \tag{B–4c}$$

$$k^0 = \mu^0\beta(\vec{\hat{k}}\cdot\vec{v}_1) \tag{B–5a}$$

$$\vec{k} = (k^1, k^2, k^3) = \vec{\hat{k}} + (u^0 - 1)(\vec{\hat{k}}\cdot\vec{v}_1)\vec{v}_1 \tag{B–5b}$$

where

$$\vec{\hat{k}}\cdot\vec{v}_1 = \sin\hat{\theta}\sin(\hat{\varphi}-\tilde{\varphi})\sin\tilde{\Psi} + \sin\hat{\theta}\cos\tilde{\theta}\cos(\hat{\varphi}-\tilde{\varphi})\cos\tilde{\Psi} - \cos\hat{\theta}\sin\tilde{\theta}\cos\tilde{\Psi} \tag{B–5c}$$

$$\lambda^0 = u^0\beta(\vec{\hat{\lambda}}\cdot\vec{v}_1) \tag{B·6a}$$

$$\vec{\lambda} = (\lambda^1, \lambda^2, \lambda^3) = \vec{\hat{\lambda}} + (u^0 - 1)(\vec{\hat{\lambda}}\cdot\vec{v}_1)\vec{v}_1 \tag{B–6b}$$

where

$$\vec{\hat{\lambda}}\cdot\vec{v}_1 = \sin\hat{\theta}\sin(\hat{\varphi}-\tilde{\varphi})\cos\tilde{\Psi} - \sin\hat{\theta}\cos(\hat{\varphi}-\tilde{\varphi})\cos\tilde{\theta}\sin\tilde{\Psi} + \cos\hat{\theta}\sin\tilde{\theta}\sin\tilde{\Psi} \tag{B–6c}$$

By means of (B–1) – (B–6), all sixteen components of the tetrapod (normalized to unity) have been specified as functions of β, $\hat{\theta}$, $\hat{\varphi}$, $\tilde{\theta}$, $\tilde{\varphi}$, and $\tilde{\Psi}$. We must now find expressions for the components of ζ^α and χ^α in terms of these variables (plus ρ and ϵ) such that the relations (7–27) are satisfied. It can be verified by direct calculation that the following are the desired expressions:

$$\zeta^\alpha = 2^{-1/4}\rho^{1/2}e^{1/2(\epsilon-\tilde{\Psi})}\hat{\zeta}^\alpha \tag{B–7a}$$

$$\chi^\alpha = 2^{-1/4}\rho^{1/2}e^{1/2(\epsilon+\tilde{\Psi})}\hat{\chi}^\alpha \tag{B–7b}$$

where

$$\hat{\zeta}^1 = [\cosh v/2 + \sinh v/2\cos\hat{\theta}]\cos\tilde{\theta}/2e^{-i\tilde{\varphi}/2} + \sinh v/2\sin\hat{\theta}\sin\tilde{\theta}/2e^{-i(\hat{\varphi}-\tilde{\varphi}/2)} \tag{B–8a}$$

$$\hat{\zeta}^2 = [\cosh v/2 - \sinh v/2\cos\hat{\theta}]\sin\tilde{\theta}/2e^{i\tilde{\varphi}/2} + \sinh v/2\sin\hat{\theta}\cos\tilde{\theta}/2e^{i(\hat{\varphi}-\tilde{\varphi}/2)} \tag{B–8b}$$

$$\hat{\chi}^1 = -[\cosh v/2 + \sinh v/2\cos\hat{\theta}]\sin\tilde{\theta}/2e^{-i\tilde{\varphi}/2} + \sinh v/2\sin\hat{\theta}\cos\tilde{\theta}/2e^{-i(\hat{\varphi}-\tilde{\varphi}/2)} \tag{B–8c}$$

$$\hat{\chi}^2 = [\cosh v/2 - \sinh v/2\cos\hat{\theta}]\cos\tilde{\theta}/2e^{i\tilde{\varphi}/2} - \sinh v/2\sin\hat{\theta}\sin\tilde{\theta}/2e^{i(\hat{\varphi}-\tilde{\varphi}/2)} \tag{B–8d}$$

where

$$v = \tanh^{-1}\beta \tag{B–8e}$$

and

$$\begin{aligned}\cosh v/2 &= \sqrt{\tfrac{1}{2}(u^0+1)}\\ \sinh v/2 &= \sqrt{\tfrac{1}{2}(u^0-1)}\end{aligned} \tag{B–8f}$$

Let us now find expressions for the variables ρ, ϵ, β, $\hat{\theta}$, $\hat{\varphi}$, $\tilde{\theta}$, $\tilde{\varphi}$, and $\tilde{\Psi}$ in terms of the spinors ζ^α

and χ^α. From (7–35) we have

$$\rho=[2(\zeta_\alpha\chi^\alpha)(\zeta_{\dot\beta}\chi^{\dot\beta})]^{\frac{1}{2}} \tag{B–9}$$

and

$$e^{i\epsilon}=\left[\frac{\zeta_\alpha\chi^\alpha}{\zeta_{\dot\beta}\chi^{\dot\beta}}\right]^{1/2} \tag{B–10}$$

In order to find expressions for the other variables, it is convenient to introduce the functions A and B defined as follows:

$$A=\frac{\zeta^1}{\sqrt{\zeta_\alpha\chi^\alpha}}+\frac{\chi^{\bar 2}}{\sqrt{\zeta_{\dot\beta}\chi^{\dot\beta}}}$$
$$=\sqrt{2(u^0+1)}\cos\tilde\theta/2e^{-i/2(\tilde\Psi+\tilde\varphi)} \tag{B–11a}$$

$$B=\frac{\zeta^2}{\sqrt{\zeta_\alpha\chi^\alpha}}-\frac{\chi^1}{\sqrt{\zeta_{\dot\beta}\chi^{\dot\beta}}}$$
$$=\sqrt{2(u^0+1)}\sin\tilde\theta/2e^{-i/2(\tilde\Psi-\tilde\varphi)} \tag{B–11b}$$

From (B–11) it follows that

$$\cos\tilde\theta=\frac{A\bar A-B\bar B}{A\bar A+B\bar B} \tag{B–12}$$

$$e^{i\tilde\varphi}=\left[\frac{\bar AB}{A\bar B}\right]^{1/2} \tag{B–13}$$

$$e^{i\tilde\Psi}=\left[\frac{\bar A\bar B}{AB}\right]^{1/2} \tag{B–14}$$

In order to find β, $\hat\theta$, and $\hat\varphi$, it is most convenient to use the following relations, which result from (7–26), (7–25), and (7–22):

$$\rho u^0=\frac{1}{\sqrt{2}}[\zeta^1\zeta^{\bar 1}+\zeta^2\zeta^{\bar 2}+\chi^1\chi^{\bar 1}+\chi^2\chi^{\bar 2}] \tag{B–15a}$$

$$\rho\mu^1=\rho\mu^0\beta\sin\hat\theta\cos\hat\varphi=\sqrt{2}R[\zeta^1\zeta^{\bar 2}+\chi^1\chi^{\bar 2}] \tag{B–15b}$$

$$\rho\mu^2=\rho\mu^0\beta\sin\hat\theta\sin\hat\varphi=-\sqrt{2}I[\zeta^1\zeta^{\bar 2}+\chi^1\chi^{\bar 2}] \tag{B–15c}$$

$$\rho u^3=\rho u^0\beta\cos\hat\theta=\frac{1}{\sqrt{2}}[\zeta^1\zeta^{\bar 1}-\zeta^2\zeta^{\bar 2}+\chi^1\chi^{\bar 1}-\chi^2\chi^{\bar 2}] \tag{B–15d}$$

Having found μ^0 from (B–15a) and (B–9), we use the following relation to find:

$$\beta=\frac{1}{u^0}[(u^0)^2-1]^{1/2} \tag{B–17}$$

From (B–15b) – (B–15d) we have

$$\tan\hat\theta=\frac{1}{u^3}[(u^1)^2+(u^2)^2]^{1/2} \tag{B–18}$$

$$\tan\hat\varphi=u^2/u^1 \tag{B–19}$$

In the solution of a magnetogas-dynamical problem of the self-excited dynamo type, we would solve the spinor equations (8–1) or (8–31) and arrive at expressions for ζ^α and χ^α as functions of time t and the spherical coordinates r, θ, and φ. We would then use the above expressions to find ρ, β, $\hat\theta$, $\hat\varphi$, $\tilde\theta$, $\tilde\varphi$, and $\tilde\Psi$ as functions of the coordinates (t, r, θ, φ).

APPENDIX C: DIFFERENTIAL SPINOR RELATIONS

Observable Quantities

Using the spinors ζ^α and x^α, and the gradient operator $\partial_{\alpha\dot\beta}$ (or ∂_j) as the basic ingredients, it is possible to construct many differential expressions. Before any physical meaning can be given to these expressions, however, it is necessary to translate them into expressions involving the observable quantities ρ, ϵ, u^j, μ^j, ω^j, and s^{jk} which are defined as follows:

$$\rho e^{i\epsilon}\equiv\sqrt{2}(\zeta_\alpha\chi^\alpha) \tag{C–1}$$

$$\rho\mu^j\equiv\sigma^j_{\alpha\dot\beta}(\zeta^\beta\zeta^{\bar\alpha}+\chi^\beta\chi^{\bar\alpha} \tag{C–2}$$

$$\rho\mu^j\equiv\sigma^j_{\alpha\dot\beta}(\zeta^\beta\zeta^{\bar\alpha}-\chi^\beta\chi^{\bar\alpha}) \tag{C–3}$$

$$\omega^j\equiv\tfrac{1}{2}\lambda_l\partial^j k^l \tag{C–4a}$$

where

$$\rho k^j\equiv\sigma^j_{\alpha\dot\beta}(\zeta^\beta\chi^{\bar\alpha}+\chi^\beta\zeta^{\bar\alpha}) \tag{C–4b}$$

$$\rho\lambda^j\equiv-i\sigma^j_{\alpha\dot\beta}(\zeta^\beta\chi^{\bar\alpha}-\chi^\beta\zeta^{\bar\alpha}) \tag{C–4c}$$

Using (C–4b) and (C–4c) in (C–4a), ω^j can also be written as follows:

$$\omega^j=I\left[\left(\frac{\chi^\beta}{\sqrt{\zeta_\alpha\chi^\alpha}}\right)\partial^j\left(\frac{\zeta_\beta}{\sqrt{\zeta_\alpha\chi^\alpha}}\right)\right] \tag{C–5a}$$

or

$$\omega^j=\tfrac{1}{2}I[(\zeta_\alpha\chi^\alpha)^{-1}(\chi^\beta\partial^j\zeta_\beta-\zeta_\beta\partial^j\chi^\beta)] \tag{C–5b}$$

The dual tensors s^{jk} and $\hat s^{jk}$ are defined as follows:

$$s^{jk}\equiv-\epsilon^{jkln}\mu_l\mu_n=-\tfrac{1}{2}\epsilon^{jkln}s_{ln} \tag{C–6}$$

$$\hat s^{jk}\equiv\tfrac{1}{2}\epsilon^{jkln}s_{ln}=u^j\mu^k-u^k\mu^j \tag{C–7}$$

The following orthogonality properties result from (C–6):

$$s^{jk}u_k=u_j s^{jk}=0 \tag{C–8a}$$

$$s^{jk}\mu_k=\mu_j s^{jk}=0 \tag{C–8b}$$

It can be shown that s^{jk} and $\hat{s}^{jk}$ are related to the spinors ζ^α and χ^α as follows:

$$\hat{s}^{jk}=R[(\zeta_\nu\chi^\nu)^{-1}g^{kl}\sigma_l^{\beta\dot{\lambda}}\sigma^j_{\dot{\lambda}\alpha}(\zeta^\alpha\chi_\beta+\chi^\alpha\zeta_\beta)] \quad \text{(C-9)}$$

$$s^{jk}=-I[(\zeta_\nu\chi^\nu)^{-1}g^{kl}\sigma_l^{\beta\dot{\lambda}}\sigma^j_{\dot{\lambda}\alpha}(\zeta^\alpha\chi_\beta+\chi^\alpha\zeta_\beta)] \quad \text{(C-10)}$$

Thus we see that the symmetric spinor expression $(\zeta^\alpha\chi_\beta+\chi^\alpha\zeta_\beta)$ is to be associated with the dual tensors $\hat{s}^{jk}$ and s^{jk}. The antisymmetric spinor expression $(\zeta^\alpha\chi_\beta-\chi^\alpha\zeta_\beta)$ corresponds to the Kronecker delta function $\zeta_j{}^k$ as shown by the following identity:

$$\zeta_j{}^k=-(\zeta_\nu\chi^\nu)^{-1}\sigma_j^{\beta\dot{\lambda}}\sigma^k_{\dot{\lambda}\alpha}(\zeta^\alpha\chi_\beta-\chi^\alpha\zeta_\beta) \quad \text{(C-11)}$$

In this appendix we shall have no need to label quantities with the (+) and (−) subscripts. When the sign of the particle charge enters an equation it will be indicated by σ, where $\sigma=\pm1$.

Gradients of Spinors

Using the definitions (C-1) to (C-5) it can be verified that the following expressions for the gradients of ζ^β and χ^β are valid identities:

$$\partial_j\zeta^\beta=\zeta^\beta[\tfrac{1}{2}(\partial_j\ln\rho+\mu_i\partial_ju^i)+i(\omega_j+\tfrac{1}{2}\partial_j\epsilon)]$$
$$-\frac{e^{i\epsilon}}{\sqrt{2}}\chi_{\bar{\alpha}}\partial_j(u^{\beta\bar{\alpha}}+\mu^{\beta\bar{\alpha}}) \quad \text{(C-12a)}$$

$$\partial_j\chi^\beta=\chi^\beta[\tfrac{1}{2}(\partial_j\ln\rho-\mu_i\partial_ju^i)-i(\omega_j-\tfrac{1}{2}\partial_j\epsilon)]$$
$$+\frac{e^{i\epsilon}}{\sqrt{2}}\zeta_{\bar{\alpha}}\partial_j(u^{\beta\bar{\alpha}}-\mu^{\beta\bar{\alpha}}) \quad \text{(C-12b)}$$

These identities can be written in a more symmetric form, as follows:

$$\partial_k\zeta^\beta=-\sqrt{2}e^{i\epsilon}\sigma_j^{\beta\dot{\lambda}}\chi_{\dot{\lambda}}[(2\rho)^{-1}\partial_k(\rho\mu^j)+i(u^j\omega_k$$
$$+\tfrac{1}{2}\mu^j\partial_k\epsilon-\tfrac{1}{2}s^{jl}\partial_ku_l)] \quad \text{(C-13a)}$$

$$\partial_k\chi^\beta=-\sqrt{2}e^{i\epsilon}\sigma_j^{\beta\dot{\lambda}}\zeta_{\dot{\lambda}}\Big[(2\rho)^{-1}\partial_k(\rho\mu^j)$$
$$+i(u^j\omega_k+\frac{1}{2}\mu^j\partial_k\epsilon-\frac{1}{2}s^{jl}\partial_ku_l)\Big] \quad \text{(C-13b)}$$

If we multiply these two identities by $\sigma_{\bar{\alpha}\beta}{}^k$, contracting the k and the β, we arrive at the following identities:

$$\partial_{\bar{\alpha}\beta}\zeta^\beta=\zeta^\beta\sigma^k_{\bar{\alpha}\beta}[\tfrac{1}{2}(\partial_k\ln\rho-\mu_j\partial_ku^j-\partial^j\hat{s}_{jk}+s_{kj}\partial^j\epsilon)$$
$$+i(\omega_k+\tfrac{1}{2}\partial^js_{jk}+\tfrac{1}{2}\hat{s}_{kj}\partial^j\epsilon)] \quad \text{(C-14a)}$$

$$\partial_{\bar{\alpha}\beta}\chi^\beta=\chi^\beta\sigma^k_{\bar{\alpha}\beta}[\tfrac{1}{2}(\partial_k\ln\rho+\mu_j\partial_ku^j+\partial^j\hat{s}_{jk}-s_{kj}\partial^j\epsilon)$$
$$-i(\omega_k+\tfrac{1}{2}\partial^js_{jk}+\tfrac{1}{2}\hat{s}_{kj}\partial^j\epsilon)] \quad \text{(C-14b)}$$

Alternative expressions for $\partial_{\bar{\alpha}\beta}\zeta^\beta$ and $\partial_{\bar{\alpha}\beta}\chi^\beta$ could be found by contracting (C-12) instead of (C-13).

Using (C-12) – (C-14), it is possible to derive the following identities:

Scalar Identities

$$\zeta^{\bar{\alpha}}\partial_{\bar{\alpha}\beta}\zeta^\beta-\chi^{\bar{\alpha}}\partial_{\bar{\alpha}\beta}\chi^\beta=\tfrac{1}{2}\partial_j(\rho\mu^j)$$
$$+iu_j[\rho\omega^j+\tfrac{1}{2}\partial_k(\rho s^{kj})+\tfrac{1}{2}\rho\hat{s}^{jk}\partial_k\epsilon] \quad \text{(C-15)}$$

$$\zeta^{\bar{\alpha}}\partial_{\bar{\alpha}\beta}\zeta^\beta+\chi^{\bar{\alpha}}\partial_{\bar{\alpha}\beta}\chi^\beta=\tfrac{1}{2}\partial_j(\rho u^j)$$
$$+i\mu_j[\rho\omega^j-\tfrac{1}{2}\partial_k(\rho s^{kj})+\tfrac{1}{2}\rho\hat{s}^{jk}\partial_k\epsilon] \quad \text{(C-16)}$$

Vector Identities

$$g^{jl}\sigma_l^{\nu\bar{\alpha}}(\chi_\nu\partial_{\bar{\alpha}\beta}\zeta^\beta-\zeta_\nu\partial_{\bar{\alpha}\beta}\chi^\beta)=$$
$$=-\frac{e^{i\epsilon}}{\sqrt{2}}\Big\{s^{jk}\Big[\rho\omega_k+\frac{1}{2}\partial^l(\rho s_{lk})\Big]$$
$$+\frac{1}{2}u^j\partial_k(\rho u^k)-\frac{1}{2}\mu^j\partial_k(\rho\mu^k)\Big\} \quad \text{(C-17)}$$
$$-\frac{ie^{i\epsilon}}{2\sqrt{2}}\Big\{\rho\partial^j\epsilon+2\hat{s}^{jk}\Big[\rho\omega_k+\frac{1}{2}\partial^l(\rho s_{lk})\Big]$$
$$+\rho s^{jk}[u^l\partial_l\mu_k-\mu^l(\partial_lu_k-\partial_ku_l)]\Big\}$$

$$g^{jl}\sigma_l^{\nu\bar{\alpha}}(\chi_\nu\partial_{\bar{\alpha}\beta}\zeta^\beta+\zeta_\nu\partial_{\bar{\alpha}\beta}\chi^\beta)=$$
$$=\frac{e^{i\epsilon}}{2\sqrt{2}}\{\rho u^k\partial_k\mu^j+\rho\omega_k[(\partial^ju^k-\partial^ku^j)$$
$$-\epsilon^{jkln}u_l\partial_n\epsilon]+\mu^j\partial_k(\rho u^k)-u^j\partial_k(\rho\mu^k)\} \quad \text{(C-18)}$$
$$-\frac{ie^{i\epsilon}}{\sqrt{2}}\Big\{\rho\omega^j+\frac{1}{2}\partial_k(\rho s^{kj})+\frac{1}{2}\rho\hat{s}^{jk}\partial_k\epsilon\Big\}$$

Tensor Identities

$$g^{jl}\sigma_l^{\beta\nu}\sigma_{\nu\alpha}{}^k(\zeta^\alpha\partial_{\lambda\beta}\zeta^\lambda-\chi^\alpha\partial_{\bar{\lambda}\beta}\chi^{\bar{\lambda}})=$$
$$=\tfrac{1}{2}g^{jk}\{\tfrac{1}{2}(\partial_i(\rho\mu^l)-iu_n$$
$$[\rho\omega^n+\tfrac{1}{2}\partial_l(\rho s^{ln})+\tfrac{1}{2}\rho\hat{s}^{nl}\partial_l\epsilon]\}$$
$$+\tfrac{1}{2}\{\tfrac{1}{2}[\partial^j(\rho\mu^k)-\partial^k(\rho\omega^j)]-\epsilon^{jkln}\mathfrak{F}_{ln}\}$$

$$-\frac{i}{4}\left\{\begin{array}{l} s^{jk}\partial_l(\rho u^l)+2\left[\rho\omega^j+\frac{1}{2}\partial_l(\rho s^{lj})+\frac{1}{2}\rho\hat{s}^{jl}\partial_l\epsilon\right]u^k \\ \qquad -2\left[\rho\omega^k+\frac{1}{2}\partial_l(\rho s^{lk})+\frac{1}{2}\rho\hat{s}^{kl}\partial_l\epsilon\right]u^j \\ +\rho u^l\partial_l s^{jk}-\rho s^j{}_l[(\partial^l u^k-\partial^k u^l)-\epsilon^{lkpq}u_p\partial_q\epsilon] \\ \qquad +\rho s^k{}_l[(\partial^l u^j-\partial^j u^l)-\epsilon^{ljpq}u_p\partial_q\epsilon] \end{array}\right\} \tag{C-19}$$

$$\begin{aligned} &g^{jl}\sigma_l^\beta{}^r\sigma^k_{r\alpha}(\zeta^\alpha\partial_{\lambda\beta}\zeta^\lambda+\chi^\alpha\partial_{\lambda\beta}\chi^\lambda)= \\ &\quad=\tfrac{1}{2}g^{jk}\{\tfrac{1}{2}\partial_l(\rho u^l)-i\mu_n, \\ &\qquad\qquad [\rho\omega^n+\tfrac{1}{2}\partial_l(\rho s^{ln})+\tfrac{1}{2}\rho\hat{s}^{nl}\partial_l\epsilon]\} \\ &\quad+\tfrac{1}{2}\{\tfrac{1}{2}[\partial^j(\rho u^k)-\partial^k(\rho u^j)]-\epsilon^{jkln}Q_{ln}\} \end{aligned}$$

$$-\frac{i}{4}\left\{\begin{array}{l} s^{jk}\partial_l(\rho\mu^l)-2\left[\rho\omega^j-\frac{1}{2}\partial_l(\rho s^{lj})+\frac{1}{2}\rho\hat{s}^{jl}\partial_l\epsilon\right]\mu^k \\ \qquad -2\left[\rho\omega^k+\frac{1}{2}\partial_l(\rho s^{lk})+\frac{1}{2}\rho\hat{s}^{kl}\partial_l\epsilon\right]\mu^j \\ +\rho\mu^l\partial_l s^{jk}-\rho s^j{}_l[(\partial^l u^k-\partial^k\mu^l)-\epsilon^{lkpq}\mu_p\partial_q\epsilon] \\ \qquad +\rho s^k{}_l[(\partial^l\mu^j-\partial^j\mu^l)-\epsilon^{ljpq}\mu_p\partial_q\epsilon] \end{array}\right\} \tag{C-20}$$

where

$$g_{jl}\sigma_{\bar{\alpha}\beta}{}^l(\zeta^{\bar{\alpha}}\partial_k\zeta^\beta-\chi^{\bar{\alpha}}\partial_k\chi^\beta)=\tfrac{1}{2}\partial_k(\rho\mu_j)+iS_{jk} \tag{C-21}$$

$$g_{jl}\sigma_{\bar{\alpha}\beta}{}^l(\zeta^{\bar{\alpha}}\partial_k\zeta^\beta+\chi^{\bar{\alpha}}\partial_k\chi^\beta)=\tfrac{1}{2}\partial_k(\rho u_j)+i\mathfrak{F}_{jk} \tag{C-22}$$

where

$$S_{jk}=\rho[u_j\omega_k+\tfrac{1}{2}\mu_j\partial_k\epsilon-\tfrac{1}{2}s_{jl}\partial_k u^l] \tag{C-23}$$

$$\mathfrak{F}_{jk}=\rho[\mu_j\omega_k+\tfrac{1}{2}u_j\partial_k\epsilon-\tfrac{1}{2}s_{jl}\partial_k\mu^l] \tag{C-24}$$

Spinor Equations

The relations (C-15) – (C-24) are just identities based on the definitions (C-1) – (C-10). By introducing spinor equations of motion, and substituting these into the above identities, we can find the fluid equations of motion in terms of the observable quantities defined in (C-1) – (C-10). The spinor equations of motion are postulated to have the following form:

$$i\hbar\partial_{\bar{\alpha}\beta}\zeta^\beta=\frac{\tilde{m}c}{\sqrt{2}}\chi_{\bar{\alpha}}-\frac{q}{c}A_{\bar{\alpha}\beta}\zeta^\beta \tag{C-25a}$$

$$i\hbar\partial_{\bar{\alpha}\beta}\chi^\beta=\frac{\tilde{m}c}{\sqrt{2}}\zeta_{\bar{\alpha}}+\frac{q}{c}A_{\bar{\alpha}\beta}\chi^\beta \tag{C-25b}$$

where

$$\tilde{m}=m(1+g/c^2+h/c^2) \tag{C-25c}$$

We shall at times impose the following constraint on the solutions of (C-25):

$$I(\zeta_\beta\chi^\beta)=0 \tag{C-26a}$$

which is equivalent to the constraint

$$\cos\epsilon=\sigma=\pm 1 \tag{C-26b}$$

or

$$\begin{aligned} &\epsilon=0 \text{ for } \sigma=+1 \\ &\epsilon=\pi \text{ for } \sigma=-1 \end{aligned} \tag{C-26c}$$

where ϵ is the phase angle common to ζ^β and χ^β and σ is the sign of the particle charge. The symbol σ will always represent either $+1$ or -1, but in much of what follows we shall not impose conditions (C-26), with the result that in such cases $\cos\epsilon\neq\sigma$. When condition (C-26) is imposed, however, we note that the product $\sigma\cos\epsilon$, which occurs in some of the following expressions, equals $+1$.

Fluid Equations of Motion

We shall first derive the fluid equations of motion without imposing the constraint (C-26). In this way it will be possible to estimate the error introduced by neglecting this constraint.

Substituting (C-25) into (C-15) and (C-16), we have from the real parts of these identities the

following relations:

$$\partial_j(\rho\mu^j) = \frac{2\rho}{\bar{\lambda}} \sin \epsilon \qquad \text{(C-27a)}$$

where

$$\bar{\lambda} = \frac{\hbar}{\tilde{m}c} \qquad \text{(C-27b)}$$

and

$$\partial_j(\rho u^j) = 0 \qquad \text{(C-28)}$$

These equations hold for each of the two charged fluids individually. (C-28) is just the statement of charge conservation, or conservation of the number of particles, for each of the two fluids taken individually.

Equation (C-27) allows us to estimate the magnitude of ϵ for the case that we solve (C-25) without imposing the constraint (C-26). We note that, if we limit ourselves to solutions in which ρ and μ^j are smoothly varying functions whose values change significantly only over distances comparable with L and in time intervals greater than L/c, where L is a length that specifies the dimensions of the system, then from (C-27) we have

$$|\sin \epsilon| = \left|\frac{\bar{\lambda}}{2\rho}\partial_j(\rho\mu^j)\right| \sim \frac{\bar{\lambda}}{L} \approx \frac{\lambda}{L} \qquad \text{(C-29)}$$

where in the last step we have replaced $\bar{\lambda}$ by λ (defined in (7–47) in terms of the free, rather than the bound, particle mass) because the gravitational and thermal energy of a particle is small compared with its rest energy. From (C-29) we see that $|\sin \epsilon|$ is very small for macroscopic systems. Thus, for such systems, we have

$$\begin{aligned} |\epsilon| &\sim \frac{\lambda}{L} \text{ for } \sigma = +1 \\ |\pi - \epsilon| &\sim \frac{\lambda}{L} \text{ for } \sigma = -1 \end{aligned} \qquad \text{(C-30)}$$

where the first case applies to positively charged particles, and the second case to negatively charged particles.

If we impose the condition (C-26), we have, instead of (C-27)

$$\partial_j(\rho\mu^j) = 0 \quad \text{for } \cos\epsilon = \sigma \qquad \text{(C-31)}$$

Substituting (C-25) into (C-18), we arrive at two real vector equations, one of which is the following:

$$\sigma\hbar\rho\omega^j = \rho(\sigma \cos \epsilon)\tilde{m}cu^j - \frac{\sigma\hbar}{2}\partial_k(\rho s^{kj}) - \\ - \frac{\sigma\hbar}{2}\rho s^{jk}\partial_k\epsilon + \frac{\rho(\sigma q)}{c}A^j \qquad \text{(C-32)}$$

In order to gain an intuitive feeling for this equation, let us examine first the second term on the right side. Using (C-6), we can derive the following identity:

$$-\frac{\sigma\hbar}{2}\partial_k(\rho s^{kj}) = \rho\tilde{m}cu^j + \sigma\hbar\rho\pi^j \qquad \text{(C-33a)}$$

where

$$\tilde{m} = -\frac{\sigma\hbar}{2c\rho}u_j\,\partial_k(\rho s^{kj}) \qquad \text{(C-33b)}$$

and

$$\pi^j = (2\rho)^{-1}\epsilon^{jklm}u_k\partial^l(\rho\mu_m) - \tfrac{1}{2}s^{jk}u^l\partial_i u_k \qquad \text{(C-33c)}$$

We note that, because of (C-8a) and the antisymmetry of ϵ^{jklm}

$$\pi^j u^j = 0 \qquad \text{(C-34)}$$

Let us evaluate $\sigma\hbar\rho\pi^j$ in the particle rest-frame:

$$\begin{aligned} \sigma\hbar\rho\vec{\pi} &= [\sigma\hbar\rho(\pi^1,\pi^2,\pi^3)]v_{-0} \\ &= \left[\vec{\nabla}\times\left(\frac{\sigma\hbar}{2}\rho\vec{\mu}\right)\right]_{v_{-0}} - \frac{\sigma\hbar}{2c^2}\dot{\rho}\vec{\mu}\times\left[\frac{d\vec{v}}{d\tau}\right]_{v_{-0}} \qquad \text{(C-35)} \\ &= \left[\vec{\nabla}\times\left(\frac{\sigma\hbar}{2}\rho\vec{\mu}\right)\right]_{v_{-0}} + \left[\rho\frac{d}{d\tau}\left(\frac{\vec{v}}{c^2}\times\frac{\sigma\hbar\vec{\mu}}{2}\right)\right]_{v_{-0}} \end{aligned}$$

If we interpret $\frac{\sigma\hbar}{2}\rho\vec{\mu}$ as the particle spin angular momentum density, then the first term on the right side of (C-35) is just the contribution to the linear momentum density arising from the spin angular momentum density. This linear momentum density is completely analogous to the equivalent current density in a magnetized body that is given by the curl of the magnetization.

In order to understand the second term on the right side of (C-35), we first note that a body whose 3-velocity is $\vec{v}$ and whose mass and angular

momentum in its rest-frame are respectively m and $\frac{\sigma\hbar}{2}\vec{\mu}$ experiences a displacement $\vec{d}$ of its center-of-mass given by the well-known relation(*)

$$\vec{d}=\frac{\vec{v}\times\left(\frac{\sigma\hbar}{2}\vec{\mu}\right)}{\breve{m}c^2} \qquad \text{(C-36)}$$

Thus (C-35) may be written

$$\sigma\hbar\rho\overset{0}{\vec{\pi}}=\left[\vec{\nabla}\times\left(\frac{\sigma\hbar}{2}\rho\vec{\mu}\right)\right]_{v=0}+\left[\rho\frac{d(\breve{m}\vec{d})}{d\tau}\right]_{v=0} \qquad \text{(C-37)}$$

The second term on the right side of (C-37) is the momentum density that results from the change with time of the mass moment of a particle.

Since $\tilde{m}$, which is defined in (C-33b), is an invariant, its value in the particle rest-frame equals its value in all frames.

Thus

$$\rho\tilde{m}=-\frac{\sigma\hbar}{2c}[\partial_k s^{k0}]_{v=0}$$
$$=-\left[\vec{\nabla}\cdot\left\{(\rho u^0)\frac{\vec{v}\times\left(\frac{\sigma\hbar}{2}\vec{\mu}\right)}{c^2}\right\}\right]_{v=0} \qquad \text{(C-38)}$$

where we have made use of (5-2b). Using (C-36) in (C-38), we have

$$\rho\tilde{m}=-[\vec{\nabla}\cdot(\rho u^0\breve{m}\vec{d})]_{v=0} \qquad \text{(C-39)}$$

Thus ρm is just the contribution to the mass density that results from the center-of-mass displacement given by (C-36).

Thus, when we impose the constraint (C-26), the vector equation (C-32) can be written in the particle rest-frame as follows:

$$\sigma\hbar\rho\dot{\omega}^0=\rho\breve{m}c-c[\vec{\nabla}\cdot(\rho u^0\breve{m}\vec{d})]_{v=0} \text{ for } \cos\epsilon=\sigma \qquad \text{(C-40a)}$$

$$\sigma\hbar\rho\dot{\vec{\omega}}=[\sigma\hbar\rho(\omega^1,\omega^2,\omega^3)]_{v=0}$$
$$=-\left[\vec{\nabla}\times\left(\frac{\sigma\hbar}{2}\rho\vec{\mu}\right)\right]_{v=0}+\left[\rho\frac{d(\breve{m}\vec{d})}{d\tau}\right]_{v=0}$$
$$+\frac{\rho(\sigma q)}{c}\vec{A} \qquad \text{(C-40b)}$$
$$\text{for } \cos\epsilon=\sigma$$

This shows us that $\sigma\hbar\rho\omega^j$ is just the rigorously correct expression for the 4-vector canonical momentum density including all the spin effects. Since the spin-dependent contribution to the right side of (C-40a) is very small compared with $\rho\breve{m}c$, the right side of (C-40a) is always positive. Because of the presence of the factor σ on the left, we have the result that the sign of $\dot{\omega}^0$ depends on the sign of the particle charge. This is exactly what we concluded from (7-14), which resulted from identifying the direction of rotation of $\vec{k}$ and $\vec{\lambda}$ about $\vec{\mu}$ with the sign of the particle charge. Thus we see that identifying $\cos\epsilon$ with σ, the sign of the particle charge, as we did in (C-26b), has as a consequence the fact that, for solutions of the spinor equations (C-25), the direction of rotation of $\vec{k}$ and $\vec{\lambda}$ about $\vec{\mu}$ (all derived from the spinors by means of (7-27)) is dependent on the sign of the particle charge. This proves the consistency of the two ways of identifying the sign of particle charge—either as $\cos\epsilon$, or as the direction of rotation of $\vec{k}$ about $\vec{\mu}$.

We can derive the equation of motion (6-6a) by substituting (C-32) into (7-44). Doing this, we have

$$-\partial^j\Phi=Mcu^j+\frac{\sigma q}{c}(A^j+a^j) \text{ for } \cos\epsilon=\sigma \qquad \text{(C-41a)}$$

where

$$M=\breve{m}+\tilde{m} \qquad \text{(C-41b)}$$

is to be regarded as the total mass per particle and

$$a^j=\frac{\hbar c}{q}(\pi^j-r^j)=\frac{q}{\alpha}(\pi^j-r^j) \qquad \text{(C-41c)}$$

where

$$\alpha=\frac{q^2}{\hbar c}=\frac{1}{137.04} \qquad \text{(C-41d)}$$

is the fine-structure constant and π^j and r^j are defined in (C-33c) and (7-42) respectively. a^j is to be regarded as an effective electromagnetic potential to be associated with the magnetic moment of the particle. It is to be noted that, whereas A^j is the same in the dynamical equations for each of the two charged fluids, a^j in (C-41a) represents the two different potentials $a^j_{(+)}$ and $a^j_{(-)}$.

We have already seen in Section III that (C-41) suffices to yield a complete dynamical description of each charged fluid for the case of given thermal, gravitational, and electromagnetic fields.

*See, for example, ref. 10, p. 172, eq. 48. (The difference in sign between this equation and (C-36) results merely from a difference in sign in the definition of velocity.)

Equation (C-39) gives what might be called a kinetic explanation of the mass increment $\tilde{m}$, based as it is on the center-of-mass displacement produced by the motion of the particle. We shall now derive an electromagnetic explanation. First, we arrive at an expression for $(\partial^j u^k - \partial^k u^j)$ by taking the curl of (C-41a):

$$(\partial^j u^k - \partial^k u^j) = u^j \partial^k \ln M - u^k \partial^j \ln M - \frac{\sigma q}{Mc^2}(F^{jk} + f^{jk}) \quad \text{(C-42a)}$$

where

$$f^{jk} = \partial^j a^k - \partial^k a^j \quad \text{(C-42b)}$$

Now we note that (C-33b) can be rewritten as follows:

$$\tilde{m} = -\frac{\sigma\hbar}{2c\rho} u_k \partial_j (\rho s^{jk}) = \frac{\sigma\hbar}{2c} s^{jk} \partial_j u_k = \frac{\sigma\hbar}{4c} s^{jk} (\partial_j u_k - \partial_k u_j) \quad \text{(C-43)}$$

where we have made use of (C-8a) and the antisymmetry of s^{jk}. Substituting (C-42) into (C-43), we have

$$\tilde{m} = \frac{q\hbar}{4Mc^3} s^{jk}(F_{jk} + f_{jk}) = \frac{q\hbar}{4Mc^3} u_j [\epsilon^{jkln} u_k (F_{ln} + f_{ln})] \quad \text{(C-44)}$$

where, in the second step, we have used (C-6). Evaluating the invariant $\tilde{m}c^2$ in the particle rest-frame, we have

$$\tilde{m}c^2 = -\left[\frac{q}{Mc}\left(\frac{\hbar}{2}\vec{\mu}\right)\cdot(\vec{B} + \vec{b})\right] \quad \text{(C-45)}$$

where we have used (2-20) and (5-6). Thus $\tilde{m}$ is the mass corresponding to the interaction energy of a magnetic dipole $\frac{q\hbar}{2Mc}\vec{\mu}$ in a magnetic field $(\vec{B} + \vec{b})$. We note that the magnitude of the gyromagnetic ratio is $\frac{q}{Mc}$, which is the same value we obtain from the Dirac Theory, except that in our case the particle mass M includes the thermal, gravitational, and dipole-interaction energy contributions to the mass, as well as the rest mass.

Magnitude of Spin-Dependent Electromagnetic Field

Now let us estimate the magnitudes of a^j and f^{jk}. If we regard the magnitudes of u^j, μ^j, and s^{jk} as being of order unity, and consider only solutions for which these quantities, as well as $\ln\rho$, change significantly only over a distance of order L and in times intervals greater than L/c, then from (C-33c) we have

$$|\pi^j| \sim \frac{1}{L} \quad \text{(C-46)}$$

Thus from (7-43), (C-46), and (C-41c), we have

$$|a^j| \sim \frac{q}{\alpha L} \quad \text{(C-47)}$$

and from (C-42b), it follows that

$$|f^{jk}| \sim \frac{q}{\alpha L^2} \quad \text{(C-48)}$$

Gauge Transformations

Now let us study the effect of a gauge transformation on equation (C-41a). Let $\tilde{A}_j$ and A^j differ only by the gauge transformation generated by the scalar function η, i.e.

$$A^j = \tilde{A}^j + \partial^j \eta \quad \text{(C-49)}$$

Substituting this into (C-41a), we have

$$-\partial^j\left(\Phi + \frac{\sigma q}{c}\eta\right) = Mcu^j + \frac{\sigma q}{c}(\tilde{A}^j + \partial^j \eta) \quad \text{(C-50)}$$

Thus we see that a gauge transformation is equivalent to adding a scalar function to the de Broglie phase function Φ (or Hamilton's Characteristic Function).

If $\tilde{\zeta}^\alpha$ and $\tilde{\chi}^\alpha$ are the solutions to the spinor equations (C-25) in which A^j is used to describe the electromagnetic field, whereas ζ^α and χ^α are the solutions resulting when the field is described by A^j, then we find from (C-50), (7-8), and (7-33), or by direct substitution of (C-49) into (C-25), that

$$\tilde{\zeta}^\beta = e^{-i\left(\frac{e}{\hbar c}\eta\right)}\zeta^\beta = e^{-i\alpha\eta'}\zeta^\beta \quad \text{(C-51a)}$$

$$\tilde{\chi}^\beta = e^{+i\left(\frac{e}{\hbar c}\eta\right)}\chi^\beta = e^{+i\alpha\eta'}\chi^\beta \quad \text{(C-51b)}$$

where

$$\eta' = \frac{\eta}{\chi} \quad \text{(C-51c)}$$

is the dimensionless form of η, and α is the fine-structure constant introduced in (C-41d). Thus

we see that a gauge transformation is equivalent to a phase change of the spinors.

From (C–50) it is evident that it is always possible to choose η so that Φ has the form

$$\Phi = -mc^2 t \tag{C–52}$$

Using (7–8), (7–33), and (C–52), we see that, for such a choice of gauge, ζ^β and χ^β may be written as follows:

$$\zeta^\beta = e^{i\left(\frac{\sigma mc^2}{\hbar}t\right)}\left[\left(\frac{\rho e^{i\epsilon}}{\sqrt{2}}\right)^{1/2}\hat{\zeta}^\beta\right] \tag{C–53a}$$

$$\chi^\beta = e^{-i\left(\frac{\sigma mc^2}{\hbar}t\right)}\left[\left(\frac{\rho e^{i\epsilon}}{\sqrt{2}}\right)^{1/2}\hat{\chi}^\beta\right] \tag{C–53b}$$

The significance of this way of writing the spinors is that it shows that they can always be written in a form such that the fast time dependence is completely contained in a phase factor whose time dependence is given by the de Broglie frequency but which has no spatial dependence. For a magnetogas-dynamical system in which all the observable quantities are time-independent, the square brackets in (C–53) are completely time independent. In such a case ζ^β and χ^β are irreducible representations of the symmetry operation consisting of a displacement of the origin of the observer's time axis. Thus it would seem that the choice of gauge that corresponds to (C–52), and hence to (C–53), has a fundamental significance that other choices, such as the Lorentz Gauge, do not enjoy.

Error Caused by Variable ϵ

The relations (C–41) assumes that the constraint (C–26) is satisfied. In order to estimate the effect of neglecting this constraint, we substitute (C–32) for arbitrary ϵ into (7–44) and obtain

$$-\tilde{v}^j\Phi = [(\sigma\cos\epsilon)\check{m} - \frac{\sigma\hbar}{2c}\mu^k\partial_k\epsilon + \tilde{m}]cu^j + \frac{\sigma q}{c}\left(A^j + a^j + \frac{\hbar}{2q}\mu^j\frac{d\epsilon}{d\tau}\right) \tag{C–54}$$

Thus the effect of not holding ϵ constant is to introduce an error $\Delta_\epsilon m$ into the mass and an error $\Delta_\epsilon A^j$ into the 4-potential, where

$$\Delta_\epsilon m = -\check{m}(1-\sigma\cos\epsilon) \sim \frac{\sigma\hbar}{2c}\mu^k\partial_k\epsilon \tag{C–55}$$

$$\Delta_\epsilon A^j = \frac{\hbar}{2q}\mu^j\frac{d\epsilon}{d\tau} = \frac{\hbar c}{2q}\mu^j u^k\partial_k\epsilon \tag{C–56}$$

Using (C–30), we have

$$|1-\sigma\cos\epsilon| \sim (\lambda\!\!\!^{-}/L)^2 \tag{C–57}$$

$$\left|\frac{\sigma\hbar}{2c}\mu^k\partial_k\epsilon\right| \sim \frac{\hbar}{2c}\left|\frac{\epsilon}{L}\right| \sim \frac{\hbar}{2c}\frac{\lambda\!\!\!^{-}}{L^2} = \frac{m}{2}\left(\frac{\lambda\!\!\!^{-}}{L}\right)^2 \tag{C–58}$$

$$|\Delta_\epsilon 1^j| \sim \frac{\hbar c}{2q}\left|\frac{\epsilon}{L}\right| \sim \frac{\hbar c}{2q}\frac{\lambda\!\!\!^{-}}{L^2} = \left(\frac{q}{2\alpha L}\right)\left(\frac{\lambda\!\!\!^{-}}{L}\right) \tag{C–59}$$

For the purposes of an estimate, we may replace $\check{m}$ in (C–55) with m. Thus from (C–55), (C–57), and (C–58), we have

$$|\Delta_\epsilon m| \sim m(\lambda\!\!\!^{-}/L)^2 \tag{C–60}$$

Using (C–47), we can write (C–59) as follows:

$$|\Delta_\epsilon A^j| \sim |a^j|(\lambda\!\!\!^{-}/L) \tag{C–61}$$

which shows immediately that, for all macroscopic dynamical systems, $\Delta_\epsilon A^j$ is *a fortiori* negligible if a^j is negligible.

In order to gain an intuitive feeling for the significance of (C–60), let us regard $\Delta_\epsilon m$ as being caused by an error $\Delta_\epsilon h$ in the specific enthalpy. Since the effective mass associated with h is $m(h/c^2)$, we have

$$|\Delta_\epsilon h| \sim c^2(\lambda\!\!\!^{-}/L)^2 \tag{C–62}$$

Using (2–2b) and (2–9), $\Delta_\epsilon h$ can in turn be expressed as an equivalent error $\Delta_\epsilon T$ in the temperature:

$$|\Delta_\epsilon T| \sim \frac{m}{k}|\Delta\epsilon h| \sim \frac{mc^2}{k}(\lambda\!\!\!^{-}/L)^2 \sim \frac{10^{-14}\ \text{cm}^2\ ^\circ\text{K}}{L^2}\ \text{for protons} \sim \frac{10^{-11}\ \text{cm}^2\ ^\circ\text{K}}{L^2}\ \text{for electrons} \tag{C–63}$$

which shows immediately that, for systems on a macroscopic scale, $\Delta_\epsilon T$ (and hence $\Delta_\epsilon m$) is completely negligible.

Precession of Particle Spin

Substituting the spinor equations of motion (C–25) into the complex identity (C–18) yields

two real vector equations. One of these is (C–32). The other is

$$u^j\partial_j\mu^k=\mu_j[(\partial^j u^k-\partial^k u^j)-\epsilon^{jkln}u_l\partial_n\epsilon] \quad \text{(C–64)}$$

where we have made use of (C–27) and (C–28).

As a preliminary to explaining the physical significance of (C–64), we first introduce the 4-vector Ω^j defined as follows:

$$\Omega^j\equiv\frac{c}{2}\epsilon^{jkln}u_k\partial_l u_n \quad \text{(C–65)}$$

The physical significance of Ω^j becomes clear when we consider its value in the fluid rest-frame:

$$\dot{\Omega}^0=0;\ \dot{\vec{\Omega}}=(\Omega^1,\ \Omega^2,\ \Omega^3)_{v=0}=\tfrac{1}{2}(\vec{\nabla}\times\vec{v})_{v=0} \quad \text{(C–66)}$$

Thus Ω^j is just the four-dimensional generalization of the local angular velocity associated with the fluid vorticity.

The following identity follows from the definition of Ω^j:

$$(\partial^j u^k-\partial^k u^j)=\frac{2}{c}\epsilon^{jkln}u_l\Omega_n+(u^j u^l\partial_l u^k-u^k u^l\partial_l u^j) \quad \text{(C–67)}$$

Thus

$$\mu_j(\partial^j u^k-\partial^k u^j)=\frac{2}{c}\mu_j(\epsilon^{jkln}u_l\Omega_n)-\frac{1}{c}u^k\mu_j\frac{du^j}{d\tau} \quad \text{(C–68)}$$

and (C–64) becomes

$$\frac{d\mu^k}{d\tau}=2\mu_j\epsilon^{jkln}u_l(\Omega_n-\frac{c}{2}\partial_n\epsilon)-u^k\mu_j\frac{du^j}{d\tau} \quad \text{(C–69)}$$

In the fluid rest-frame we have

$$\frac{d\dot{\mu}^0}{d\tau}=\frac{1}{c}\left(\vec{\mu}\cdot\frac{d\vec{v}}{d\tau}\right)_{v=0} \quad \text{(C–70a)}$$

$$\frac{d\dot{\vec{\mu}}}{d\tau}=2\left[\left(\vec{\Omega}-\frac{c}{2}\vec{\nabla}\epsilon\right)\times\vec{\mu}\right]_{v=0} \quad \text{(C–70b)}$$

Equation (C–70a) is just an identity. This is most easily seen by noting that the last term on the right side of (C–69) can be transformed as follows:

$$-u^k\mu_j\frac{du^j}{d\tau}=+u^k v_j\frac{d\mu^j}{d\tau} \quad \text{(C–71)}$$

Making this transformation in (C–69), (C–70a) becomes the identity

$$\frac{d\dot{\mu}_0}{d\tau}=\left(u^0 u_j\frac{d\mu^j}{d\tau}\right)_{v=0}=\left(\frac{d\mu}{d\tau}\right)_{v=0} \quad \text{(C–72)}$$

As a preliminary to interpreting (C–70b), we substitute (C–42) into (C–65) and obtain

$$\Omega^j=-\frac{\sigma q}{4Mc}\epsilon^{jkln}u_k(F_{ln}+f_{ln}) \quad \text{(C–73)}$$

In the fluid rest-frame this becomes

$$\dot{\Omega}^0=0;\quad \dot{\vec{\Omega}}=-\frac{\sigma q}{2Mc}(\dot{\vec{B}}+\dot{\vec{b}}) \quad \text{(C–74)}$$

which is just the Larmor Condition we met in (2–44). Thus (C–73) is the relativistic statement of the Larmor Condition.

If we substitute (C–74) into (C–70b), we obtain

$$\frac{d\left(\frac{\sigma\hbar}{2}\dot{\vec{\mu}}\right)}{d\tau}=\left(\frac{q\hbar}{2Mc}\dot{\vec{\mu}}\right)\times(\dot{\vec{B}}+\dot{\vec{b}})+\frac{\sigma\hbar c}{2}\dot{\vec{\mu}}\times\dot{\vec{\nabla}}\epsilon \quad \text{(C–75)}$$

Neglecting for the moment the second term on the right, which vanishes when we impose the constraint (C–26), we see that this equation describes the precession in its own rest-frame of a particle having spin angular momentum $\frac{\sigma\hbar}{2}\dot{\vec{\mu}}$ and a magnetic dipole moment $\frac{q\hbar}{2Mc}\dot{\vec{\mu}}$ in a magnetic field $(\dot{\vec{B}}+\dot{\vec{b}})$. These values correspond to the gyromagnetic ratio $\frac{\sigma q}{Mc}$.

The second term on the right corresponds to an additional precession whose angular velocity is $-c\nabla\epsilon$. With the same basic assumptions used in making previous estimates, we find

$$\frac{\left|\frac{\sigma\hbar c}{2}\dot{\vec{\mu}}\times\nabla\epsilon\right|}{\left|\frac{q\hbar}{2Mc}\dot{\vec{\mu}}\times(\dot{\vec{B}}+\dot{\vec{b}})\right|}\sim\frac{\hbar c}{q|\dot{\vec{B}}+\dot{\vec{b}}|L^2} \quad \text{(C–76)}$$

In the case of a system of laboratory dimensions, we can take $L\sim 1$ cm and $|\dot{\vec{B}}|\sim 1$ gauss. We have already seen that for such a case $|\dot{\vec{b}}|\ll|\dot{\vec{B}}|$, and so $\dot{\vec{b}}$ can be neglected. Thus we find that the ratio (C–76) of the precession velocities is of order 10^{-7}. For larger L or $\dot{\vec{B}}$, the ratio would be even smaller. Thus, for macroscopic systems, the spurious pre-

cession resulting from not holding ϵ constant (i.e. not imposing the constraint (C–26)) is completely negligible compared with the real precession given by the first term on the right side of (C–75).

Finally we note that, substituting the spinor equations (C–25) into (C–19), we obtain from the imaginary antisymmetric part of this equation

$$u^l\partial_l s^{jk} = s^j{}_l[(\partial^l u^k - \partial^k u^l) - {}^{lkpq}u_p\partial_q\epsilon] - s^k{}_l[(\partial^l u^j - \partial^j u^l) - \epsilon^{ljpq}u_p\partial_q\epsilon] \quad \text{(C–77)}$$

where we have made use of (C–28) and (C–32). (C–77) is equivalent to (C–64), except that the spin has been represented by the tensor s^{jk} instead of the vector μ^j. The 3-vector equation obtained by writing the space-like components of (C–77) in the particle rest-frame is just (C–70b), which has already been discussed.

Stress-Energy Tensor

Substituting the spinor equations of motion (C–25) into (C–21), the imaginary part of the resulting equation yields

$$I[\partial_{\bar\alpha\beta}(\zeta^{\bar\alpha}\partial_k\zeta^\beta - \chi^{\bar\alpha}\partial_k\chi^\beta)] = \partial^j\mathfrak{F}_{jk} = \frac{\rho c}{\hbar}\cos\epsilon\partial_k\check m + \frac{\rho q}{\hbar c}u_l\partial_k A_l \quad \text{(C–78)}$$

where $\mathfrak{F}_{jk}$ is defined in (C–23). As a preliminary to explaining this relation, let us define a 4-vector p^j as follows:

$$p^j = \sigma\hbar\omega^j - \frac{\sigma q}{c}A^j \quad \text{(C–79)}$$

Since $\sigma\hbar\rho\omega^j$ has been interpreted as the total (or canonical) particle momentum, including spin effects and the contribution of the external electromagnetic field, p^j is just the inertial part of the particle momentum, including spin effects. From (C–32) we have

$$p^j = (\sigma\cos\epsilon)\check m c u^j - \frac{\sigma\hbar}{2\rho}\partial_k(\rho s^{kj}) - \frac{\sigma\hbar}{2}\hat s^{jk}\partial_k\epsilon \quad \text{(C–80)}$$

Using (C–33), this can also be written

$$p^j = [(\sigma\cos\epsilon)\check m + \tilde m]cu^j + \sigma\hbar\Pi^j - \frac{\sigma\hbar}{2}\hat s^{jk}\partial_k\epsilon \quad \text{(C–81)}$$

Now let us define the tensor T_{jk} as follows:

$$T_{jk} \equiv \sigma\hbar c\mathfrak{F}_{jk} - \sigma q\rho u_j A_k = c\rho\left[u_j p_k + \frac{\sigma\hbar}{2}\mu_j\partial_k\epsilon - \frac{\sigma\hbar}{2}s_{jl}\partial_k u^l\right] \quad \text{(C–82)}$$

From (C–82), (C–21), and (C–23) we have

$$T^{jk} = \sigma\hbar c I[\sigma^j_{\bar\alpha\beta}(\zeta^{\bar\beta}\partial^k\zeta^\beta - \chi^{\bar\alpha}\partial^k\chi^\beta)] - \sigma q[\sigma^j_{\bar\alpha\beta}(\zeta^{\bar\alpha}\zeta^\beta + \chi^{\bar\alpha}\chi^\beta)A^k \quad \text{(C–83)}$$

which expresses T^{jk} directly in terms of the spinors ζ^α and χ^α.

Substituting (C–82) into (C–78) and using (C–28), we have

$$\partial^j T_{jk} = \rho(\sigma\cos\epsilon)\partial_k(\check m c^2) + \sigma q\rho(\partial_k A_l - \partial_l A_k)u^l \quad \text{(C–84)}$$

For the case in which condition (C–26) is satisfied, (C–84) becomes

$$\partial^j T_{jk} = \rho\partial_k(\check m c^2) + \sigma q\rho\mathfrak{F}_{kl}u^l = k_k \qquad \text{for } \cos\epsilon = \sigma \quad \text{(C–85)}$$

where k_k is the force density defined in (6–16).

We note that (C–85) is identical with (6–15) except that the contraction on the left side of (C–85) involves the first index, instead of the second as in (6–15). We shall now show that $\partial^j T_{jk} = \partial^j T_{kj}$ (which is not obvious, since T_{jk} is not symmetric).

Substituting the spinor equations of motion into the identity (C–19), we have from the real antisymmetric part of the resulting equation the following relation:

$$\tfrac{1}{2}[\partial^j(\rho\mu^k) - \partial^k(\rho\mu^j)] - \epsilon^{jkln}\mathfrak{F}_{ln} = -\frac{q\rho}{\hbar c}\epsilon^{jkln}u_l A_n \quad \text{(C–86)}$$

Contracting this with ϵ_{pqjk} and using (C–82), we arrive at the following:

$$\frac{1}{c}(T_{jk} - T_{kj}) = -\frac{\sigma\hbar}{2}\epsilon_{jkln}\partial^l(\rho\mu^n) \quad \text{(C–87)}$$

Because the divergence of the right side vanishes, we have the desired result:

$$\partial^j T_{jk} = \partial^j T_{kj} \quad \text{(C–88)}$$

Thus (C–85) is identical with (6–15), and we are justified in interpreting T_{jk} as the stress-energy tensor for the fluid.

Substituting (C–80) into (C–82), we have for the case $\cos\epsilon = \sigma$

$$T^{jk} = \rho\check m c^2 u^j u^k - \frac{\sigma\hbar c}{2}u^j\partial_l(\rho s^{lk}) - \frac{\sigma\hbar c}{2}\rho s^{jl}\partial_k u_l \qquad \text{for } \cos\epsilon = \sigma \quad \text{(C–89)}$$

Comparing with (6–11), we have

$$t^{jk} = -\frac{\sigma\hbar c}{2}[u^j\partial_l(\rho s^{lk}) + \rho s^{jl}\partial_k u_l] \quad \text{for } \cos\epsilon = \sigma \qquad \text{(C–90)}$$

where t^{jk} is the spin-dependent part of T^{jk}.

As discussed in Section V, the fluid energy density is given by T^{00}; the fluid energy flux density $U^j{}_{fl}$ (designated by $U^j{}_{(\pm)}$ in Section V) is given by cT^{0j}; and the fluid momentum density $G^j{}_{fl}$ (designated by $G^j{}_{(\pm)}$ in Section V) is given by $\frac{1}{c}T^{j0}$. It can be shown that in the fluid rest-frame these quantities assume the following form:

$$\dot{T}^{00} = \rho(\check{m}+\tilde{m})c^2 \qquad \text{for } \cos\epsilon = \sigma \qquad \text{(C–91)}$$

$$\vec{\dot{U}}_{fl} = c^2\rho\left[\frac{d(\check{m}\vec{d})}{d\tau}\right]_{v=0} + c^2\left[\vec{\nabla}\times\rho\left(\frac{\sigma\hbar}{2}\vec{\mu}\right)\right]_{v=0} \quad \text{for } \cos\epsilon = \sigma \qquad \text{(C–92)}$$

$$\vec{\dot{G}}_{fl} = \rho\left[\frac{d(\check{m}\vec{d})}{d\tau}\right]_{v=0} \quad \text{for } \cos\epsilon = \sigma \qquad \text{(C–93)}$$

where $\vec{d}$ is defined in (C–36). These expressions make the role of particle spin intuitively evident. The first term on the right side of (C–92) is the energy flux arising from the change in displacement of the center-of-mass of an accelerating particle having spin. $\vec{\dot{G}}_{fl}$ is just the corresponding momentum density. The second term on the right side of (C–92) is the energy flux associated with the linear momentum density arising from the spin angular momentum. If this term (divided by c^2) were to appear in $\vec{\dot{G}}_{fl}$ as well, it would, when we came to calculate the angular momentum density $\vec{r}\times\vec{G}_{fl}$ associated with T^{jk}, give rise to an angular momentum contribution that had its origin in the particle spin. The fact that this term does not appear in $\vec{\dot{G}}_{fl}$ is consistent with the fact that we treat the spin angular momentum density $\vec{S}$ as a separate contribution that is not included in $\vec{r}\times\vec{G}_{fl}$.

Angular Momentum

From (C–85) and (C–88), we obtain the following:

$$(T^{jk} - T^{kj}) = -\partial_l(x^jT^{kl} - x^kT^{jl}) + (x^jk^k - x^kk^j) \qquad \text{(C–94)}$$

Substituting this into (C–87), we have

$$\partial_l M^{jkl} = \frac{1}{c}(x^jk^k - x^kk^j) \qquad \text{(C–95a)}$$

where

$$M^{jkl} = S^{jkl} + \frac{1}{c}(x^jT^{kl} - x^kT^{jl}) \qquad \text{(C–95b)}$$

and

$$S^{jkl} = -\sigma\frac{\hbar}{2}\epsilon^{jkln}(\rho\mu_n) \qquad \text{(C–95c)}$$

As shown in Section V, (C–95a) and (C–95b) constitute the differential form of the statement of conservation of angular momentum, whose integral form is

$$\frac{d}{dt}\int_{v3}(\vec{S} + \vec{r}\times\vec{G}_{fl})dV_3 = \int_{v3}(\vec{r}\times\vec{k})dV_3 \qquad \text{(C–96a)}$$

where from (C–95c) we have

$$\vec{S} \equiv (S^{230}, S^{310}, S^{120}) = \frac{\sigma\hbar}{2}\rho\vec{\mu} = (\rho u^0)\left(\frac{\sigma\hbar}{2}\frac{\vec{\mu}}{u^0}\right) \qquad \text{(C–96b)}$$

The integrand on the left in (C–96a) is the total angular momentum density, and the integrand on the right is the torque density. Since $\vec{S}$ is the spin angular momentum density, and (ρu^0) is the particle density in the observer's reference frame, we see from (C–96b) that $\frac{\sigma\hbar}{2}\frac{\vec{\mu}}{u^0}$ is to be interpreted as the spin angular momentum of a single particle as seen in the observer's frame.

REFERENCES

1. W. M. Elsasser, Rev. Mod. Phys. 28, 135 (1956)
2. T. G. Cowling, *Magnetohydrodynamics* (Interscience Publishers, New York, 1957), Chapter 5
3. E. Bullard and H. Gellman, Phil. Trans. Roy. Soc. A247, 213 (1954)
4. G. E. Backus, Ann. of Phys. 4, 372 (1958)
5. A. Herzenberg, Phil. Trans. Roy. Soc. A250, 543 (1958)
6. T. Takabayasi, Prog. Theor. Phys. Suppl. 4 (1957)
7. F. Halbwachs, *Theorie Relativiste des Fluides à Spin* (Gauthier—Villars, Paris, 1960)
8. R. Schiller, Phys. Rev. 128, 1402 (1962)
9. Z. Grossmann and A. Peres, Phys. Rev. 132, 2346 (1963)
10. C. Moller, *The Theory of Relativity* (Oxford University Press, New York, 1952)
11. B. L. van der Waerden, Nachr. Akad. Wiss. Göttingen, Math.—physik. K1. (1929), p. 100
12. O. Laporte and G. E. Uhlenbeck, Phys. Rev. 37, 1380 (1931)
13. W. L. Bade and H. Jehle, Rev. Mod. Phys. 25, 714 (1953)

N64-23929

RESULTS FROM THE IMP-I RETARDING POTENTIAL ANALYZER*

G. P. SERBU

Presented are some preliminary experimental results obtained with a planar geometry geometry retarding potential analyzer which was flown on the IMP-I Satellite. The plasma energy spectrum for both ions and electrons was measured in the range from 0 to 100 electron volts. Charged particle density measurements have been obtained continuously from 1000 km to 30 earth radii.

The results show a sharp decrease of about an order of magnitude in charged particle density at about $4.5R_E$, similar to the decrease deduced from whistler observations. The electrons exhibited thermal energies for geocentric distances less than 4.5 R_E. The average electron energy then increased gradually to values above 100 electron volts at about 8 R_E. The observed satellite potential was less than 1 volt positive.

INTRODUCTION

Measurement of the ionized plasma enveloping the earth have been reported on by numerous investigators to date. The region to 30 earth radii has been extensively traversed by a number of satellites, beyond this distance measurements have been made from the Soviet Lunik probes and the American Mariner series.

Primarily the energy spectrum of positively charged particles, presumably protons, has been investigated, and only a few measurements on the electron spectrum below 10 keV are available to date.

Presented are some preliminary experimental results obtained with a retarding potential experiment designed to measure the number density and the energy of ions and electrons in the energy range below 100 eV at distances to 30 earth radii.

2. INSTRUMENTATION

The retarding potential analyzer gives information on the arrival direction and energy distribution of both ions and electrons up to 100 eV. The detector is a charged particle trap of planar geometry which is programmed with appropriate voltages such that the number of charged particles of either polarity can be measured separately as a function of a given energy interval.

The charged particle trap is a circular cup 7 cm in diameter mounted such that the aperture is flush with the satellite skin, and so that it looked out at right angles to the spin axis of the satellite.

The outside aperture is a hole of 5 cm^2 in area over which is stretched a fine wire mesh of 95 percent transmission to light. Surrounding the aperture is a tungsten surface of 35 cm^2 area which is electrically connected to the aperture grid. Spaced parallel to the aperture and 4 mm behind it is the retarding grid. The collector is a disk of tungsten which is, parallel to, and 4 mm behind the retarding grid. The geometry and size of the aperture and the relative spacings provide a view angle for the trap of 5 steradians. Figure 1 shows the total viewing direction of the trap as it was spin-modulated at the satellite spin rate of 22 revolutions per minute. Nearly all directions are viewed during one revolution, with the sun vector lying within 20° of the normal to the trap once each revolution.

The collector current is measured by a logarithmic amplifier whose analog output is presented directly to telemetry. The electrometer sensitivity is from 10^{-11} to 10^{-6} amperes of either polarity. An internal current calibrator is used to check the electronics in flight.

*Published as *Goddard Space Flight Center Document X-615-64-109*, May 1964.

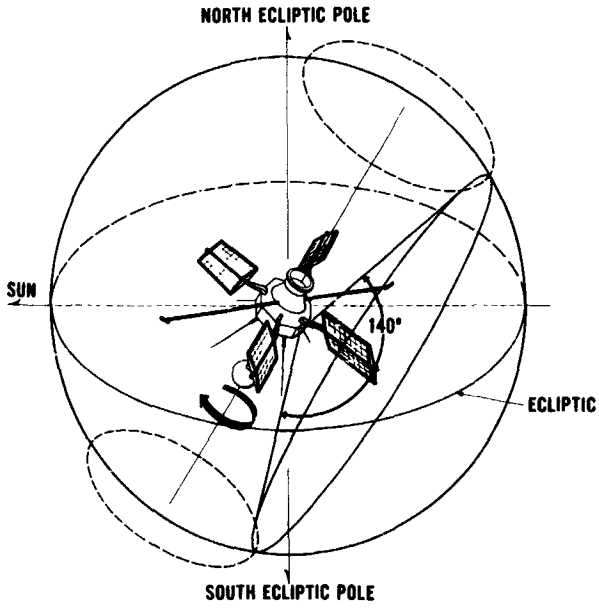
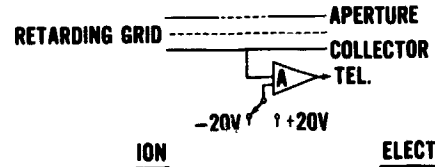


FIGURE 1.—Location and viewing sphere of the retarding potential analyzer aboard the IMP–1 satellite.



	ION		ELECTRON	
	LOW	HI	LOW	HI
APERTURE	−30V	−100	+30V	+100
RETARDING	0-+28V	0-+100V	0--28V	0--100V
COLLECTOR	−20V	−20V	+20V	+20V

APERTURE AREA=5cm²

FIGURE 2.—Schematic diagram of sensor and the various programmed voltages.

Figure 2 shows a schematic diagram of the sensor and also the sequence of voltages programmed to the various trap elements in order to define the four modes of operation: low (30 eV) and high (100 eV) ion spectrum and the low (30 eV) and high (100 eV) electron spectrum. A mechanical programmer which operates in synchronism with the telemetry system, is used to switch the voltages to the trap. As an example, the low energy (30 eV) ion spectrum is done as follows: the aperture grid and guard electrode are maintained at −30 volts with respect to the satellite skin, thus all electrons with energy less than 30 eV are excluded from the trap and all positive ions are accelerated through the aperture. The retarding grid voltage is changed in fifteen equal steps between zero volts and plus 28 volts. A current measurement is made and telemetered at each retarding potential step. Only those ions with energies in excess of the retarding potential will pass through the retarding grid and impinge on the collector. In this manner a current voltage characteristic is obtained for ions with energy below 30 eV.

During the ion measurement photoemission current from the collector appear as a positive current with an expected magnitude of about 10^{-8} amperes. As the collector views the sun the photoemission current amplitude will vary as the cosine of the angle between the sun vector and the trap normal and it will be repetitive with the spin cycle, thus the photocurrent can be corrected for in the data. A highly accurate solar aspect sensor is used to find the precise time that the sun vector lies in the plane normal to the trap.

The electron mode of operation is such that the positive 20-volt potential on the collector suppresses all photoelectron emission from the collector. However, photoelectrons can be emitted from the grids and will appear as a negative current, the expected value being 5.10^{-10} amperes. Grid photoemission currents can also be easily corrected for in the electron data.

The 30 eV and 100 eV ion spectra are obtained first and then the trap voltages are reversed and the 30 eV and 100 eV electron spectra are obtained. The entire ion and electron sequence is repeated once every 10 minutes, each individual 15 step spectrum analysis is done within 5.4 seconds.

3. RESULTS

Figure 3 shows a 30 eV electron spectrum which was obtained at 2.2 earth radii (R_e) geocentric distance. The data points are the measured current values, plotted on a logarithmic scale as a function of the known retarding potential voltage. The resultant semilogarithmic plot is a current-voltage characteristic which yields information on the mean energy and the density of electrons. The observed two orders of magnitude decrease in collector current with two volts negative retardation suggests a maxwellian distribution of electron velocities. A second component or high-energy "tail" to the distribution is noted beyond 2 volts.

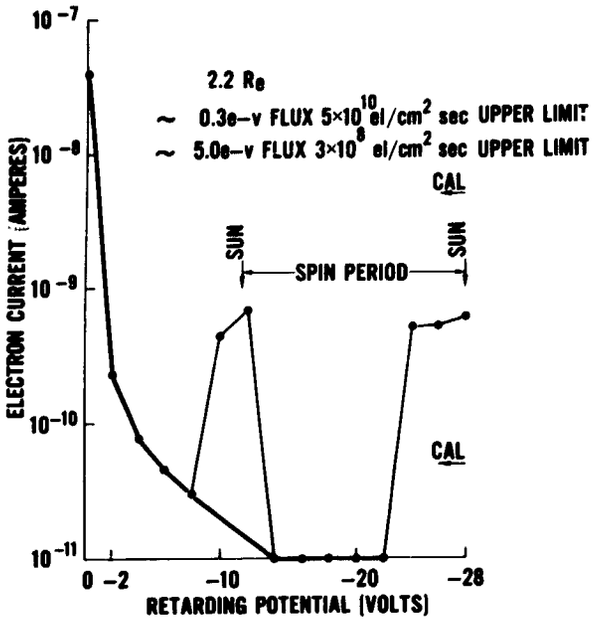


FIGURE 3.—The 30 eV electron spectrum for 2.2 R_e geocentric.

The primary slope yields a value of 0.3 eV or an equivalent electron temperature of 3500°K. From the governing equation:

$$I_- = \tfrac{1}{4} N_e \bar{V}_- eA\, e^{-ve/kT_e} \qquad (1)$$

where I_- is the meaured current, N_e the electron density, $\bar{V}_-$ most probable velocity, A area, v retarding potential, e electronic charge, k Boltzmann constant and T_e the electron temperature. We compute the electron density if A, the effective area is known. Low energy electrons which normally would not impinge on the aperture are deflected into the aperture along the accelerating +30 volt electric field lines; thus an uncertainty in the effective collection area exists. Using the geometric aperture area we compute a flux of 5×10^{10} electrons/cm² sec, this flux could be a factor of 10 too large due to the collection area uncertainty. Using equation (1) a flux of 5.10^{10} el/cm² sec with a mean energy of 0.3 eV at 2.2 R_e; if maxwellian in distribution, represents an electron density of 3.5×10^{3} el/cm³. The high energy component of 5.0 eV has an upper flux value of 3.10^{8} el/cm² sec. It has been suggested by Hanson (1) that at 7000 km, approximately this altitude, a 10^{8} el/cm² sec flux of 5 eV photoelectrons exists. These photoelectrons are formed at altitudes above 300 km and diffuse along magnetic field lines; Mariani (2) extended the work of Hanson and describes the role that these photoelectrons may play in the explanation of the "equatorial" or "geomagnetic" anomaly. Since these electrons are constrained to follow field lines the flux will drop off with the cube of the distance; thus beyond 3 R_e this flux should be below the minimum sensitivity of this experiment.

In Figure 3 the current to the trap decreases below the minimum sensitivity at retarding potentials greater than −12 volts, thus fewer than 1×10^{7} electrons/cm² sec exist with energy in excess of 12 eV. The two photoemission peaks correlate well with the indicated sun position. The magnitude of the photoemission current is consistent with the value of 5×10^{-10} amperes computed on the basis of solar illumination of the 95 percent transparent grid.

Satellites, in general, are charged negative* owing to the high velocity of electrons compared with either the velocity of the satellite or the thermal velocity of ambient ions. However, at very high altitudes, and in interplanetary space, where the electron concentrations are small, a satellite might tend to have a net positive charge due to the photoelectric effect (Ja L Al'Pert et al, 5).

A direct measurement of the net charge, or the resultant potential difference between satellite and plasma as a consequence of the accumulated charge, can be based on observed current-voltage characteristics, Loeb (6); such as the one in Figure 3. Since a positive potential will accelerate ambient electrons to the satellite the measured electron current will not be retarded until an opposing voltage of equal magnitude to that of the satellite potential is reached. At that potential the characteristic will change slope and a further increase in negative voltage will retard the electron current. It is evident from the data of Figure 3 that a break in the curve could exist for potentials more negative than +2 volts; therefore, the satellite potential is not more than 1 volt positive, more probably it is only a few tenths of a volt negative. There is not evidence throughout the magnetosphere that the satellite potential is ever positive.

The 30 eV electron spectrum obtained at 4.6 R_e geocentric is shown in Figure 4. The slope of

*Jastrow and Pearse (3), Beard and Johnson (4)

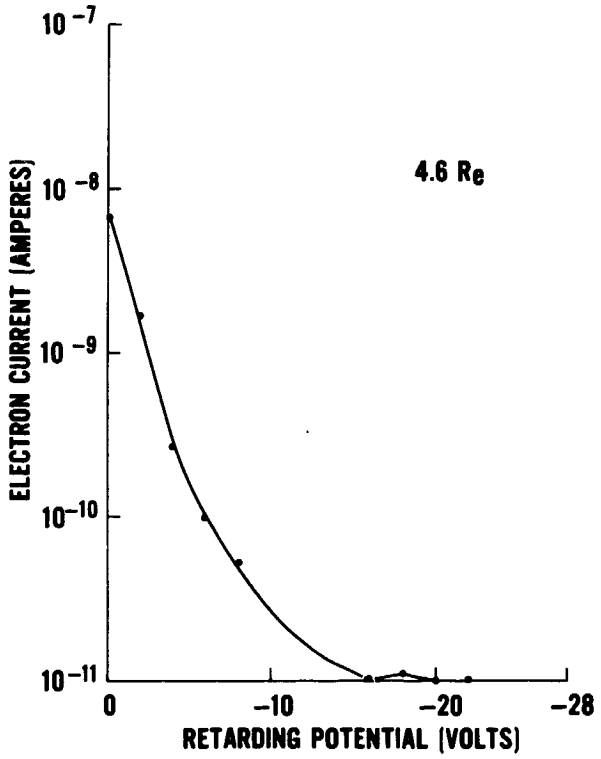


FIGURE 4.—The 30 eV electron spectrum for 4.6 R_e geocentric.

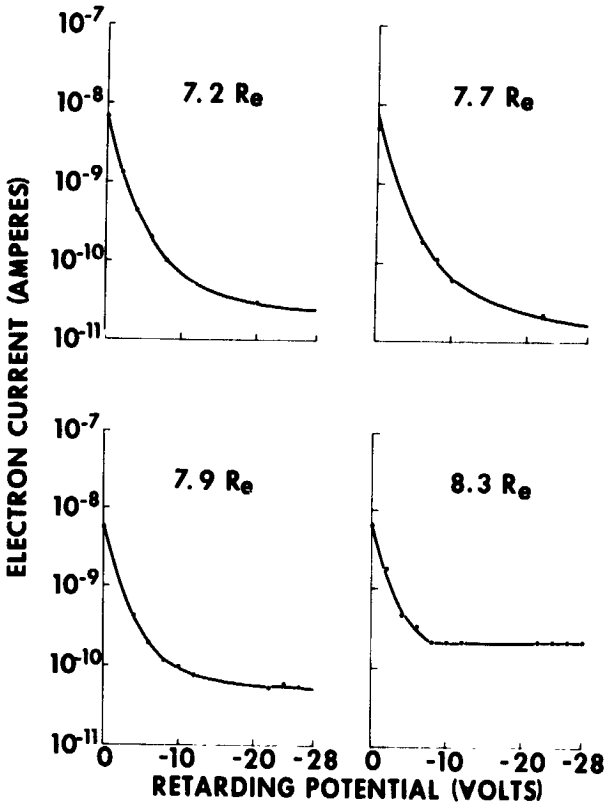


FIGURE 5.—A successive series of 30 eV electron spectra depicting the transit into a region characterized by a residual current in the trap, due to an omnidirectional flux of electrons.

this characteristic yields an electron energy value of 1.4 eV, or an equivalent gas temperature of 16,400°K. A maximum flux of 9×10^9 electrons/cm² sec is computed from the data, assuming the effective area to be equal to the geometric area. Converting to energy, Figure 4 data yields, 2.2×10^{-2}ergs/cm²sec, thus when we compare this to the value of 1 to 10ergs/cm²sec for particles with $E\geq200$ eV obtained by the Explorer XII Cadmium Sulphide Total Energy Detector (CSDTE) Freeman et al (7), it is apparent that a large number of low energy electrons, $E\leq1.4$ eV, exist in the second radiation belt, however, their energy content represents only a small fraction of the total particle energy Gringauz et al (8) infers from the Soviet charge particle traps that in this region an electron flux of 10^9 electrons/cm²sec with energies below 200 eV is present.

Figure 5 shows a consecutive series of 30 eV electron spectra obtained between 7.2 and 8.3 R_e. It is apparent from all spectra that a residual negative current component is measured even at the maximum retardation voltage of 28 volts. Similar data, obtained at 9.2 R_e, is shown in more detail in Figure 6. The residual current amplitude is not roll modulated, thus at 8.2 R_e, on the sunlit side of earth, the satellite crossed into a region of omnidirectional electrons with energy in excess of 100 eV at a flux value of 1.5×10^8 electrons/cm²scc. Gringauz et al (9, 10, 11) observed, with Lunik 1 and also with Lunik 2 and Mars 1, a flux of ($1-2\times10^8$el/cm²sec) at these distances with an energy larger than 200 eV, this region has been labeled by him as "the outermost belt of charged particles."

Figure 7 shows the 100 eV ion and electron spectra obtained at 9.0 and 9.1 R_e. The omnidirectional current residual in both traps at 100 volt retardation indicates that both ions and electrons are present in the outermost belt of charged particles with energies in excess of 100 eV. The data of Figure 7 may also be used to illustrate that the net current flow to the satellite is essentially zero. At zero retardation, the current flow to the collector should be representative of any given area on the surface of the satellite, thus in order to maintain a constant satellite to plasma

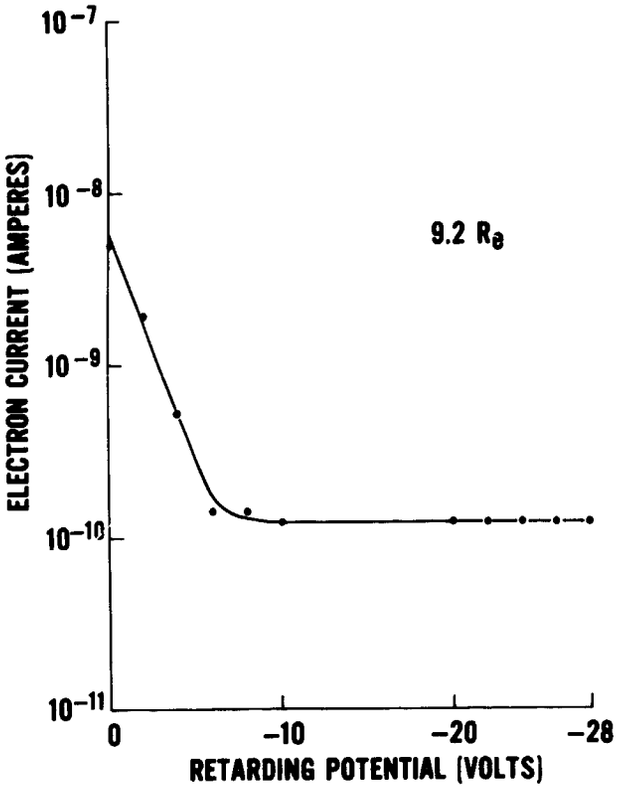


FIGURE 6.—The 30 eV electron spectrum for 9.2 R_e geocentric.

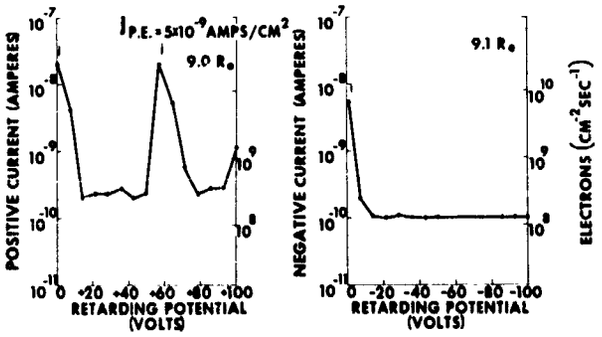


FIGURE 7.—The 100 eV ion and electron spectra obtained at 9.0 R_e and 9.1 R_e geocentric.

potential it must be shown that the net current flow to the satellite is zero. The photoemission current constitutes a positive current flow, i.e. electrons flowing outward from the satellite, which acts over the illuminated cross sectional area of the satellite. The negative current is, on the other hand, effective over the entire area of the satellite velocity, and so electrons diffuse inward from all directions. In order to have a net zero current flow the ratio of the photocurrent to electron current must be equal to 4 which is the inverse ratio of the areas to which these current flows are effective. From Figure 7 the current ratios at zero volt retardation are seen to be 4; thus the trap currents at zero retardation are inferred to be representative of the overall current flow to the satellite and we deduce that the equilibrium satellite potential in sunlight is stabilized near zero volts.

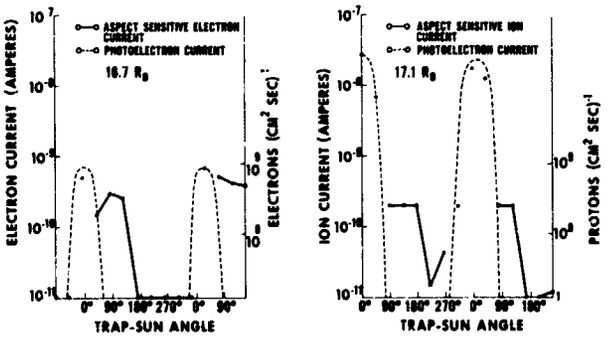


FIGURE 8.—The ion and electron currents plotted as a function of trap-sun angle orientation. These responses typify the data obtained at distances beyond 16.2 R_e geocentric on the first IMP–1 orbit.

The data in Figure 8 is representative of the trap currents at distances beyond 16.2 R_e geocentric. Positive and negative currents have been plotted as a function of orientation relative to the sun. This presentation was chosen since attempts at plotting the currents as a function of retarding voltages were fruitless. After subtracting out data points which are due to photoemission, the dashed curve, a residual net current in both traps, is evident at 90 degree clockwise rotation from the sun, in the anti-solar direction the currents to the trap drop below 1×10^{-11} amperes. Since the retarding potentials do not affect the trap currents it is evident that the negative current measurement represents a net negative charge entering the trap with energies in excess of 100 eV, and that the positive current measurement represents a sum of net positive charge plus secondary electron emission. Assuming a secondary electron emission efficiency of 100 percent these observations are consistent with the computation that what is seen beyond 16.2 R_e, i.e. in the solar wind region is a net flux of both ions and electrons of the order of 10^8 particles/cm²sec with energies in excess of 100 eV.

4. CONCLUSIONS

An overall summary of these observations may be made with the use of Figure 9, which is a plot of the electron flux as a function of geocentric distance. The three plotted curves were computed from the negative current in the trap found in the 0 to −5 volt retardation interval, the −5 to −10 volt and the −100 volt interval.

The flux of electrons with energies $0 \leq E \leq 5$ eV shows a steep decrease with distance between 3 and 5 R_e. This type of electron drop-off is in agreement with whistler observations, Carpenter (12) refers to the decrease as the "equatorial knee" and points out that the equatorial knee moves further out; i.e. beyond 3 R_e, during magnetically quiet periods. Out to about 3 R_e we note that the flux in the energy interval $5 \leq E \leq 10$ eV is decreasing with distance. This energy interval contains the previously mentioned Hanson photoelectrons, evident in Figure 2. Furthermore the drop off in flux evident in Figure 9 lends credulity to our belief that in this region we observe a flux of trapped photoelectrons produced at altitudes above 300 km.

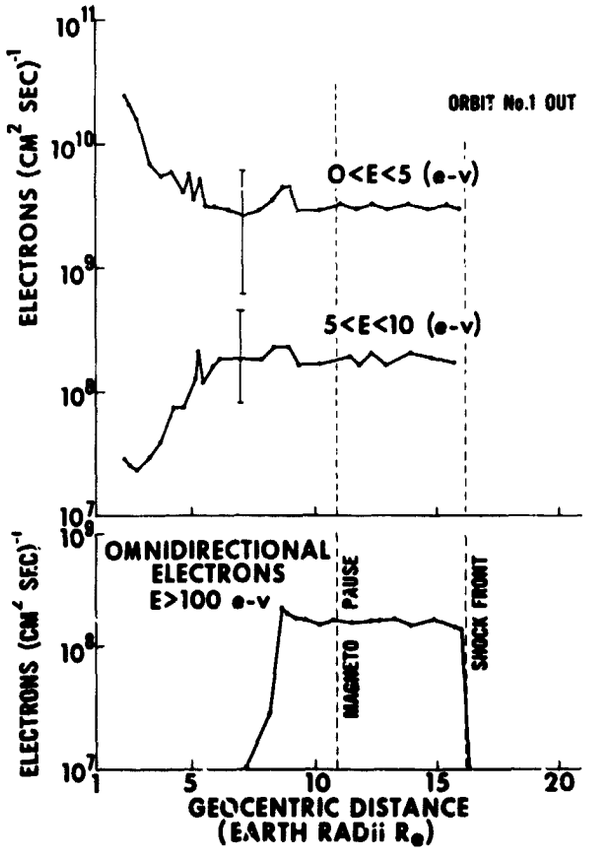


FIGURE 9.—Plot of the electron flux as a function of geocentric distance. Three electron energy intervals are shown.

Beyond 3 R_e an increase in the flux of electrons with energy $5 \leq E \leq 10$ eV is noted, at 7 R_e this flux is comparable to the flux of lower energy electrons. The error bars in the figure reflect the previously discussed uncertainty in the "effective collection area." Thus the overall picture is one of an initial sharp decrease in thermal electron concentration accompanied by an increase in the average energy of electrons, such that, at some 7 R_e geocentric there are nearly as many electrons with energy greater than 5 eV as there are those with energy below 5 eV. At 8.2 R_e geocentric, on the sunlit side of earth, with a payload-earth-sun angle of 25 degrees, a flux of 2×10^6 electrons/cm² sec omnidirectional electrons with energy greater than 100 eV is observed. There is no significant variation in this flux rate across the magnetosphere; the magnetosphere boundary was determined through the IMP-I magnetometer data of Ness (12). Beyond 16.2 R_e the flux of these omnidirectional particles drops off sharply. Whereas, crossing of the magnetosphere boundary is not evident from the trap data, the crossing of the shock front is. On the earth side of the shock front the trap currents are governed by the retarding potential voltages. Outside the shock front the trap currents are governed by the trap-sun angle orientation.

The recent controversy Gringauz (13), Van Allen (14), concerning the observations of particles in the outermost belt of charged particles and whether or not these particles are trapped is not resolved on the basis of these IMP observations; since it is possible to explain the observed constant flux across the magnetosphere boundary either on the basis of some unknown mechanism for solar wind penetration as suggested by Gringauz or on the basis that the total trapped flux inside the boundary just equals the flux outside the boundary, a possibility not excluded by the measurements of Explorer XII and IV. Frank et al (15).

5. ACKNOWLEDGMENTS

Messrs. W. J. Archer and J. Luce were responsible for the instrumentation associated with this

experiment. Also acknowledged are the helpful discussions with Mr. R. E. Bourdeau.

6. REFERENCES

1. HANSON, W. B., "Electron Temperatures in the Upper Atmosphere", Space Research III Proceedings of the Third International Space Sciences Symposium, pp. 282–302, North-Holland Publishing Company, Amsterdam, 1963.
2. MARIANI, F., "Pitch Angle Distribution of the Photoelectrons and Origin of the Geomagnetic Anomaly in the F2 Layer", J. G. R., Vol. 69, No. 3, Feb. 1, 1964.
3. JASTROW, R., and PEARSE, C. A., "Atmospheric Drag on the Satellite", J. G. R., Vol. 62, No. 3, Sept., 1957.
4. BEARD, D. B., and JOHNSON, F. S., "Charge and Magnetic Field Interaction with Satellites", J. G. R., Vol. 65, No. 1, Jan., 1960.
5. AL'PERT, JA. L., GUREVIC, A. V., and PITAEVSKIJ, L. P., "Effects Due to an Artificial Satellite in Rapid Motion through the Ionosphere or the Interplanetary Medium", Space Science Reviews, Vol. 11, No. 5, Nov., 1963, pp. 680–748.
6. LOEB, L. B., "Basic Processes of Gaseous Electronics", University of California Press, Berkeley and Los Angeles, 1955, pp. 331–371.
7. FREEMAN, J. W., VAN ALLEN, J. A., and CAHILL, L. J., "Explorer XII Observations of Magnetospheric Boundary and the Associated Solar Plasma on Sept. 13, 1961. J. G. R., Vol. 68, No. 8, April, 1963.
8. GRINGAUZ, K. I., BALANDINA, S. M., GORDOVSKII, G. A., and SHYUTTE, N. M., "On the Results of Tests with Three-Electrode Charged Particle Traps in the Second Radiation Belt and in the Outermost Belt of Charged Particles", Planetary Space Science, 1964, Vol. 12, pp. 175–179. Pergamon Press Ltd., Printed in Northern Ireland.
9. GRINGAUZ, K. I., BEZRUKIRKH, V. V., OZEROV, V. D., and RYBCHINSKII, R. E., "A Study of Interplanetary Ionized Gas, Energetic Electrons and Solar Corpuscular Radiation Using Three-Electrode Charged Particle Traps on the Second Soviet Cosmic Rocket", DOKL. AKAD NAUK SSSR 131, 1301 (1960).
10. GRINGAUZ, K. I., KURT, V. G., MOROZ, V. I., and SHKLOVSKII, I. S., "Ionized Gas and Fast Electrons in the Neighborhood of the Earth and in Interplanetary Space" [ISKUSSTVENNYE SPUTNIKI ZEMLI (Artificial Satellites of the Earth)] No. 6, p. 108, IZD. AKAD. NAUK. SSSR, Moscow, 1961.
11. GRINGAUZ, K. I., BEZRUKIRKH, V. V., MUSATOV, L. S., RUBCHINSKII, R. E., and SHERONOVA, S. M., "Measurements Made in the Earth's Magnetosphere by Means of Charged Particle Traps Aboard the Mars 1 Probe", Space Research Proc. Intern. Space Science Symp. 4th Warsaw, 1963. North-Holland Publishing Company, Amsterdam, June, 1963.
12. NESS, N. F., (Private Communication)
13. GRINGAUZ, K. I., "Remarks on Papers by J. W. Freeman, J. A. Van Allen, and L. J. Cahill, Explorer 12 Observations of the Magnetosphere Boundary and the Associated Solar Plasma on September 13, 1961, and by L. A. Frank, J. A. Van Allen and E. Macagno, "Charged-Particle Observations in the Earth's Outer Magnetosphere", J. G. R. Vol. 69, No. 5, March 1, 1964.
14. VAN ALLEN, J. A., "Remarks on Accompanying Letter by K. I. Gringauz", J. G. R., Vol. 69, No. 5, March 1, 1964.
15. FRANK, L. A., VAN ALLEN, J. A., and MACAGNO, E., "Charged-Particle Observations in the Earth's Outer Magnetosphere", J. G. R., Vol. 68, No. 12, June 15, 1963.

N64-28084

EFFECT OF COMPRESSIBILITY, ROTATION AND MAGNETIC FIELD ON THE DRAG ON A SPHERE OSCILLATING IN A CONDUCTING, VISCOUS MEDIUM*

M. P. SINGH†

I. INTRODUCTION

The role of rotation and magnetic fields in cosmical fluid dynamics is well known. Recently, some attention has been given by several authors (1–7) on the interaction of Coriolis forces and Lorentz forces on flow phenomena.

It is generally accepted (1, 2) that hydromagnetic flow in the earth's liquid core is somehow responsible for the main geomagnetic field, and therefore a theory of the dynamics of core motions is required in order to understand a plausible theory of the earth's magnetic field. It has been suggested that the near coincidence between the geographic and geomagnetic poles is the result of the strong influence of Coriolis forces, due to the earth's rotation, on motions in the core.

So much of meteorology depends ultimately upon the dynamics of a revolving fluid. The large-scale and moderate motions of the atmosphere are greatly affected by the rotation of the earth. Kelvin is said to have pointed out (8) that rotation confers on a fluid certain properties resembling those of an elastic solid, and a rotating fluid can transmit waves. In the case of an infinite liquid, rotating as a rigid body about an axis, the amount of energy possessed by the liquid is infinite and it is of great interest to know how small disturbances propagate in such a liquid.

To understand some of these phenomena, it will be interesting to study the flow of a rotating fluid around elementary bodies. We consider here small oscillations of a sphere in a compressible, viscous, electrically conducting and rotating medium in the presence of a uniform magnetic field. In addition to the above mentioned interests, this problem may have some applications in connection with the transmission of sound by fog (see 9, p. 659).

The classical problem of the oscillation of a sphere in a viscous fluid was first considered by Stokes (9, p. 643). The drag force experienced by the sphere is (in dimensionless form)‡

$$X_0 = -6\pi u_s\left(\frac{1}{R_e}+\frac{1}{\sqrt{2R_e}}\right)-2\pi\frac{du_s}{dt}\left(\frac{1}{3}+\frac{3}{\sqrt{2R_e}}\right). \quad (1)$$

It is the purpose of the present investigation to determine the effects of the magnetic field, compressibility, coriolis forces and their mutual interaction on the drag and other physical variables. Because of the motion of the fluid in the magnetic field, an associated electrical field is produced which, according to Ohm's law sets up electrical currents in the fluid if the latter is a conductor. The interaction of these currents with the magnetic field then produces a body force which must be included in the Navier-Stokes equations for the motion of the fluid. The effect of this body force is to inhibit the motion of the fluid across the lines of force. The viscous effect gives rise to viscous dissipative waves, the magnetic effect is responsible for Alfvén waves while the compressibility and rotation produce sound and the so-called Taylor waves respectively. Their mutual interaction will make the situation even more complicated. In the next section, we make several assumptions to make the problem mathematically more amenable.

*Published as *Goddard Space Flight Center Document X-640-64-143*, May 1964.

†Dr. Singh is a National Academy of Sciences—National Research Council Postdoctoral Resident Research Associate.

‡The dimensionless quantities and parameters in (1) are explained in section II.

II. FUNDAMENTAL EQUATIONS AND FORMULATION OF THE PROBLEM

Let the sphere oscillate along the axis of the rotating fluid which is taken to be the x-axis. The origin is at the mean position of the center of the sphere. Let the uniform magnetic field be $H^0\underline{i}$. The equations of motion of a compressible, viscous, unbounded fluid rotating with a constant angular velocity $\Omega\underline{i}$, referred to a rotating frame of reference are (in m.k.s. units)

$$\rho\frac{d\underline{v}}{dt}+2\rho\underline{\Omega}\times\underline{v}+\rho\underline{\Omega}\times(\underline{\Omega}\times\underline{r}) = -\nabla p+\mu \text{ curl } \underline{H}\times\underline{H}+\rho\nu\nabla^2\underline{v}+\frac{\rho\nu}{3}\nabla \text{ div } \underline{v}, \quad (2)$$

where ρ is the density, $\underline{v}$ the velocity of the fluid, $\underline{H}$ is total magnetic field, p is the pressure, r is the radial coordinate: $r^2=y^2+z^2$, ν the kinematic viscosity (assumed constant) and μ the magnetic permeability (assumed constant).

The equation of continuity is

$$\text{div } (\rho\underline{v})+\frac{\partial\rho}{\partial t}=0. \quad (3)$$

The physical equation (the conduction of heat is neglected) is

$$\frac{1}{p}\frac{dp}{dt}=\frac{\gamma}{\rho}\frac{d\rho}{dt},$$

(where γ denotes the ratio of two specific heats) which can be written as (9, p. 654)

$$p=p^0+\rho^0 c^2 s, \quad (4)$$

where c is the velocity of sound in the absence of viscosity, p^0, ρ^0 represent pressure and density in the undisturbed state and s denotes the condensation

$$\rho=\rho^0(1+s). \quad (5)$$

Maxwell's equations with the usual notation for electromagnetic quantities are

$$\text{(i) curl } \underline{H}=\underline{J}, \qquad \text{(ii) div } \underline{H}=0,$$
$$\text{(iii) curl } \underline{E}=-\mu\frac{\partial\underline{H}}{\partial t}, \qquad \text{(iv) div } \underline{E}=0, \quad (6)$$
$$\text{(v) } \underline{J}=\sigma[\underline{E}+\mu\underline{v}\times\underline{H}],$$

where the electrical conductivity σ is treated as constant.

From (i), (iii) and (v) we obtain

$$\frac{\partial\underline{H}}{\partial t}-\text{curl } (\underline{v}\times\underline{H})=\frac{1}{\mu\sigma}\nabla^2\underline{H}. \quad (7)$$

As in the classical problem, the convective terms in (2) will henceforth be neglected. This is justified if

$$R_s=a|v_s|/\nu\ll 1, \quad (8)$$

where v_s is the velocity of the sphere and a is its radius.

Further linearization of (2) and (3) is possible by means of (5) which yields

$$\rho^0\frac{\partial\underline{v}}{\partial t}+2\rho^0\underline{\Omega}\times\underline{v}+\underline{\Omega}\times(\underline{\Omega}\times\underline{r}) = -\nabla p+\mu \text{ curl } \underline{H}\times\underline{H}+\rho^0\nu\nabla^2\underline{v}+\frac{\rho^0\nu}{3}\nabla \text{ div } \underline{v} \quad (9)$$

and

$$\text{div } \underline{v}+\frac{\partial s}{\partial t}=0. \quad (10)$$

These equations will be now reduced to a dimensionless form by referring the length to a, the radius of the sphere, the velocity to $a\lambda$ ($\lambda/2\pi$ is the frequency of the oscillation), the time to $1/\lambda$, the magnetic field to H^0, the electric field to $\mu a\lambda H^0$, the current density $\underline{J}$ to H^0/a and the pressure to $\rho^0 a^2\lambda^2$. The equations (6), (7) then become

$$\text{(i) curl } \underline{H}=\underline{J}, \qquad \text{(ii) div } \underline{H}=0,$$
$$\text{(iii) curl } \underline{E}=\frac{\partial\underline{H}}{\partial t}, \qquad \text{(iv) div } \underline{E}=0, \quad (11)$$
$$\text{(v) curl } \underline{H}=R_m[\underline{E}+\underline{v}\times\underline{H}]$$

and

$$\nabla^2\underline{H}=R_m\left[\frac{\partial\underline{H}}{\partial t}-\text{curl } (\underline{v}\times\underline{H})\right] \quad (12)$$

where

$$R_m \text{ (Magnetic Reynolds number)}=a^2\lambda\mu\sigma. \quad (13)$$

In addition, we introduce the following parameters:

$$R_e \text{ (Reynolds number)}^*=a^2\lambda/\nu \quad (14a)$$

*It should be noted that we have used $a\lambda$ as velocity in the definition of the Reynolds number. There still is another important dimensionless parameter α/a, where α is the amplitude of the oscillatory motion of the sphere. It is assumed that this quantity is small compared to unity.

The velocity of sphere is $v_s\,\underline{i}=\alpha\,\lambda\,e^{\Omega t}\,\underline{i}$, or in dimensionless form, $u_s\,\underline{i}=\alpha/a\,e^{it}\,\underline{i}$.

The Reynolds number for the flow might also have been defined as

$$R_s=a\alpha\lambda/\nu\ (=R_e\,\alpha/a).$$

It is necessary that

$$R_s\ll 1,$$

whereas the requirement for R_e is less severe.

$$\beta\ (\text{magnetic pressure number})=\frac{\mu H^{02}}{\rho^0 a^2\lambda^2}, \tag{14b}$$

the ratio of the magnetic pressure to dynamic pressure, and

$$\delta=2\Omega/\lambda, \tag{14c}$$

the inverse of the so-called Rossby number $R_o=\lambda/(2\Omega)=1/\delta$. In terms of these parameters, equations (9) and (10) reduce to

$$\frac{\partial \underline{v}}{\partial t}+\delta\underline{i}\times\underline{v}=\beta\ \text{curl}\ \underline{H}\times\underline{H}+\frac{1}{R_e}\nabla^2\underline{v}-\left(c^2+\frac{i}{3R_e}\right)\nabla s$$

and

$$\text{div}\ \underline{v}=-\frac{\partial s}{\partial t} \tag{16}$$

where velocity of sound c is now in non-dimensional form referred to $a\lambda$, and if we assume that

$$R_o\gg 1, \tag{17}$$

so as to neglect second and higher powers of δ, the centrifugal force in (9) being of 0 (δ^2) is dropped* in (15).

The boundary conditions on $\underline{v}$ are $\underline{v}=u_s\,\underline{i}$ on the surface of the sphere and is zero at infinity. If the conductivity σ' of the sphere† is comparable to the conductivity σ of the fluid, we require that $\underline{H}$ should be continuous across the surface of the sphere, while at infinity, $\underline{H}=\underline{i}$.‡

We next make two important assumptions:

(1) The magnetic Reynolds number R_m is a small parameter. We can then find a perturbation solution for small R_m. This is a valid assumption in most practical problems and this technique has been adopted in several discussions (see for example, Ludford [10] and Tamada [11]).

(2) The magnetic pressure is comparable to dynamic pressure, i.e. we assume $\beta=0(1)$.

We assume an expansion in powers of R_m:

$$\begin{aligned}\underline{v}&=\underline{v}_o+R_m\underline{v}_1+\cdots\\ \underline{H}&=\underline{H}_o+R_m\underline{H}_1+\cdots\\ \underline{E}&=\underline{E}_o+R_m\underline{E}_1+\cdots\\ s&=s_o+R_m s_1+\cdots\end{aligned} \tag{18}$$

and obtain a set of equations from (11), (12), (15) and (16) with appropriate boundary conditions. It should be noted that some of the complications which arise from boundary conditions in magneto-gas-dynamics are conveniently absent in this problem.

III. SOLUTION OF THE RESULTING EQUATIONS

Zeroth Order Equations

(i) curl $\underline{H}_o=0$, div $\underline{H}_o=0$, that is the magnetic field to this order is independent of the fluid velocity. The same equations hold inside the sphere and the boundary conditions are that $\underline{H}_o=\underline{i}$ at infinity and it is continuous across the surface of the sphere. The solution is, therefore,

$$\underline{H}_o\equiv\underline{i}. \tag{19}$$

(ii) The electric field $\underline{E}_o$ satisfies curl $\underline{E}_o=0$, div $\underline{E}_o=0$, both in the fluid and the sphere. The continuity of the tangential electric field on the surface of the sphere and vanishing of $\underline{E}_o$ at infinity implies

$$\underline{E}_o\equiv 0. \tag{20}$$

(iii) The zeroth order flow fields and pressure fields are unaffected by the magnetic field and are given by§

$$(\nabla^2+h^2)\underline{v}_0=-i\left(\frac{k^2-h^2}{k^2}\right)\nabla s_0-\delta R_e\underline{v}_0\times\underline{i} \tag{21}$$

*Notice that for large values of radial distance, our neglecting centrifugal force may not be justified even though δ^2 is negligible. In the Incompressible case, it presents no difficulty since it can be absorbed with pressure term. However, in Compressible case, this cannot be done. But since we are primarily interested in the effects of Coriolis forces on the motion, following the standard procedure in such cases (see for example Hide and Roberts 1960 [2] and Talwar 1964 [5]), centrifugal force may be ignored.

†The discussion will be valid even if the sphere is an insulator. We assume that the permeabilities of the sphere and the fluid are equal.

‡These conditions imply that the tangential components of electric field will be continuous across the sphere and that $\underline{E}=0$ at infinity.

§The boundary condition $\underline{v}=u_s\,\underline{i}=\frac{\alpha}{a}e^{it}\,\underline{i}$ on the surface of the sphere and the linearity of the system of differential equations imply that the only time dependence of the physical variables is a factor e^{it}.

and

$$\text{div } \underline{v}_0 = -i\, s_0 \tag{22}$$

where

$$k^2 = \frac{3R_e}{4i + 3R_e c^2} \tag{23}$$

$$h^2 = -i\, R_e. \tag{24}$$

As in the non-rotating case, we have here axial symmetry in that the physical variables are independent of the azimuthal coordinate φ.* However, the azimuthal components of velocity and magnetic field will be non-zero here in contrast to the non-rotating case in which v_φ and H_φ were zero.

From (21) and (22) it follows that $v_{cx} = 0(1)$, $v_{or} = 0(1)$ and $v_{o\varphi} = 0(\delta)$ and hence

$$(\nabla^2 + k^2) s_o = 0(\delta^2). \tag{25}$$

It follows that the coupled equations (21) can be simplified in this case. In fact, in view of (17) and (25), v_{cx} and v_{or} satisfy identically the same system of equations and boundary conditions as in the non-rotating case. The solution is therefore (9, p. 656)

$$\text{(i) } s_0 = A \times f_1\,(kR) \tag{26}$$

$$\text{(ii)} \underline{v}_{-0} = \frac{i}{k^2} \nabla s_0 + \underline{v}^c_{-0}$$

where the first term in (ii) is a particular integral of (21) and (22) and the last term

$$\underline{v}^c_{-0} = B\left[2 f_0(hR) \nabla x - h^2 R^5 f_2(hR) \nabla \frac{x}{R^3}\right], \tag{27}$$

is the complementary function. (R, Θ, φ)** are the spherical polar coordinates. The function $f_n\,(\zeta)$, defined as

$$f_n(\zeta) = \left(-\frac{1}{\zeta}\frac{d}{d\zeta}\right)^n \frac{e^{-\zeta}}{\zeta}, \tag{28}$$

represents spherical waves with rapidly diminishing amplitude—the one with argument $\zeta = hR$ denotes boundary layer effects (viscous) while the one with argument $\zeta = kR$ essentially represents compressibility effects (sound waves). The constants A and B are given by

$$A = \frac{-ik^2 h f_1'(h) u_s}{2k f_1(h) f_1'(k) + h f_1'(h)[k f_1'(k) + f_1(k)]},$$
$$B = \frac{k f_1'(k) u_s}{3[2k f_1(h) f_1'(k) + h f_1'(h)\{k f_1'(k) + f_1(k)\}]}. \tag{29}$$

Now v_{0_φ} can be determined in a straight-forward manner from (21). In fact, exactly similar arguments can be applied while determining $\underline{H}_1$ and $\underline{v}_1$ to show that only azimuthal components H_{1_φ} and $v_1{}^\varphi$ are affected by rotation. In view of this decoupling of the system of equations, the rotational effects, which only generate azimuthal components of velocity and magnetic field, are discussed in Appendix and henceforth, we would only consider non-rotating case.

First Order Equations

(i) The first order magnetic field $\underline{H}_1$ satisfies

$$\nabla^2 \underline{H}_1 = -i s_0 \underline{i} - \frac{\partial \underline{v}_0}{\partial x},$$
$$\text{div } \underline{H}_1 = 0,$$

whose solution is

$$\underline{H}_1 = \underline{H}^c_1 + \frac{i}{k^2} s_0 \underline{i} + \frac{i}{k^4} \nabla \frac{\partial s_0}{\partial x} + \frac{1}{h^2} \frac{\partial \underline{v}_0{}^c}{\partial x} \tag{30}$$

where the first term is the complementary function

$$\underline{H}^c_1 = C \nabla \frac{P_2(\cos\theta)}{R^3} \tag{30a}$$

and the remaining terms represent a particular integral.

On the other hand, inside the sphere, curl $\underline{H}_1 = 0$, div $\underline{H}_1 = 0$, which shows that the sphere acts like an insulator up to this order. It follows that

$$\underline{H}_1 = D \nabla [R^2 P_2(\cos\theta)] \tag{31}$$

inside the sphere. The constants C and D in (30a) and (31) are determined by the requirement that $\underline{H}_1$ is continuous across the surface of the sphere.

(ii) Finally, the first order flow fields and pressure fields are given by

$$\left.\begin{aligned} (\nabla^2 + h^2) \underline{v}_1 &= \beta R_e \left[\nabla H_{1x} - \frac{\partial \underline{H}_1}{\partial x}\right] - i\left(\frac{k^2 - h^2}{k^2}\right) \nabla s_1, \\ \text{div } \underline{v}_1 &= -i s_1 \end{aligned}\right\} \tag{32}$$

*(r, φ, x) are the cylindrical coordinates.

**$x = R \cos\theta$, $r = R \sin\theta$.

from which it follows that

$$(\nabla^2+k^2)s_1=\beta k^2\left[\frac{i}{k^2}\frac{\partial^2 s_0}{\partial x^2}+is_0+\frac{\partial v_{0x}^c}{\partial x}\right]^* \quad (33)$$

We have therefore

$$s_1=s_1^c+\frac{i\beta}{2}\times\frac{\partial s_0}{\partial x}-\frac{Ai\beta}{2}\times f_0(kR)+\frac{\beta k^2}{k^2-h^2}\frac{\partial v_{0x}^c}{\partial x} \quad (34)$$

where the first term

$$s_1^c=E_1R\cdot P_1(\cos\theta)\,f_1(kR)+E_2R^3\cdot P_3(\cos\theta)f_3(kR) \quad (35)$$

is the complementary function of (33) and the remaining terms are a particular integral; and

$$v_1=v_1^c-\frac{\beta}{2k^2}\left\{\nabla\left(x\frac{\partial s_0}{\partial x}\right)-\frac{2}{h^2-k^2}\nabla\frac{\partial^2 s_0}{\partial x^2}\right\}-\frac{\beta}{k^2}\nabla s_0+$$
$$+\frac{\beta h^2}{k^2(h^2-k^2)}\frac{\partial s_0}{\partial x}\underline{i}+\frac{\beta A}{2k^2}\nabla[xf_0(kR)]+\frac{i}{k^2}\nabla s_1^c- \quad (36)$$
$$-\frac{i\beta}{2}\times\frac{\partial v_0^c}{\partial x}-\frac{i\beta}{2}\frac{k^2+h^2}{k^2-h^2}v_{0x}^c\underline{i}.$$

Here the first term

$$v_1^c=F_1\left[2f_0(hR)\nabla x-h^2R^5f_2(hR)\nabla\frac{x}{R^3}\right]$$
$$+F_2\left[4f_2(hR)\nabla\{R^3P_3(\cos\theta)\}\right. \quad (37)$$
$$\left.-3h^2R^9f_4(hR)\nabla\frac{P_3(\cos\theta)}{R^4}\right]$$

is the complementary function of (32) and the remaining terms are a particular integral. The constants E_1, E_2, F_1 and F_2 are determined by the requirement that $\underline{v}_1=0$ at $R=1$.

The following points are worth noting in our solution:

(a) The terms (like $f_n(hR)$) containing the coefficients B, F_1 and F_2 represent the compressibility effects on viscous waves.

(b) The terms (like $f_n(kR)$) containing the coefficients A, E_1 and E_2 represent viscous effects on sound waves.

(c) The terms corresponding to the coefficients C and D represent the potential part of the magnetic field.

*v_{0x}^c is the x-component of the complementary function v_0^c given by (27).

At this point, we would discuss in some more detail the case when the fluid is incompressible.

IV. INCOMPRESSIBLE CASE

The solution in this case can be obtained from the above in the limit as $c\to\infty$ (i.e., $k\to 0$), $s\to 0$ and $c^2s\to p$. (38)

It is clear that we will still have terms (a) and (c) and the corresponding coefficients will be in the limit:

$$B\to\frac{u_s}{2f_0(h)},\quad F_1\to-\frac{\beta Bi}{10}\left[5+h^2\frac{f_1(h)}{f_0(h)}\right],$$
$$F_2\to-\frac{\beta Bih^2}{35}[4+f_1(h)/f_2(h)] \quad (39)$$
$$C\to\frac{h^2B}{5}[f_2(h)-2f_3(h)]\text{ and }D\to-\frac{u_s}{10}.$$

Evidently waves (sound) represented by (b) will no longer be present and the corresponding terms from (b) can be obtained by the limiting process (38):

$$p_0=c^2s_0\to A_1\frac{P_1(\cos\theta)}{R^2}$$

$$p_1=c^2s_1\to\left[\frac{D_1}{R^2}+\beta\left\{\frac{6}{5}BRf_1(hR)-\frac{2}{5}\frac{A_1i}{R^2}\right\}\right]P_1(\cos\theta)$$
$$+\left[\frac{D_2}{R^4}-\beta\left\{\frac{6}{5}Bh^2R^3f_3(hR)\right.\right.$$
$$\left.\left.+\frac{3}{5}\frac{A_1i}{R^2}-\frac{3C}{R^4}\right\}\right]P_3(\cos\theta)$$

$$\underline{v}_0:\ \frac{i}{k^2}\nabla s_0\to A_1i\nabla\frac{P_1(\cos\theta)}{R^2}$$

$$\underline{v}_1:\ -\frac{\beta}{2k^2}\left\{\nabla\left(x\frac{\partial s_0}{\partial x}\right)-\frac{2}{h^2-k^2}\nabla\frac{\partial^2 s_0}{\partial x^2}+2\nabla s_0\right.$$
$$\left.-\frac{2h^2}{h^2-k^2}\frac{\partial s_0}{\partial x}i-A\nabla(xf_0(kR))\right\},$$
$$+\frac{i}{k^2}\nabla s_1^c\to i\nabla\left[D_1\frac{P_1(\cos\theta)}{R^2}+D_2\frac{P_3(\cos\theta)}{R^4}\right],$$
$$+3i\beta C\nabla\frac{P_3(\cos\theta)}{R^4}+\frac{\beta}{h^2}\nabla\frac{\partial^2 p_0}{\partial x^2}$$
$$-\frac{\beta}{10}\frac{\partial}{\partial x}\left[\frac{1}{R^3}\nabla\left(R^5\frac{\partial p_0}{\partial x}\right)-R^2\nabla\frac{\partial p_0}{\partial x}\right]$$

$$\underline{H}_1:\ \frac{i}{k^2}s_0i+\frac{i}{k^4}\nabla\frac{\partial s_0}{\partial x}\to\frac{A_1i}{5}\left[\frac{1}{R^3}\nabla(R^2P_2(\cos\theta))-\right.$$
$$\left.-R^2\nabla\frac{P_2(\cos\theta)}{R^3}\right] \quad (40)$$

where

$$A_1 = -ih^2 B f_2(h),$$

$$D_1 = -\frac{\beta B h^2}{10 f_0(h)}[3f_1^2(h) - f_0(h) f_2(h)],$$

$$D_2 = -\frac{3}{5}\beta B\left[f_0(h) - 5f_1(h) + h^2 \cdot \frac{f_1(h) f_3(h)}{f_2(h)}\right].$$

Next, we can write down the total drag experienced by the sphere*

$$X = -\pi u_s\left[6\left\{\frac{1}{R_e} + \frac{1}{\sqrt{2R_e}}\right\} + \frac{\beta R_m}{5}\left(2 + \frac{3}{\sqrt{2R_e}}\right)\right] - \pi\frac{du_s}{dt}\left[2\left\{\frac{1}{3} + \frac{3}{\sqrt{2R_e}}\right\} - \frac{3\beta R_m}{5\sqrt{2R_e}}\right]. \quad (41)$$

The terms in the curly brackets represent the non-magnetic effect, while the last term in each square bracket represents a hydromagnetic effect. The first square bracket gives a frictional force varying as the velocity: the first term is the Stokes drag for the classical problem in which the sphere undergoes motion of translation; the second term is the oscillatory effect in non-magnetic case; the third term is hydromagnetic, non-viscous, oscillatory effect and the last term is hydromagnetic viscous effect due to oscillation. The magnetic field tends to increase the resistance.

The second square bracket gives the correction to the inertia of the sphere. This amounts to the fraction

$$\frac{1}{2} + \frac{9}{2\sqrt{2R_e}}\left(1 - \frac{\beta R_m}{10}\right)$$

of the mass of fluid displaced, instead of ½ as in the frictionless, non-magnetic case. The effect of the magnetic field is to decrease the apparent mass of the fluid.

Although the derivations strictly do not hold for this case, we may consider the value $R_e \to \infty$ while at the same time keeping $R_e \ll 1$. We then find

$$X = -2\pi\left[\frac{1}{3}\frac{du_s}{dt} + \frac{1}{5}\beta R_m u_s\right],$$

a result which can be checked in a straightforward manner from the corresponding hydromagnetic inviscid problem. The non-magnetic part provides the inertia effect and the magnetic part contributes to the frictional part of the drag.

When the period $\frac{2\pi}{\lambda}$ is made infinitely long, the drag reduces to the Stokes drag in the classical problem in which the sphere moves uniformly in a straight line. This is understandable because as $\lambda \to 0$, the uniform magnetic field H^0 should also approach zero in view of

$$\beta = \frac{\mu H^{0^2}}{\rho^0 a^2 \lambda^2} = 0(1)$$

and so the magnetic effect is absent in this limit.

V. DISCUSSION

First we will discuss the conditions under which our analysis can be expected to give a close approximation to the true flow.

We have neglected the displacement current in comparison with the conduction current, which is justified if

$$\frac{a^2\lambda^2}{L^2} \ll 1, \quad (42)$$

which follows from the order of magnitudes of the various terms in the Maxwell equation (6, *i*) where L is the velocity of light. This is satisfied by all hydromagnetic flow problems in the laboratory. An associated condition, that the excess charge may be neglected, is also satisfied if (42) holds.

As in the classical problem, the convective terms have been neglected. This is justified if† $\alpha \ll a$ (43)

(where α is the amplitude of the oscillation and a is the radius of the sphere), provided

$$R_e = \frac{a^2\lambda}{\nu} = 0(1), \quad (44)$$

which follows from (15), remembering that $s \to 0$, $c^2 s \to p$ in the Incompressible case. For common

*There will be no contribution of the Maxwell stress to the order we are interested in, as there are no currents on the surface of the sphere:

$$\underline{J} = R_m[\underline{E}_0 + \underline{v}_0 \times \underline{H}_0]_{s=1} = 0.$$

†This condition justifies the omission of convective terms both in zeroth order and first order approximations as can be seen by the orders of magnitudes of various terms involved in the momentum equation.

liquids in hydromagnetic experiments, $\nu = 0(10^{-6})$ in m.k.s. units and hence we must make

$$a^2\lambda = 0(10^{-6}). \tag{44a}$$

Next, in the zeroth order approximation of the momentum equation, we neglected the Lorentz force $\underline{J}\times\underline{B} = \beta$ curl $\underline{H}\times\underline{H}$. This is justified if

$$\beta = \frac{\mu H^{0^2}}{\rho^0 a^2 \lambda^2} = 0(1) \tag{45}$$

and

$$R_m = a^2\lambda\mu\sigma \ll 1. \tag{46}$$

In the case of liquid Sodium and Mercury, for example, $\mu\sigma = 0(1)$ and so (46) follows in view of (44a). Condition (45) implies that magnetic pressure is at most comparable to dynamic pressure, which holds for weak field, say 50 gauss or less. In short, our analysis in valid for the oscillations of small globules in the presence of weak magnetic fields.

The following table gives some typical values of these parameters for Mercury, Liquid Sodium and Saturated Salt Water at 25°C when

$$a = .05 \text{ cm}, \quad B^0 = 25 \text{ gauss}, \quad \lambda = 25.$$

	Liquid Sodium	Saturated Salt Water At 25° C	Mercury
R_e	3	6	60
β	25	25	2
R_m	11×10^{-5}	2×10^{-10}	85×10^{-5}

Finally, to study the nature and properties of the solution, it will be convenient to introduce the stream functions Ψ and Φ (where

$$v_z = -\frac{1}{r}\frac{\partial\Psi}{\partial r}, \quad v_r = \frac{1}{r}\frac{\partial\Psi}{\partial x}, \quad H_z = -\frac{1}{r}\frac{\partial\Phi}{\partial r}, \quad H_r = \frac{1}{r}\frac{\partial\Phi}{\partial x}.$$

For the sake of brevity, we write the results here in a shorter form:

$$\Psi = e^{it}\sin^2\theta\left[\frac{a_1(R_e)}{R} + a_2(R_e, R)e^{-(1+i)\sqrt{Re/2}(R-1)}\right]^* + e^{it}\sin^2\theta\left[\frac{a_3(R_e)}{R} + \frac{a_4(R_e)}{R^3} + a_5(R_e, R)e^{-(1+i)\sqrt{Re/2}(R-1)}\right]$$
$$+ e^{it}\sin^2\theta\cos^2\theta\left[\frac{a_6(R_e)}{R} + \frac{a_7(R_e)}{R_3} + a_8(R_e, R)e^{-(1+i)\sqrt{Re/2}(R-1)}\right] \tag{47}$$

(where the terms under the bar represent the non-magnetic part);

$$\Phi = e^{it}\cos\theta\sin^2\theta\left[b_1(R_e) + \frac{b_2(R_e)}{R^2} + b_3(R_e, R)e^{-(1+i)\sqrt{Re/2}(R-1)}\right] \tag{48}$$

As $R_e \to \infty$, $a_1 \to -\frac{1}{2}\frac{\alpha}{a}$, $a_3 \to -\frac{3}{20}i\beta R_m\frac{\alpha}{a}$, $a_4 \to \frac{3}{20}i\beta R_m\frac{\alpha}{a}$, $a_6 \to \frac{3}{4}i\beta R_m\frac{\alpha}{a}$, $a_7 \to -\frac{3}{4}i\beta R_m\frac{\alpha}{a}$, $b_1 \to \frac{1}{4}R_m\frac{\alpha}{a}$

and

$$b_2 \to -\frac{3}{20}R_m\frac{\alpha}{a}. \tag{49}$$

If Φ_s is the corresponding function for the magnetic lines of force inside the sphere,

$$\Phi_s = \frac{1}{10}\frac{\alpha}{a}R_m R^3\cos\theta\sin^2\theta\, e^{it}, \tag{50}$$

which is current-free and hence up to this approximation, the sphere acts like an insulator.

From (47) and (48) it follows that there are three types of terms in the solution:

(1) *Wave-like* (corresponding to the appearance of a boundary layer), effective in a small distance from the surface of the sphere and behaving like $\exp[-(1+i)\sqrt{R_e/2}(R-1)]$. The nature of boundary layer solution is essential the same in hydromagnetics as in a non-magnetic case. The term $\exp[-(1+i)\sqrt{R_e/2}(R-1)]$ represents a

*The wave-like term $\exp[-(1+i)\sqrt{R_e/2}(R-1)]$ comes from the functions $f_n(hR)$ defined by (28).

wave of vibrations propagated from the boundary of the sphere with the velocity $\sqrt{2/R_e}$, but with rapidly diminishing amplitude, the falling off within a wave length being in the ratio $e^{-2\pi}$, or 1/535. The linear magnitude $\sqrt{2/R_e}$, important in all problems of oscillations in incompressible viscous fluids, indicates the extent to which the effects of viscosity penetrate into the fluid. In the case of Mercury its value is .018 $P^{\frac{1}{2}}$ cms, where P is the period of oscillation in seconds. For liquid Sodium, the corresponding value is .079 $P^{\frac{1}{2}}$.

(2) *Irrotational*

$$\left(\frac{\sin^2\theta}{R}, \frac{\cos\theta \sin^2\theta}{R^2}\right), \text{ and}$$

(3) *Rotational*

$$\left(\frac{\sin^2\theta}{R^3}, \frac{\sin^2\theta \cos^2\theta}{R}, \frac{\sin^2\theta \cos^2\theta}{R^3}, \text{in}^2\theta \cos\theta\right).$$

In the non-magnetic case, the motion in the flow field is essentially irrotational at large distances, and it is an interesting result that the magnetic field generates vorticity.

In the case of a magnetic field, the term $\frac{\cos\theta \sin^2\theta}{R^2}$ is current free, while the term $\cos\theta\sin^2\theta$ is responsible for currents at large distances.

From (49) it follows that for large distances, the disturbance in the present problem differs only in amplitude and phase from the one generated by the oscillation of a sphere in a frictionless fluid in the presence of a magnetic field.

The diagrams 1, 2 and 3 illustrate the effects of various parameters on the flow fields. Figure 1 shows how the drag on the sphere varies as it

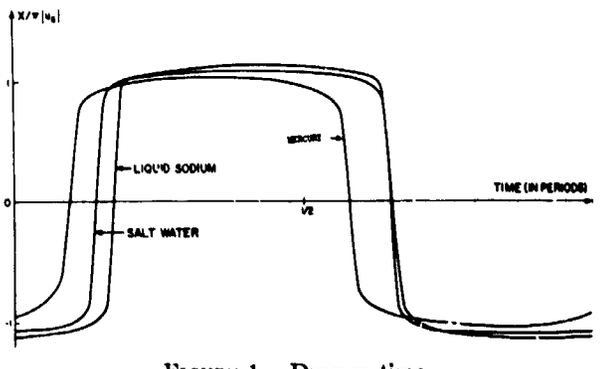


FIGURE 1.—Drag vs time.

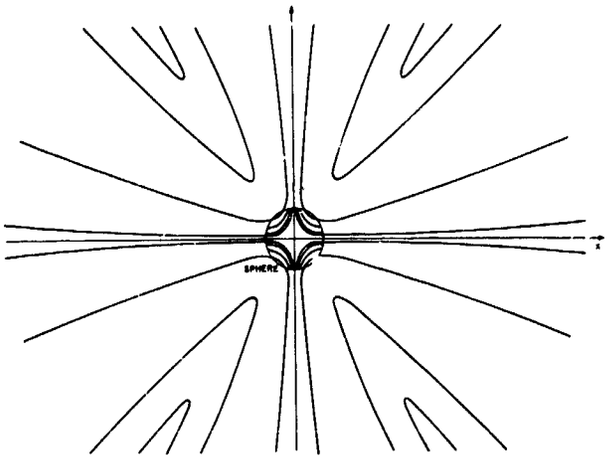


FIGURE 2(a).—Magnetic lines of force (perturbed) at .95th of the oscillation (mercury).

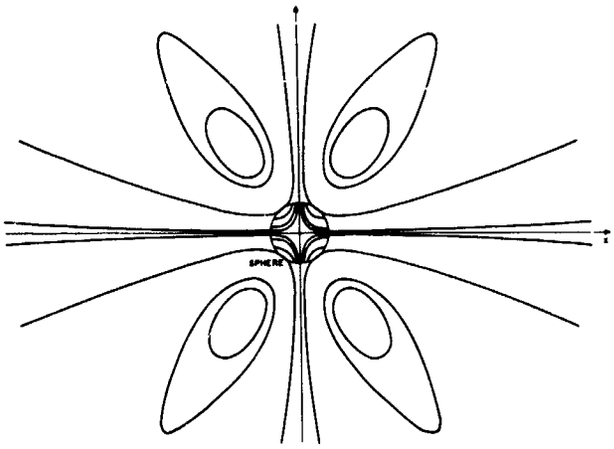


FIGURE 2(b).—Magnetic lines of force (perturbed) at .95th of the oscillation (liquid sodium.)

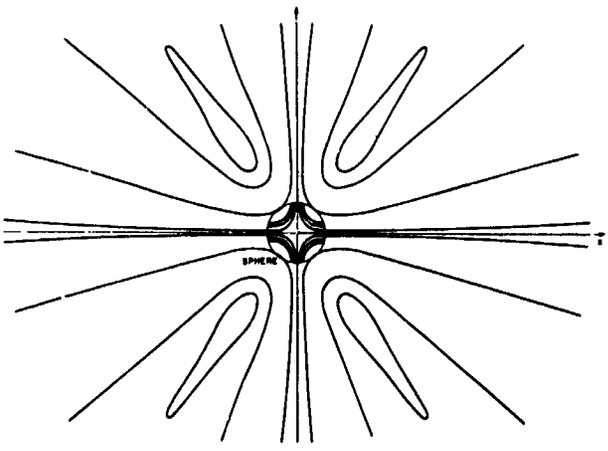


FIGURE 2(c).—Magnetic lines of force (perturbed) at .95th of the oscillation (salt water).

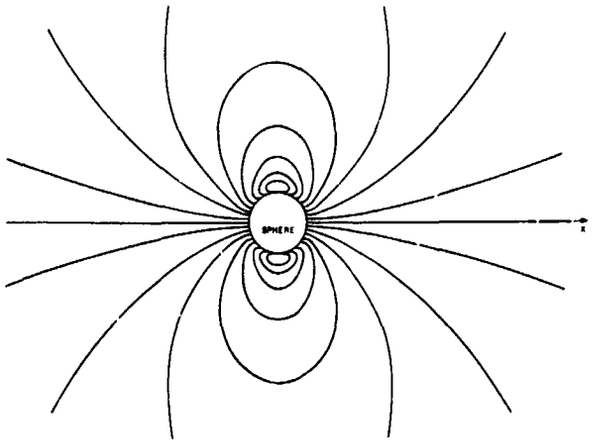


FIGURE 3(a).—Streamlines pattern (non-magnetic case) at the completion of an oscillation (mercury).

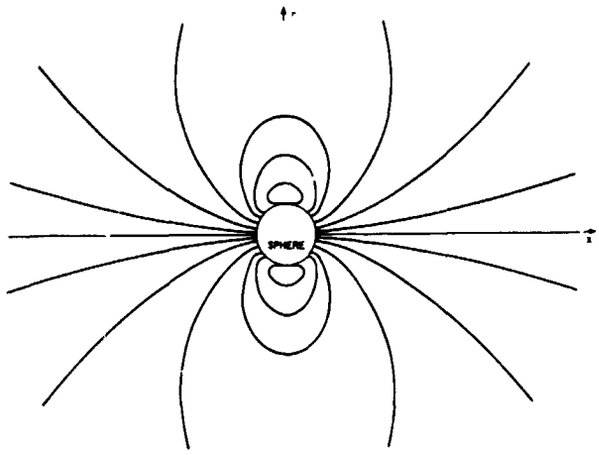


FIGURE 3(b).—Streamlines pattern (magnetic case: $B^\circ = 25$ gauss) at the completion of an oscillation (mercury).

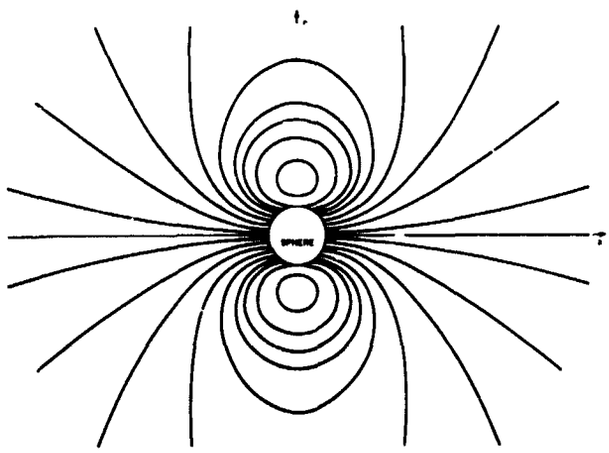


FIGURE 3(c).—Streamlines pattern (non-magnetic case) at the completion of an oscillation (liquid sodium).

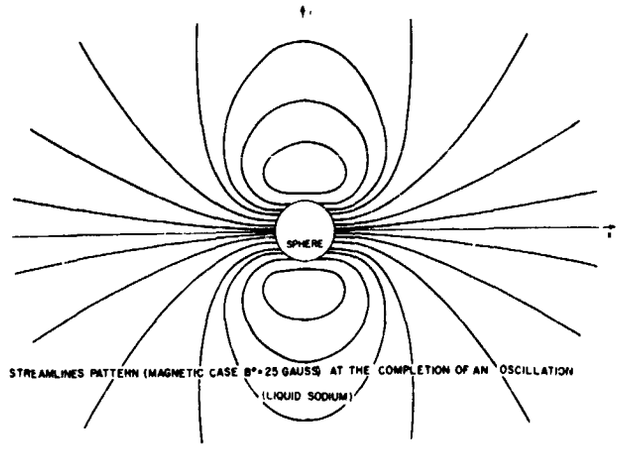


FIGURE 3(d).—Streamlines pattern (magnetic case: $B^\circ = 25$ gauss) at the completion of an oscillation (liquid sodium).

oscillates. It also shows the influence of conductivity and viscosity on the drag. Figure 2 shows the effect of Reynolds number R_e on magnetic lines of force. The viscosity tends to contract the lines of force. On the other hand, Figure 3 shows the stretching effect of the magnetic field on the stream lines (compare Figure 3 (*a*) and (*b*); (*c*) and (*d*)). Viscosity also tends to produce similar effects on the flow (compare Figure 3 (*a*) and (*c*); (*b*) and (*d*)).

APPENDIX

As pointed out earlier, the only effect of rotation in this problem is that azimuthal components of $\underline{v}$ and $\underline{H}$ are affected in that they are non-zero here in contrast to the non-rotating case in which v_φ and H_φ are zero.

To determine $v_{0\varphi}$, we use (21) which gives

$$\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}-\frac{1}{r^2}+h^2\right]v_{0\varphi}=\delta R_e v_{0r} \tag{A1}$$

and the boundary conditions are

$$v_{0\varphi}=0 \text{ at } R=1 \text{ and at infinity.} \tag{A2}$$

The solution is (using (26))

$$v_{0\varphi}=B_1\times rf_2(hR)+\frac{\delta R_e}{h^2-k^2}\frac{i}{k^2}\frac{\partial s_0}{\partial_r}-\frac{3i\delta}{2}Bh^2\times rf_1(hR) \tag{A3}$$

where the first term is the complementary function and the remaining terms are a particular

integral. The constant B_1 can be chosen in such a way that the boundary condition $v_{0\varphi}=0$ at $R=1$ is satisfied.

Next, $H_{1\varphi}$ can be determined by

$$\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}-\frac{1}{r^2}\right]H_{1\varphi}=-\frac{\partial v_{0\varphi}}{\partial x}, \tag{A4}$$

while inside the sphere

$$\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}-\frac{1}{r^2}\right]H_{1\varphi}=0 \tag{A5}$$

and we require the continuity of $H_{1\varphi}$ across the surface of the sphere.

The solution is

$$H_{1\varphi}=B_2\frac{P_1^1(\cos\theta)}{R^2}+B_3\frac{P_3^1(\cos\theta)}{R^4}+\frac{\delta R_e i}{k^4(h^2-k^2)}\cdot\frac{\partial^2 s_0}{\partial\times\partial r}$$
$$-\frac{3i\delta B}{2}\left[\frac{2}{h^2}\frac{\partial^2}{\partial x^2}+x\frac{\partial}{\partial x}+1\right]rf_1(hR)+$$
$$+\frac{B_1}{h^2}\frac{\partial}{\partial x}[xrf_2(hR)], \tag{A6}$$

and inside the sphere

$$H_{1\varphi}=C_1R\cdot P_1^1(\cos\theta)+C_2R^3\cdot P_3^1(\cos\theta). \tag{A7}$$

The first two terms in (A6) are the complementary function and the remaining terms are a particular integral. The constants B_2, B_3, C_1 and C_2 are determined by the requirement that $H_{1\varphi}$ is continuous at $R=1$.

Finally, $v_{1\varphi}$ is determined by

$$\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}-\frac{1}{r^2}+h^2\right]v_{1\varphi}=\beta R_e v_{0\varphi}+\delta R_e v_{1r} \tag{A8}$$

and the boundary conditions $v_{1\varphi}=0$ at $R=1$ and at infinity. (A9)

The solution is

$$v_{1\varphi}=G_1R^2\cdot P_2^1(\cos\theta)f_2(hR)+G_2R^4\cdot P_4^1(\cos\theta)f_4(hR)+$$
$$+\delta R_e\left[-\frac{\beta}{2k^2}\left\{\frac{4h^2-2k^2}{(h^2-k^2)^2}\frac{\partial s_0}{\partial r}+\frac{x}{h^2-k^2}\frac{\partial^2 s_0}{\partial x\partial r}-\right.\right.$$
$$-\frac{4}{(h^2-k^2)^2}\frac{\partial^3 s_0}{\partial r\partial x^2}-\frac{A}{h^2-k^2}\left(\times\frac{\partial f_0(kR)}{\partial r}-\right.$$
$$\left.\left.-\frac{2}{h^2-k^2}\frac{\partial^2 f_0(kR)}{\partial r\partial x}\right)\right\}-\frac{3}{8}B\beta i\times r\left\{f_1(hR)-\right.$$
$$\left.-f_0(hR)+h^2x^2f_2(hR)\right\}+\frac{i}{k^2(h^2-k^2)}\frac{\partial s_1^c}{\partial r}-$$
$$-\frac{3}{2h^2}(h^2F_1-4F_2)xrf_1(hR)+$$
$$+\frac{15}{4}F_2xrf_3(hR)(-4x^2+3r^2)\Big]-\frac{B_1\beta_i}{2}xrf_1(hR) \tag{A10}$$

where the first two terms are the complementary function and the remaining are a particular integral. The constants G_1 and G_2 are determined by the requirement that $v_{1\varphi}=0$ at $R=1$.

It is clear from the above solution that there will be no contribution of Maxwell and Viscous stresses (produced due to rotation) on the drag—a result in agreement with our previous conclusions (see Singh, 1964 [12]).

Since the wave-like terms $f_n(\zeta)$ are damped heavily, it follows that in non-rotating compressible case, the flow will be essentially current free and irrotational at large distances. Our solution above shows that the effect of rotation will be to produce vorticity and current which again agrees with our conclusions in (12).

Concluding, it should be noted that the nature of wave propagation remains essentially similar in both rotating and non-rotating cases.

ACKNOWLEDGMENTS

The author is highly indebted to Prof. J. M. Burgers for going through the manuscript and for various helpful suggestions. This work was performed while the author held a NASA Resident Research Associateship of the National Academy of Sciences—National Research Council.

REFERENCES

1. R. Hide, "Hydrodynamics of the Earth's Core," in *Physics and Chemistry of the Earth*, L. H. Ahrens, K. Rankama, and S. K. Runcorn, Editors (Pergamon Press, New York, 1956), Vol. 1.
2. R. Hide and P. H. Roberts, "The Origin of the Earth's Magnetic Field," in *Physics and Chemistry of the Earth*, L. H. Ahrens, K. Rankama, and S. K. Runcorn, Editors (Pergamon Press, New York 1960), Vol. 4; ———, *Rev. Mod. Phys.*, *32* (1960), 799.

3. B. Lehnert, *Astrophys. J., 119* (1954), 647; ———, *Astrophys. J., 121* (1955), 481.
4. S. Chandrasekhar, *Proc. Roy. Soc. (London), A225* (1954), 173; ———, *Proc. Roy. Soc. (London), A237;* (1956), 476; ———, *Proc. Am. Acad. Arts Sci. U. S., 86* (1957), 372.
5. S. P. Talwar, *Jour. Fluid Mech., 9* (1960), 581; ———, NASA X-640-64-39 (1964).
6. J. N. Tandon and S. P. Talwar, *Nuclear Fusion, 3* (1963), 75.
7. S. D. Nigam, and P. D. Nigam, *Proc. Roy. Soc. (London), A272* (1963), 529.
8. G. I. Taylor, *Proc. Roy. Soc. (London), A102* (1922), 180.
9. H. Lamb, *Hydrodynamics*, New York, Dover Publ. (1932).
10. G. S. S. Ludford, *Rev. Mod. Phys., 32* (1960), 1000; ———, *ZAMM, 43* (1963), 9.
11. K. Tamada, *Phys. Fluids, 5* (1962), 817.
12. M. P. Singh, NASA X-640-64-142, (1964).

ON OSCILLATING SPHERE IN A ROTATING VISCOUS FLUID*

M. P. SINGH†

The moderate and large-scale motions of the atmosphere are greatly influenced by the vorticity of the earth's rotation. In the case of an infinite liquid, rotating as a rigid body about an axis, the amount of energy possessed by the liquid is infinite and it is of great interest to know how small disturbances propagate in such a liquid. To understand some of these phenomena, it will be interesting to study the flow of a rotating fluid around elementary bodies.

Here we consider small oscillations of a sphere in an incompressible, viscous fluid rotating with a constant angular velocity. For flow in the absence of rotation, the problem was first considered by Stokes [Lamb, 1932] on the assumption that the Reynolds number $R_s = a|V_s|/\gamma$ was negligibly small (so as to discard convective terms), where V_s is the velocity of the sphere, a is the radius of the sphere and γ the kinematic viscosity. It is the purpose of the present investigation to determine the effect of the coriolis forces on the flow pattern.

Let the sphere oscillate along the axis of the rotating fluid which is taken to be the x-axis. The origin is at the mean position of the center of the sphere. The equations of motion of an incompressible, viscous, unbounded liquid rotating about x-axis with a constant angular velocity $\Omega\underline{i}$ referred to a rotating frame of reference are

$$\frac{\partial \underline{v}}{\partial t} + \underline{v}\cdot\nabla\underline{v} + 2\underline{\Omega}\times\underline{v} + \underline{\Omega}\times(\underline{\Omega}\times\underline{r}) = -\frac{1}{\rho}\nabla p + \gamma\nabla^2\underline{v} \quad (1)$$

where ρ is the density, $\underline{v}$ the velocity of the fluid, p is the pressure and r is the radian coordinate:

$$r^2 = y^2 + z^2.$$

The equation of continuity is

$$\operatorname{div}\underline{v} = 0.$$

As in the classical problem, the convective terms in (1) will henceforth be neglected. This is justified if

$$R_s = a|V_s|/\gamma << 1.$$

The equations (1) and (2) will be reduced to a non-dimensional form by referring the length to a, the radius of the sphere, the velocity to $a\lambda$ ($\lambda/2\pi$ is the frequency of the oscillation), the time to $1/\lambda$ and the pressure to $\rho a^2\lambda^2$. Then we have

$$\frac{\partial \underline{v}}{\partial t} = -\nabla p + \delta\underline{v}\times\underline{i} + \frac{1}{R_e}\nabla^2\underline{v} \quad (4)$$

and

$$\operatorname{div}\underline{v} = 0 \quad (5)$$

where

$$R_e(\text{Reynolds number})\ddagger = a^2\lambda/\gamma \quad (6)$$

and

$\delta = 2\Omega/\gamma$, the inverse of the so-called Rossby number

$$R_0 = \frac{\lambda}{2\Omega} = \frac{1}{\delta} \quad (7)$$

and the centrifugal force $-\delta^2\nabla r^2/8$ has been absorbed in the pressure term.

*Published as *Goddard Space Flight Center Document X-640-64-143*, May 1964.

†Dr. Singh is a National Academy of Sciences—National Research Council Postdoctoral Resident Research Associate.

‡It should be noted that we have used $a\lambda$ as velocity in the definition of the Reynolds number. There still is another important dimensionless parameter ϵ/a, where ϵ is the amplitude of the oscillatory motion of the sphere. It is assumed that this quantity is small compared to unity. The velocity of sphere is $V_s\underline{i} = \epsilon\lambda e^{\lambda t}\underline{i}$, or in dimensionless form: $U_s\underline{i} = \epsilon/a\, e^{it}\underline{i}$. The Reynolds number for the flow might have also been defined as $R_e = a\epsilon\lambda/\gamma$ ($= R_e\,\epsilon/a$). It is necessary that $R_e << 1$, whereas the requirement for R_e is less severe.

The boundary conditions on $\underline{v}$ are that $\underline{v}=U_s\underline{i}$ at the sphere $R=1$

(where $R^2=r^2+x^2$) and $\underline{v}=0$ at infinity. (8)

As in non-rotating case, we have here axi-symmetry in that the physical variables are independent of the azimuthal coordinate*Φ. However, the azimuthal component of velocity v_Φ is non-zero here in contrast to the non-rotating case in which v_Φ is zero.

We next make an important assumption in that the Rossby number R_0 is large so as to neglect second or higher powers of δ. (9)

From (4) and (5) it follows that $v_x=0(1)$, $v_r=0(1)$ and $v_\Phi=0(\delta)$, hence

$$\nabla^2 p=0(\delta^2). \tag{10}$$

It follows that the coupled equations (4) can be simplified in this case. In fact, in view of (9) and (10), v_x and v_r satisfy identically the same system of equations and boundary conditions as in non-rotating case. The solution is therefore [Lamb, 1932]

$$p=A\frac{P_1(\cos\theta)}{R^2} \tag{11}$$

and

$$\underline{v}=Ai\nabla\frac{P_1(\cos\theta)}{R^2}+2Bf_0(hR)\nabla(RP_1(\cos\theta))$$
$$-Bf_2(hR)h^2R^5\nabla\frac{P_1(\cos\theta)}{R^2} \tag{11}$$

where (R, θ, Φ) are the spherical polar coordinates† and

$$h^2=-iR_\epsilon. \tag{12}$$

The function $f_n(\xi)$, defined as

$$f_n(\xi)=\left(-\frac{1}{\xi}\frac{d}{d\xi}\right)^n\frac{e^{-i\xi}}{\xi} \tag{13}$$

essentially represents boundary layer effects, while the constants A and B are given by

$$A=-ih^2f_2(h)\,B,\quad B=U_s/2f_0(h). \tag{14}$$

Finally, v_Φ remains to be determined, which satisfies

$$\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}-\frac{1}{r^2}+h^2\right]v_\Phi=\delta R_\epsilon v_r \tag{15}$$

and the boundary conditions

$$v_\Phi=0 \text{ at } R=1 \text{ and at infinity.} \tag{16}$$

The solution of (15) and (16) is

$$v_\Phi=-\frac{3}{2}\delta BR_\epsilon xrf_1(hR)+3\delta A\frac{xr}{R^5}+cxrf_2(hR) \tag{17}$$

where

$$c=3B\delta R_\epsilon\left[1+\frac{f_1(h)}{2f_2(h)}\right]. \tag{18}$$

To summarize, the following conclusions are noteworthy:

(i) The azimuthal component of velocity is non-zero in contrast to non-rotating case in which $v_\Phi=0$.

(ii) The rotating effect generates vorticity in that for large distances, the flow pattern is irrotational in non-rotating case. In rotating case, the azimuthal component of velocity is responsible for vorticity for large R.

(iii) The wave-like terms $f_n(hR)$ (corresponding to a boundary layer) appearing in the solution, effective in a small distance from the surface of the sphere and behaving like $\exp[-(1+i)\sqrt{R_\epsilon/2}(R-1)]$ are essentially similar in nature in rotating and non-rotating cases.

(iv) The drag on the sphere remains unchanged (due to rotation) to the order δ since v_x, v_r and p remain unchanged to that order.

In a forthcoming paper, the above discussion will be extended to a compressible fluid in the presence of a magnetic field.

ACKNOWLEDGMENTS

The author would like to thank Prof. S. I. Pai and Dr. S. P. Talwar for various helpful suggestions and many fruitful discussions. This work was done while the author held a Resident Research Associateship of the National Academy of Sciences—National Research Council.

REFERENCE

LAMB, H., *Hydrodynamics*, Dover Publ., New York (1932) P. 643.

*(r, Φ, x) are the cylindrical coordinates.
†$z=R\cos\theta$, $r=R\sin\theta$.

See N65-24112

VARIATIONS OF GEOMAGNETIC ACTIVITY WITH LUNAR PHASE*

HAROLD L. STOLOV† AND A. G. W. CAMERON

An analysis of 31 years of K_p data suggests a variation of geomagnetic disturbance with lunar phase. A general increase in geomagnetic activity of about 4 percent begins after full moon and lasts for seven days. A general decrease in geomagnetic activity of about 4 percent is found for the seven days preceding full moon. A study of randomized data indicates that the probability that these variations should have occurred by chance is less than 5 percent. The effect is found to be associated with the K_p data derived from periods of geomagnetic quiet conditions; it is not evident in the data from disturbed periods.

INTRODUCTION

Reports of lunar influence on a number of geophysical variables have appeared in the literature. *Bradley et al.* [1962] and *Adderley and Bowen* [1962] claim to find a lunar influence on rainfall. *Bigg* (1963*a*] presents evidence for a lunar modulation of freezing nucleus concentrations measured in the lower atmosphere. *Bowen* [1963] reports variations in the incoming meteor rate with lunar phase, and *Adderley* [1963] claims a lunar effect on the amount of atmospheric ozone. *Dorman and Shatashvili* [1961] present evidence for a lunar diurnal variation in the intensity of the neutron component of cosmic radiation dependent on lunar phase. Acceptable physical mechanisms for the various lunar modulation effects are not easy to find.

The search for a better understanding of these surprising lunar influences led *Bigg* [1963*b*, *c*] to re-examine the earlier work of *Sucksdorff* [1956] on influence of the moon on geomagnetic disturbances. Several confusing and contradictory claims have been made in previous lunar-geomagnetic activity studies:

1. That there is a distinct decrease of geomagnetic activity, averaging 12 percent during new moon, and a less marked decrease of 7 percent during full moon. Geomagnetic activity is 7 percent higher than average during first quarter [*Sucksdorff*, 1956].

2. That there is a minimum of geomagnetic disturbance at new moon, 20 percent lower than any other minimum on the curve, and a broad maximum of geomagnetic activity shortly after full moon [*Bigg*, 1963*b*].

3. That there is a decrease in the frequency of magnetic storms at new moon and a tendency for magnetic storms to occur preferentially near first and third quarters [*Bigg*, 1963*b*].

4. That the decrease in the frequency of magnetic storms at new moon applies only to the gradual-commencement type of storm. The sudden-commencement storms show an increased frequency at new moon [*Bigg*, 1963*c*].

Bartels [1963] shows that the conclusions stated in paragraph 3 are questionable, since randomizing the data yields variations of size comparable to those in the lunar data.

It is the purpose of this paper to determine, by further investigation, whether genuine influences of the moon on geomagnetic activity may be present. It is suggested herein that correlations with lunar phase do exist in the geomagnetic data for the last 31 years; they may prove valuable in seeking to understand the solar wind-magnetosphere interactions. After the present study was completed another study was published [*Bell and Defouw*, 1964] in which similar conclusions were reached. Since the present method of analysis is quite different from that of Bell and Defouw, it is thought to be of interest to present it also.

*Published in *Journal of Geophysical Research*, 69(23):4975–4982, December 1, 1964.

†Goddard Space Flight Center and City College of the City University of New York.

PROCEDURE

There are now available K_p indices for more than 31 years, consisting of more than 90,000 well-determined values of standard quality. They represent a detailed history (8 values per day) of the disturbance variations of the geomagnetic field over nearly three sunspot cycles. The scientific community is indebted to the late Professor J. Bartels and his collaborators for their dedicated efforts in the compilation of this series. The K_p index is a world-wide measure of geomagnetic activity expressed on a quasi-logarithmic scale and based on the measurements at a fixed group of 12 observatories in upper middle geomagnetic latitudes. The K_p index is traditionally taken to be a measure of the velocity of the solar corpuscular stream. The Mariner 2 measurements of *Snyder et al.* [1963] establish with a considerable degree of certainty that the classical view is well taken.

Our search for a lunar influence on K_p began with the establishment of a suitable reference level with which to compare the magnetic activity at individual times during the lunar cycle. The mean value of K_p for each lunar synodic period was taken; this is called $\bar{K}_p$. The difference between each K_p value during the lunar month and $\bar{K}_p$ was found. This difference, divided by $\bar{K}_p$ and multiplied by 100, is the percent departure of magnetic disturbance from the reference level.

$$\% \text{ departure} = 100(K_p - \bar{K}_p)/\bar{K}_p$$

Percent departure was found for all the lunar synodic periods contained in the 31 years of K_p data. Mean values of the percent departures (8 per day) that were taken during the course of the lunar cycle are shown in Figure 1. Figure 1 is the output of a plotter where each vertical marker represents one of the mean values of per cent departure. Figure 1 shows the following features: (*a*) A broad maximum of geomagnetic activity ($\approx$4 percent) begins a half day after full moon and lasts nearly seven days. (*b*) A broad minimum of geomagnetic activity ($\approx$4 percent) exists for about seven days preceding full moon. (*c*) There is nothing at new moon but random statistical fluctuations.

These results are generally unexpected, in view of the emphasis given to the new moon position by previous investigators, who have usually suggested some type of 'shadowing' mechanism at new moon.

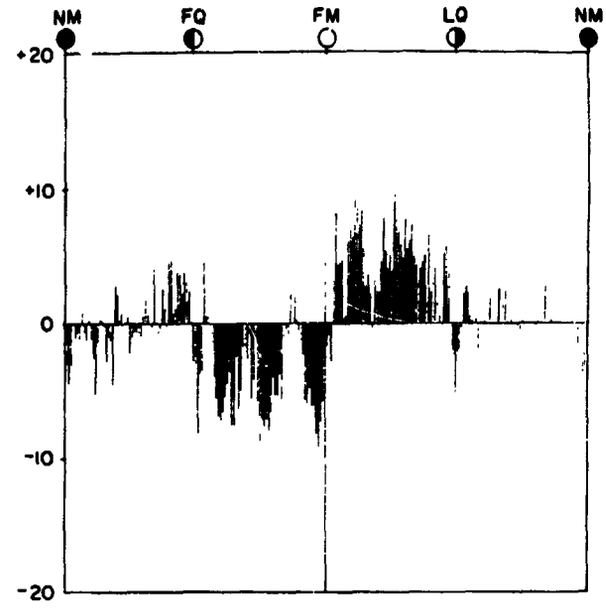


FIGURE 1.—Mean values of the percent departures of K_p as a function of lunar phase (31 years of K_p data).

Several additional studies were performed to examine the above conclusions.

1. To test the significance of the departures observed before and after full moon, and to determine if they might have originated in random fluctuations, a series of artificial lunar periods of 25, 26, 27, 28, 29, 30, 31, and 32 days were run through the same computer program using all the 31 years of K_p data in each case. In all these random data there were five cases when essentially clean departures of about 4 percent existed for periods of about three days, and three cases when departures of about 3 percent existed for periods of about four days. The normal departure lasts only one day or less. As examples of the random data, Figures 2 and 3 are presented for the 27-day and 29-day runs, respectively. Figures 2 and 3 show more than one random negative departure comparable to the meaningless departure at new moon in Figure 1. Many more were found in the runs that have not been reproduced. There is nothing like the seven-day departures of Figure 1 in any of this set of random data.

2. It is not clear what statistical laws we should expect the frequency distribution of departures from the null axis to follow. Accordingly,

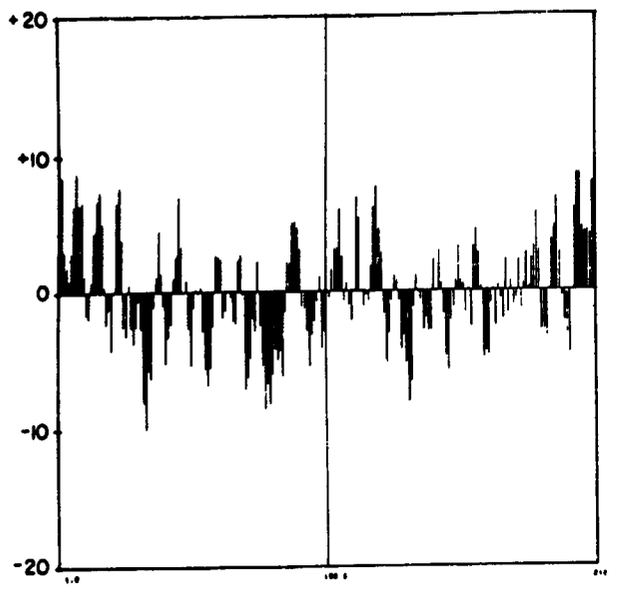


FIGURE 2.—Mean values of the percent departure of K_p for 27-day periods (31 years of K_p data).

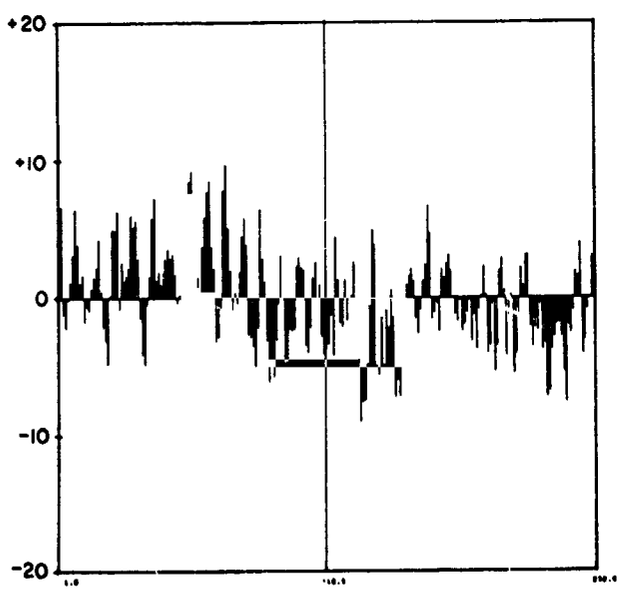


FIGURE 3.—Mean values of the percent departure of K_p for 29-day periods (31 years of K_p data.)

we have determined the distribution function empirically by a systematic randomization of the data used. This randomization was accomplished in the following manner. Each lunar month used here contains 241 three-hour intervals, 120 on each side of the three-hour interval during which full moon occurs. Each point of Figure 1 is an average performed separately for each three-hour interval in the superposition of lunar months. Therefore 120 fictitious superpositions of the lunar months were made by shifting the positions of the three-hour intervals of each lunar month, those pieces of data shifted off one end of the month being shifted in on the other end. In general, in the mth superposition the data of the nth lunar month were shifted $2mn$ positions. In this way 120 randomized diagrams equivalent to Figure 1 were obtained. Two studies were made of the properties of these diagrams.

Let us define an excursion length as the number of three-hour intervals adjacent to one another in which the departure from zero has the same sign (positive or negative). The number distribution of the excursion lengths was found by summing all those in the 120 randomized cases and dividing by 120; this distribution is shown by the histogram in Figure 4. Also potted in Figure 4 is the actual distribution of excursion lengths in Figure 1 (denoted by crosses). The probability that a single case will contain an excursion length equal to or

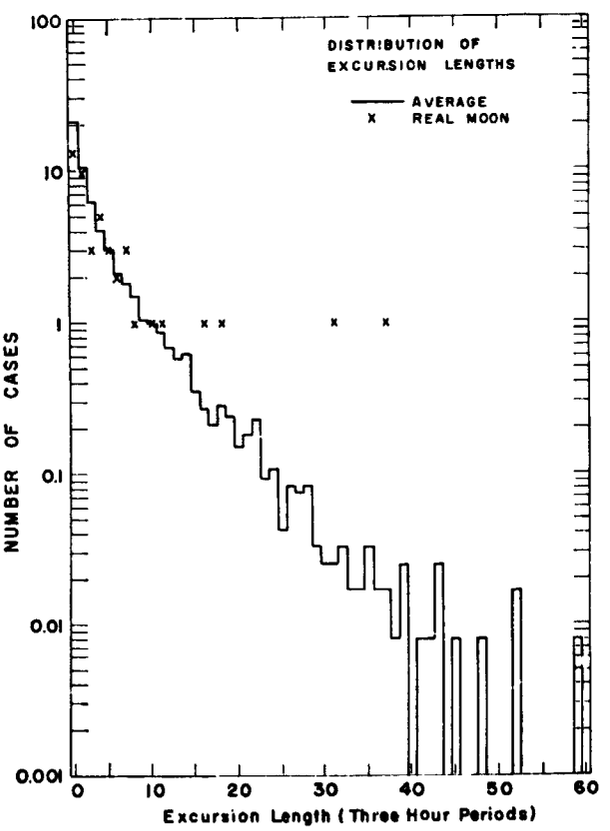


FIGURE 4.—Frequency distribution of excursion lengths calculated from the 120 averages of randomized lunar periods, compared to the distribution found for Figure 1.

greater than some value can be determined by summing the histogram upward from the point under consideration. Figure 1 contains two large excursion lengths. If the probabilities of their occurrence in the same case can be considered independent, then the probability of their occurring by chance turns our to be 0.035.

Let us define an excursion area as an excursion length multiplied by the average percent departure during the excursion. The distribution of excursion areas was calculated in the same way as described above for excursion lengths; the results are shown in Figure 5. Once again the crosses represent the data of Figure 1. If the two high points of the Figure 1 data have independent probabilities of occurrence, then the probability that a case like Figure 1 would arise by chance is 0.0074. Since it is possible that the presence of one large excursion area might slightly enhance the probability of occurrence of somewhat larger excursion areas associated with the excursions of the opposite sign, it is possible that the above value may slightly underestimate the true probability of occurrence of two large excursion areas in a single case.

In any event it seems safe to state that the probability of occurrence of Figure 1 by chance is less than 5 percent, the level of probability below which a distribution is usually considered to have statistical significance.

3. To examine further the effects observed before and after full moon, two separate runs of per cent departures were made, each for eight years of data only. The first was for the eight years of the quiet sun consisting of K_p indices from 1932; 1933, 1934, 1942 (last half only), 1943, 1944, 1953, 1954, and 1955 (first half only). The second was for the eight years of the disturbed sun consisting of K_p indices from 1937, 1938, 1939 (first half only), 1947, 1948, 1949, 1957, 1958, and 1959 (first half only). Figures 6 and 7 represent the results of this study. For the quiet sun (Figure 6), when fluctuations of K_p are expected to be small, the previously observed effects occur as found in Figure 1. For the disturbed sun (Figure 7), when random fluctuations of K_p are expected to be large, the effects about full moon exhibit a considerably more randomized pattern.

4. Instead of using the Zurich relative sunspot numbers to distinguish those periods of strong solar plasma flow from the quieter periods, the

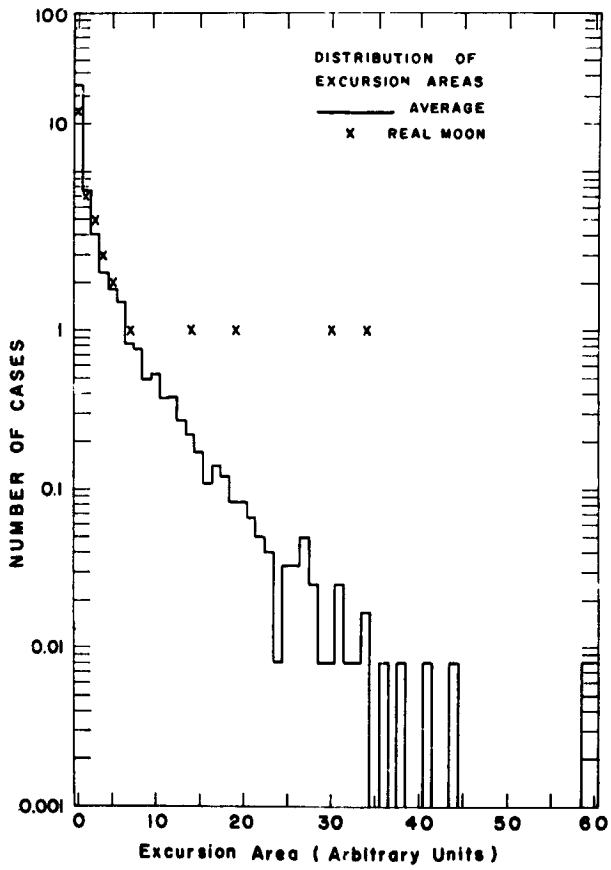


FIGURE 5.—Frequency distribution of excursion areas calculated from the 120 averages of randomized lunar periods, compared to the distribution found for Figure 1.

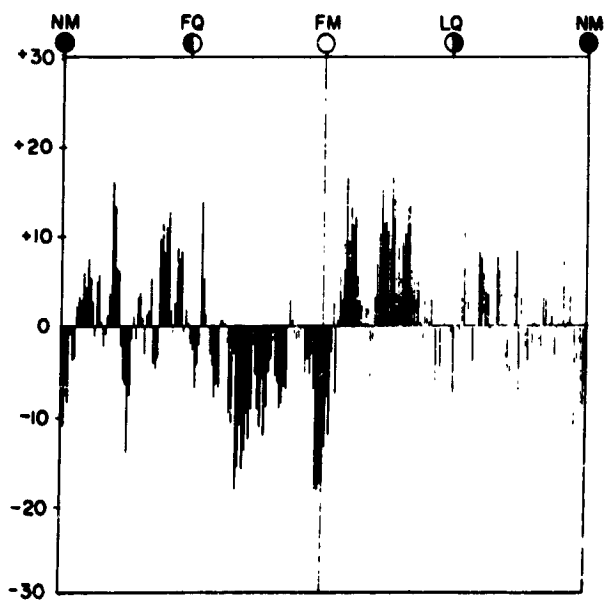


FIGURE 6.—Mean values of the percent departure of K_p as a function of lunar phase (8 quiet sun years only).

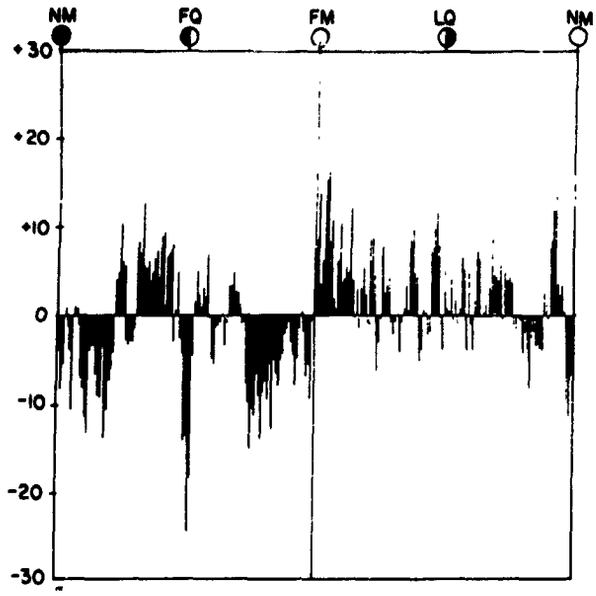


FIGURE 7.—Mean values of the percent departure of K_p as a function of lunar phase (8 disturbed sun years only).

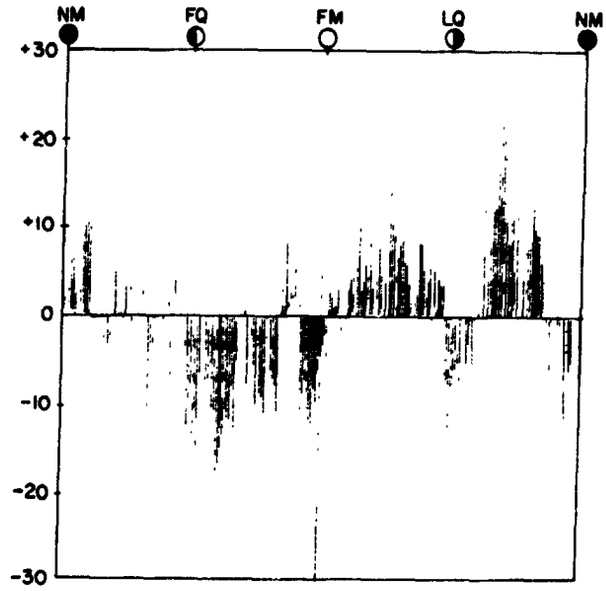


FIGURE 9.—Same as Figure 6 ($\bar{K}_p$ values of the next to the lowest quarter only).

$\bar{K}_p$ values themselves might be used. In this study the 382 lunar periods were divided into four groups on the basis of different levels of $\bar{K}_p$. All four groups were run through the same per cent departure analysis. The results for the lowest to the highest $\bar{K}_p$ groups are shown in Figures 8, 9, 10, and 11, respectively. In Figure 8 the random fluctuations of K_p are expected to be small, and the observed lunar effects of Figure 1 are seen to be

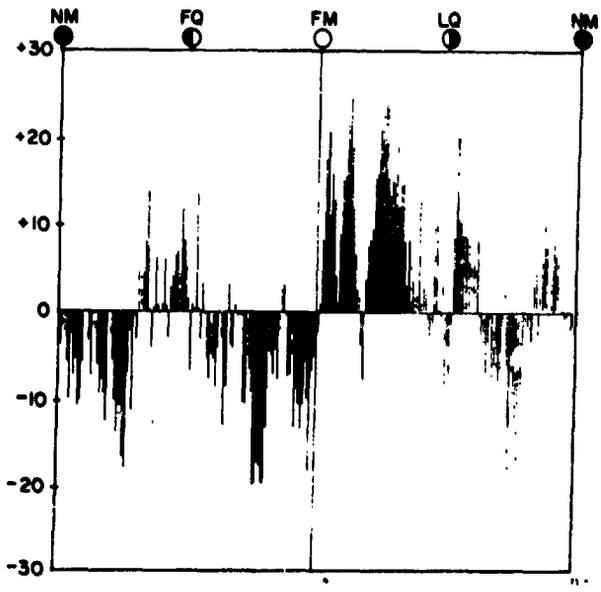


FIGURE 8.—Mean values of the percent departures of K_p as a function of lunar phase ($\bar{K}_p$ values of the lowest quarter only).

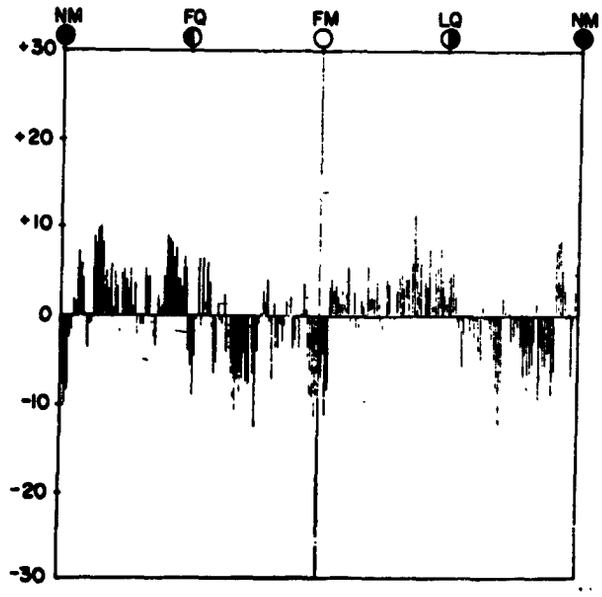


FIGURE 10.—Same as Figure 6 ($\bar{K}_p$ values of the next to the highest quarter only).

present. In Figure 11 the random fluctuations of K_p are expected to be large, and the effects at full moon are not evident. The figures for the two intermediate groups represent a gradation between Figures 8 and 11. The lunar periods assembled to form Figure 8 are not necessarily those of Figure 6, nor are those of Figure 11 the same as the lunar periods of Figure 7. In fact, they are not the same in more than half the cases. This

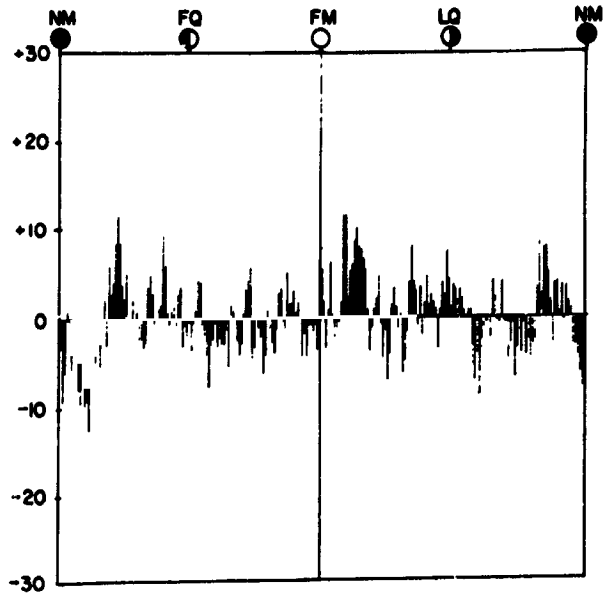


FIGURE 11.—Same as Figure 6 ($\bar{K}_p$ values of the highest quarter only).

is not surprising when we remember the occurrence of disturbed periods in low sunspot years or quiet periods in high sunspot years.

Bartels [1964] has previously reached the conclusion that there is no evidence for a lunar effect on geomagnetic activity during disturbed periods. His analysis was based upon the daily $\bar{K}_p$ sums of 30 or greater during 1932–1961 and on magnetic character figures C_i of 1.3 or higher for 1884–1961. Our analysis presented above is fully in agreement with his conclusion.

DISCUSSION

It has been shown that for the last 31 years geomagnetic activity has been on the average ≈4 per cent smaller for the period beginning after lunar first quarter and lasting until just after full moon and that it has been on the average ≈4 percent greater from just after full moon until last quarter.

Similar perturbations of the general level of the K_p index have been found in an analysis by *Michel et al.* [1964]. These authors note that magnetic perturbations persist for about three days, and hence they do not consider their longer effects near full moon as more than statistical fluctuations.

Another technique that can be applied to this problem is that of autocorrelation analysis, to determine whether there is any regularity to the K_p pattern associated with a lunar month of 29.5 days. It is well known that such effects are associated with the 27-day period of rotation of the sun. We appreciate very much receiving permission to report that Shapiro and Ward (private communication, 1964) have applied the technique of 'power spectrum' analysis to compute a single spectrum from 80 years of C_i data. In a general region of enhanced power centered on a frequency corresponding to the solar rotation period, this spectrum shows two peaks, the principal one near 27 days, and another peak, somewhat smaller, near 29.5 days. They have sufficient resolution in the spectrum between 27 and 29 days to determine that two distinct peaks are present. Shapiro and Ward consider the physical reality of the peak at 29.5 days to be uncertain at this time.

Any physical mechanism to account for the variations of geomagnetic activity about full moon must be sought in the interactions between the moon and the tail of the geomagnetic cavity formed by the magnetosphere embedded in the flow of solar wind.

ACKNOWLEDGMENT

It is a pleasure to acknowledge the able assistance of Mr. Bernard Goldstein, who programmed and carried out the numerical computations on the IBM 7094 computer and S-C 4020 plotter.

REFERENCES

ADDERLEY, E. E., The influence of the moon on atmospheric ozone, *J. Geophys. Res.*, *68*, 1405–1408, 1963.

ADDERLEY, E. E., and E. G. BOWEN, Lunar component in precipitation data, *Science*, *137*, 749–750, 1962.

BARTELS, J., Behauptete einflusse des mondes auf die erdmagnetische unruhe und ihr statistischer hintergrund, *Naturwissenschaften*, *50*, 592, 1963.

BARTELS, J., Statische hintergrunde für geophysikalische synchronisierungs-versuche und kritic an behaupteten mond-einflussen auf die erdmagnetische aktivitat, *Nachr. Akad. Wis. Goettingen*, *2*, *Math. Physik. Kl. 1963* (23), 333–356, 1964.

BELL, B., and R. J. DEFOUW, Concerning a lunar modulation of geomagnetic activity, *J. Geophys. Res.*, *69*, 3169–3174, 1964.

BIGG, E. K., A lunar influence on ice nucleus concentrations, *Nature*, *197*, 172–173, 1963*a*.

BIGG, E. K., The influence of the moon on geomagnetic disturbances, *J. Geophys. Res.*, *68*, 1409–1413, 1963*b*.

BIGG, E. K., Lunar and planetary influences on geomagnetic disturbances, *J. Geophys. Res.*, *68*, 4099–4104, 1963*c*.

BOWEN, E. G., A lunar effect on the incoming meteor rate, *J. Geophys. Res.*, *68*, 1401–1403, 1963.

BRADLEY, D. A., M. A. WOODBURY, and G. W. BRIER, Lunar synodical period and widespread precipitation, *Science*, *137*, 748–749, 1962.

DORMAN, L. I., and L. KH. SHATASHVILI, Lunar diurnal variation of the neutron component of cosmic radiation and the question of its origin, *Geomagnetizm Aeronomiya, 1*, 663–670, 1961.

MICHEL, F. C., A. J. DESSLER, and G. K. WALTERS, A search for correlation between K_p and the lunar phase, *J. Geophys. Res., 69*(19), October 1, 1964.

SNYDER, C. W., M. NEUGEBAUER, and U. R. RAO, The solar wind velocity and its correlation with cosmic-ray variations and with solar and geomagnetic activity, *J. Geophys. Res., 68*, 6361–6370, 1963.

SUCKSDORFF, E., The influence of the moon and the inner planets on the geomagnetic activity, *Geophysica Helsinki, 5*(2), 95–106, 1956.

PROPAGATION OF HYDROMAGNETIC WAVES IN THE MAGNETOSPHERE*

MASAHISA SUGIURA

Characteristics of waves in a two-component cold plasma are reviewed. Using the Clemmow-Mullaly-Allis diagram, the topological types of the wave-normal surfaces are shown. A consistent system of labeling the modes, initially given by Allis, is explained. Reversal in the polarization in the electric field is examined, and all the modes in which the reversal occurs are specified. There is no polarization reversal in ULF to VLF waves in the magnetosphere. The lower hybrid resonance frequency in the magnetosphere is discussed.

The equations of motion for an electromagnetic ray are derived. Defining the action for the ray in analogy with that for a particle in classical mechanics, the principle of least action is proved. It is shown that if the dispersion relation is homogeneous in the wave vector and the frequency, the principle of least action implies the principle of least time, i.e., Fermat's principle. When the principle of least time holds, as is the case with Alfvén compressional waves, the trajectory of a ray can be determined from a variational equation, from which the problem can be formulated in Hamiltonian form. For the axially symmetric case, the generalized momentum conjugate to the azimuthal coordinate is a constant of motion. Using this relation, "allowed" and "forbidden" regions are defined, when a set of initial conditions for the ray is given. This method is applied to a model magnetosphere with a dipole magnetic field. It is shown that the accessibility of hydromagnetic rays originating from the boundary of the magnetosphere to the earth is greatly limited. For a distorted magnetosphere the canonical equations for a hydromagnetic ray are integrated by a numerical method. Typical trajectories in the equatorial plane are shown, and the effects of the deformation of the dipole field on the ray trajectories are discussed.

1. INTRODUCTION

Hydromagnetic waves can be derived by a low-frequency approximation in a general scheme of treating plasma waves. For a complete understanding of the characteristics of hydromagnetic waves it is desirable to review plasma waves in general for the entire range of frequency and for all possible values of plasma parameters.

Clemmow and Mullaly (1954) presented a comprehensive study of the dependence of the phase refractive index n, derived by the Appleton-Hartree approximation, on the angle θ which the direction of phase propagation makes with that of the magnetic field. These authors developed a systematic method of investigating the topological genera of the $(n\theta)$ surface. Allis (1959), and Allis, Buchsbaum, and Bers (1963) modified the scheme by using the wave-normal surface which is obtained by inverting the (n,θ) surface about the origin. Stix (1962) has given an excellent summary of the wave-normal surface topology for waves in a cold plasma. In Sections 2.1 to 2.9 we will follow Stix's representation.

Stix (1962) called a diagram showing topological genera of the wave-normal surfaces for various regions in plasma parameter space the "Clemmow-Mullaly-Allis diagram," or, in short, the "CMA diagram." The CMA diagram for an idealized two-component cold plasma is presented in Section 2.4, and the characteristics of modes for different regions in plasma parameter space are reviewed. The labeling of modes by the sense of polarization, i.e., "right-handed" or "left-handed," for the propagation parallel to the magnetic field, or by "ordinary" or "extraordinary" mode according as the refractive index for the

*Published as *Goddard Space Flight Center Document X-612-64-240*; and presented at the Symposium on Ultra Low Frequency Electromagnetic Fields (from 30 c/s to 0.001 c/s), Boulder, Colorado, August 17–20, 1964.

propagation perpendicular to the magnetic field is independent or dependent on the magnetic field is explained using the CMA diagram.

In Section 2.6 the wave-normal surface topology in those regions in the CMA diagram that are relevant to ULF and VLF waves in the magnetosphere is studied.

Reversal of the polarization in the electric field that may occur for v not equal to 0 is examined in detail (Section 2.7), and the lower hybrid resonance frequency in the magnetosphere is presented (Section 2.8.)

In Sections 3.1 to 3.3 geometrical electromagnetics, or ray theory, is formulated, and in Sections 3.4 to 3.8 the theory is applied to the propagation of hydromagnetic waves in the magnetosphere.

Analogy between the Hamiltonian form of classical mechanics and geometrical electromagnetics is demonstrated (Sections 3.2, 3.3, 3.4), and validity of Fermat's principle in geometrical electromagnetics is examined (Section 3.3). The action is defined for a ray, and the principle of least action is established. It is then shown that if the dispersion relation is homogeneous in $\vec{k}$ and ω, the principle of least action implies the principle of least time, i.e., Fermat's principle. This is the same result as that obtained by Weinberg (1962) by the eikonal theory.

By formulating the ray theory in Hamiltonian form it is shown (Section 3.5) that when the magnetic field and the plasma are axially symmetric, and when the wave-normal surface is spherical, the generalized momentum conjugate to the azimuthal coordinate is a constant of motion. Applying this result to the propagation of hydromagnetic waves in a dipole model magnetosphere (Section 3.5), we will define, for a set of initial conditions of a ray, "forbidden" and "allowed" regions for the ray in the same manner at Störmer (1953) did for a charged particle moving in a dipole magnetic field. It will be shown that because of the Alfvén velocity maximum at an altitude of several thousand kilometers above ground the accessibility of hydromagnetic rays generated in the outer regions of the magnetosphere to the immediate vicinity of the earth is very limited.

When the magnetic field is not axially symmetric, the canonical equations must be integrated. In Section 3.8 we will present examples of hydromagnetic ray trajectories computed by numerical method for a model magnetosphere that takes into account the distortion of the dipole field by solar wind. Trajectories in the equatorial plane alonge are presented in Section 3.8; more detailed discussions will be given in a separate paper.

2. PROPAGATION OF HYDROMAGNETIC WAVES

2.1 The Dispersion Relation

The dispersion relation for waves in a cold plasma in a uniform magnetic field has been given by Åström (1950), Sitenko and Stepanov (1957) and Allis (1959). Here we only outline the derivation of the dispersion relation; for the details, see, e.g., Stix (1962)

We assume that in the unperturbed stat the magnetic field $\vec{B}_0$ and the plasma are static and uniform, and we take $\vec{B}_0$ and the quantities characterizing the plasma to be zero-order quantities. The perturbation $\vec{B}$ in the magnetic field, the electric field $\vec{E}$, the current and the particle velocities are considered to be first-order variables, and the relevant equations are all linearized by ignoring second- and higher-order terms. The first-order variables are assumed to change as $\exp i\,(\vec{k}\cdot\vec{r}-\omega t)$.

Considering the displacement current to be the sum of that in vacuum and the plasma current, and by Fourier analysis the electric displacement $\vec{D}$ can be expressed in the form

$$\vec{D}=\vec{K}\cdot\vec{E}$$

thus defining the dielectric tensor $\vec{K}$.

The expression for $\vec{D}$ and Maxwell's equations yield the following dispersion relation in terms of the refractive index n:

$$An^4-Bn^2+C=0 \qquad (1)$$

where

$$
\begin{aligned}
A &= S\sin^2\theta + P\cos^2\theta \\
B &= RL\sin^2\theta + PS(1+\cos^2\theta) \\
C &= PRL \\
S &= \tfrac{1}{2}(R+L) \\
R &= 1-\sum_k(\Pi_k^2/\omega^2)/(1+\epsilon_k\Omega_k/\omega) \\
L &= 1-\sum_k(\Pi_k^2/\omega^2)/(1-\epsilon_k\Omega_k/\omega) \\
P &= 1-\sum_k\Pi_k^2/\omega^2 \\
\Pi_k^2 &= 4\pi n_k q_k^2/m_k \\
\Omega_k &= |q_k|B_0/(m_k c) \\
\epsilon_k &= q_k/|q_k|
\end{aligned}
\tag{2}
$$

Here the subscript k refers to the particle of type k whose mass and charge are m_k and q_k, respectively, and c is the velocity of light; π_k and Ω_k are the plasma frequency and the cyclotron frequency for the k-th constituent; ϵ_k merely specifies the sign of the charge.

2.2 The Wave-Normal Surface

When all the plasma parameters and the frequency ω are specified, n is a function of θ, the angle between the static magnetic field $\vec{B}_0$ and the wave vector $\vec{k}$. Denoting the unit vector in the direction of $\vec{k}$ by $\vec{k}$, $\vec{k}/n$ gives the phase velocity, measured in units of c, in the direction of phase propagation.

We define the wave-normal surface by revolving about the direction of the magnetic field $\vec{B}_0$ the locus of the tip of the vector $\vec{k}/n$ when θ is changed from 0 to π; since the magnetic field and the plasma are assumed to be uniform, at each point in space the wave-normal surface is axially symmetric about the direction of the magnetic field. Thus the topological nature of the wave-normal surface is determined by the dependence of n on θ.

From the dependence of the coefficients A and B in (1) on θ as indicated in (2) it is obvious that the wave-normal surface is symmetric with respect to the plane through the origin and normal to the magnetic field; namely, $n(\theta)=n(\pi-\theta)$.

The solutions of (1) are

$$n^2=(B\pm F)/(2A) \tag{3}$$

where

$$F^2=(RL-PS)^2\sin^4\theta+P^2(R-L)^2\cos^2\theta$$

There are two branches in n^2, and for real values of θ, n is either pure real or pure imaginary.

The solutions for n^2 for propagation at $\theta=0$ (parallel to the magnetic field) and $\theta=\frac{1}{2}\pi$ (perpendicular to the magnetic field) reduce to the following simple expressions.

For $\theta=0$,

$$n^2=R \tag{4}$$

$$n^2=L \tag{5}$$

and for $\theta=\frac{1}{2}\pi$,

$$n^2=RL/S \tag{6}$$

$$n^2=P \tag{7}$$

It can be shown that for $\theta=0$, the polarization of the electric field for the branch with $n^2=R$ is circular and right-handed and that for the branch $n^2=L$ also circular but left-handed.

For $\theta=\frac{1}{2}\pi$, the phase velocity for the branch with n^2 given by (6) depends on the magnetic field, whereas that for the branch represented by (7) is independent of the magnetic field.

These characteristics at $\theta=0$ and $\theta=\frac{1}{2}\pi$ will be used later for labeling the modes.

To study the shape of the wave-normal surface it is important to know where n^2 becomes 0 or infinity. For instance, if n^2 crosses 0 from positive to negative, then n becomes imaginary and hence the wave becomes evanescent.

Equation (1) shows that $n^2=0$ is a solution only if $C=0$, namely, if

$$P=0, \text{ or } R=0, \text{ or } L=0. \tag{8}$$

Then the phase velocity is infinite.

The other case, $n^2=\infty$ occurs when $A=0$; from (2) we see that this happens when θ satisfies the relation

$$\tan^2\theta=-P/S \tag{9}$$

Thus, if P and S are of opposite sign, then at θ satisfying (9) and also at $\pi-\theta$, n^2 is infinite and the phase velocity is zero. When this happens, n^2 is zero at no other real θ. At $\theta=0$, n^2 is ∞

when $S=\pm\infty$, and at $\theta=\frac{1}{2}\pi$, n^2 is ∞ when $S=0$.

A circumstance in which $n^2=0$ is called a cutoff and that in which $n^2=\infty$ a resonance (Allis, 1959). Allis calls the resonances at $\theta=0$ and $\theta=\frac{1}{2}\pi$ the principal resonances.

From (4) and (5) we see that at $\theta=0$, $n^2=\pm\infty$, or $L=\pm\infty$. The definitions in (2) show that the case $R=\pm\infty$ corresponds to electron cyclotron resonance and the case $L=\pm\infty$ to ion cyclotron resonance. Indeed, the polarization of the electric field in each case can be shown to be in the same direction as that of gyration of the respective particle.

The resonance at $\theta=\frac{1}{2}\pi$, corresponding to $S=0$, is called the hybrid resonance; there are two branches in the solution to $S=0$, and according to the frequency, the two resonances are called the upper and the lower hybrid resonances. The lower hybrid resonance will be discussed in Section 2.8.

In the limit $\theta\to 0$ and $P\to 0$, a resonance may or may not result, depending upon the path of approach to this double limit (Stix, 1962).

The topological characteristics of the wave-normal surfaces can be classified into three categories. If n^2 is positive for all real values of θ, the wave-normal surface is topologically equivalent to a sphere.

As we have seen, if P and S are of opposite sign, there is a resonance at θ_{res} satisfying (9). In this case, if n^2 is positive for $0\leqslant\theta\leqslant\theta_{res}$, and for $\pi-\theta_{res}\leqslant\theta\leqslant\pi$ and negative for $\theta_{res}<\theta<\pi-\theta_{res}$, the wave-normal surface is equivalent to a dumbbell-shaped lemniscoid. If the sign of n^2 in these regions is reversed, the wave-normal surface is equivalent to a wheel-shaped lemniscoid.

2.3 Plasma Parameter Space

To determine the wave-normal surfaces for the two branches of n^2 certain parameters of the plasma must be specified. Equations (1) and (2) indicate that the wave-normal surfaces are uniquely determined when the ratios π_k/ω and Ω_k/ω (and ϵ_k) are specified for all k. Thus, if we imagine $2k$-dimensional space with $2k$ mutually orthogonal axes representing π_k/ω and Ω_k/ω, the wave-normal surfaces for the two branches of n^2 can be assigned to every point in this plasma parameter space. However, if we specify fixed parameters, such as ratios of number densities of different types of particles, and mass ratios, the minimum number of coordinate axes required is reduced.

To be specific in our representation we now limit ourselves to a neutral cold plasma consisting of electrons and hydrogen ions. We denote the plasma frequencies and the cyclotron frequencies for electrons and ions by affixing subscript e or i to π and Ω.

The quantities R, L, and P defined in (2) are then reduced to much simpler expressions:

$$R=1-\alpha/[(1+\Omega_i/\omega)(1-\Omega_e/\omega)] \tag{10}$$

$$L=1-\alpha/[(1-\Omega_i/\omega)(1+\Omega_e/\omega)] \tag{11}$$

$$P=1-\alpha \tag{12}$$

where

$$\alpha=(\pi_e^2+\pi_i^2)/\omega^2 \tag{13}$$

Since $\Omega_e/\Omega_i=m_i/m_e\equiv\mu$, is a fixed parameter, we see that two-dimensional parameter space, i.e., a plane, suffices for our purpose. Following Allis (1959) and Stix (1962) we take Ω_e^2/ω^2 as ordinate and α as abscissa; we are only concerned with the quadrant bounded by the positive axes.

2.4 The Clemmow-Mullaly-Allis Diagram

We now divide this plasma parameter plane into bounded areas by curves representing the principal resonances and cutoffs. To make the bounding curves reasonably separated from each other it is helpful to take the ion-to-electron mass ratio μ to be smaller than its actual value. Following Allis and Stix, we take μ to be 4 for our illustrative purpose.

By examining the wave-normal surfaces in the manner described in the preceding Section it can be shown that the topological type of each of the wave-normal surfaces for the two branches remains the same throughout each of the bounded areas.

The CMA diagram constructed in this way is shown in Figure 1. In each of the bounded areas the topological types of the wave-normal surfaces for the two branches are indicated. Wherever only one wave-normal surface is drawn, the mode corresponding to the other branch is evanescent, and in the area to the right of the curve $L=0$ and

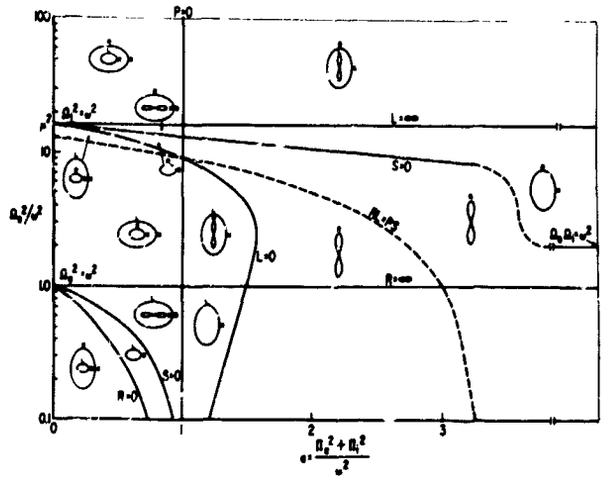


FIGURE 1.—The CMA diagram for a two-component cold plasma, showing the topological characteristics of the wavenormal surfaces.

below the horizontal line $R=\infty$ both branches are evanescent.

The curve for $RL-PS=0$ is drawn in with broken lines. This curve represents neither cutoff nor resonance, but it proves to be useful in labeling the modes.

As we have seen in Section 2.2 the polarization in the electric field is either right-handed or left-handed at $\theta=0$. The wave-normal surfaces in Figure 1 are labeled R or L on top of each sketch according as the polarization at $\theta=0$ is right-handed or left-handed, respectively.

For the propagation across the magnetic field, i.e., $\theta=\frac{1}{2}\pi$, we found that n^2 for one of the two branches depends on the magnetic field and that n^2 for the other branch is independent of the magnetic field. Allis (1959) termed the mode with n^2 dependent on the magnetic field "extraordinary" mode and the mode with n^2 independent on the magnetic field "ordinary" mode. We follow Allis' nomenclature, and wherever the propagation in this direction ($\theta=\frac{1}{2}\pi$) is possible, the symbol 0 or X is attached to the right of each sketch of the wave-normal surface, according as the mode is "ordinary" or "extraordinary."

This system of labeling the modes seems to be most systematic and is recommended for the general use to eliminate confusion that has existed in the past.

As we shall see in Section 2.9, for frequencies much less than the ion cyclotron frequency the wave-normal surface for the mode labeled with R and X in Figure 1 becomes a sphere, that is, n^2 is independent of θ. For this reason Åström (1950) called this mode "ordinary" and the other "extraordinary," and Åström's nomenclature has been used widely in the literature dealing with hydromagnetic waves. However, this labeling is not consistent with Allis' system. Since the spherical wave-normal surface in question is merely an approximation valid only for frequencies well below the ion cyclotron frequency, Allis' system seems to be preferable.

As is evident in Figure 1, a right-handed (or left-handed) mode at $\theta=0$ may be either an ordinary or extraordinary mode at $\theta=\frac{1}{2}\pi$, and hence labeling the modes with only one label R or L, or 0 or X is not adequate.

It is also noted that the polarization in the electric field may reverse its direction in some regions in the plasma parameter plane (Stix, 1962). This problem will be discussed in detail in Section 2.7.

The two modes can be distinguished by still another labeling method. The two modes are labeled "fast" or "slow" by comparing the size of the wave-normal surfaces. There can be no crossing of the wave-normal surfaces for the two branches so that the labeling with "fast" or "slow" mode can be made unambiguously. Wherever only one branch is nonevanescent, it is reasonable to assume that the wave-normal surfaces change continuously in the plasma parameter plane and to label that mode consistently with the naming of the corresponding mode in the neighboring bounded areas where a comparison of two modes is possible.

2.5 A Model Magnetosphere

In Section 2.6 a CMA diagram for ULF and VLF waves in the magnetosphere will be presented. The model magnetosphere to be used for the CMA diagram and also for all the discussions of hydromagnetic waves in the later Sections is described in this Section.

We approximate the earth's magnetic field by a dipole except in Section 3.8 where the distortion of the dipole field due to solar wind is taken into account.

The electron density distribution in the magnetosphere adopted here is based on the recent determination by Liemohn and Scarf (1964) using

nose whistlers. Among the electron density distributions which these authors considered to give self-consistent results, we adopt the simplest distribution, namely, the model in which the electron density is inversely proportional to the distance from the earth's center. Their results apply to the region of the magnetosphere approximately from 3 to 5 earth-radii. We assume that this inverse cube law for the electron density holds in regions below and above these altitudes. To be precise, we assume that the electron density varies as $N_0(a/r)^3$ on the equatorial plane from 15,000 km geocentric distance to the boundary of the magnetosphere which is taken to be at $10a$; here a is the radius of the earth and N_0 is taken to be 1.41×10^4 electrons/cm^3 (Liemohn and Scarf, 1964).

For altitudes below the bottom limit of the above distribution (15,000 km geocentric distance) we base our model on that given by Dessler, Francis, and Parker (1960), but to ensure continuity of the electron density we apply smoothing. In so doing the region below 15,000 km was divided into two regions and in each region the electron density was expressed in a power series. At the boundary between the two regions and at 15,000 km geocentric distance, the electron density and its first derivative with respect to radial distance are made continuous. The density was expressed analytically for the convenience for the numerical calculations required later in the ray treatment.

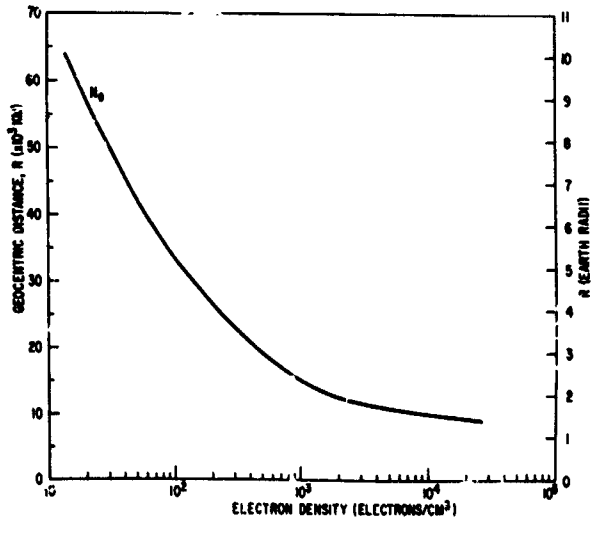


FIGURE 2.—The electron density distribution in the model magnetosphere used in this paper.

Figure 2 shows the electron density distribution constructed in the manner described above and used throughout this paper. The distribution given in Figure 2 refers to that in the equatorial plane, but we assume that the electron density is a function of radial distance alone.

Figure 3 shows the electron plasma frequency and the electron cyclotron frequency as functions of geocentric distance. The former is assumed to be spherically symmetric, but the latter, of course, varies with latitude.

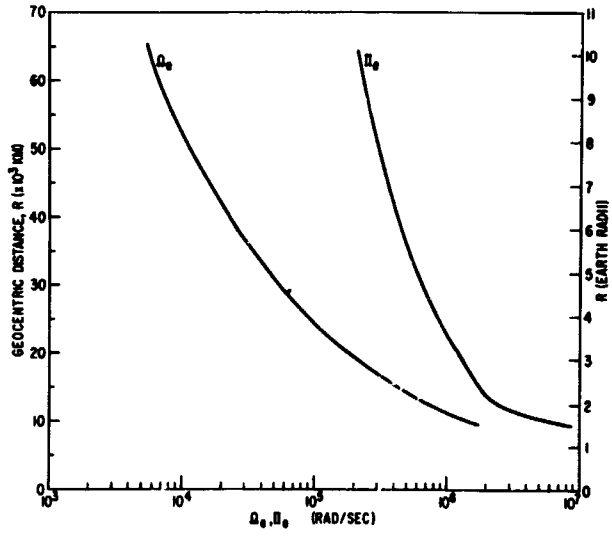


FIGURE 3.—The electron cyclotron frequency and the electron plasma frequency as functions of distance from the earth's center.

2.6 The CMA Diagram for ULF and VLF Waves In the Atmosphere

Having obtained in Section 2.3 the whole view of the CMA diagram for an idealized plasma in which the ion-to-electron mass ratio is taken to be 4, we now ask what regions in the plasma parameter plane are relevant to propagation of ULF and VLF waves in the magnetosphere. We will now take the actual value for the hydrogen-ion-to-electron mass ratio μ.

In Section 2.3 we took α and Ω_e^2/ω^2 as coordinates, but the conditions that are of interest to us now make it more convenient to use π_e/ω as abscissa and Ω_e/ω as ordinate.

A CMA diagram for ULF and VLF with these coordinate axes is shown in Figure 4. The diagram can be used in two ways. If the frequency ω is specified, the change in the plasma parameters

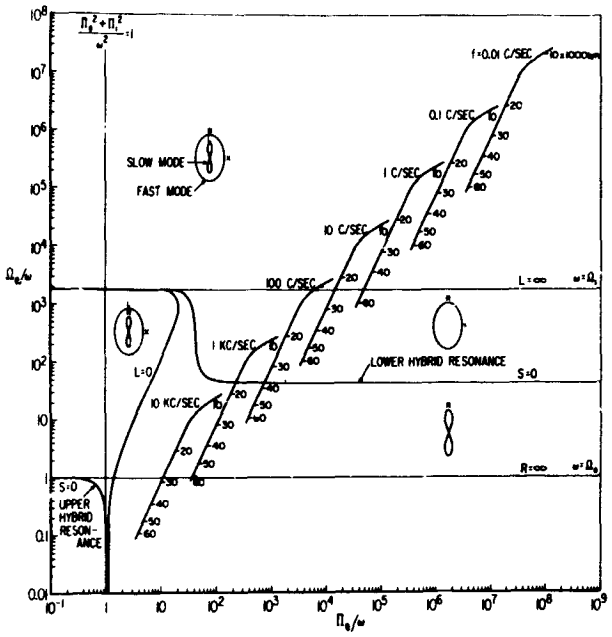


FIGURE 4.—The CMA diagram for ULF and VLF waves in the model magnetosphere.

with distance from the earth's center (for instance, on the equatorial plane) can be represented by a curve along which the radial distance is marked.

Alternatively, if a position in the magnetosphere is specified, a continuous change in the wave frequency at that position can be represented by a continuous curve in the CMA diagram.

For the sake of convenience we take the former representation. In Figure 4, curves are drawn for frequencies from 0.01 c/sec to 10 kc/sec. Since the plasma parameter variations with radial distance are the same for all frequencies, curves for different frequencies can be obtained from one of them by merely displacing it parallel to a fixed straight line.

Since the magnetic field varies with latitude, the curves drawn in Figure 4 for the equatorial plane will be changed at higher latitudes, but the changes are only slight because of the logarithmic scale in the diagram, and the general features which are discussed below will not be altered appreciably.

First, we observe that for waves with frequencies below about 1 c/sec, two modes are possible. One has a wave-normal surface topologically equivalent to a sphere; its electric field has right-handed circular polarization at $\theta=0$, and the mode is "extraordinary" at $\theta=\frac{1}{2}\pi$; this mode is the "fast" mode. The other mode, which is the "slow" mode, has a wave-normal surface topologically equivalent to a dumbbell-shaped lemniscoid, and its electric field has left-handed circular polarization at $\theta=0$. In the hydromagnetic approximation the former mode corresponds to Alfvén compressional wave and the latter to Alfvén shear wave. It is the latter mode that Alfvén (1942) originally derived by treating plasma as a conducting fluid.

In the frequency range approximately from 1 c/sec to 100 c/sec, ion cyclotron resonance takes place at some altitude, and above this altitude only the fast mode can propagate. (We are concerned here only with altitudes above several hundred kilometers above ground level.)

For each frequency in the range from several tens cycles/sec to several kilocycles/sec, the lower hybrid resonance (at $\theta=\frac{1}{2}\pi$, and $S=0$) is encountered at a certain altitude, and above that altitude the wave-normal surface is transformed to a dumbbell-shaped lemniscoid; propagation across the magnetic field becomes impossible. The mode prevailing at frequencies above the lower hybrid resonance frequency is the "whistler" mode.

As we go to still higher frequencies we reach electron cyclotron resonance, and the whistler mode is destroyed. If we cross the cutoff $L=0$ from right to left in Figure 4, we have one or two modes, according as the frequency is above or below the electron cyclotron resonance. In the latter case the slow mode represents the whistler mode.

2.7 Polarization Reversal

In this Section we examine polarization at θ not equal to zero. In Section 2.4 we already mentioned the possibility of reversal of polarization.

Stix (1962) showed that for one of the branches the polarization of the electric field changes direction of rotation at θ satisfying the relation

$$\sin^2\theta = P/S \qquad (14)$$

Clearly, for this reversal of polarization to take place at real θ, P and S must be of the same sign and $|P/S| \leqslant 1$. Even these conditions are both satisfied, the reversal may occur in an evanescent branch. Thus it is worthwhile examining the problem in detail.

From the condition that P and S are of the same sign we can eliminate about one half of the

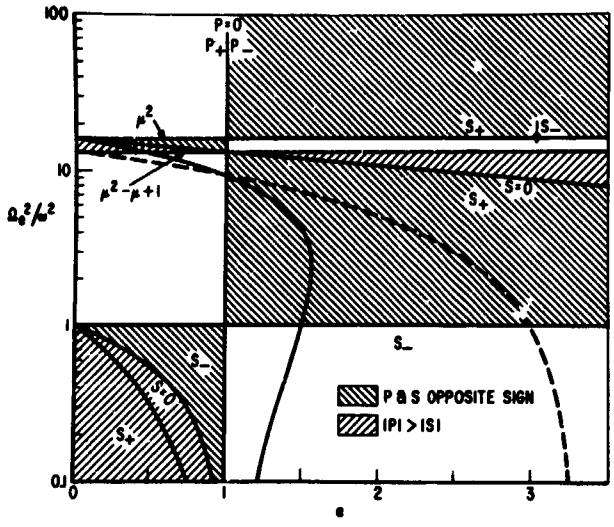


FIGURE 5.—Areas in the plasma parameter plane in which the polarization reversal can occur at θ satisfying (14).

bounded areas in the CMA diagram. By the second condition, namely, that $|P/S| \leqslant 1$, parts of the remaining areas are further eliminated. In Figure 5, areas where polarization reversal cannot occur are shaded; signs of P and S are indicated by a small symbol + or −. The coordinates in Figure 5 are the same as in Figure 1, and the ion-to-electron mass ratio μ is again taken to be 4; this is just for schematical representation, and our discussion in this Section applies to the case in which μ is taken to be the actual ion-to-electron mass ratio. The horizontal line at $\Omega_e^2/\omega^2 = \mu^2$ corresponds to $\Omega_i^2 = \omega^2$, i.e., ion cyclotron resonance. Another horizontal line at $\Omega_e^2/\omega^2 = \mu^2 - \mu + 1$ plays an important role in the following discussions. This comes about from the fact that $P - S$ has the factor $\Omega_e^2/\omega^2 - (\mu^2 - \mu + 1)$ in the numerator. For $\mu > 1$, $\mu^2 - \mu + 1$ is less than μ^2 and is greater than 1. It can be shown that $\mu^2 - \mu + 1$ is the ordinate (Ω_e^2/ω^2) for the intersection of $P = 0$ and $S = 0$, and that the curve $RL - PS = 0$ intersects the Ω_e^2/ω^2 axis at $\mu^2 - \mu + 1$.

From Figure 5 it is already clear that the reversal can never occur in magnetosphere in waves with frequencies below the ion cyclotron frequency.

Next we examine in which branch the polarization reversal takes place, if it does at all, without limiting ourselves to the conditions of our immediate interest.

By examining the polarization of the electric field it can be shown that if the polarization reversal occurs, then it does so in the branch that comes from the positive sign in (3) when $RL + PS - 2S^2$ is negative, and in the branch from the negative sign in (3) when this quantity is positive.

It is, therefore, instructive to locate the solution to the equation

$$RL + PS - 2S^2 = 0 \tag{15}$$

The left-hand side of (15) can be written as follows:

$$\begin{aligned} RL + PS - 2S^2 = {} & \alpha x^2 (\mu^2 - x^2)^{-2} (1 - x^2)^{-2} [x^6 \\ & - \{2\mu^2 - (1+\alpha)\mu + 2\} x^4 + \{\mu^4 - (1+\alpha)\mu^3 + 3\mu^2 \\ & - (1+\alpha)\mu + 1\} x^2 - \mu^2 \{(1-2\alpha)\mu^2 - (1 \\ & - 3\alpha)\mu + 1 - 2\alpha\}] \end{aligned} \tag{16}$$

where $x = \Omega_e/\omega$

The expression on the right-hand side of (16) can also be written as

$$\begin{aligned} & -\alpha x^2 (\mu^2 - x^2)^{-2} (1 - x^2)^{-2} [(x^2 - \mu^2) \{x^2 - (\mu^2 - \mu \\ & + 1)\} (x^2 - 1) + \alpha\mu \{(x^2 - \mu^2)(x^2 - 1) + 2\mu(\mu - 1)^2\}] \end{aligned}$$

Though the latter expression is convenient to determine α when x^2 is given, we will use (16) to obtain the solution to (15). From (16) we see that $x^2 = 0$ is a solution of (15), which, however, is of no interest to us.

We now examine the solution of the equation that is obtained by equating the content of the square brackets in (16) to zero. Since the equation so obtained is cubic in x^2, we can determine the number of real roots by examining the discriminant. This method is helpful in locating the solutions.

For $\alpha \gg \mu$ and x^2 not very much greater in order of magnitude than μ, x^2 becomes independent of α, and we have two positive roots:

$$x^2 = \mu^2(1 - 2/\mu) \quad \text{or} \quad 2\mu$$

thus, giving two positive roots for x

$$x = \mu(1 - 2/\mu)^{\frac{1}{2}} \quad \text{or} \quad (2\mu)^{\frac{1}{2}}$$

Numerically, these are approximately 1835.50 and 60.64, respectively.

There are three positive real roots (x^2 satisfying (15)) for positive α less than 0.175. Between

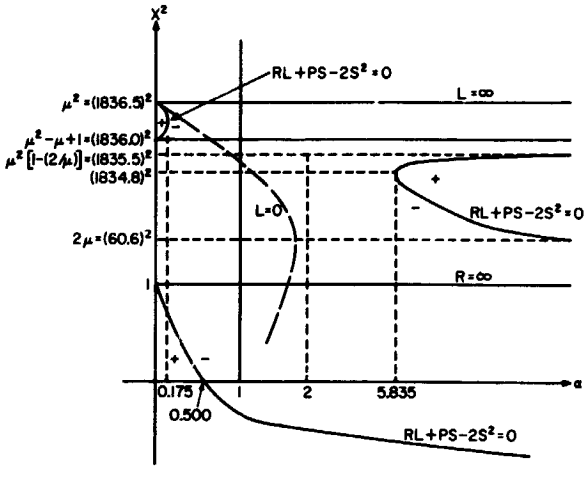


FIGURE 6.—Illustrating the solutions to $RL+PS-2S^2=0$.

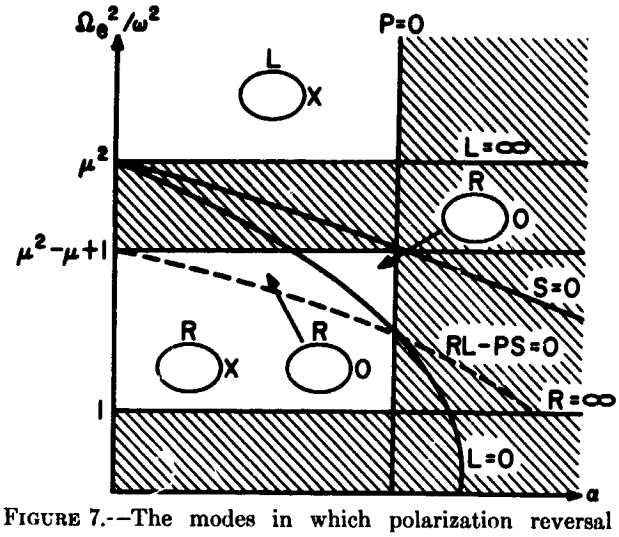


FIGURE 7.—The modes in which polarization reversal occurs at θ satisfying (14).

0.175 and 5.835 there is only one negative real root. For α greater than 5.835 there are two positive real roots and one negative real root. (The numbers quoted are accurate to ± 0.005.)

The curves representing (15) are schematically shown in Figure 6; the sign for $RL+PS-2S^2$ is also indicated for areas separated by the curves.

We now combine the results in Figures 5 and 6. In so doing we observe that the intersection of $R=\infty$ and $L=0$ is at $\alpha=2(1-1/\mu)<2$. We see that all the unshaded areas in Figure 5 where the polarization reversal can occur are in the regions where $RL+PS-2S^2$ is negative. Thus the reversal occurs in the branch resulting from the plus sign in (3).

We reach the conclusion that there is no polarization reversal in the nonevanescent modes to the right of the vertical line $\alpha=1$, and that for α less than 1 the polarization reversal occurs in the L–X mode for $\omega^2<\Omega_i^2$ and in the R–O or R–X mode for $\Omega_i^2(1-1/\mu+1/\mu^2)<\omega^2<\Omega_e^2$.

The mode in which the polarization reversal occurs are indicated in Figure 7, which summarizes the discussions given in this Section.

We conclude that for ULF and VLF waves ni the magnetosphere the polarization reversal at θ satisfying (14) does not occur.

2.8. The Lower Hybrid Resonance

We see from (6) that for propagation across the magnetic field, i.e., for $\theta=\frac{1}{2}\pi$, there is resonance at $S=0$. As is indicated in Figures 1 and 4, there are two such resonances, and the one at lower frequency is called the lower hybrid resonance and the other at higher frequency the upper hybrid resonance.

Approximate expressions for the two hybrid resonance frequencies can be obtained by ignoring terms of order μ^{-1} compared with unity in the dispersion relation for $\theta=\frac{1}{2}\pi$. Stix (1962) gives for the lower hybrid resonance frequency ω_{LH}

$$1/\omega_{LH}^2=1/(\Omega_i^2+\Pi_i^2)+1/\Omega_i\Omega_e \qquad (17)$$

and for the upper hybrid resonance frequency ω_{UH}

$$\omega_{UH}^2=\Omega_e^2+\Pi_e^2 \qquad (18)$$

In the frequency range we are concerned here and for the plasma parameters appropriate to the magnetosphere the lower hybrid resonance frequency can be approximated to a good accuracy by the geometric mean of the ion and electron cyclotron frequencies:

$$\omega_{LH}=(\Omega_i\Omega_e)^{\frac{1}{2}} \qquad (19)$$

A clear physical picture for this resonance was given by Auer, Hurwitz, Jr., and Miller (1958).

The lower hybrid resonance frequency in the magnetosphere at different distances from the earth's center is given in Table 1.

2.9. The Hydromagnetic Approximation

When the wave frequency ω is well below the ion cyclotron frequency Ω_i, the quantities R and

TABLE 1.—*The lower hybrid resonance frequency f_{LH} at various distances from the center of the earth.*

r	f_{LH}
km	c/sec
10,000	5,260
15,000	1,560
20,000	657
25,000	337
30,000	195
35,000	123
40,000	82
45,000	58
50,000	42
55,000	32
60,000	24

L in (10) and (11) are simplified and both R and L can be approximated by $1+\gamma$, where $\gamma=4\pi\rho c^2/B_0^2$; here $\rho=n_im_i+n_em_e$, i.e., the plasma density. With these approximations for R and L the dielectric tensor defined in Section 2.1 becomes diagonal. The component K_1 of the dielectric tensor perpendicular to the magnetic field than takes the well-known form (Spitzer, 1956)

$$K_1=1+\gamma$$

If α is large, as is the case with ULF waves in the magnetosphere, P in (12) can be approximated by $-\alpha$. With these approximations for R, L, and P the dispersion relation is simplified and readily factored, giving the following two modes. One of these modes gives

$$n^2=1+\gamma \tag{20}$$

and the other

$$n^2\cos^2\theta=1+\gamma \tag{21}$$

In the first mode n^2 is independent of θ. For this reason Åström (1950) called this mode "ordinary" mode, and the second "extraordinary" mode. However, a closer examination shows that the first mode corresponds to the branch labeled R–X in Figure 4 and the second to that labeled L. Hence the exact expression for n^2 for the first mode depends on the magnetic field for $\theta=\frac{1}{2}\pi$. The problem of labeling the modes has already been discussed in Section 2.4.

By studying the ion velocity the mode corresponding to (20) can be shown to represent *compressional* wave, and the other mode (21) *shear* wave.

From a comparison of phase velocity for these two modes one finds that the compressional mode is the fast mode and that the shear mode is the slow mode.

From the continuity of the wave-normal surface we see that the Alfvén compressional mode (or the fast hydromagnetic mode) and the whistler mode belong to the same family. The Alfvén shear mode (or the slow hydromagnetic mode) disappears at the ion cyclotron frequency.

When $\gamma\gg1$ the phase velocities for the compressional and shear modes recede to V_A and V_A cos θ, respectively, where $V_A=B_0/>\sqrt{4\pi\rho}$, which is usually called the Alfvén velocity. The condition that $\gamma\gg1$ is equivalent to the condition that $V_A/c\ll1$.

These approximations for the phase velocities can be readily obtained by a fluid-dynamical treatment of plasma with infinite conductivity and by neglecting displacement current.

3. RAY THEORY

3.1. Introduction

So far we have reviewed the wave modes in a cold plasma by studying the wave-normal surfaces. The basis for the study was the dispersion relation. We now investigate propagation of hydromagnetic waves from a somewhat different angle by constructing a ray theory. As the plane wave theory described in the preceding Sections has limitations in its application, the ray theory developed in the following Sections is valid only under certain conditions. However, just as the general discussions of wave modes are useful in understanding the propagation of waves, the theoretical study of the behavior or rays may be helpful in understanding some of the electromagnetic phenomena occurring in the magnetosphere.

Although the theoretical discussions in Sections 3.2 and 3.3 are of general nature, the theory is applied in Sections 3.4 to 3.8 to propagation of hydromagnetic waves in the magnetosphere.

3.2. The Equations of Motion of a Ray

The basis of the theory is that the ray propagates with the group velocity. Hines (1951) and

Auer, Hurwitz, Jr., and Miller (1958) have shown that the constructive interference maximum of a wave packet moves with the group velocity. We also assume that the mode in which the waves constituting the wave packet propagate is the same throughout its motion in the anisotropic medium. Coupling between different modes is assumed not to take place.

Denoting the position of a ray by $\vec{r}$ and the group velocity by $\vec{v}_g$, we write the equation of motion for the ray as follows.

$$\dot{\vec{r}}=\vec{v}_g \tag{22}$$

where the dot signifies differentiation with time t. The group velocity $\vec{v}_g$ is given by

$$\vec{v}_g=\partial\omega/\partial\vec{k} \tag{23}$$

The wave frequency ω and the wave vector $\vec{k}$ are related by the dispersion relation which we write

$$D(\vec{k},\ \omega,\ \vec{r},\ t)=0 \tag{24}$$

We introduce a parameter τ along the path of the ray. Then, using (23) and (24), the equation of motion (22) can be written in the following form.

$$\frac{d\vec{r}/d\tau}{dt/d\tau}=-\frac{\partial D/\partial\vec{k}}{\partial D/\partial\omega} \tag{25}$$

From our assumption D is constant along the trajectory, and hence we have

$$\delta D=\left(\frac{\partial D}{\partial\vec{k}}\cdot\frac{d\vec{k}}{d\tau}+\frac{\partial D}{\partial\omega}\frac{d\omega}{d\tau}+\frac{\partial D}{\partial\vec{r}}\cdot\frac{d\vec{r}}{d\tau}+\frac{\partial D}{\partial t}\frac{dt}{d\tau}\right)=0 \tag{26}$$

For (26) to hold for any τ the content in the parentheses must be zero. We group the four terms in the paretheses in (26) as follows:

$$\left(\frac{\partial D}{\partial\vec{k}}\cdot\frac{d\vec{k}}{d\tau}+\frac{\partial D}{\partial\vec{r}}\cdot\frac{d\vec{r}}{d\tau}\right)+\left(\frac{\partial D}{\partial\omega}\frac{d\omega}{d\tau}+\frac{\partial D}{\partial t}\frac{dt}{d\tau}\right)=0 \tag{27}$$

If each of the two factors in (27) is zero, the required condition (26) is satisfied. Equations (25) and (27) are satisfied by the following set of equations:

$$d\vec{r}/d\tau=\partial D/\partial\vec{k} \tag{28}$$

$$d\vec{k}/d\tau=\partial D/\partial\vec{r} \tag{29}$$

$$dt/d\tau=\partial D/\partial\omega \tag{30}$$

$$d\omega/d\tau=\partial D/\partial t \tag{31}$$

If D is independent of time, the last equation (31) states that ω is conserved along the ray trajectory. Third equation (30) gives the relation between time t and parameter τ.

The above set of equations clearly indicates an analogy between the ray theory and classical mechanics. The wave vector $\vec{k}$ and the frequency ω of a ray play the roles of the momentum and the energy of a particle, respectively. However, this analogy does not enable us to formulate the ray theory in Hamiltonian form. This is because the motion of a ray corresponds to that of a particle of zero mass, and hence it is expected that the Lagrangian for the ray is identically zero. In the end of the following Section we will show that the Lagrangian indeed vanishes.

However, when Fermat's principle is valid the problem can be formulated in Hamiltonian form. Hence we now examine under what conditions Fermat's principle is consistent with our formulation.

3.3. Fermat's Principle

Weinberg (1962) showed that when the *eikonal* is stationary, the principle of least time, i.e., Fermat's principle, holds if the dispersion relation is homogeneous in $\vec{k}$ and ω In this Section we drive the same condition by pursuing the formulation in the preceding Section.

In classical mechanics the principle of least action holds when the Hamiltonian is conserved. We limit ourselves to the case when the frequency ω is conserved along the path, that is, D is conserved along the path.

We first define the action A for the ray by the integral

$$A=\int_{t_1}^{t_2}\vec{k}\cdot\dot{\vec{r}}dt \tag{32}$$

where the dot means, as before, differentiation with t.

Next, we calculate the variation ΔA, where the Δ-variation differs from the δ-variation appearing in virtual displacement; in the latter, time is kept unchanged, whereas in the former the process involves a change dt (see, e.g., Goldstein, 1951). For any function f of $\vec{r}$ and t, the Δ-variation of f is

$$\Delta f=d\tau\left(\frac{\partial f}{d\tau}+\dot{f}\frac{dt}{d\tau}\right)=\delta f+\dot{f}\Delta t \tag{33}$$

In particular,

$$\Delta\vec{r}=\delta\vec{r}+\dot{\vec{r}}\Delta t \tag{34}$$

Applying the Δ-variation to (32) we have

$$\Delta A=\int_{t_1}^{t_2}\delta(\vec{k}\cdot\dot{\vec{r}})dt+\vec{k}\cdot\dot{\vec{r}}\Delta t\Big|_{t_1}^{t_2} \tag{35}$$

The first integral can be calculated in the following way:

$$\int_{t_1}^{t_2}\delta(\vec{k}\cdot\dot{\vec{r}})dt=\int_{t_1}^{t_2}\delta\vec{k}\cdot\dot{\vec{r}}dt+\int_{t_1}^{t_2}\vec{k}\cdot\delta\dot{\vec{r}}dt=\int_{t_1}^{t_2}\delta\vec{k}\cdot\dot{\vec{r}}dt$$

$$-\int_{t_1}^{t_2}\dot{\vec{k}}\cdot\delta\vec{r}dt+k\cdot\delta\vec{r}\Big|_{t_1}^{t_2}=\int_{t_1}^{t_2}(\delta\vec{k}\cdot\dot{\vec{r}}-\dot{\vec{k}}\cdot\delta\vec{r})dt-\vec{k}\cdot\dot{\vec{r}}\Delta t\Big|_{t_1}^{t_2}$$

Here we used the fact that the order of the δ- and *dot*-operation can be interchanged, and partial integration was performed on the second integral on the right-hand side of the first line. To obtain the last result we used (34) and the condition that $\Delta\vec{r}=0$ at the end points.

Thus (35) reduces to

$$\Delta A=\int_{t_1}^{t_2}(\delta\vec{k}\cdot\dot{\vec{r}}-\dot{\vec{k}}\cdot\delta\vec{r})dt$$

Using (22), (23), (38), (29), and (30), the integrand can be transformed as follows:

$$\delta\vec{k}\cdot\dot{\vec{r}}-\dot{\vec{k}}\cdot\delta\vec{r}=-(\partial D/\partial\omega)^{-1}\left(\delta\vec{k}\cdot\frac{\partial D}{\partial\vec{k}}+\delta\vec{r}\cdot\frac{\partial D}{\partial\vec{r}}\right)$$

$$=-(\partial D/\partial\omega)^{-1}\delta D=0$$

Thus we have proved that

$$\Delta A=0$$

Namely, in our system the principle of least action holds if the action is defined by (32). Having proved this principle we go back to the expression for ΔA given in (35). If

$$\vec{k}\cdot\dot{\vec{r}}=\text{constant}\neq 0 \tag{36}$$

then (35) reduces to

$$\Delta(t_2-t_1)=0$$

which implies the principle of least time, or Fermat's principle. The condition (36) can be rewritten as follows:

$$-(\vec{k}\cdot\partial D/\partial\vec{k})/(\partial D/\partial\omega)=\text{constant} \tag{37}$$

A sufficient condition for (37) to hold is that D is homogeneous in $\vec{k}$ and ω, because if D is homogeneous in $\vec{k}$ and ω, i.e., if

$$D(\alpha\vec{k},\ \alpha\omega)=\alpha^n D(\vec{k},\ \omega)$$

Euler's homogeneity equation becomes

$$\vec{k}\cdot\frac{\partial D}{\partial\vec{k}}+\omega\frac{\partial D}{\partial\omega}=nD=0 \tag{38}$$

Thus the constant in (37) takes the value of ω.

The homogeneity equation (38) can be transformed into the form

$$\vec{v}_g\cdot\frac{\vec{k}}{\omega}=1$$

which implies the equality of the group and phase velocities.

It is pointed out here that the action (32) is the same as the eikonal S in Weinberg's formulation, and that the principle of least action derived here is equivalent to the principle of stationary S in Weinberg's eikonal theory.

For the Alfvén compressional mode the dispersion relation is homogeneous in $\vec{k}$ and ω, and hence Fermat's principle holds. As a matter of fact, for this mode the group velocity is equal to the phase velocity. However, for the shear mode the dispersion relation is not homogeneous in the components of $\vec{k}$, and thus Fermat's principle does not hold. As has been noted by Weinberg (1962), the application of Fermat's principle by Francis, Green and Dessler (1959) is justified.

In concluding this Section a remark is made on the Lagrangian. In classical mechanics the Lagrangian L is related to the Hamiltonian by

$$L=\vec{p}\cdot\dot{\vec{r}}-H$$

where $\vec{p}$ is the momentum.

If we define the Lagrangian for the ray by

$$L=\vec{k}\cdot\dot{\vec{r}}-\omega$$

then from (36) our Lagrangian is identically zero, confirming our expectation expressed in Section 3.2. Thus the ray theory cannot be constructed in Hamiltonian form using the Lagrangian defined above.

3.4 A Ray Theory for Modes with Isotropic Phase Velocity

In this and the following Sections we discuss the case in which the phase velocity is isotropic. We specifically study propagation of a hydromagnetic ray. For the Alfvén compressional mode the group velocity is the same as the phase velocity. Hence we simply refer to these velocities by the single term the Alfvén velocity.

Fermat's principle states that the motion of a ray from a point P_1 to another point P_2 is such that the variation of the line intergral $\int_{P_1}^{P_2} ds/V$ for fixed P_1 and P_2 is zero, i.e.,

$$\delta \int_{P_1}^{P_2} ds/V(r,\theta,\phi) = 0 \qquad (39)$$

where $V(r.\theta.\phi)$ is the Alfvén velocity. We use a spherical coordinate system (r,θ,ϕ) in this and following Sections; when we refer to the magnetosphere the origin of the spherical coordinate system is taken to coincide with the earth's center. In the preceding Sections θ was the angle which $\vec{k}$ makes with the magnetic field, but in the rest of the paper θ is the polar angle.

We write (39) in the form

$$\delta \int_{t_1}^{t_2} \frac{(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\dot{\phi}^2 \sin^2\theta)^{\frac{1}{2}}}{V(r,\theta,\phi)} dt = 0 \qquad (40)$$

where t_1 and t_2 are the times when the ray is at P_1 and P_2, and where the dot means differentiation with respect to time t.

Equation (40) is formally the same as the equation expressing Hamilton's principle for a system whose Lagrangian is equal to the integrand in (40). Thus we take the integrand of (40) as the Lagrangian of our system, and define the generalized momenta conjugate to r, θ, ϕ by

$$p_k = \partial L/\partial \dot{q}_k \qquad (k = 1, 2, 3)$$

where it is understood that the subscript k refers to r, θ, ϕ components and that q_k $(k = 1, 2, 3)$ represents r, θ, ϕ, respectively.

The Hamiltonian H of the system is

$$H = \sum_k p_k \dot{q}_k - L$$

$$= V(r, \theta, \phi)^2[p_r^2 + p_\theta^2/r^2 + p_\phi^2/(r^2\sin^2\theta)] - 1$$

Then the canonical equations can readily be formed. Stegelmann and von Kenschitzki (1964) proceeded to integrate the cononical equations numerically.

3.5. Axially Symmetric Case: *Allowed and Forbidden Regions for a Ray*

It is obvious that if the Alfvén velocity is independent of ϕ, the Hamiltonian does not contain ϕ explicity; thus, ϕ is a *cyclic* coordinate. It follows that the conjugate momentum p_ϕ is a constant of motion. From the definition of p_ϕ we immediately obtain the equation

$$(r^2\dot{\phi}\sin^2\theta)/V(r,\theta)^2 = \text{constant} \qquad (41)$$

This equation, of course, is the canonical equation for ϕ with p_ϕ constant.

Since $\phi = V d\phi/ds$, where s is path length along the trajectory, (41) can be written as follows:

$$(R^2/V) d\phi/ds = \alpha \qquad (42)$$

where

$$R = r \sin\theta \qquad (43)$$

and where α is a constant.

We define an angle χ by the equation

$$R\, d\phi/ds = \sin\chi \qquad (44)$$

so that χ is the angle between the tangent to the ray in the direction of its motion and the meridian plane. As can be seen in (44) the sign of χ is taken such that χ is positive when ϕ increases as the ray advances. Figure 8 illustrates the angle χ.

Using χ, (42) reduces to

$$(R/V) \sin\chi = \alpha \qquad (45)$$

For rays belonging to α, (45) gives the angle χ as a function of r and θ.

Since $-1 \leqslant \sin\chi \leqslant 1$, we have the relation

$$= 1 \leqslant \alpha V/R \leqslant 1 \qquad (46)$$

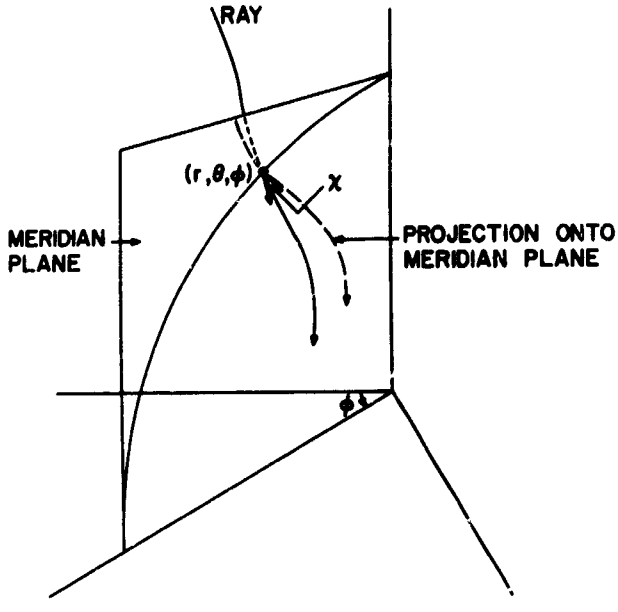


FIGURE 8.—Illustrating the angle χ.

Thus, given the value of α, (46) defines the "allowed" region for the rays belonging to α. Areas outside the allowed region are forbidden to these rays.

The constant α can be taken as the initial condition specified by r_0, θ_0, and χ_0:

$$\alpha = (R_0/V_0) \sin \chi_0 \qquad (47)$$

where R_0 and V_0 are the values of R and V at r_0, θ_0.

It should be noted that the angle χ_0 does not specify the initial direction of a ray completely, except in the equatorial plane. To specify the initial direction of a ray completely one more angle is needed (Section 3.8); however, we only need χ_0 for the time being.

When the initial conditions r_0, θ_0, and χ_0 are given, the *allowed* and *forbidden regions* can be mapped out using (46) without integrating the equations of motion.

The problem is formally identical with that in Störmer's work on the motion of a charged particle in a dipole magnetic field (Störmer, 1955). In Störmer's problem also the Hamiltonian does not contain ϕ explicitly.

The meaning of (44) becomes more explicit if we limit ourselves to rays confined in the equatorial plane. The Alfvén velocity V is now a function of r alone, and χ is the angle which the tangent to the ray trajectory makes with the radial direction. Let χ be measured positively from the inward radial direction toward the direction of increasing ϕ (which is taken to be eastward) and let χ vary from $=\pi$ to π; thus rays with $-\frac{1}{2}\pi < \chi < \frac{1}{2}\pi$ are inbound and those with $|\chi| > \frac{1}{2}\pi$ are outbound, and when $\chi = \pm\frac{1}{2}\pi$, a ray is tangent to the circle of radius r.

The angle χ at r is related to its initial value χ_0 at the point source at r_0 by

$$\sin \chi = (r_0/r)(V/V_0) \sin \chi_0$$

If $V(r)$ is a maximum at $r = r_m$ and is a monotonically decreasing function of r with increasing r, then for $r_0 > r > r_m$, the inequality $|\chi| > |\chi_0|$ always holds. This simply implies the obvious result that the ray is bent away from the region of high Alfvén velocity.

The Alfvén velocity is a maximum at several thousand kilometers altitude and decreases both above and below this level (Dessler, 1958). The Alfvén velocity increases again near the F_2 peak. but we are not concerned here with the propagation of hydromagnetic waves in the ionosphere. We only consider the ray trajectories above 600 km altitude.

It is of interest to find the critical initial angle $\chi_{0,c}$ at which an inbound ray from a point source at a great distance is reflected away from the earth at the region ($r = r_m$) of the Alfvén velocity maximum. This critical initial angle can be determined by

$$|\sin \chi_{0,c}| = (r_m/r_0) \quad (V_0/V_m)$$

where V_m is the value of V at r_m.

For a rough estimate, taking $r_0 = 60{,}000$ km, $r_m = 10{,}000$ km, and $V_0 = 400$ km/sec, $V_m = 2{,}000$ km/sec, we obtain $|\sin \chi_{0,c}| = 1/30$, and hence $|\chi_{0,c}|$ is about 1.9°. There are uncertainties in the distribution of Alfvén velocity in the magnetosphere, but $\chi_{0,c}$ is not likely to be changed greatly as more accurate information on the Alfvén velocity distribution becomes available.

3.6. A Further Remark on Ray Tracing in the Equatorial Plane

In Section 3.5, time t used for the variable in the variation equation. However, any one of

the three coordinates can be used as the variable in place of t.

If we choose ϕ as the variable, and if we confine ourselves to the equatorial plane, the integrand L in the variation equation reduces to

$$L(r, r') = (r'^2 + r^2)^{\frac{1}{2}} / V(r)$$

where the prime represents differentiation with respect to ϕ.

Considering this function as the Lagrangian, Lagrange's equation of motion is

$$\frac{d}{d\phi}\left(\frac{\partial L}{\partial r'}\right) - \frac{\partial L}{\partial r} = 0$$

This is the equation used by Francis, Green and Dessler (1959).

Denoting the momentum conjugate to r by p, the Hamiltonian for the system is given by

$$H = pr' - L$$

where

$$p = \partial L / \partial r'$$

It is understood that H is expressed as a function of r and p. Then the Hamiltonian does not contain Φ explicitly. Thus the Hamiltonian is a constant of motion. We immediately arrive at the equation:

$$r' \frac{\partial L}{\partial r'} - L = \text{constant}$$

This is the equation which Francis, Green and Dessler (1959) derived mathematically and used for their calculation of the transit time for the ray, and which Dessler, Francis, and Parker (1960) used for their two dimensional ray tracing.

3.7. Hydromagnetic Rays in the Magnetosphere: Axially Symmetric Case

Using the method described in Section 3.5, we will now investigate the accessibility of hydromagnetic rays originating from the magnetospheric boundary to the vicinity of the earth. We use the same model magnetosphere as the one presented in Section 2.5; for the magnetic field we approximate the geomagnetic field by a centered dipole.

As was shown in Section 3.5, when the position $(r_0\theta_0)$ of the point source and the initial value of χ of the ray are specified, we can determine the allowed and forbidden regions by (46) and (47).

We place the point source at the distance of 10 earth-radii from the earth's center, i.e., $r_0 = 10a$, where a is the earth's radius. and we determine allowed and forbidden regions for $\theta_0 = 30°$, 60°, and 90°; the last value of θ_0 places the source on the equatorial plane.

Typical diagrams showing the allowed and forbidden regions are presented in Figure 9. In the Figure, forbidden regions are indicated by patches and the open areas represent allowed regions. The patched circle in the center represents the earth and the poisition of the point source can be at any of the four arrows on the great circle whose radius is ten times that of the earth. All the diagrams are symmetric with respect to the equator, and the three dimensional allowed (or forbidden) region can be obtained by rotating each diagram about the vertical axis through the center, namely, the dipole axis.

For very small values of χ_0 (well below 1°, say), that is, when the initial direction of the ray deviates from the meridian plane only by a small angle, the ray can reach the earth's vicinity except directly above the poles. As χ_0 increases the two

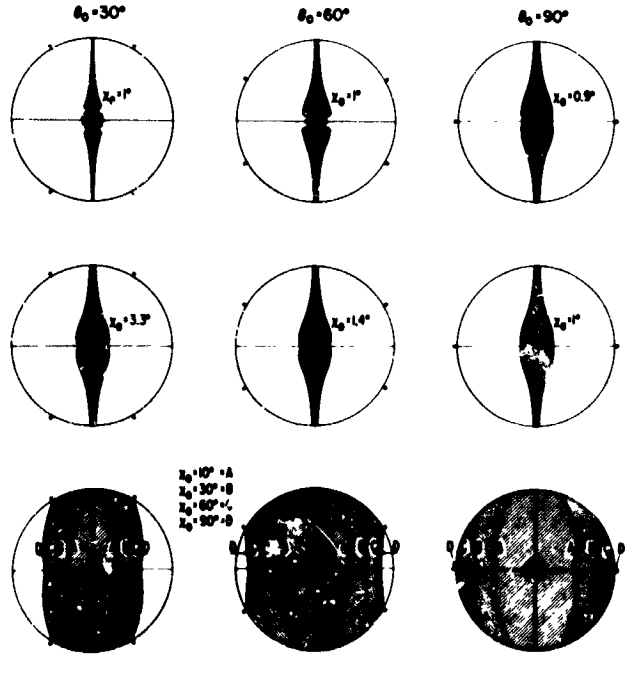


FIGURE 9.—Allowed and forbidden regions for a hydromagnetic ray in the magnetosphere with a dipole field.

forbidden regions around the axis, one in each hemisphere, become larger and extend to lower latitudes near the altitude of the Alfvén velocity maximum. When χ_0 reaches some critical value, the tips of the northern and southern forbidden regions touch each other on the equatorial plane. There is an allowed region between the ionosphere and the altitude at which the joining of the two forbidden regions takes place on the equatorial plane, but this inner allowed region is not accessible to the ray coming from outside.

When χ_0 exceeds the critical value, the earth is completely immersed in a forbidden region and the outer allowed region is more and more pushed outward, and finally, as χ_0 tends to 90° its inner surface approaches some limiting surface which intersects the large sphere (of radius r_0) at $\theta=\theta_0$ and $\theta=\pi-\theta_0$. When θ_0 is 90°, the allowed region for $\chi_0=90°$ degenerates to a circle of radius r_0 on the equatorial plane.

The critical value of χ_0 becomes smaller as θ_0 increases; for $\theta_0=30°$, 60°, and 90°, the critical χ_0 is 3.3°, 1.4°, and 0.9°, respectively. In Figure 9 the diagrams for these critical circumstances are included.

For $\chi_0=10°$, 30°, 60°, and 90° the forbidden regions are indicated in one diagram for each θ_0.

We conclude that the earth and its immediate vicinity are remarkably well protected from the hydromagnetic rays generated in the outer regions of the magnetosphere. This feature has been shown by Stegelmann and von Kenschitzki (1964) with their results from numerical ray tracing.

3.8. Hydromagnetic Rays in the Distorted Magnetosphere: Axially Asymmetric Case

The magnetosphere is contained in a cavity in the streaming solar plasma (Cahill and Amazeen, 1963; Ness et al., 1964). A number of workers have attempted to theoretically determine the shape of the boundary of the magnetosphere (for reference, see a review paper by Beard, 1964).

In this Section we only briefly discuss the effect of the distortion of the magnetosphere on the propagation of hydromagnetic waves.

In the absence of symmetry we have to write down the equations of motion and solve them by some numerical method.

For the sake of convenience we multiply the Hamiltonian given in Section 3.4 by the factor $\frac{1}{2}$. With this Hamiltonian the canonical equations are

$$\dot{r}=V^2p_r$$

$$\dot{\theta}=V^2p_\theta/r^2$$

$$\dot{\phi}=V^2p_\phi/(r^2\sin^2\theta)$$

$$\dot{p}_r=-(1/V)\partial V/\partial r+(V^2/r^3)(p_\theta^2+p_\phi^2/\sin^2\theta$$

$$\dot{p}_\theta=-(1/V)\partial V/\partial\theta+(V^2/r^2)p_\phi^2\cos\theta/\sin^3\theta$$

$$\dot{p}_\phi=-(1/\mathrm{V})\partial V/\partial\phi \qquad (48)$$

These equations are not completely independent. The Hamiltonian of the system is identically zero, and we have

$$p_r^2+p_\theta^2/r^2+p_\phi^2/(r^2\sin^2\theta)=1/V^2 \qquad (49)$$

Using (49) one of the variables can be eliminated from the set of equations (48). But it is found convenient to use (49) as a check in the numerical ray tracing calculation.

We define the direction of the tangent to the ray trajectory at a point P by two angles χ and η. The angle χ is the same as that defined in the preceding Section, and η is the angle which the tangent to the projection of the trajectory onto the meridian plane makes with the radial direction, namely,

$$\sin\chi=(d\phi/ds)\,r\sin\theta$$

$$\cos\eta=-(dr/ds)\sec\chi$$

We denote the initial values of χ and η at (r_0, θ_0, ϕ_0) by χ_0 and η_0, respectively.

The initial values of the momenta can be written in terms of r_0, θ_0, ϕ_0, χ_0, η_0, and $V(r_0, \theta_0, \phi_0)$.

For the deformed geomagnetic field we take the model proposed by Mead (1964). For the dipole field we take $g_1^0=-0.31$ gauss, and for the additional field due to the deformation we take

$$g_1^0=-0.2515/r_b^3 \text{ gauss}$$

$$g_2^1=0.1215/r_b^4 \text{ gauss}$$

where g's are well-known Gauss coefficients in the spherical harmonic expansion of the magnetic field, and where r_b is the distance, measured in earth-radii, from the earth's center to the boundary of the magnetosphere at the subsolar point. Here r_b is taken to be 10 earth-radii.

A computer program has been developed to integrate the equations of motion (48) with r_0, θ_0, ϕ_0, χ_0, and η_0 as the initial conditions. For integration the Runge-Kutta method was used.

In this paper, trajectories on the equatorial plane alone are discussed.

Figure 10 shows typical examples of the trajectories in the equatorial plane. The position of the point source is place at 10 earth-radii regardless of the longitude. This assumption is made because the location of the magnetospheric boundary is not well known on the dark side of the magnetosphere, and because with a fixed r_0 we can compare trajectories starting from sources at different longitudes more directly. Here longitude, ϕ, is measured eastward from the midnight meridian; in Figure 10 the midnight meridian is towards the left and the longitude increases counterclockwise.

In the lower half of Figure 10, trajectories starting at $\phi_0=0°$, 45°, 135°, and 180°, all with $\chi_0=0$ (i.e., rays directed initially towards the origin) are shown. In the upper half, trajectories with their initial position at $\phi_0=90°$ with $\chi_0=0$, 1°, 3°, 4°, 5°, 6°, and 7° are drawn.

The effect of the distortion of the magnetosphere on the ray trajectories can be described by saying that hydromagnetic rays tend to be "blown" towards the direction away from the sun. This is because the magnetic field is more compressed on the sunlit side of the magnetosphere that on its dark side, thus increasing the Alfvén velocity in the region facing the sun.

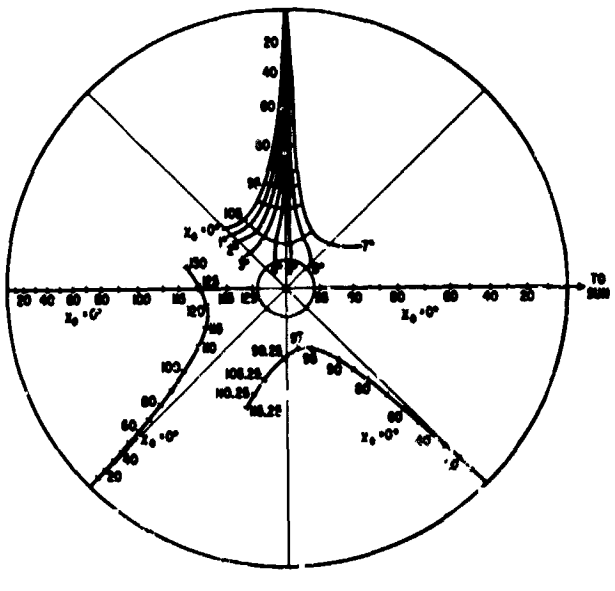


FIGURE 10.—Illustrating ray trajectories in the equatorial plane in the distorted magnetosphere.

However, the results presented in this Section should be interpreted with caution. Although the distortion of the geomagnetic field is taken into account, possible charges in the plasma density distribution associated with the distortion are not considered here. Appreciable asymmetry may be introduced in the plasma density, but no observational data are available as yet that indicate such an effect.

The magnetic field configuration on the dark side of the magnetosphere has also not as yet been established. Since the trajectories are sensitive to large scale magnetic field patterns, the actual trajectories in the magnetosphere may be different from those shown in this paper. Our purpose is to demonstrate qualitative characteristics of ray trajectories in the magnetosphere. However, for the model used in this Section the calculations are made as accurately as possible within the practical limitations.

It is observed in Figure 10 that if the point source is not in the meridian containing the sun (i.e., $\phi_0=0°$ or 180°), the ray directed initially towards the origin does not reach the earth. For the ray to reach the earth it must start with χ_0 slightly greater than 0. If χ_0 becomes too large, the ray is bent back at the region of Alfvén velocity maximum. This circumstance is shown for $\phi_0=90°$ in the upper half of Figure 10. At this ϕ_0, rays with $\chi_0=3.0°$ and less are "blown" towards the back of the magnetosphere. At χ_0 between 3.0° and 3.2° the ray begins to penetrate into the immediate vicinity of the earth, and this condition prevails till χ_0 reaches a value a little less than 6.5°. Beyond this latter angle the ray is again reflected away from the earth at the region of Alfvén velocity maximum.

In Figure 10 the time in seconds is indicated along the trajectories. It is of interest to compare the transit time from r_0 to some altitude near the earth for the trajectory in the midnight meridian with the corresponding transit time for the trajectory in the noon meridian. In Figure 10 the altitude of the point marked 95 seconds in the noon meridian and that of the point marked 129 seconds in the midnight meridian are within 4[illegible]

km from each other. The mean of the two altitudes is 1,363 km above the earth's surface. Thus the difference in the transit time from 10 earth-radii to this altitude is about 34 seconds.

However, as has already been mentioned, caution should be exercised in applying this result to any actual events that occur in the magnetosphere.

In concluding the discussions of the ray theory the following remarks are made. The hydromagnetic approximation is based on the condition that $\omega \ll \Omega_i$. Thus for the most part of the magnetosphere this approximation is good below 10 c/sec (Figure 4). There is another limitation to the ray theory, namely, that the wavelength be short compared with the dimension under consideration. We put this condition in the form $\omega \gg V/L$, where L is the typical scale length. If we take L to be the smallest value of the radius of curvature for the ray trajectories, then the minimum frequency may be set at about 1 c/sec. Thus, roughly speaking, the ray theory is applicable to propagation of hydromagnetic waves of frequencies about 1 to 10 c/sec.

A more extensive study of the ray trajectories in the magnetosphere and their physical implications will be reported later. For instance, the efficiency of energy transfer from solar winds to the ionosphere via hydromagnetic waves is considerably reduced by the limited accessibility of the hydromagnetic waves to the immediate vicinity of the earth.

A theoretical study of geometrical hydromagnetics based on a classical hydromagnetic fluid has been made by Bazer and Hurley (1963); their paper includes comprehensive reference to the literature on the subject.

4. CONCLUSIONS

We reviewed the possible modes of waves in a two-component cold plasma using the Clemmow-Mullaly-Allis diagram. A systematic method of labeling the modes was explained.

The modes relevant to propagation of ULF and VLF waves in the regions of plasma parameter space representing the conditions in the magnetosphere were reviewed. For frequencies below the ion cyclotron frequency there are two modes: the *fast mode* with right-handed circular polarization for propagation along the magnetic field and with phase velocity dependent on the magnetic field for propagation across it, and the *slow mode* with left-handed circular polarization for propagation along the magnetic field. Waves in the latter mode do not propagate across the magnetic field. Above the ion cyclotron frequency, only the fast mode represents propagating wave, and above the lower hybrid resonance frequency this mode becomes the whistler mode.

Reversal of polarization (in the electric field) depending upon the direction of phase propagation with reference to the direction of the magnetic field was discussed. We concluded that there is no such reversal in polarization in ULF and VLF waves in the magnetosphere.

The hydromagnetic approximation was examined and its relation to the more exact treatment was indicated.

In this paper we only discussed propagation of waves in a collisionless plasma. When the thermal motions of electrons and ions are included, the waves found in a cold plasma are modified. The modifications are often only slight, but in a hot plasma new modes are introduced which have no counterpart in a cold plasma. There are two such modes in a relatively low frequency range. They are ion acoustic waves and electrostatic ion cyclotron waves. These waves were not discussed in this paper; the reader is referred to discussions on these waves made, for instance, by Spitzer (1956), Bernstein (1958) and Stix (1962).

In the latter half of this paper, we changed the line of approach, and formulated a ray theory. The equations of motion of a ray were derived from a simple postulate that a ray moves with the group velocity. The action of the ray was defined in analogy with classical mechanics, and the principle of least action was proved. It was shown that the principle of least action takes the form of the principle of least time when the dispersion relation is homogeneous in the wave vector $\vec{k}$ and the frequency ω.

For the case in which the wave-normal surface is spherical, a ray theory was formulated in Hamiltonian form. In the axially symmetric case the generalized momentum conjugate to the azimuthal coordinate becomes a constant of motion. Using this relation, allowed and forbidden regions were defined for a hydromagnetic ray in the magnetosphere with the magnetic field approximated by that of a dipole. It was shown that a ray

originating from the magnetospheric boundary can reach the ionosphere only if the deviation of the initial direction of the ray from the meridian plane is small.

When the distortion of the geomagnetic field due to solar wind is taken into account, the ray trajectories in the magnetosphere are appreciably altered from those in a dipole field.

In spite of the limitations in its application the hydromagnetic ray theory for the magnetosphere should provide a guide towards a more complete understanding of propagation of hydromagnetic waves in the magnetosphere.

ACKNOWLEDGMENTS

I wish to thank Mr. W. F. Cahill for his valuable suggestions on the numerical calculations in this paper. I am indebted to Mr. W. H. Mish, Mrs. S. J. Hendricks, and Mrs. P. J. Connor for their helpful assistance in preparing the computer programs and performing the calculation.

REFERENCES

Alfvén, H. (1942), On the existence of electromagnetic-hydromagnetic waves, Arkiv f. Mat., **29 B.** No. 2; Existence of electromagnetic-hydromagnetic waves, Nature, **150.** 405–406.

Allis, W. P. (1959), Waves in a plasma, Sherwood Conf. Contr. Fusion, Gatlinburg, Apr. 27–28, p. 32, TID–7582.

Allis, W. P., S. J. Buchsbaum, and A. Bers (1963), Waves in anisotropic plasmas (MIT Press, Cambridge, Mass.).

Åström, E. O. (1950), On waves in an ionized gas, Arkiv f. Fysik, **2.** No. 42, 443–457.

Auer, P. L., H. Hurwitz, Jr., and R. D. Miller (1958), Collective oscillations in a cold plasma, Phys. Fluids, **1.** No. 6, 501–514.

Bazer, J., and Hurley, J. (1963), Geometrical hydromagnetics, J. Geophys. Res., **68.** No. 1, 147–174.

Beard, D. B. (1964), The solar wind geomagnetic field boundary, Rev. Geophys., **2.** No. 2, 335–365.

Bernstein, I. B. (1958), Waves in a plasma in a magnetic field, Phys. Rev., **109.** No. 1, 10–21.

Cahill, L. J., and P. G. Amazeen (1963), The boundary of the geomagnetic field, J. Geophys. Res., **68.** No. 7, 1835–1843.

Clemmow, P. C., and R. F. Mullaly (1955), Dependence of the refractive index in magneto-ionic theory on the direction of the wave normal, "Physics of the Ionosphere: Report of Phys. Soc. Conf., Cambridge, Sept. 1954", 340–350, Phys. Soc., London.

Dessler, A. J. (1958), The propagation velocity of world-wide sudden commencements of magnetic storms., J. Geophys. Res., **63.** No. 2, 405–408.

Dessler, A. J., W. E. Francis, and E. N. Parker (1960), Geomagnetic Storm sudden commencement rise times, J. Geophys. Res., **65.** No. 9, 2715–2719.

Francis, W. E., M. I. Green, and A. J. Dessler (1959), Hydromagnetic propagation of sudden commencements of magnetic storms, J. Geophys. Res., **63.** No. 10, 1643–1645.

Goldstein, H. (1951), Classical mechanics (Addison-Wesley Press, Inc., Cambridge, Mass.).

Hines, C. O. (1951), Wave packets, the Poynting vector, and energy flow: Part II—group propagation through dissipative isotropic media, J. Geophys. Res., **56.** No. 2, 197–206.

Liemohn, H. B., and F. L. Scarf (1964), Whistler determination of electron energy and density distributions in the magnetosphere, J. Geophys. Res., **69.** No. 5, 883–904.

Mead, G. D. (1964), Deformation of the geomagnetic field by the solar wind, J. Geophys. Res. **69.** No. 7, 1181–1195.

Ness, N. F., C. S. Scearce, and J. B. Seek (1964), Initial results of the IMP–I magnetic field experiment, NASA-Goddard Space Flight Center, X–612–64–95, Apr. 1964.

Sitenko, A. G., and K. N. Stepanov (1957), On the oscillations of an electron plasma in a magnetic field, Soviet Phys. JETP, **4.** No. 4, 512–520.

Spitzer, L., Jr., (1956), Physics of fully ionized gases (Interscience Publishers, Inc., New York, N. Y.).

Stegelmann, E. J., and C. H. von Kenschitzki (1964), On the interpretation of the sudden commencement of geomagnetic storms, J. Geophys. Res., **69.** No. 1, 139–155.

Stix, T. H. (1962), The theory of plasma waves (McGraw-Hill Book Co., Inc., New York, N. Y.).

Störmer, C. (1955), The polar aurora, Part II, Chs. I and II Oxford Univ. Press, Oxford).

Weinberg, S. (1962), Eikonal method in magnetohydrodynamics, Phys. Rev., **126.** No. 6, 1899–1909.

N64-88481

HYDROMAGNETIC STABILITY OF THE MAGNETOSPHERE BOUNDARY*

31963

S. P. TALWAR†

1. INTRODUCTION

It is now generally established that a continual expansion of the solar corona leads to a solar wind which blows radially into space (Parker, 1963). The solar plasma interacts with the geomagnetic field confining it in a region called the magnetosphere. The magnetosphere is asymmetrical in form, and has the shape of a tear drop, having a "tail" pointing away from the sun. An important question about which there has been some controversy in literature is regarding the stability of the interface between the solar wind and the geomagnetic field. Dungey (1955) and Parker (1958) concluded on theoretical grounds that the interface is unstable. Dessler (1961) on the other hand, contends, on the basis of transient geomagnetic activity data obtained at the earth's surface from magnetometers, that the interface is stable. If the latter inference be correct, theories of aurora, magnetic storms, Van Allen radiation which make use of the concept of turbulent solar injection need to be re-examined.

The purpose of the present paper is to investigate the hydromagnetic stability of a model of the interface between the solar wind and the magnetosphere. The solar wind on impinging upon the magnetosphere boundary, slips past it, leading to a situation where Kelvin-Helmholtz instability should play an important role. We regard both the solar beam and the magnetosphere medium to be uniform, infinitely extended plasmas of zero dissipation and having a planar interface. Further, we assume that the solar wind and the magnetosphere medium carry homogenous constant magnetic fields which may be unequal, leading to a current sheet at the interface.

We shall regard the solar wind and the magnetosphere medium as continuum fluids. This assumption should be reasonable as the proton Larr or radius in interplanetary magnetic field can be shown to be small compared with typical dimension characterising the magnetosphere (Axford, 1962).

2. EQUATIONS OF THE PROBLEM

Consider a system of cartesian axes with the z-direction vertical. Suppose that a plane surface of discontinuity of tangential velocity (vortex sheet) and of magnetic field (current sheet) exist at the common, interface ($z=0$) between two semi-infinitely extended homogenous non-dissipative plasmas. Let the uniform velocities and the uniform magnetic fields (both in the horizontal λ- phase) be $\underline{U}_1$, $\underline{U}_2$ and $\underline{H}_1$, $\underline{H}_2$ respectively for $z<0$ and $z>0$ regions. Let ρ_1 and ρ_2 denote the uniform densities of the lower and the upper plasma characterised by sound speeds C_1 and C_2 respectively.

The steady state of the configuration requires

$$\nabla p = 0 \tag{1}$$

for either medium. Here p denotes the plasma pressure. Further the equilibrium of the interface is governed by

$$p_1+\frac{H_1^2}{8\pi}=p_2+\frac{H_2^2}{8\pi} \tag{2}$$

To investigate the stability of the initial steady state, let us consider the effect of a small velocity field disturbance u with components u, v, w in the x-, y-, z- directions.

*Published as *Goddard Space Flight Center Document X-640-64-38*, February; 1964 also published in *Journal of Geophysical Research*, 69 (13):-2707-2713, July 1, 1964.

†Goddard Space Flight Center, on leave from Department of Physics and Astrophysics, Delhi University, India. This work was done while the author held a Senior Resident Research Associateship of National Academy of Sciences.

We write,

$$\begin{aligned}
\underline{v} &= [U_{0x}, U_{0y}, o] + \underline{u}(x,y,z,t) \\
\rho &= \rho_0 + \delta\rho(x,y,z,t) \\
p &= p_0 + \delta p(x,y,z,t) \qquad (3) \\
\underline{H} &= [H_{0x}, H_{0y}, o] + \underline{h}(x,y,z,t)
\end{aligned}$$

where $\delta\rho$, δp, and $\underline{h}$ denote perturbations in density, pressure and magnetic field respectively and the suffix "O" refers to equilibrium values. The perturbations are assumed to be of first order of smallness so that powers higher than the first and their mutual products are neglected.

Assuming the components of the perturbations to vary with x, y, z, t, as (some functions of z). exp. $[ik_x z + ik_y y + nt]$ we may write the linearised hydromagnetic equations as,

$$\zeta_0(n+i\underline{k}\cdot\underline{U}_0)u = -ik_x\delta p - \frac{H_{0y}}{4\pi}(ik_x h_y - ik_y h_x) \qquad (4)$$

$$\zeta_0(n+i\underline{k}\cdot\underline{U}_0)v = -ik_y\delta p + \frac{H_{0x}}{4\pi}(ik_x h_y - ik_y h_x) \qquad (5)$$

$$\zeta_0(n+i\underline{k}\cdot\underline{U}_0)w = -D\delta p - \frac{H_{0x}}{4\pi}(Dh_x - ik_x h_z) + \frac{H_{0y}}{4\pi}(ik_y h_z - Dh_y) \qquad (6)$$

$$(n+i\underline{k}\cdot\underline{U}o)\delta\rho + -\rho_0\nabla\cdot\underline{u} \qquad (7)$$

$$Sp = c^2\delta\rho \qquad (8)$$

$$(n+i\underline{k}\cdot\underline{U}o)h_x = i\underline{k}\cdot\underline{H}_0 u - H_{0x}\nabla\cdot\underline{u} \qquad (9)$$

$$(n+i\underline{k}\cdot\underline{U}o)h_y = i\underline{k}\cdot\underline{H}_0 u - H_{0y}\nabla\cdot\underline{u} \qquad (10)$$

$$(n+i\underline{k}\cdot\underline{U}o)h_z + i\underline{k}\cdot\underline{H}_0 w \qquad (11)$$

and

$$ik_x h_x + ik_y h_y + Dh_z = o \qquad (12)$$

where D stands for d/d_z

Let us now eliminate some of the variables and derive an equation: for, say, the vertical component w of velocity perturbation vector. Multiply equations (4) and (5) by ik_x and ik_y respectively and add to obtain,

$$\zeta_0(n+i\underline{k}\cdot\underline{U}_0)(\nabla\cdot\underline{u} - Dw) - \frac{ik_y H_{0z}}{4\pi}(ik_x h_y - ik_y h_x) + \frac{ik_x H_{0y}}{4\pi}(ik_x h_y - ik_y h_z) = k^2\delta p \qquad (13)$$

where

$$k^2 = k_x^2 + k_y^2 \qquad (14)$$

Eliminate δp from equations (6) and (13) to get,

$$\begin{aligned}
&-\zeta_0(n+i\underline{k}\cdot\underline{U}_0)(D^2-k^2)w + \zeta_0(n+i\underline{k}\cdot\underline{U}_0)D\nabla\cdot\underline{u} \\
&+\frac{H_{0y}}{4\pi}[ik_x\{ik_x Dh_y - ik_y Dh_x\} + k^2\{Dh_{xy} - ik_y h_z\}] \\
&-\frac{H_{0x}}{4\pi}[ik_y\{ik_x Dh_y - ik_y Dh_x\} \\
&\qquad -k^2\{Dh_x - ik_x h_z] = 0 \qquad (15)
\end{aligned}$$

The magnetic field terms in equations (15) reduce to, on using the equations (9) $-$(12),

$$-\frac{(\underline{H}_0\cdot\underline{k})^2}{4\pi(n+i\underline{k}\cdot\underline{U}_0)}(D^2-k^2)w \qquad (16)$$

The perturbations equation (15) is written as,

$$\left[1+\frac{(\underline{k}\cdot\underline{V}_0)^2}{(n+i\underline{k}\cdot\underline{U}_0)^2}\right](D^2-k^2)w - D\nabla\cdot\underline{u} = 0 \qquad (17)$$

where $\underline{V}_0$ denotes the Alfvén velocity vector $\underline{H}_0/\Delta\pi\rho_0)^{\frac{1}{2}}$. The expression for $\nabla\cdot\underline{u}$ is now obtained from the equations (7), (8), (9), (10), and (13) as,

$$\begin{aligned}
&\nabla\cdot\underline{u}[c^2k^2 + (n+i\underline{k}\cdot\underline{U}_0)^2 + (\underline{k}\times\underline{V}_0)^2] \\
&\quad = (n+i\underline{k}\cdot\underline{U}_0)^2 Dw - \underline{k}\cdot\underline{V}_0(\underline{V}_0\times\underline{k})\zeta \qquad (18)
\end{aligned}$$

where ζ is written for $(ik_x v - ik_y u)$.

Let us now evaluate ζ by eliminating δp from equations (4) and (5). We get

$$\zeta[(n+i\underline{k}\cdot\underline{U}_0)^2 + (\underline{k}\cdot\underline{V}_0)^2] = -\nabla\cdot\underline{u}\ \underline{k}\cdot\underline{V}_0(\underline{k}\times\underline{V}_0) \qquad (19)$$

combining (18) and (19), we obtain

$$\begin{aligned}
&\nabla\cdot\underline{u}[\{(n+i\underline{k}\cdot\underline{U}_0)^2 + c^2k^2\}\{(n+i\underline{k}\cdot\underline{U}_0)^2 + (\underline{k}\cdot\underline{V}_0)^2\} \\
&\quad + (n+i\underline{k}\cdot\underline{U}_0)^2(\underline{k}\times\underline{V}_0)^2] \\
&\quad = (n+i\underline{k}\cdot\underline{U}_0)^2\{(n+i\underline{k}\cdot\underline{U}_0)^2 + (\underline{k}\cdot\underline{V}_0)^2\}Dw \qquad (20)
\end{aligned}$$

Substituting equation (20) in equation (17 we obtain

$$\begin{aligned}
&D^2w[\{(n+i\underline{k}\cdot\underline{U}_0)^2 + (\underline{k}\cdot\underline{V}_0)^2\}\{(n+i\underline{k}\cdot\underline{U}_0)^2 + c^2k^2\} \\
&\quad + (n+i\underline{k}\cdot\underline{U}_0)^2(\underline{k}\times\underline{V}_0)^2 - (n+i\underline{k}\cdot\underline{U}_0)^4] \\
&\quad = k^2w[\{(n+i\underline{k}\cdot\underline{U}_0)^2 + (\underline{k}\cdot\underline{V}_0)^2\}\{(n+i\underline{k}\cdot\underline{U}_0)^2 \\
&\quad + c^2k^2\} + (n+i\underline{k}\cdot\underline{U}_0)^2(\underline{k}\times\underline{V}_0)^2] \qquad (21)
\end{aligned}$$

as the equation determining w.

3. BOUNDARY CONDITIONS AND DISPERSION RELATION

For a configuration of two superposed uniform plasmas slipping past each other at the horizontal interface $z=0$, the respective solutions vanishing $at z=\pm\infty$, of equation (21) are written as,

$$w_1=A_1 e^{m_1 z} \quad (z<0)$$

and (22)

$$w_2=A_2 e^{-m_2 z} \quad (z>0)$$

where

$$m_j^2=k^2\frac{[n_j^2+(\underline{k}\cdot\underline{V}_j)^2][n_j^2+k^2c_j^2]+n_j^2(\underline{k}\times\underline{V}_j)^2}{[\{n_j^2+(\underline{k}\cdot\underline{V}_j)^2\}\{n_j^2+k^2c_j^2\}+n_j^2(\underline{k}\times\underline{V}_j)^2-n_j^4]} \quad (j=1,2) \tag{23}$$

Here n_j is written for $n+i\underline{k}\cdot\underline{U}_j$ for brevity.

At the common interface we have to satisfy the following boundary conditions,

(i) the normal component of velocity is continuous. This leads to

$$w_1-U_{1x}\frac{\partial\xi}{\partial x}-U_{1y}\frac{\partial\xi}{\partial y}=w_2-U_{2x}\frac{\partial\xi}{\partial x}-U_{2y}\frac{\partial\xi}{\partial y}=\frac{\partial\xi}{\partial\zeta} \tag{24}$$

where ζ denotes the small displacement of the interface. Using equation (22) we get

$$A_2=A_1\frac{n_2}{n_1} \tag{25}$$

(ii) The normal component of magnetic field is continuous.

One can easily verify that this condition is automatically satisfied as a consequence of (i).

(iii) The normal stress should be continuous across the interface. This means that

$$\delta p_1-\delta p_2+\frac{1}{4\pi}[\underline{H}_1\cdot(\underline{h})_1-\underline{H}_2\cdot(\underline{h})_2]=0 \tag{26}$$

Making use of equations (11), (12), (13), (20), and (25), and simplifying, the above equation (26) gives,

$$\zeta_2 m_2[n_2^2+(\underline{k}\cdot\underline{V}_2)]\frac{[\{n_2^2+(\underline{k}\cdot\underline{V}_2)^2\}(n_2^2+k^2c_2^2)+n_2^2(\underline{k}\times\underline{V}_2)^2-n_2^4]}{[\{n_2^2+(\underline{k}\cdot\underline{V}_2)^2\}(n_2^2+k^2c_2^2)+n_2^2(\underline{k}\times\underline{V}_2)^2]}$$

$$+\rho_1 m_1[n_1^2+(\underline{k}\cdot\underline{V}_1)^2]\frac{[\{n_1^2+(\underline{k}\cdot\underline{V}_1)^2\}(n_1^2+k^2c_1^2)+n_1^2(\underline{k}\times\underline{V}_1)^2-n_1^4]}{[\{n^2_1+(\underline{k}\cdot\underline{V}_1)^2\}(n_1^2+k^2c_1^2)+n_1^2(\underline{k}\times\underline{V}_1)^2]}=0 \tag{27}$$

Substituting the expressions for m_1, m_2 as given by equation (23), we finally obtain the dispersion relation in the following form,

$$\zeta_2[n_2^2+(\underline{k}\cdot\underline{V}_2)^2]\frac{[\{n_2^2+(\underline{k}\cdot\underline{V}_2)_2^2\}(n^2+k^2c_2^2)+n_2^2(\underline{k}\times\underline{V}_2)^2-n_2^4]^{1/2}}{[\{n_2^2+(\underline{k}\cdot\underline{V}_2)^2\}(n_2^2+k^2c_2^2)+n_2^2(\underline{k}\times\underline{V}_2)^2]^{1/2}}$$

$$+\rho_1[n_1^2+(\underline{k}\cdot\underline{V}_1)^2]\frac{[\{n_1^2+(\underline{k}\cdot\underline{V}_1)^2\}(n_1^2+k^2c_1^2)+n_1^2(\underline{k}\times\underline{V}_1)^2-n_1^4]^{1/2}}{[\{n_1^2+(\underline{k}\cdot\underline{V}_1)^2\}(n_1^2+k^2c^2)+n_1^2(\underline{k}\times\underline{V}_1)^2]^{1/2}}=0 \tag{28}$$

The equation (28) is the characteristic equation for n, the parameter determining the stability of the configuration. The above equation is rather unwieldly for discussion in the general case. We shall therefore discuss some special cases, assuming $\underline{H}_1$, $\underline{H}_2$, $\underline{U}_1$, $\underline{U}_2$ to be parallel vectors. This assumption should not be too bad to make as the solar wind drags the solar magnetic field with it during radial expansion.

4. DISCUSSION OF RESULTS

(A) Propagation Transverse to the Direction of Streaming ($k_x=0$, $k_y=k$)

If the streaming plasmas carry the same magnetic field (i.e., $H_1=H_2$) and are characterised by the same γ (ratio of the two specific heats) the equilibrium of the interface (equation (2)) gives

$$\zeta_1 C_1^2+\zeta_2 C_2^2 \tag{29}$$

Also

$$\zeta_1 V_1^2+\zeta_2 V_2^2 \tag{30}$$

The dispersion equation, when the wave vector is non-vanishing in direction transverse to streaming alone, reduces to

$$\zeta_2^2\left[1-\frac{U_p^2}{c^2+V^2}\right]=\zeta_1^2\left[1-\frac{U_p^2}{c_2^2+V_2^2}\right] \tag{31}$$

where U_p is the phase velocity of disturbance, given by $\frac{i\omega}{k}$.

The configuration of two slipping plasmas is, therefore stable, as can be easily verified, for perturbations characterised by $k_x=O$, $k_y=k$. The perturbation, having a non-vanishing wave number k_x along the direction of streaming may, however, bring about instability. Let us therefore investigate the question of the stability of the configuration for the other extreme case namely $k_x=k$, $k_y=O$.

(B) Propagation Parallel to the Direction of Streaming ($k_x=k$, $k_y=0$)

In this particular case, the general dispersion equation (28) reduces to,

$$\zeta_2\left[\frac{V_2^2-(U_2-U_p)^2}{c_2^2-(U_2-U_p)^2}\right]^{1/2}[c_2^2V_2^2-(U_2-U_p)^2(c_2^2+V_2^2)]^{1/2}=-\zeta_1\left[\frac{V_1^2-(U_1-U_p)^2}{c_1^2-(U_1-U_p)^2}\right]^{1/2}[c_1^2V_1^2-(U_1-U_p)^2(c_1^2+V^2)]^{1/2} \tag{32}$$

where c_1, c_2 and V_1, V_2 are related by equations (29) and (30) respectively. If both c_1 and c_2 tend to fininity, the above equation gives the dispersion formula for Kelvin-Helmholtz instability for incompressible fluids in hydromagnetics. This is the same as obtained and discussed in earlier papers (Talwar 1961, 1962).

To simplify the discussion still further let us take $U=-U_2=+U_1$. The equation (32) is then written as,

$$U_p^6(1-\delta)-2UU_p^5(1+\delta)-U_p^4(1-\delta)[U^2+(1+\delta)c_2^2]+4UU_p^3[U^2(1+\delta)+c_2^2(1+\delta^2)]$$
$$+U_p^2(1-\delta)\left[-U^4-6U^2c_2^2(1+\delta)+\frac{2c_2^4V_2^2}{c_2^2+V_2^2}\delta\right]+U_p\left[-2U^5(1+\delta)+4U^3c_2^2(1+\delta^2)-\frac{4Uc_2^4V_2^2}{(c_2^2+V_2^2)}\delta(1+\delta)\right]$$
$$+(1-\delta)\left[U^6-c_2^2U^4(1+\delta)+\frac{2U^2\delta c_2^4V_2^2}{c_2^2+V_2^2}\right]=0 \tag{33}$$

Here $\delta=\zeta_2/\zeta_1$ and we have made use of equations (29) and (30). The equation (33) is a sixth degree equation in phase velocity and to insure that the configuration of two slipping plasmas is stable also for perturbations along the direction of streaming, we require that all the roots of the above equation should be real. In a particularly simple case where we take $\delta=1$ so that different layers of the same plasma fluid permeated with a uniform horizontal magnetic field have a relative tangential velocity, the equation (33) gives,

$$U_p^4-2U_p^2(C^2+U^2)+\left[U^4-2U^2C^2+\frac{2C^4V^2}{C^2+V^2}\right]=0 \tag{34}$$

where now $C_1=C_2=C$ and $V_1=V_2=V$.

Equation (33) gives the following condition for the stability of the configuration

$$U>C\left[1+\left(\frac{C^2-V^2}{C^2+V^2}\right)^{1/2}\right]^{1/2} \tag{35}$$

Here U denotes the relative speed between two layers of the plasma characterised by the sound speed C and hydromagnetic speed V. In the absence of magnetic field we get the condition for stability as $U>\sqrt{2}C$. We thus conclude that whereas a tangential discontinuity in a liquid is unstable for all non-zero speeds in the absence of magnetic field, it is no longer so unless the relative speed is smaller than a certain critical value in a compressible fluid. In the presence of magnetic fields the stability of the vortex sheet in compressible plasmas is improved (as also is the case with

infinitely conducting liquids) in that the critical relative speed below which the configuration is unstable is decreased as a consequence of the magnetic field. Alternatively, we could conclude that any relative speed between two layers of plasma can be stabilized by a strong enough magnetic field given by,

$$V_*^2 = \frac{U^2C^2(2C^2 - U^2)}{(U^2 - C^2)^2 + C^4} \qquad (36)$$

Figures 1 and 2 respectively show the dependence of critical tangential velocity U^* in units of C as a function of V^2/C^2 and that of critical Alfvén speed against Mach number U/C.

The above results, it is felt, should have a useful application in regard to determining whether the interface of the magnetosphere with the solar wind is stable or not. Observations (Gringauz et. al. (1960); Bridge et. al. (1961)) show that the solar wind has a speed $\sim 250-400$ km/sec. and density $\sim$ 10 protons per c.c. under quiet solar conditions. When the sun is active these values may mount to $\sim 10^3$ km/sec. with densities $\sim$ 300 per c.c. The speed of the incoming solar wind is therefore supersonic as it hits the magnetosphere. As the solar wind streams along the magnetosphere boundary, the speed of solar plasma relative to the medium within the magnetosphere boundary is likely to diminish in value. This may be on account of internal motions within the magnetosphere (Gold, 1959). The internal convective motions (roughly along the magnetosphere boundary) may arise as a result of viscous-like interaction between the magnetosphere and the solar wind (Axford and Hines, 1961). If this relative speed is continuously diminished as the solar wind sweeps round from the day side to the night side of the magnetosphere, it may very well happen that the relative speed at some place on the interface becomes less than the critical value given by equation (35) and the part of the interface thereafter becomes unstable. This situation leading to the instability of the magnetosphere—solar wind interface is likely to arise towards the tail of the magnetosphere and under quiet solar conditions when the speed of the incoming solar wind is less anyway. A verification of the above ideas is possible if it can be ascertained that the characteristic effects of instability, e.g., turbulence, high level geomagnetic fluctuations, irregularities of ionisation density are observed during comparatively quiet solar conditions and more so towards the night side.

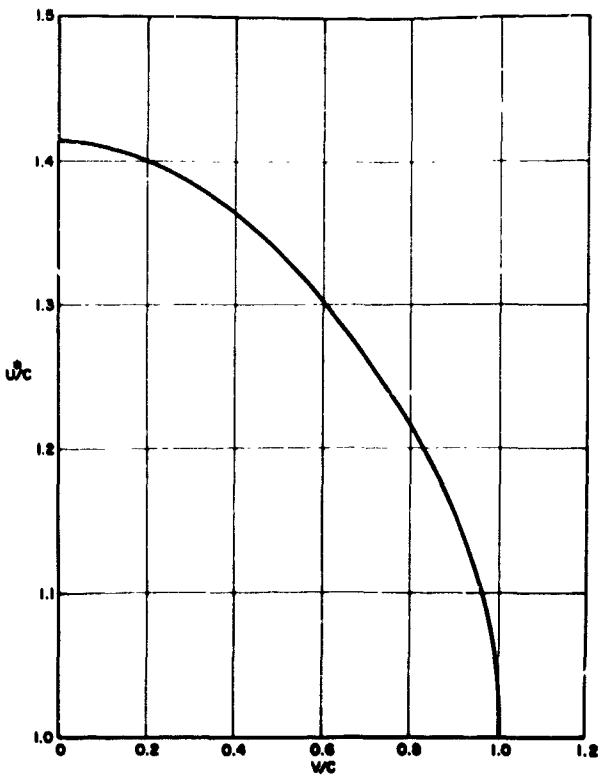


FIGURE 1.

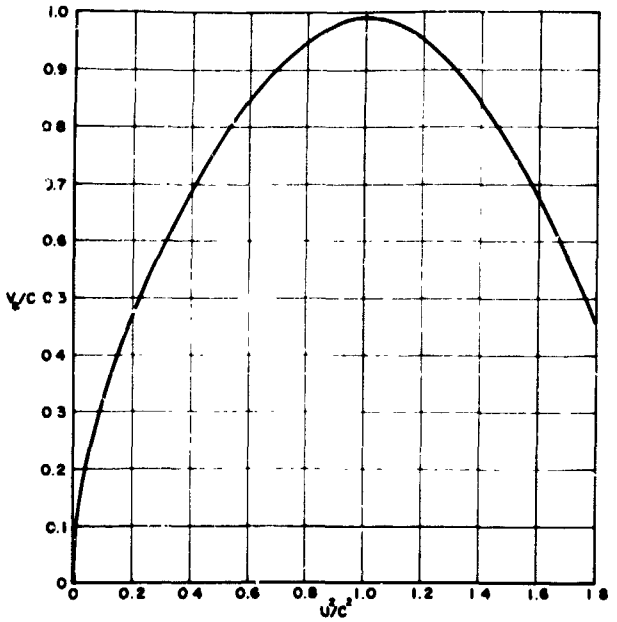


FIGURE 2.

ACKNOWLEDGMENTS

The author is grateful to Dr. Gilbert Mead for helpful suggestions.

REFERENCES

Axford, W. I., (1962), J Geophys. Research 67, 3791.

Axford, W. I., and Hines, C. O. (1961), Can. J. Phys. 39, 1433.

Bridge, H. S., Dilworth, C., Lazarus, A. J., Kyon, E. F., Rossi, B., and Scherb, F. (1961), Proc, Kyoto IAU Conference on Cosmic Rays and Earth Storms.

Dessler, A. J., (1961), J. Geophys. Res. 66, 3587.

Dungey, J. W., (1955), The Physics of the Ionosphere, The Physical Society, London, p. 229.

Gold, T. (1959), J. Geophys. Res. 64, 1219.

Gringauz, K. I., Bezrukikh, V. D., Ozerov and Rybchinskii, R. E. (1960), Dokl. Acad. Nauk SSSR, 131, 1301.

Parker, E. N., (1958), Phys. Fluids, 1, 171.

Parker, E. N., (1963), Interplanetary Dynamical Processes (Interscience Publishers).

Talwar, S. P. (1961), Proc. Nat. Inst. Sci. (India) 27, 263.

Talwar, S. P., (1962), Indian J. Phys. 36, 351.

N 64-31570

CERENKOV SCINTILLATION COUNTER MEASUREMENTS OF THE INTENSITY AND MODULATION OF LOW RIGIDITY COSMIC RAYS AND FEATURES OF THE GEOMAGNETIC CUTOFF RIGIDITY*

W. R. WEBBER† AND F. B. McDONALD

The primary cosmic radiation has been measured with a Cerenkov scintillation counter at various times in the solar cycle at locations having nominal cutoff rigidities of 0.8, 1.6, and 2.7 bv. Results are used here to delineate spectral features of low rigidity protons and α particles and to examine the rigidity and charge dependence of the eleven-year solar molulation. Details of particle transmission near the geomagnetic cutoff are also examined. The measurements indicate: (1) The observed nonsharp cutoff appears to be primarily due to expected transmission effects in a dipole-like field and to instrumental effects, and the actual cutoff is quasi-sharp in the Störmer sense. This is true even though the observed cutoff may be reduced to half that to be expected from the internal field, presumably because of external current systems flowing in the magnetosphere and its boundary. (2) The rigidity dependence of the eleven-year modulation of protons and that of α particles are closely similar above $\sim$1.2 bv. None of the presently proposed models for this modulation agree with our results. The implications of this lack of agreement are discussed, and it is concluded that certain general features of particle motion in the interplanetary medium not considered in present models may be of importance. One such general feature may be diffusive deceleration in the radially expanding interplanetary medium, which is shown to have the same effect as a static heliocentric electric field. (3) The differential spectrum of galactic cosmic rays near sunspot minimum at about 1-bv rigidity is either flat or falling slightly toward lower rigidities. The measurements in the rigidity range below 1 bv are compared with recent observations in this range.

INTRODUCTION

A detailed knowledge and comparison of the galactic cosmic ray proton and α-particle spectra at low rigidities (<2 bv) at various times in the solar cycle is essential in any attempt to select one of the various solar mechanisms proposed in recent years to account for the modulation of the primary radiation. In addition, a study of these spectra at sunspot minimum, when solar modulation effects are small, should provide information about the origin of cosmic rays and their propagation to the earth.

A number of difficulties arise in obtaining and interpreting such spectra. These are:

1. Secondary proton albedo effects, both upward-moving (splash) and downward-moving (reentrant), may contribute a significant background of low energy protons that do not reflect the true primary spectrum.

2. Low rigidity protons are directly produced in the atmosphere above the detector.

3. Solar cosmic rays, both protons and α particles of low rigidities frequently present near the earth, are superimposed on the galactic cosmic ray spectrum.

4. Where the geomagnetic cutoff is greater than instrumental cutoff effects, it may seriously modify the shape of the measured spectrum, not only because of the uncertainty of what geomagnetic cutoff to apply to a particular location, but also because the transmission of particles through the earth's field is not experimentally observed to have a sharp cutoff.

Early (1955–1959) in the solar cycle now being completed, *McDonald* [1956] and *McDonald and*

*Published in *Journal of Geophysical Research*, 69(15):3097–3114, August 1, 1964.

†University of Minnesota, Minneapolis.

Weber [1959] extensively surveyed the spectra and galactic cosmic ray spectral changes at low rigidities. The details of this survey are important for comparison with more recent measurements of the proton and α-particle spectra, and particularly with measurements made near the preceding sunspot minimum in 1954 and the one approaching in 1965.

The McDonald and Webber measurements of interest here were made during 1955–1956 at Minneapolis, where the problems connected with the geomagnetic cutoff and its effects on measured spectra are severe. In fact, these spectral measurements were used originally to define the geomagnetic cutoff existing at Minneapolis in an era when such measurements were of great value in understanding the influence of the non-dipole terms in the earth's field on the cutoff [*McDonald and Webber*, 1959].

At that time it was not clear whether the spectra of protons and α particles measured were the spectra actually existing outside the earth's magnetic field, although it was obvious that features of the low rigidity end of the primary proton spectrum were being measured. We noted that the measured spectra were not consistent with the observations of *Neher* [1957] and of *Winckler and Anderson* [1957] who, from their intensity-altitude data and studies of the latitude effect at balloon altitudes above 51°N, found evidence of a larger flux of low energy particles at that time in the solar cycle (see also *Webber* [1962]). It was also not clear how the cutoff influenced the character of the solar modulation as deduced from several flights at the same latitude, although the cutoff was assumed to operate the same way on all flights. Thus a simple ratio of differential intensities at different times should be a measure of the true solar modulation effect.

With this background information, several observers have recently reported measurements of the low rigidity ends of the primary proton and α-particle spectra. These measurements, made since the last of the McDonald and Webber flights in 1959, and showing features of the spectra during the period of decreasing solar activity, have attracted attention particularly to the problems associated with measuring and interpreting the low rigidity end of the proton spectrum. They have, in fact, provided a picture that is inconsistent with the earlier McDonald and Webber results if these latter measurements are interpreted as the ture primary proton spectra outside the earth's field.

Vogt [1962] and *Meyer and Vogt* [1963] recently reported intensities of low rigidity protons during 1960 and 1961 which considerably exceed those measured earlier by McDonald and Webber. The measurement by McDonald and Webber was made close to solar minimum at a time when the intensity of high rigidity (>1.5 bv) protons was actually greater. Such a modulation would appear to be inconsistent with the accepted pictures for the modulation of primary cosmic rays in a solar environment. Vogt and Meyer have interpreted this excess as possibly due to a steady solar particle emission at low rigidities or as due to protons ejected by large flares and stored for long periods in inter-planetary space. Other measurements made during this later period with emulsions and counters [*Fichtel et al.*, 1964, *Bryant et al.*, 1962] suggests a proton spectrum somewhere between that of Vogt and Meyer and the earlier McDonald and Webber results.

To aid in the understanding of solar modulation effects and to focus attention on the problems associated with the measurement and interpretation of the low rigidity proton spectrum, we report here the results of a balloon flight that carried a Cerenkov scintillation counter detector above International Falls in September 1956. Preliminary results of this flight have been reported elsewhere [*Webber*, 1962]; more detailed results are presented here, and their implications are discussed. The fact that this flight was made when the primary cosmic ray intensity had decreased only slightly from its sunspot minimum value and where the geomagnetic cutoff is appreciably lower than at Minneapolis are of great importance in interpreting these data. In fact, the data are used to determine the transmission features of the earth's field at rigidities near the cutoff value at quiet times near Minneapolis, as well as to determine more appropriate values for the low rigidity part of the primary spectra of protons and α particles near sunspot minimum. In particular, it will be argued that interpreting the spectrums measured by McDonald and Webber at Minneapolis as the spectra of galactic particles existing outside the earth's field (or, what is

equivalent, comparable to measurements at higher geomagnetic latitudes) is not correct.

EXPERIMENTAL RESULTS AND COMPARISON WITH OTHER MEASUREMENTS

The instrumentation used at International Falls is identical to that used in all other flights of this series, including those at Minneapolis, and has been described in detail [*McDonald*, 1956; *McDonald and Webber*, 1959]. Briefly, it consisted of a three-element telescope containing a NaI crystal scintillation counter, a lucite Cerenkov detector, and a Geiger counter tray (Figure 1). The NaI crystal and the Geiger counter tray are the defining elements of the telescope. For each particle which traverses these elements, the outputs from the Cerenkov counter and the NaI crystal are recorded. A notation is also made when a telescope event is accompanied by the triggering of more than one counter in the tray or of one of the systems of guard counters. The data, which are recorded on a continuously moving film with a superimposed time base, consist of a pulse height from the NaI crystal and the Cerenkov detector for each event. The energy calibration is obtained by measuring the most probable energy loss for α particles and protons at minimum ionization, $(E_a)_{min}$ or $(E_p)_{min}$, in both detectors. The selected intervals of ionization or Cerenkov loss are thus defined in terms of $(E_a)_{min}$ or $(E_p)_{min}$ and are matched to the corresponding kinetic energy and corrected for energy loss by ionization to the top of the atmosphere.

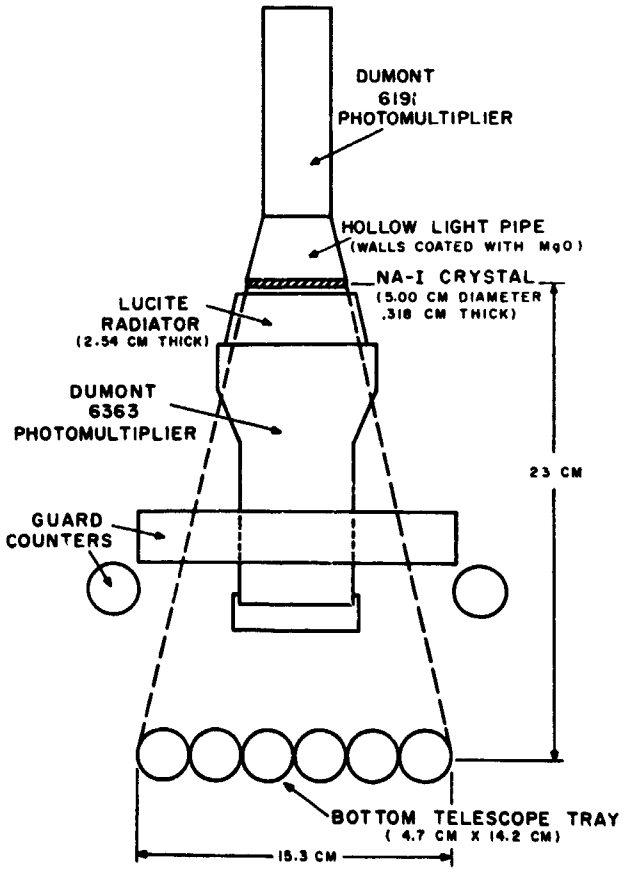


FIGURE 1.—The basic Cerenkov scintillation counter telescope.

The corrections necessary to deduce a true intensity of primary particles at the top of the atmosphere have been discussed in detail by *McDonald and Webber* [1959]. Making these corrections, particularly those for secondary protons produced in the atmosphere and for splash and re-entrant albedo, is greatly facilitated by examining the pulse height distributions in five-minute intervals as the balloons rise to altitude. The intensities of protons and α particles in each energy interval can thus be obtained as a function of altitude. The relatively large geometric factor of the telescope used (7.2 g/cm^2) and high altitude attained (5–6 g/cm^2) greatly enhances this method of analysis.

The same two instruments, unchanged in any way, were used for all flights in the series. A standard set of corrections was applied to all flights, this set of corrections having evolved over the series of flights at a variety of latitudes. This standardization makes possible direct comparison of measured results, in particular, changes in intensity, with a high degree of accuracy, although absolute intensities could be subject to systematic errors.

The flight to be discussed here was launched on September 7, 1956, at International Falls, Minnesota. The average floating altitude of the equipment was 6.5 g/cm^2 (the telescope itself had an additional 8.8 g/cm^2 of material) at a geomagnetic latitude of $59.5 \pm 0.2°$, with a corresponding Störmer cutoff rigidity of 0.8 ± 0.1 bv, according to the most accurate estimates based on the earth's internal field [*Quenby and Wenk*, 1962]. Pertinent details about this flight and others that we will use in our comparison are shown in Table 1.

Preliminary examination of the data from the flight of September 7, 1956, while it was at

TABLE 1.—*Details of Balloon Flights*

Date	Average Altitude, g/cm²	Average Cutoff Rigidity, bv	Bihourly Rate			
			Mt. Wash.	Climax	Mt. Norikura	Huancayo
September 7, 1956; International Falls, Minn.	6.5	0.8 ±0.1	2340 −1.8%	6430 −1.3%	4155 −0.8%	
August 21, 1956; Minneapolis, Minn.	5.5	1.5 ±0.1	2302 −8.1%	6356 −9.0%	4122 −3.2%	5380 −3.2%
June 2, 1959; International Falls, Minn.	5.0	0.85±0.1	2112 +2.4%	5770 +2.8%	3994 +0.3%	5208 +0.8%
July 8, 1961 [*Fichtel et al.*, 1960, 1964]		0.2	2165 −1.6%	5934 −1.7%	4003 0.0	5250 +0.5%
August 6–8, 1961 [*Meyer and Vogt*, 1962, 1963]	5.0	0.2	2130	5840	4002	5278

altitude revealed two unusual decreases of approximately 20 per cent in the counting rate. The rate after the second decrease remained at the lower value, within experimental errors, for the rest of the flight. The indications were that two of the six Geiger counters in the tray had ceased to function, thereby reducing the effective area of telescope. Because of this possible uncertainty in the geometric factor, the data from this flight were not reported in the earlier analysis [*McDonald and Webber*, 1959].

A careful analysis of events that occurred during this flight revealed the following situation. The equipment performed normally while the balloon was rising to its peak altitude. Shortly after it reached peak altitude, the counting rate suddenly changed by 15 to 20 per cent within about 15 minutes. This rate remained constant for a few hours, then the counting rate dropped another 15 to 20 percent within 15 minutes. The rate remained at this lower level (65 to 70 percent of the initial rate) until the end of the flight. These rate changes were consistent with the fact that 2 of the 6 Geiger counters in the telescope tray had ceased to operate. Such a failure would be expected to affect only the geometric factor of the telescope, not the pulse height distribution. Examination of pulse height distributions before and after the changes in counting rate revealed no change in the distribution of pulse sizes. Upon recovery of the instruments, it was found that two Geiger counters were inoperative. Thus the data are believed to represent a true measure of the spectrum of incident particles. A more accurate geometric factor for the telescope was deduced independently by comparing the counting rates of the relativistic protons and α particles with those obtained under normal conditions when

TABLE 2.—*Proton and α-Particles Intensities Measured on September 7, 1956*

Particle and Rigidity Range, bv	Integral Intensity, particles/m² ster sec	Differential Intensity, particles/m² ster sec Mv
Protons		
>1.2	2030 ± 70	
1.0–1.2		0.63 ±0.05
0.83–1.0		0.60 ±0.06
0.72–0.83		0.60 ±0.06
0.64–0.72		0.46 ±0.06
0.58–0.64		0.52 ±0.09
0.54–0.58		0.40 ±0.12
α Particles		
>4.5	86 ± 6	
3.0–4.5		0.029 ± 0.03
2.5–3.0		0.043 ± 0.05
2.0–2.5		0.071 ± 0.07
1.8–2.0		0.100 ± 0.09
1.6–1.8		0.123 ± 0.12
1.4–1.6		0.106 ± 0.10
1.2–1.4		0.088 ± 0.09
>1.2	290 ± 20	

the sea level neutron monitor rates were at approximately the same level as for the September 1956 flight. This further confirmed the fact that the telescope geometric factor during the flight was 65–70 percent of the original value.

The differential intensities of primary protons in the rigidity range 0.54 to 1.20 bv and α particles in the range 1.20 to 4.5 bv deduced at the top of the atmosphere on September 7, 1956, are shown in Table 2 and in Figure 2. Also listed in Table 2 are certain integral intensity values for protons and α particles. The α-particle data in Figure 2 have been multiplied by 7.2 so that they can be directly compared with the proton values. These results are compared with results of two other flights in this general series; for this reason the data from the August 21, 1956, flight launched at Minneapolis are shown in Figure 2. The data from this flight have frequently been used as a measure of the intensity of low rigidity protons existing near the earth at sunspot minimum. These two flights are used to determine the features of the cutoff at Minneapolis.

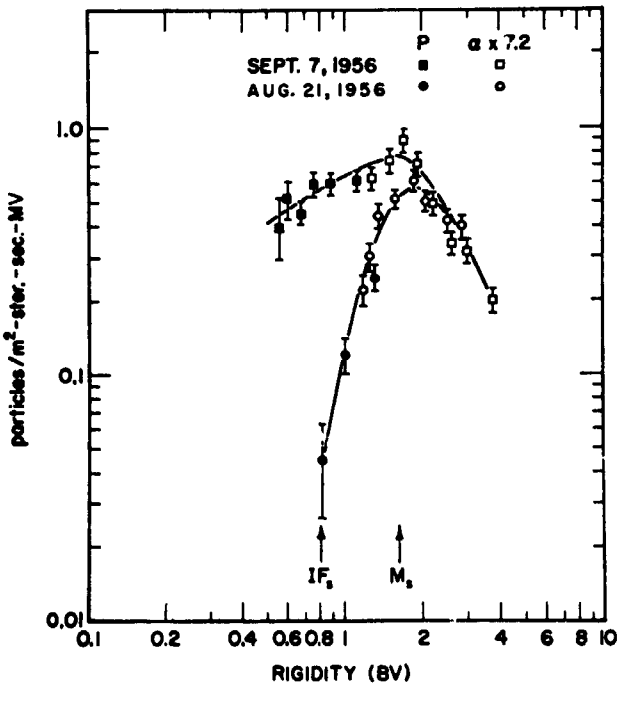


FIGURE 2.—Differential proton and α-particle intensities measured at International Falls on September 7, 1956, and at Minneapolis on August 21, 1956. Nominal Störmer cutoff rigidities at locations of flights are shown by arrows. Alpha-particle intensities are multiplied by 7.2.

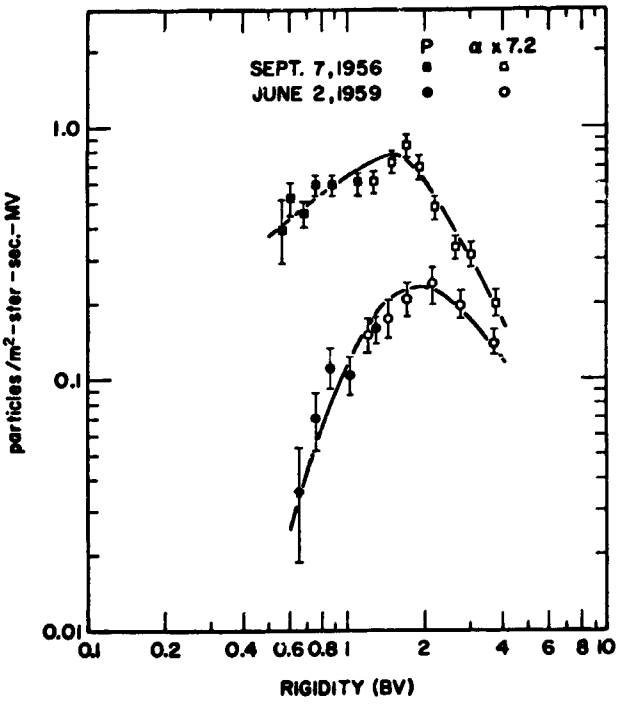


FIGURE 3.—Comparison of differential proton and α-particle intensities measured on two flights at International Falls: September 7, 1956, and June 2, 1959. Alpha-particle intensities are multiplied by 7.2.

In Figure 3 the data from a flight on June 2, 1959, also made at International Falls, are presented. The results of the International Falls flights in 1956 and 1959, when high latitude neutron monitors recorded a 10 percent decrease in counting rate, are compared, and features of the eleven-year modulation of primary cosmic rays near the earth are discussed.

FEATURES OF THE GEOMAGNETIC CUTOFF

During the two weeks between August 21 and September 7, 1956, the primary intensity, as measured by high latitude neutron monitors, changed by only 1.5 percent. Any major difference in the spectra of protons and α particles measured on these two dates might then reasonably be attributed to a difference in cutoff rigidity at the two locations. It is believed that these data were not affected by the very small solar proton event on August 31, 1956 [*McCracken*, 1959]. In particular, we might at this time expect to define certain features of the cutoff rigidity existing near Minneapolis. Examination of Figure 2 reveals a sharp divergence of the two spectra at

rigidities less than 2 bv, although the two spectra are identical above this rigidity within experimental errors. This divergence is noted for both protons and α particles and is grossly unlike any solar modulation process observed during this series of measurements. If the data are adjusted for the modulation that would be expected to accompany a 1.5 percent change in the counting rate of a high latitude neutron monitor, and if we take the ratio of the differential intensities measured near Minneapolis and near International Falls, the results shown in Figure 4 are obtained. If these results are to be interpreted as the effect of a geomagnetic cutoff, then, indeed, they do not fit the usual concept of a relatively sharp cutoff. However, before we discuss this in greater detail, a few preliminary remarks should be made.

Neither flight, at peak altitude, was exactly over Minneapolis or International Falls. The Minneapolis flight floated over southern Minnesota at an average cutoff rigidity of 1.6 ± 0.1 bv according to *Quenby and Wenk* [1962]. The International Falls flight floated over southern Ontario at an average cutoff of 0.8 ± 0.1 bv. Thus the cutoff difference is actually much greater than the difference between Minneapolis (1.35 bv) and International Falls (0.90 bv).

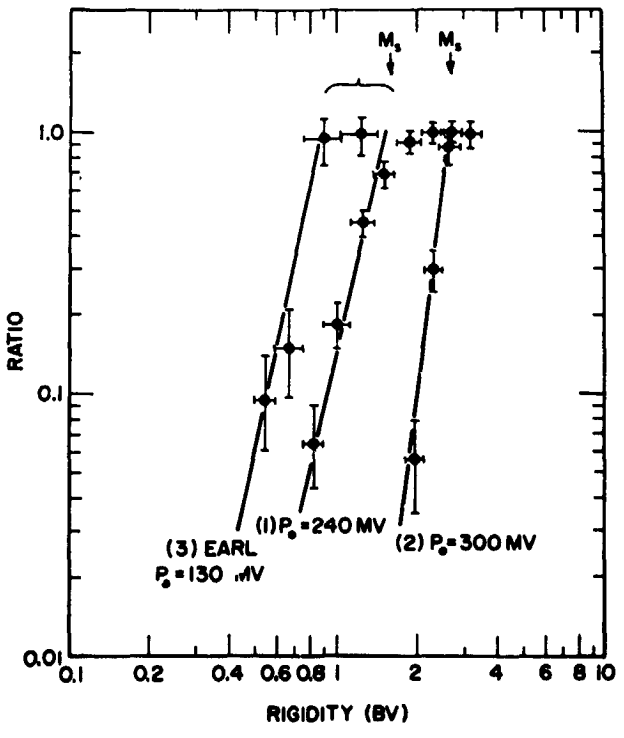


FIGURE 4.—Fraction of particles transmitted through the earth's field as a function of ridigity near the Störmer cutoff rigidity at various latitudes.

Observations at other periods in the sunspot cycle have revealed that the effective geomagnetic cutoffs associated with geomagnetic activity are frequently lower than those deduced only from the earth's internal field [*Winckler et al.*, 1961]. This may be true at geomagnetically quiet times as well [*Webber*, 1962; *Kellogg and Winckler*, 1961]. Such a lowered cutoff may be caused partly by a ring current of trapped protons having energies less than 1 Mev precessing in the earth's field. In view of the present understanding of these trapped protons, such a ring current may be a more or less persistent feature of the earth's magnetosphere during most, if not all, of the sunspot cycle. From a recent study covering the period 1959–1961, *Webber* [1963] has found that the average effective quiet time cutoff rigidity near Minneapolis is about 0.7 of that to be expected from the internal field only. If we associate this cutoff with the Störmer cutoff, it might be expected that a cutoff of about 1.1 bv would be appropriate for the Minneapolis flight and about 0.55 for the International Falls flight.

Turning to the interpretation of the data presented in Figure 4, we have represented the so-called transmission of the earth's field as a function of rigidity by the formula

$$T = \exp\left[-\frac{(P_c - P)}{P_0}\right]$$

where P_c is the rigidity, $T = 1$ (not to be confused with the Störmer cutoff), and P_0, the shape parameter of the cutoff, equals 260 ± 30 Mv in the present measurement. Such a representation is not claimed to have any physical meaning and is used merely to facilitate comparison of these results with those of *Earl* [1962] and *Freier* [1962], who have measured the transmission near the cutoff at Minneapolis during the incidence of solar protons in a magnetically active period.

On the basis of Störmer theory and Liouville's theorem, it is known that in a given direction at a given instant the transmission of particles through the earth's field drops sharply from 1 to zero at a value known as the Störmer cutoff rigidity. In any experiment, however, idealized conditions cannot be attained. Further, there are experimental errors inherent in the measurements of

rigidity, and, finally, there are transmission effects in the earth's field itself (penumbral effects), all of which tend to smear out what in reality may be a sharp Störmer cutoff. Keeping in mind the earlier discussion of the variability of the Störmer cutoff itself, we shall attempt to evaluate some of the effects that could smear the cutoff and see if they are compatible with the measurements shown in Figure 4. These effects include:

1. The penumbral effects in a dipole field. For values of the Störmer constant $\gamma < 0.93$, all orbits are allowed and relatively simple; for $\gamma > 1.00$, all orbits are forbidden. For $0.93 \leq \gamma \leq 1.00$, the orbits are very complex, and it is necessary to delineate the allowed and forbidden values of γ in this range which correspond to allowed and forbidden rigidities in a particular direction only. The width of this region of partial transmission corresponding to the above range of γ is 15 percent of the Störmer cutoff rigidity; e.g., in the vertical direction $P_s = 14.9 \cos^4\lambda/\gamma^2$. Although it is generally assumed that there is nearly perfect transmission in this penumbral region at latitudes greater than 50°, a careful study of the orbit calculations of *Störmer* [1955], *Schwartz* [1959], and *Kasper* [1959] reveals a systematic and clearly defined series of forbidden γ in the region of latitudes greater than 55°. These values of forbidden γ corresponds to perhaps 6 or 8 very narrow rigidity ranges in this penumbral region. These rigidity ranges become narrower, and some of them may vanish altogether, at higher latitudes. In a typical detector with a resolution of about ± 0.05 bv, these forbidden intervals (generally about 0.01 bv wide or less) will smear the cutoff rigidity slightly. This effect is shown in curve 1 of Figure 5 for a nominal Störmer cutoff of 1.1 bv and a detector resolution of ± 0.05 bv.

2. The experimental arrangement is such that the particles arrive at the detector over a finite range of angles. In our measurement the opening angle is $\pm 30°$ from the vertical, an angle that is also typical for the acceptance angle of emulsions and other counter devices. Such an angle introduces a spread of cutoffs that can be deduced from

$$P_s = \frac{59.2 \cos^4 \lambda}{[1+(1-\sin \epsilon \cos \phi \cos^3 \lambda)^{1/2}]^2}$$

where λ is the geomagnetic latitude, ϕ the azimuth measured from east, and ϵ the zenith angle. For a Störmer cutoff of 1.1 bv and an opening angle of $\pm 30°$ from the vertical, this effect introduces extremes of ± 65 Mv from the mean cutoff.

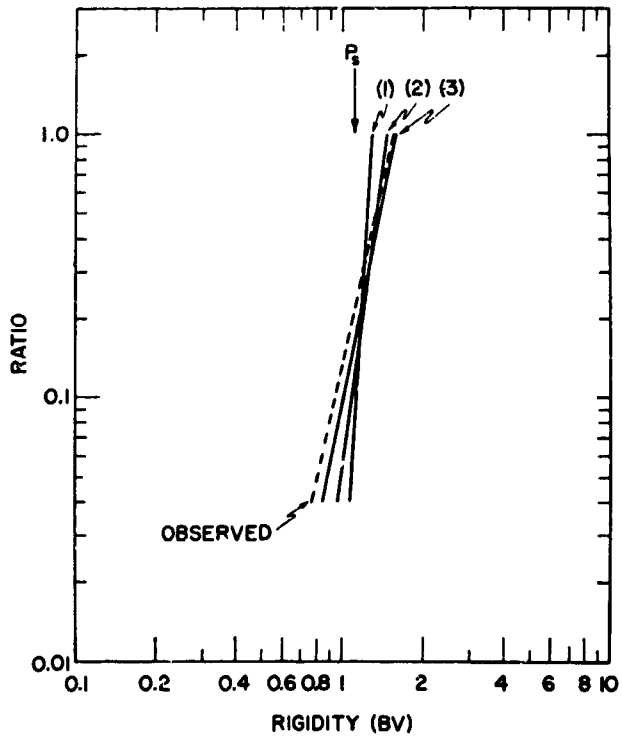


FIGURE 5.—Comparison of observed and predicted transmission of particles through the earth's field at rigidities near the Störmer cutoff rigidity.

3. In a balloon flight of a number of hours' duration, a drift of ± 25 miles from the mean geomagnetic cutoff is typical. Since the data are usually averaged over the whole flight, the drift may introduce a spread of ± 60 Mv at a cutoff of 1.1 bv. Curve 2 of Figure 5 shows the superposition of the aforementioned three effects, which are essentially geomagnetic.

In a very long observation period, a time-varying cutoff could seriously influence the measurements. Since the data were obtained during quiet periods, this effect is not considered.

There are also effects which are instrumental or which occur in the atmosphere above the detector. These include (1) Landau fluctuations in energy loss, both in the detector and in the atmosphere above the detector; (2) fragmentation of heavier nuclei into protons (or α particles) in the atmosphere above the detector (this effect [*Fowler et al.*, 1957] will principally contribute particles below the geomagnetic cutoff for protons

and α particles); and (3) experimental uncertainty in energy or rigidity measurements.

Finally, a small intensity of undentified deuterons or tritons or, in the case of α particles, unidentified He^3 nuclei would serve to diffuse the cutoff.

It is difficult to evaluate these effects analytically. In this study it is estimated that they would introduce a minimum uncertainty in rigidity of ± 10 percent. A superposition of all the effects discussed above is shown in curve 3 of Figure 5. Comparison of this curve with the dotted curve representing the experimental data suggests that we can interpret the vertical cutoff at Minneapolis on August 21, 1956, as a relatively sharp Störmer cutoff at 1.0 to 1.1 bv. These measurements do not exclude a diffuse cutoff over the interval 0.9 to 1.1 bv.

To extend our study of the measured transmission of particles near the cutoff rigidity, we shall examine data from another flight. This was made from Moberly, Missouri, and floated at an average cutoff rigidity of 2.7 ± 0.1 bv [*Quenby and Wenk*, 1962]. The data from this flight, which have been published by *McDonald* [1959] and *Webber* [1962], are consistent with an effective cutoff rigidity of 2.25 ± 0.1 bv. This cutoff was determined by measuring the point of a sudden break in the α-particle differential spectrum. A determination of the primary spectrum from an International Falls flight is available for a time, one year later, when the counting rate of high latitude sea level monitors was within 1 percent of the rate during the Missouri flight. Thus, if the ratio of differential intensities measured in Missouri and at International Falls is taken, the transmission curve shown in Figure 4 is obtained. The shape parameter of the cutoff in this case is given by $P_0 = 300$ Mv. Again, this is compatible with a quasi-sharp cutoff at 2.2 to 2.3 bv and a superposition of the instrumental effects discussed earlier.

Finally, these results (Figure 4) on the shape of the cutoff can be compared with those of *Earl* [1962] obtained near Minneapolis during the course of a solar cosmic ray event on September 4, 1960. Earl's measurements with a cloud chamber are consistent with a shape parameter $P_0 \sim 100$ Mv and a sharp Störmer cutoff at 0.55–0.60 bv. The cutoff was much below normal at Minneapolis at that time because of strong magnetic activity. In addition, because of the high intensity and steep spectrum of the solar particles, Earl could actually observe the change in cutoff during periods of less than one hour. The value of the shape parameter Earl obtained using solar cosmic rays is somewhat less than that obtained using galactic cosmic rays. However, this is to be expected for a number of reasons, one being, of course, that in the short time interval Earl used little balloon drift could occur. Earl's measurements also suggest that a better parameter for defining the experimental shape of the cutoff might be a relative shape parameter P_0/P_s, where P_s is the effective Störmer cutoff. For Earl's work this ratio is 0.18, and for our measurements 0.23 and 0.14 at 1.1 and 2.3 bv, respectively.

We have shown that our results can be explained by a sharp Störmer cutoff, and we believe that an intrinsic diffuse Störmer cutoff greater than ± 0.1 bv would manifest itself in our measurements. At Minneapolis this corresponds to an energy range for protons of 370 to 520, and we cannot rule out a diffuse cutoff over this interval.

Further, it appears that such a quasi-sharp cutoff exists, even though the cutoff may be reduced considerably below that appropriate to the internal field. The mechanisms responsible for the lowered cutoffs are thus limited to those that preserve motion in the earth's field in the Störmer sense. Specifically, such a cutoff would rule out an energy-dependent spatial diffusion through a disordered earth's outer field to positions where Störmer (or Alfvén) trajectories could be followed to the surface of the earth.

Before concluding this discussion of the features of the geomagnetic cutoff, it might be well to ask what, if any, are the effects of a cutoff on the measured spectrum of primary protons at International Falls presented in this paper. To obtain a useful answer to this problem, we recall that slightly south of Minneapolis, where the predicted Störmer cutoff is 1.6 bv, the measured transmission in the earth's field began to deviate from 1 about 1.5 ± 0.1 bv, and the data are consistent with a sharp Störmer cutoff of 1.0–1.1 bv. Most models for the change in cutoff due to external field systems suggest that the fractional change should not be a strong function of latitude between Minneapolis and International Falls. Thus, for

similar conditions, the effective Störmer cutoff during the International Falls flight might be expected to be approximately 0.5 bv and the effective transmission might begin to deviate from 1 at approximately 0.6–0.7 bv. If this happens, the experimental points below 0.7 bv may not represent the true primary spectrum near the earth. We shall consider the implications of this point in the next section.

THE ELEVEN-YEAR MODULATION OF THE GALACTIC COSMIC RAYS AND THE PRIMARY SPECTRUM AT LOW ENERGIES

Although it is possible that the spectrum measured at International Falls does not represent the true primary spectrum below 0.7 bv, we believe that it is possible to compare two flights at this location to obtain useful information on the modulation of the low rigidity primary radiation during the sunspot cycle. This information should be accurate at rigidities above 0.7 bv and below this rigidity as well, provided that the features of the cutoff have not changed between or during the two flights. The measurements we wish to compare were made on balloon flights at International Falls on September 7, 1956, and June 2, 1959, when the counting rates of high latitude neutron monitors decreased by about 10 percent because of the eleven-year variation in solar activity. The differential spectra measured on these two flights are shown in Figure 3. The difference in these two spectra is believed to represent the features of the eleven-year modulation acting during this period. For convenience when comparing these curves with various theoretical models, we have taken the ratio of the differential spectra measured on the two flights and present this information in Figure 6. This ratio, then, should represent the fractional modulation during this period of the galactic cosmic ray spectrum as a function of rigidity in the solar environment. At least two features of the modulation are of great importance: (1) the form of the fractional modulation as a function of rigidity, and (2) the relative fractional modulation of protons and α particles at the same rigidity. Taking the second feature first, we note that the rigidity region in which the proton and α-particle differential intensity measurements overlap is extremely small, from 1.15 to 1.30 bv. In this rigidity range,

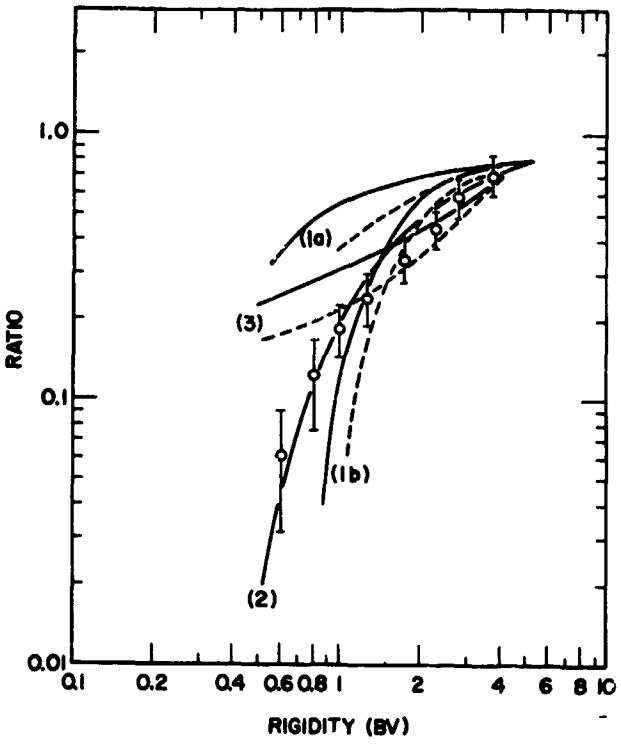


FIGURE 6.—The observed modulation of low rigidity primary cosmic rays during the solar cycle. Circles give the ratio of proton and α-particle intensities measured on June 2, 1959, to those measured on September 7, 1956, at International Falls. Curves 1a and 1b are two possible modulations according to PARKER [1962], curve 2 is the modulation expected according to ELLIOT [1960], and curve 3 is the modulation to be expected upon passage through an electric field of 400 Mv.

during a period when the differential intensity decreased to approximately 20 percent of its sunspot minimum value, there is no difference in the behavior of protons and α particle with an experimental error of approximately—15 percent on the individual measurements. Considering rigidities greater than 1.3 bv, we can make less definitive statements on the behavior of protons and α particles up to at least 4.5 bv by comparing their integral intensities measured above selected rigidities at various times in the sunspot cycle, in the series of flights with the Cerenkov scintillation counter. These data have been examined in detail by *Webber* [1962], who has concluded that the proton and α-particle variations in this rigidity range due to the eleven-year sunspot cycle are consistent within ± 15 percent.

To see how restrictive our data on the rigidity dependence of the fractional modulation and on the relative fractional modulation of protons and

α particle are in terms of a model for the long-term, or eleven-year, modulation of the galactic cosmic rays, let us investigate a few of the recent models that, at least to a first approximation, appear to predict a modulation similar to that observed. These include (1) the solar wind model of *Parker* [1962], (2) the general solar dipole field model of *Elliot* [1960], and (3) the heliocentric electric field deceleration model suggested first by *Nagashima* [1953] and more recently *Ehmert* [1960]. We shall compare the predictions of these models on the rigidity dependence of the eleven-year variation with our observations in some detail, beginning with Parker's solar wind model. The eleven-year variation we obtained corresponds to the quiet day intensity decrease discussed by Parker. For the fractional change in intensity as a function of rigidity, Parker obtains

$$\frac{\Delta j}{j_0}=\frac{4\Omega r \sin\theta}{W(1+4\gamma\Omega r^2 \sin\theta u_0/VaW)}$$

a modulation not unlike that behind the blast wave which Parker suggests causes the Forbush decrease. In the above expression, Ω is the solar angular velocity, r the distance to which the spiral field surrounding the sun is assumed to extend, θ the angle between the lines of force and a radius vector from the sun, γ a geometrical factor probably less than 1, V the average velocity of the solar wind, W the velocity of the cosmic ray particle, a an inner reference distance near the base of the corona, and u_0 a drift factor depending on the characteristics of the assumed general solar (dipole) field. In essence, the decreased intensity is a result of two effects: an outward sweeping of the cosmic ray particles by the solar wind, which dominates at low rigidities and produces a fractional modulation $\Delta j/j_0 \sim 1/W$, and a drift in a latitude across field lines from the polar regions into the equatorial plane (caused by gradients in the equatorial plane), which dominates at high rigidities and produces a fractional modulation

$$\frac{\Delta j}{j_0}\sim\left(\frac{Z}{A}\right)\frac{1}{K}$$

where a is taken to be about $2r$, $\sin\theta\sim 1$, $\Omega=2.7\times 10\alpha^{-6}$ rad/sec,

$$u_0=\frac{1\times 10^6}{Z}\,\frac{K(K+2)}{K+1}$$

and $V\sim 700$ km/sec near solar maximum. The other quantities are effectively unknowns and are determined by the modulation itself. We note that, as $\gamma\to 0$, $\Delta j/j_0\sim 1/W$ and sweeping alone dominates. The modulation is constant at high rigidities and is in disagreement with experiment. The greater the value of γ, the lower the rigidity at which the drift term begins to dominate in the modulation.

First let us try to fit the data near the upper end of our measurements at 5 bv. Then we require that $\Delta j/j_0=0.22$, which means $K/W>0.22$, and $r>40$ AU. Suppose we take $r=65$ AU; then $K/W=0.35$, and $\Delta j/j_0=0.35/[1+120(A/Z)v]$ at 5 bv. For $A/Z=1$, this give $v=5\times 10\alpha^{-3}$, and the rigidity at which the two effects are equal is approximately 7 bv. The form of the modulation, which is essentially velocity dependent, is shown as curve 1*a* in Figure 6. Notice the separation of protons and α particles above 1 bv and the small amount of modulation below 2 bv, both of which are not in agreement with our results.

On the other hand, let us fit the observations at, say, 1 bv. Then $\Delta j/j_0=0.82$, which gives $r\sim 130$ AU. If we also require that $\Delta j/j_0=0.22$ at 5 bv, then $v=1.5\times 10\alpha^{-2}$ and the rigidity at which the two effects are equal is about 2 bv. This is a somewhat better fit and is shown as curve 1*b* in Figure 6. It still, however, shows a separation of protons and α particles above 1 bv which is greater than the experimental evidence indicates. It also predicts $\Delta j/j_0>0.99$ at $P\sim 0.9$ bv, which is clearly outside the range of experimental uncertainty.

In effect our results on the eleven-year variation do not agree with the modulation expected with Parker's solar wind model. Since the presence of a solar wind in the interplanetary environment has been clearly established [*Neugebauer and Snyder*, 1962], it would appear that effects other than those considered by Parker in this admittedly simplified treatment are important to the decrease of cosmic ray intensity. Later we shall discuss what some of these effects might be.

Turning now to the general solar dipole field model of *Elliot* [1960], we see that the fractional modulation can be expressed in the form

$$\frac{\Delta j}{j_0}=\left\{1-\left[\frac{r_e^2 P}{45M_s}\right]^{\alpha(P)}\right\}$$

where

$$\alpha(P)=\left[\left\{1+4K'\left(\frac{P_0+P}{P}\right)^2\right\}^{1/2}-1\right]\Big/4$$

In this equation M_s is the dipole moment of the sun, P_0 is a quantity that is a measure of the strength of scattering centers and that gives a rigidity dependent scattering mean free path, and K_1 is a quantity that reflects the loss of particles into the solar surface. A study of the equation reveals the following effects of these three parameters: increasing or decreasing M_s (keeping α constant) increases or decreases the fractional modulation by a constant amount at each rigidity; increasing or decreasing M_s and decreasing or increasing α to obtain the same modulation at, say 5 bv will, respectively, slightly flatten or steepen $\Delta j/j_0$ versus P. Varying M_s, however, changes the rigidity dependence of the modulation very little.

The parameters K_1 and P_0 are not independent. If we keep the modulation constant at 5 bv and increase K_1 (corresponding to greater absorption by the sun) and decrease P_0, the $\Delta j/j_0$ versus P is flattened. If we decrease K_1 and increase P_0, the rigidity dependence of the fractional modulation is steepened. The best fit to the data, which also corresponds to the observed 5 percent decrease in high latitude neutron monitor counting rate, is shown as curve 2 in Figure 6; it is obtained with $M_s=1.8\times10^{35}$ gauss cm^3, $K=0.15$, and $P_0=2.4$ bv. This would correspond to a mean free path $\lambda_f\sim10^{11}$ cm (independent of rigidity) of the particles in the diffusing region near the solar equatorial plane.

Notice that this model gives excellent agreement with the observed rigidity dependence of the fractional modulation. It also predicts the same modulation for protons and α particles (in agreement with experiment), primarily because of the influence of the general solar dipole field and the application of Störmer theory to the arrival of particles near the earth. The agreement with rigidity dependence is a result of having three parameters available for adjustment. The physical reality of these parameters is not always clear. In particular, it appears that K_1 is not calculated in a consistent fashion. While we cannot reject Elliot's model on the basis of cosmic ray evidence alone (indeed, it gives remarkably good agreement), we feel it has not yet been demonstrated that such a model can predict the required cosmic ray decrease near the earth when reasonable ideas of diffusion are used in the evaluation of the constant K_1.

Finally, we compare our results with the heliocentric static electric field modulation. Here the intensity change comes about because of energy loss and the application of Liouville's theorem, which says that $D=J/P^2B$ is constant during the motion of the particles (D=density of particles in phase space, B=velocity). The form of the change in primary spectrum is given by

$$\frac{dj}{dE_e}(E-V)=\frac{dj}{dE_\infty}(E)\left\{\frac{(E-V)(E-V+2)}{E(E+2)}\right\}$$

where E and V are in bev, (dj/dE_e) $(E-V)$ is the differential intensity at the earth of particles with kinetic energy $(E-V)$, and (dj/dE_∞) (E) is the differential intensity of particles with energy E at large distances from the earth ($V=0$).

Setting $\Delta j/j_0=0.22$ at 5 bv gives the potential $V=400$ Mv and the fractional modulation shown in curve 3 of Figure 6. Note again the separation of protons and α particles and the modulation at low rigidities. The predicted proton—α-particle difference in fractional modulation reaches about 1.5 at low rigidities and is about 1.3 at rigidities between 5 and 1.5 bv where our proton and α-particle measurements overlap. Our results indicate that the modulation is the same in this region to within 10–15 percent. The flattening of the modulation at low rigidities (primarily determined by the shape of the unmodulated spectrum at slightly higher rigidities) is not in accordance with our measurements.

In summary, our results are not in agreement with the concept of a static heliocentric electric field, although experimental uncertainties do not permit us to regard this as a very powerful statement.

None of the three models discussed above for the long-term, or eleven-year depression of galactic cosmic ray intensity appears to agree satisfactorily with detailed measurements of the rigidity and charge dependence of the charges occurring during a particular interval of the recent cycle. The most fundamental difficulty seems to be that the

models predict a pronounced splitting of the proton and α-particle modulation, even above 1 bv, although this splitting is not observed above $\sim$1 bv. Also, the fact that the models cannot explain the observed rigidity dependence of the fractional modulation for protons alone suggests that perhaps some important feature of particle motion is not being considered in these models.

A number of other models have been proposed, and the problem seems mainly to be one of selecting which features of these models must be considered in understanding the eleven-year modulation. In this connection it is useful to use as a guide the recent direct measurements of the interplanetary field configuration and the recent detailed measurements relating to the propagation of the solar-flare-produced cosmic rays.

The direct measurements of the plasma [*Neugebauer and Snyder*, 1962] and magnetic field [*Coleman et al.*, 1962] on Mariner 1 are consistent with the concept that the interplanetary medium is moving outward, at relatively quiet times, with a velocity of about 400 km/sec. In addition, the Imp 1 data show the direction of the magnetic field to be primarily in the equatorial plane (N. F. Ness, private communication, 1964), and there is clear evidence from the Mariner 1 data of magnetic field and plasma variations of a character that would produce randomizing of the motions of cosmic ray particles. Such evidence points to a diffusion-like picture for at least part of the eleven-year modulation. The success with which the diffusion picture has been applied to the propagation features of solar cosmic rays [*Bryant et al.*, 1962; *Hofmann and Winckler*, 1963; *Webber*, 1963] also gives support to this idea.

Such a model has been invoked to explain the eleven-year variation by *Parker* [1958], who solved the steady state diffusion equation with convection,

$$\nabla\cdot(Vj) = \tfrac{1}{3}\nabla\cdot(\lambda\nabla(wj)$$

with some rather simple assumptions about the behavior of V, the outward velocity (convection) of the interplanetary medium, and λ, the effective mean free path between collisions (j is the density of particles, w their velocity). With the assumptions that no energy change occurs and that V is constant and radial, two solutions can be found: (1) If the features of the individual scattering centers are such that $lB \gg \rho B$ (l = scale of scattering centers, ρ = radius of curvature of a particle in field B), then

$$j,(r, E)/j(\infty, E) = \exp^{-const/W}$$

and (2) if the features of the individual scattering centers are such that $lB \ll \rho B$, then

$$j_2(r, E)/j(\infty, E) = \exp^{-const/WP^2}$$

where P is the particle rigidity. This simple form is obtained if we assume only one scale of inhomogeneities (i.e., l = constant), and the constants in the above equations are then determined by values chosen for l, B, V, and the dimensions of the disordered region around the sun [*Parker*, 1958]. Both solutions should be joined at some rigidity roughly equivalent to $PB \approx lB$; thus there will be a kink in the actual modulation owing to the simplifying assumptions about the scattering process [*Parker*, 1962].

Let us compare the experimental data with such a model. First we can use the best estimates of the average values B and V from space probe data; e.g., $B \sim 1\times10^{-4}$ G (in irregularities), and $V \sim 700$ km/sec (near solar maximum).

The value of B determines the point at which the two solutions are joined as $\mathbf{P}\sim1/35$. For the value of $l \sim 2\times10^{11}$ cm used by *Parker* [1958], $\mathbf{P}\sim6$ bv. The modulation at low rigidities is thus about $1/W$, and we have seen in discussing Parker's solar wind model that this is much too flat (curve 1*a*, Figure 5). We must thus attempt to deduce a more reasonable effective value of l. Evidence collected by *Webber* [1963] on the propagation of solar cosmic rays suggests that a possible break in the effective mean free path λ occurs at a rigidity of about 1.6 bv. Below this rigidity λ seems to be roughly a constant $\sim2\times10^{12}$ cm, whereas above this λ appears to vary with energy, although its energy dependence cannot be determined. If we take 1.6 bv as the point at which the scattering process changes character in Parker's simple description, then l becomes $\sim5\times10^{10}$ cm (recall that in Lilliot's model a value of $1\sim3\times10^{10}$ is necessary). The modulation will change from a form $\sim l/W$ to a form $\sim l/WP^2$ at this rigidity. It is in fact very similar to the modulation obtained with Parker's solar wind model, except that in this case it falls off more rapidly at high energies. Therefore, it is a rather poor fit to our

results at both ends. In fact, such a strong rigidity dependence at high energies is clearly incompatible with the long-term modulation observed by neutron monitors.

In spite of these difficulties, it seems that we should seriously consider the features of a diffusion-like model for the eleven-year modulation based on the solar cosmic ray observation. The theory as presently developed is in a most elementary form. The scattering process has been treated in a very simple way. In addition, the variation of λ with r and a distribution of scattering center sizes must be considered. The possible energy change in such a medium has not yet been considered in a workable form. We recognize that the modulation resulting from this diffusive deceleration in an expanding medium is much like that from a static electric field. It may be significant that the electric field modulation, with only a single adjustable parameter, can adequately fit the data over a very wide range and in fact predicts the dependence of the modulation observed at high energies (>10 bv).

The relation between a diffusive deceleration picture and a static electric field is even more intimate than is suggested by the similarity in the modulations. It can be reduced to the fundamental collision of a particle with a magnetic inhomogeneity which, in the case of the expanding medium, results in the inverse Fermi effect. The acceleration (or deceleration) in an individual scattering is caused in the last analysis by the induced electric field $\mathbf{E}=-(1/C)\ \mathbf{U}\times\mathbf{H}$, which is due to the velocity $\mathbf{U}$ of the scattering systems relative to the rest system in the solar system (e.g., the earth). This general feature of accelerating (or decelerating) mechanisms has been pointed out by *Parker* [1958]. *Laster et al.* [1962] have applied the diffusive deceleration picture to expanding magnetized clouds moving outward from the sun and suggest that these effects may dominate in the production of the Forbush decrease. We shall illustrate the basic idea by a specific example and note that the change in energy of the particle after reflection from a moving inhomogeneity is simply $\Delta\epsilon=Ze\int\mathbf{E}\cdot\mathbf{dS}$, where $\mathbf{dS}$ is measured along the path of the particle. Consider a particle moving toward a scattering center with velocity $\mathbf{V}$, the scattering center having a velocity $\mathbf{U}$, small compared with $\mathbf{V}$, and in the same direction as $\mathbf{V}$. If we assume the particle is spiraling around $\mathbf{H}$ and moving into a region of increasing field, $\mathbf{E}$ will oppose $\mathbf{V}$. Then $\Delta\epsilon=2Ze(\mathbf{E})r\sin\theta$, where $r=\epsilon V\gamma/ZeHc$, $\epsilon=mc^2$, $\theta=(\pi/2)-\psi$, and ψ is the pitch angle of a particle. Thus $\Delta\epsilon/\epsilon=-UV/c^2$, a decrease in total energy which is precisely equal to that obtained using the conservation of energy and momentum in an elastic collision with a moving wall. This change occurs only in the longitudinal component of the momentum; the transverse component first increases and then decreases to its initial value during the course of the collision.

The summation of many such collisions will have the equivalent effect of a net decelerating potential without the necessity of a very large uniform electric field near the sun. It is true that, at a particular location in the expanding region, particles will be found with a distribution of energy losses as opposed to the constant energy loss in the static field case; however, the effective modulation should be very similar and should be deduced by the approach of *Laster et al.* [1962].

We believe that, in trying to disentangle the effects of diffusion, energy loss, convection, drift, and other possible general characteristics of particle motion (all of which probably exist to some extent in the interplanetary region), it is important to obtain and study detailed information on the modulation of the galactic cosmic radiation, including measurements of the form of the modulation for different charge components not only at the earth but also at other locations in the solar system. Our results suggest that no single effect dominates over the extended rigidity range, although we feel that a general diffusion process, in which some form of diffusive deceleration occurs, may be an important factor producing the eleven-year modulation.

Finally, it is worth while to compare our results with those of other observers, particularly those for the period after 1959 when our measurements ceased. Comparing α particles first, we note that in the period 1955–1959 the low rigidity differential α-particle spectra over the range 1 to 3 bv measured by the Cerenkov scintillation counter technique and photographic emulsions are in essential agreement within the experimental errors, usually approximately 20 percent. Furthermore, the intensity of these low rigidity α particles is

uniquely related to the counting rates of high latitude neutron monitors during this period.

In Figure 7 we show the differential α-particle spectrum measured on July 7, 1961, by *Fichtel et al.* [1964] using emulsions. The lower dashed curve is the spectrum that would be predicted by the Cerenkov scintillation counter for July 7, 1961, on the basis of the observed neutron monitor intensities and the regression curve established during the 1955–1959 period. Thus it appears that the recovery of this component is proceeding in a manner compatible with the features established during the period 1955–1959.

Turning now to the protons, we note that no measurement comparable to ours is available on low rigidity protons (e.g. <2 bv) during the period 1955–1959. Recently a number of measurements have been made on low rigidity protons

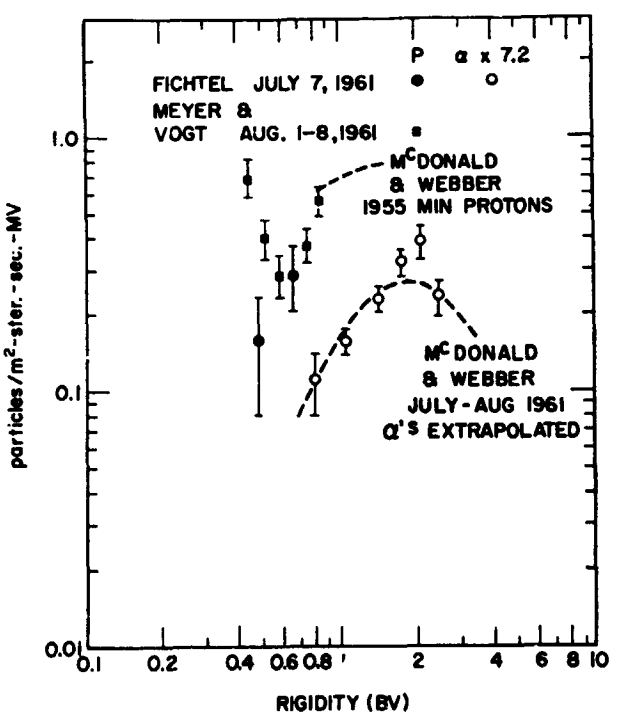


FIGURE 7.—Comparison of predicted proton and α-particle intensities according to regression curves of McDONALD and WEBBER [1962] (dashed lines) and those measured at comparable times during July and August 1961 by other observers. Note the apparent excess of low rigidity protons as measured by these observers and the good agreement for the α-particles (lower dashed curve). Upper dashed curve refers to proton intensity reported in this paper and measured at a time when the neutron monitor rate was near its sunspot minimum value and about 10 percent higher than at the time of the July-August proton measurements shown.

by both counter techniques and emulsions. Comparable measurements were made with these techniques during July–August 1961 [*Meyer and Vogt*, 1963; *Fichtel et al.*, 1964]. Their results, shown in Figure 7, can be directly compared with the α-particle spectrum also given in Figure 7 (see also Table 1). These results cover the range 0.4–0.8 bv and appear to be in general agreement, except that they appear to diverge at low rigidities. Since they are at rigidities less than 0.8 bv, the region of comparison with our result is not large, although some reasonable and interesting comparisons can be made. First, although the rigidity ranges for protons and α particles do not overlap in any one experiment, it appears that, for a reasonable extension of the proton spectrum measured below 0.8 bv in 1961, there is a definite excess of protons relative to the measurements of α particles above 1 bv reported here and to the emulsion measurements as well. Such an excess might be expected in a number of models for the long-term modulation, although the relative excess of protons as a function of rigidity is quite different from that which would be predicted by any of these models. Second, the intensity of protons measured by *Meyer and Vogt* [1963] or *Fichtel et al.* [1964] at 0.8 bv is comparable to our intensity at this rigidity measured at a time when the high latitude neutron monitor rate was 9 percent higher than in the July–August 1961 period. We do not believe the intensity we have measured at this rigidity is depressed because of geomagnetic cutoff effects. If this discrepancy is real, it may indicate the presence of an unusual modulation of galactic protons at low rigidities during the period of decreasing solar activity or the possibility that excess solar protons at low rigidities are contained in the interplanetary field for long periods [*Vogt*, 1962; *Meyer and Vogt*, 1963]. It will take further definitive measurements on the low rigidity proton and α-particle spectra to decide these problems.

REFERENCES

BRYANT, D. A., T. L. CLINE, U. D. DESAI, and F. B. McDONALD, Explorer 12 observations of solar cosmic rays and energetic storm particles after the solar flare of September 28, 1961, *J. Geophys. Res.*, 67, 4983–5000, 1962.

COLEMAN, P. J., L. DAVIS, E. J. SMITH, and C. P. SONNETT, Interplanetary magnetic fields, *Science*, 138, 1099–1100, 1962.

EARL, J. A., Cloud-chamber observations of solar cosmic rays over Minneapolis on September 4, 1960, *J. Geophys. Res.*, 67, 2107–2118, 1962.

EHMERT, L., Electric field modulation, *Space Res.*, 1, 1000–1008, 1960.

ELLIOT, H. Cosmic-ray intensity variation and the interplanetary magnetic field, *Phil. Mag.*, 5, 601, 1960.

FICHTEL, C. E., D. E. GUSS, G. R. STEVENSON, and C. J. Waddington, Cosmic-ray hydrogen and helium during a solar quiet time in July, 1961, *Phys. Rev.*, 133, B818, 1964.

FOWLER, P. H., C. J. WADDINGTON, P. S. FREIER, J. NAUGLE, and E. P. NEY, The low energy end of the cosmic ray spectrum of alpha-particles, *Phil Mag.*, 2, 293, 1957.

FREIER, P. S., The measurement of the cutoff rigidity at Minneapolis using solar protons and α particles from July 10, 1959, flare, *J. Geophys. Res.*, 67, 2617–2626, 1962.

HOFMANN, D. J., and J. R. WINCKLER, Simultaneous balloon observations at Fort Churchill and Minneapolis during solar cosmic ray events of July, 1961, *J. Geophys. Res.*, 68, 2067–2098, 1963.

KASPER, J. E., The earth's simple shadow effect on cosmic radiation, *Nuovo Cimento, Suppl.* 9, 1–26, 1959.

KELLOGG, P. J., and J. R. WINCKLER, Cosmic-ray evidence for a ring current, *J. Geophys. Res.*, 66, 3991–4002, 1961.

LASTER, H., A. M. LENCHEK, and S. F. SINGER, Forbush decreases produced by diffusive deceleration mechanism in interplanetary space, *J. Geophys. Res.*, 67, 2639–2644, 1962.

MCCRACKEN, K. G., The production of cosmic radiation by a solar flare on August 31, 1956, *Nuovo Cimento*, 13(6), 1074–1080, Sept. 16, 1959.

MCDONALD, F. B., Direct determination of primary cosmic-ray alpha-particle energy spectrum by new method, *Phys. Rev.*, 104, 1723, 1956.

MCDONALD, F. B., Primary cosmic-ray intensity near solar maximum, *Phys. Rev.*, 116, 462, 1959.

MCDONALD, F. B., and W. R. WEBBER, Proton component of the primary cosmic radiation, *Phys. Rev.*, 115, 194, 1959.

MCDONALD, F. B., and W. R. WEBBER, A study of the rigidity and charge dependence of primary cosmic ray temporal variations, *Intern. Conf. Cosmic Rays Earth Storm, J. Phys. Soc. Japan*, 17, *Suppl.* A2, 1962.

MEYER, P., and R. VOGT, Primary cosmic rays and solar protons, 2, *Phys. Rev.*, 129, 2275–2279, 1963.

NAGASHIMA, K., The world wide variation of cosmic ray intensity by the electromagnetic field, *J. Geomag. Geoelec.*, 5, 141, 1953.

NEHER, H. V., Cosmic rays near the north geomagnetic pole in the summers of 1955 and 1956, *Phys. Rev.*, 107, 588, 1957.

NUEGEBAUER, M., and C. W. SNYDER, The mission of Mariner 2: Preliminary observations, *Science*, 138, 1095, 1962.

PARKER, E. N., Origin and dynamics of cosmic rays, *Phys. Rev.*, 109, 1328, 1958.

PARKER, E. N., Sudden expansion of the corona following a large solar flare and the attendant magnetic field and cosmic-ray effects, *Astrophys. J.*, 133, 1014, 1962.

QUENBY, J. J., and G. J. WENK, Cosmic ray threshold rigidities and the earth's magnetic field, *Phil. Mag.*, 7, 1457, 1962.

SCHWARTZ, M. M., Penumbra and simple shadow cone of cosmic radiation, *Nuovo Cimento, Suppl.* 9, 27, 1959.

STÖRMER, CARL, *The Polar Aurora*, Clarendon Press, Oxford, 1955.

VOGT, R., Primary cosmic-ray and solar protons, *Phys. Rev.*, 125, 366, 1962.

WEBBER, W. R., Time variations of low rigidity cosmic rays during the recent sunspot cycle, *Prog. Elem. Particle Cosmic Ray Phys.*, 6, 77, 1962.

WEBBER, W. R., Effective cutoff rigidities for solar cosmic rays in the earth's field and the features of field and current systems external to the earth, *Trans. Am. Geophys. Union*, 44, 62, 1963.

WINCKLER, J. R., and K. A. ANDERSON, High altitude cosmic-ray latitude effect from 51° to 65°N geomagnetic latitude, *Phys. Rev.*, 108, 148, 1957.

WINCKLER, J. R., P. D. BHAVSAR, and L. PETERSON, The time variations of solar cosmic rays during July 1959 at Minneapolis, *J. Geophys. Res.*, 66, 995–1022, 1961.

VI. PLANETOLOGY

ON THE EARLY HISTORY OF THE EARTH*

WILLIAM L. DONN†, BERTRAM D. DONN AND WILBUR G. VALENTINE‡

According to current astrophysical models of the sun, solar luminosity has increased with time at a theoretical rate depending on the nature of the model and the "zero-age" of the sun. The determination of the earliest occurrence of running water on the earth's surface may establish a calibration point on the luminosity curve. Our investigation of the nature and age of the oldest rocks suggests that erosion by running water occurred at least 4 billion years ago. Calculation of the evolution of terrestrial surface temperatures based on the most favorable solar model and current estimates of the composition of the early atmosphere indicate that either some revision to the model is necessary or that the earth and sun must be older than 4.5 and 5 billion years, respectively.

The rate of weathering of bedrock and resulting erosion vary directly with temperature, atmospheric moisture content and organic acids from vegetation. Since these all increased with geologic time, it is suggested that bedrock erosion, which would have been slow initially, must also have increased with time. With the presumption of an interaction between sedimentary load and geologic activity, the duration of geosynclinal cycles also should have decreased with time.

The oldest rocks so far dated (about 3.5 billion years) are pegmatites, granites and granite gneiss complexes all of which represent terminal processes in respective geologic cycles, the beginnings of which are inferred at about 4 billion years ago.

The presence of a substantial sialic crust at this time is indicated by the compositions of both igneous rocks and metasediments.

The origin of the continents is examined in the light of these geologic data and inferences, and on the assumption of a pregeologic accumulation of the earth by accretion from cold, solid objects, probably in the form of whisker-like crystals. Theories of origin of continents by segregation from the mantle cannot be reconciled with the time scale provided by calculations from the thermal history of the earth and the currently accepted age of 4.5 billion years. An alternative hypothesis is discussed according to which the continents originated by impact of extra-terrestrial material with subsequent modification by geological processes and mantle admixture.

INTRODUCTION

As the dating of the oldest rocks in the earth's crust is pushed further back in time, the interpretation of early geologic processes begins to merge with the realm of astrophysical thought relating to the origin of the sun and earth and the rest of the solar system. One of the most significant developments in astrophysics, in recent years has been the clear demonstration that stars must be forming continuously. Involved in this scheme is the mechanism of star formation and evolution together with an estimate of stellar ages. In the borderland between astronomy and geology, geochemistry, geophysics and astrophysics provide information and insight in the investigation of the problem of the origin of the earth and solar system. And in the realm of geology, the dating of rocks of great antiquity provides the necessary data for the interpretation of processes in the primeval earth. These areas of investigation must ultimately yield a consistent pattern for they deal with connected and continuous stages in the development of the sun and earth.

Although any current attempt at a synthesis must be a hazardous one, we believe the attempt is now justified in view of the progress in these related areas. And although the analysis may be only partly successful, it may nevertheless indicate both consistencies and inconsistencies in models and theories, and also suggest where vital boundary conditions exist for fitting theories and observations. A suggestion of this nature was made by Schwarzchild (1958) when he asked whether the predicted increase in luminosity of

*Published in the *Geological Society of America Bulletin*, 76:287–306, March 1965.

†Lamont Geological Observatory and the City College of New York.
‡Brooklyn College.

the sun could have had observable geological effects. An important, but brief examination of this point was made by Ringwood (1960).

As part of this investigation we prefer to reverse the thinking of Schwarzchild and Ringwood by determining what bearings earliest observable geologic processes have on theoretical prediction of increasing solar luminosity. This in turn has further implications related to many facets of the primeval earth.

The analysis in this report is divided into three major parts. In the first, a summary of the current background relating to the origin of the sun and earth is given. In the second part, a summary of the data of the oldest rocks is presented with analyses which are the basis for the astronomical and geological conclusions in the third and final section.

ACKNOWLEDGMENTS

We are especially grateful to Professors J. L. Kulp and A. Poldevaart of Columbia University for making available unpublished data on some of the oldest rocks from east Africa and western United States and for the helpful discussion they provided. We also wish to thank Sir Harold Jeffreys and Dr. Kurt Lowe for their reading of, and related comments on the entire manuscript and Dr. Walter Bucher for his helpful discussion on certain of the arguments. The research at the Lamont Geological Observatory was supported by a research gift from the U. S. Steel Foundation.

ASTRONOMICAL BACKGROUND

Origin of Solar Type Stars

Astronomical evidence indicates that stars are forming continuously from clouds of dust and gas as elaborated by Burbidge and Burbidge (1958).

Under suitable conditions (Hoyle, 1960; Spitzer, 1963; Cameron, 1962) these clouds may convert kinetic energy either to radiant energy which is lost to space, or to energy of molecular dissociation and ionization resulting in contraction of the cloud. These processes would lead to increasing temperature and density with the cloud becoming luminous at some stage. Ultimately, the central temperature and density would reach a point where nuclear reactions occur releasing energy at a more rapid rate and finally causing contraction to cease. The initial nuclear energy phase is relatively long in the life history of a star so that most stars have observable features characteristic of this phase. In the Russell-Hertzsprung diagram, which shows stellar luminosities plotted against color or temperature, there is a marked concentration in a narrow region called "the main-sequence" (Struve, 1950; Arp, 1958; Schwarzchild, 1958). Stars in this phase of development, such as the sun, are therefore referred to as "main-sequence" stars. During their evolution they deviate progressively away from the main sequence, but the sun, which is at least 5 billion years old still lies within this group indicating it has undergone little evolutionary change.

In general, stars form as very localized multiple systems (van de Kamp, 1953) with binaries predominating. On a larger scale they seem to form primarily in groups with members numbering in the hundreds. Theories of star formation, reviewed by Burbidge and Burbidge (1959), are not detailed enough to account for these features except in a general way. Whatever the processes of star formation may be, it is believed that main-sequence stars like the sun are surrounded initially by clouds of gas and dust. Evidence to support this view arises from the interpretation (Herbig, 1962) of the T Tauri class of variable stars as being in process of formation. These stars cover a range of spectra which includes that of the sun.

It is not an unreasonable hypothesis that under some circumstances, particularly for single stars, a planetary system may develop out of the surrounding nebulosity. This is the reasoning behind the current theories of the origin of planetary systems (Hoyle, 1960; Cameron, 1962).

In order to interpret both chemical analyses of meteorites and of the earth's crust, and radioactive dating measurements, a knowledge of the primordial composition is required. A thorough discussion of abundances of the elements has been given by Suess and Urey (1956). However, their discussion did not take into account the now well established result that element distributions differ among star groups in the galaxy (Schwarzschild, 1958; Burbidge and Burbidge, 1958; Aller, 1958).

The theory of element formation in stars starts with a pure hydrogen galaxy from which the

individual stars form. Nuclear processes supply stellar energy and in the process convert hydrogen to heavier elements. At a late stage of development stars must eject a considerable fraction of their masses into interstellar space as in the case of supernovae. The resulting nebulosity leads to a second generation of stars enriched in heavier elements. The sun appears to belong to a still later cycle in which the concentration of heavier elements has been increased still further.

Age and Evolution of the Sun

The age of the sun places limits on the time scale for formation of the earth. Initial luminosity and subsequent evolutionary changes for the sun are primary factors determining the temperature of the cloud from which the earth formed and subsequently the earth's surface temperature.

The major initial evolutionary change in a star is its increasing luminosity. Calculations for the sun, first carried out by Schwarzschild (1958) and his associates showed that with an initial composition of 75–80 percent hydrogen, an increase of about 40 percent in luminosity would have occurred in about 5 billion years.

In general the age of the sun has been assumed to be about 5 billion years (Schwarzchild, 1958) on the basis that it is not much older or younger than the earth (4.55 b.y., Patterson, 1956).

Haselgrove and Hoyle (1959) have attempted to deduce the sun's age from the time it would take a homogeneous star of the sun's mass and initial composition to evolve to the present sun. Once the sun became a main-sequence star the conversion of hydrogen to helium in its central regions would destroy the homogeneity by producing a growing core depleted in hydrogen. Hoyle's calculations start with an initial sun composed of 75 percent hydrogen and having a hydrogen to helium mass ratio of 3, a value also adopted by Suess and Urey (1956). Although the rate of helium formation in the galaxy is an uncertain quantity, determinations of the hydrogen to helium mass ratio (Aller, 1958) for luminous hot stars formed within the past 10^8 years vary from 2 to 5. Because these stars are much younger than the sun, they should contain relatively less hydrogen and more helium.

The "zero-age" sun would have had about 0.7 of the present luminosity. With the adopted composition, 5.3 billion years would have been required to attain the present luminosity. The rate of stellar evolution is greater at lower hydrogen concentrations (Hoyle, 1959). Thus, Hoyle's age of 5.3 b.y. (billion years) would be increased somewhat if the initial hydrogen content of the sun were greater than 75 percent, the value for recently formed stars.

Figure 1 shows the luminosity variation of the sun according to Hoyle's calculations. These

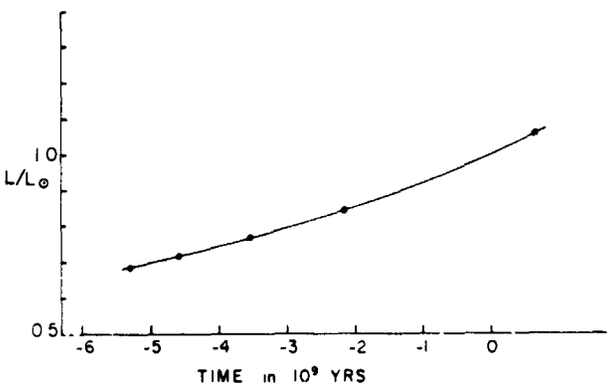


FIGURE 1.—The variation of solar luminosity with time in terms of the ratio of the past luminosity (L) to the present luminosity (L.) (Based on the calculations of Haselgrove and Hoyle, 1959).

results are in general agreement with those of Schwarzschild (1958) and Sears (1959). An increase in solar luminosity by about 30 percent since its formation must have occurred. The geological implications of this effect first called attention to by Schwarzschild (1958) and later Ringwood (1961) will be considered in a later section. Possible implications for solar evolution based on the earliest geological record will also be discussed later.

Origin of the Earth

In the last two decades or so, much effort has been devoted to the problem of the origin of the planets and meteorites. Although no comprehensive, generally accepted theory has appeared, some important conclusions regarding the conditions of formation seem to have become firmly established.

Conditions of accumulation.—Quite conclusive evidence for the formation of the earth and

meteorites from solids at low temperatures has been presented by Brown (1952) and Urey (1957 a, b).

Brown compared ratios of terrestrial to cosmic abundances of the noble gases, Ne, A, Kr, and Xe, with similar ratios for the more reactive gases, H_2O, CO_2, N_2 and O_2. For the latter group the ratios of terrestrial to cosmic abundances are much higher than for the inert noble gases. Because all of the above molecules have comparable molecular weights, gas phase fractionation in a gravitational field could not have played an important part in their accumulation. The reactive gases must therefore have been retained chemically in solids until the earth was large enough to hold them in gaseous form. A consequence of this reasoning is that the atmosphere and hydrosphere must be entirely of secondary origin.

Brown suggests that H_2O was probably in the form of hydrated minerals. This would have limited temperatures to values below which such hydrates are stable.

Urey (1957a) carried out a somewhat similar analysis for less volatile elements and confirmed the low temperature solid accumulation hypothesis. Zinc, cadmium, arsenic, and mercury are relatively volatile or form volatile compounds in a mixture having cosmic proportions of the elements. As no evidence for the depletion of these elements exists, they must have also been captured in the form of solid particles. Urey concludes that temperatures during the formation of the earth were less than a few hundred degrees centigrade. However, temperatures could not have been extremely low or the noble gases would have condensed out or have been adsorbed on the solid particles and their depletion would have been less than observed. Hence, the temperature of accumulation seems limited to a range that approximates present surface temperature ranges.

Mechanism of planet accumulation.—The major attempts to explain the formation of the four inner or terrestrial planets from solids are those by Urey (1957a, b, 1958) and the Soviet School (Schmidt, 1955, 1958; Levin, 1957, 1959). A discussion of the subsequent development of meteorites after accumulation has been given by Anders and his associates (Fish et. al., 1960; Anders and Goles, 1961). As the surface temperature, composition and structure of the growing planet depend upon the accumulation mechanism, this process has an important bearing upon the structure of the earth and the primitive crust.

The mechanism for causing more or less spherical, solid grains to aggregate into large masses has been obscure in all attempts to form large objects. A recent proposal (Donn and Sears, 1963) points out that in the solar nebula condensing particles should have included a large proportion of irregularly shaped grains in the form of plates, filaments and various combinations. These particles would have the properties associated with minute perfect crystals or whiskers, i.e., mechanical strength and elasticity several orders of magnitude greater than for bulk material of the same composition. The shape and strength of these grains provide a means for them to cohere upon collision. The resulting aggregate would be a low density porous object and an efficient collector of all particles striking it.

Composition of accreting material.—The composition of the particles from which the earth formed depended upon the temperature of the nebula and its history. In Urey's two-stage process the fragmented solids were of metallic or silicate composition. If temperatures were low enough, some of the volatile components of the gas would have condensed also. A difficulty with this scheme which Urey (1957a) has emphasized is the mechanism of accumulation of the primary and secondary objects into the planets.

In general one would expect solid grains condensing in the solar nebula to form whiskery particles as noted by Donn and Sears (1963). The condensation of nebular material into spherical particles would have required very special conditions such as proposed by Wood (1963) for the origin of chondrules.

The water in the earth may have resulted from the accretion of ice grains. This would require quite low temperatures in the region of the earth as proposed by Schmidt (1955) and Opik (1962). Brown (1952) and Latimer (1950) proposed that the water was included in hydrated minerals. Hydrated silicates with their complex crystal structure could form from a melt (Urey, 1957a), but this hardly seems possible from the vapor phase. The means of accumulation could have been as simpler hydrated molecules (MgOH, NH_4OH), as ice particles, or as adsorbed water on

silicate grains (Urey, 1952a).

A rather thorough discussion of chemical reactions in the solar nebula has been given by Urey (1952b). Latimer (1950) and Eucken (1944) have also considered this problem. At low temperatures, stable, non-volatile grains are SiO_2, Fe_2SiO_4, Mg_2SiO_4, FeS, and Fe_3O_4 as well as various metallic oxides. Among stable volatile constituents are H_2O, NH_3 and NH_4OH. Presumably the original grains had much the above composition.

Temperature of an accreting earth.—The temperature of the accreting earth was a dominant factor in determining the composition and structure of the earth and the rate of attaining its ultimate characteristics. A complete discussion becomes too involved to include here, but a survey of analyses of such temperatures is presented. The following factors contributed to the heating of the earth as it formed: (1) temperature of the solar nebula, (2) conversion of potential and kinetic energy of particles to heat, (3) release of chemical energy of metastable species stored at low temperatures, (4) work done in compressing material from initial low density, and (5) radioactive decay.

Because of the requirement of a low temperature accumulation the first factor was negligible. Initially process 2 was inefficient because of low impact velocity and small mass. As the earth became larger both factors in 2 became more effective and at some stage a high surface temperature may have developed. However as the terminal stage approached with the earth near full size the rate of growth must have declined, approaching zero as accumulation was completed. This feature has not been taken into account in any of the temperature determinations (ter Haar and Wergeland, 1947; Benfield, 1950, etc.) although it was pointed out by Levin (1957). The net effect was that thermal radiation combined with a slow rate of conversion of kinetic or potential energy to heat would have prevented high surface temperature in the final stages of formation. Local high temperatures probably occurred from occasional impacts at that stage because the accumulating grains would have aggregated into larger objects (Öpik, 1962; Donn and Sears, 1963) which would have generated high temperatures in the impact zone. Urey's lunar calculation (1952, p. 37) suggests that rapid cooling would occur in these locally heated zones.

In a low temperature nebula, chemical heating (process 3) may have been important for objects much smaller than the earth or during the initial stages of growth. When the earth was nearly full size chemical energy would be released immediately on impact and its effect included in the impact heating which would be small at this stage.

Process 4, the compressional heating, has been investigated by Lubimova (1958) and MacDonald (1959) whose results show that the temperature rise in the center would have been only a few hundred degrees. In fact, Lubimova concludes that the rise from the combined effects of impact, radioactive heating and compression would only have been 1200 degrees.

Most calculations of temperatures arising from radioactive decay, process 5, do not apply because in the present case we are concerned with an earth of essentially uniform composition. MacDonald (1959) found for such circumstances that melting temperatures for a metallic phase would be attained 0.5 to 2×10^9 years after aggregation. Differentiation of a core would then occur a long time after aggregation. MacDonald does not consider mantle-crust differentiation but it is apparent that it would take longer yet if it could occur at all.

In summary, present ideas concerning the origin and accumulation processes are far from definitive but certain conclusions can be drawn regarding the characteristics of the earth at the termination of the astronomical phase of its development—the accumulation of its complete mass. The earth's composition was basically that of a uniform mixture of relatively cold solid grains during accumulation. Such accumulation requires the subsequent heating and segregation of the interior for the development of the present structure. By the terminal stage the central region was probably well in the process of heating, although the outer portion must still have retained its primeval composition and structure. At the time of the earliest determinable geological processes the data of the oldest rocks shows that the crust was already well-differentiated. In the following sections we will try to connect the astronomical with the geological phases by a further interpretation of the oldest rocks.

THE OLDEST ROCKS

The other important evidences that must be considered in forming an opinion about early conditions on the earth are the characteristics of the oldest rocks that have been discovered. Regions in which the rocks of greatest known antiquity occur are southeastern Africa, the Kola Peninsula in the USSR, North America from central Canada to northwestern United States and Western Australia. (Although the oldest date known to us was reported by Charbert for a migmetite from French Guiana, the determination —3.93 b.y.—is considered too pulminary for inclusion in this discussion.)

Southeast Africa

The Sebakwian System represents one of the oldest groups of rocks found in the continent of Africa. The micas from a pegmatite cutting this system near Gwelo have been dated by the K–Ar method (Kulp and Poldervaart, personal communication) at 3.44 b.y. which must be regarded as a terminal date for the Sebakwian. We emphasize that this, and all K-Ar dates must be considered as minimal owing to the possibility of escape of argon. Although the conclusions that follow later are based on the experimental values, a revision to greater ages may be justified in the future.

According to the description given by MacGregor (1951), the Sebakwian rocks outcrop in a general northeast-southwest belt from near Mt. Darwin to Filabusi in eastern Southern Rhodesia. However, based on more recent observations, Poldervaart (Personal Communication) considers that only the rocks near Gwelo are authentic Sebakwian. The Gwelo rocks include high grade metamorphics such as granulites, whose original lithology is uncertain, feldspathic or arkosic quartzites and dirty sandstones with lenses of conglomerate.

Resting with high angular unconformity on the Sebakwian System is the Bulawayan System (MacGregor, 1941), which consists of volcanic greenstones with interbedded algal limestones at the base, followed by arkoses, graywackes and ferruginous cherty quartzite. At the Bikita Mine, a pegmatite intrusive into this sequence has been dated at 2.6 to 2.8 b.y. (Kulp and Poldervaart, Personal Communication). All of these rocks have been complexly folded and metamorphosed and "invaded" by granite batholiths.

Rocks of an antiquity equivalent to that of the Sebakwian occur to the south in Swaziland. As described by Hunter (1957), the basement complex designated the Swaziland System, is composed of complexly folded gneisses, granulites and migmatites. The lowest member of the system contains granodioritic gneisses with augen of feldspar and quartz, and granulites with intercalated quartz schists. The only clearly recognizable metasediments occur as lenses of quartzite. The paragneisses are predominantly light-colored, being composed essentially of quartz and feldspars. These rocks grade into darker varieties containing hornblende and pyroxene. The nature and composition of the presumed metasediments described here and in further detail by Hunter suggest strongly an origin from a sialic source.

The granites formed in the metamorphism of the Swaziland System have been dated tentatively at 3.44 b.y. by Allsopp, Roberts, Schreiner and Hunter (1962), who used the *Rb-Sr* method on the entire rock sample. This date which must also be regarded as a terminal date for the Swaziland cycle, is weak because it is based on a rock with a low radiogenic *Sr* content. But it is strengthened by the stratigraphic position below a younger granite with an average age of 3.07 b.y.

Dates from rocks in the Belgian Congo approaching this antiquity have been obtained. Ages of 3.31 to 3.43 b.y. have resulted from the method of modern lead in galena (Holmes and Cahen, 1957). These dates extend the region of oldest rocks northward from the exposures in Southern Rhodesia.

The greatest age so far obtained was determined from a charnokite granite collected by A Poldervaart to the southeast of Lake Victoria in northern Tanganyika from a hill representing the remnants of a thrust sheet (A. Poldervaart personal communications). A *K-Ar* age of 3.6 b.y. was determined at Lamont (Kulp and Engles, 1963). Although this age is consistent with that from other rocks of greatest antiquity in southeast Africa, it might be considered as tentative because of the low potassium content of about 2 percent whereas the more reliable *K-Ar* ages are usually based on a potassium content greater than 6 percent.

Kola Peninsula

The early Precambrian rocks in the Kola Peninsula are highly metamorphosed and complexly folded. Although the age relations are not entirely clarified, a fairly organized picture can be interpreted from recent Soviet publications from which the following descriptions are taken (Vinogradov and Tugarinow, 1961, Lazko, 1962; Pryatkina, 1960; Makiyevshii and Nikolayeva, 1962; Polkanov and Gerling, 1959; Kharatonov, Ozhinskii, Volodin and Polfyerov, 1961).

The oldest rocks are the Katarchean System consisting of a lower biotite gneiss complex (Kola Series) and an upper series composed of a variety of gneisses and amphibolites. This group is now recognized as being older than the lower Archean with which it was formerly included.

The lower Katarchean is a geosynclinal pile which extends eastward from the Kola Fjord (north-northeast of Murmansk). The major rock types are biotite, gneisses and garnetiferous gneisses. According to Maciyevskii and Nikolayeva (1962) the oldest biotite gneisses, which are restricted to the crests of anticlines, often contain dark-colored amphibolitic slates and magnetic quartzites. The younger garnetiferous gneisses contain more alumina and less iron oxide and magnesia.

A *K-Ar* age determinations of 3.44 to 3.48 b.y. by Polkanov and Gerling (1960) were made on biotites from a biotite gneiss of the Kola series in the Veronya River valley. Kouvo and Kulp (1961) have confirmed these dates. We again emphasize that the date refers to the metamorphism of the rocks, hence it is terminal for the lower Katarchean. The dates for the termination of the upper Katarchean, given by Polkanov and Gerling (1960) and Pryatkina (1960), fall between 2.6 and 2.8 b.y.

It is also worthy of emphasis that the presence of the quartzite lenses in the lower Katarchean or Kola Series must indicate the existence of a sialic parent rock from which the lower Katarchean sediments were derived. This will be discussed further in a later section.

The metamorphic rock sequence in the southern part of the Kola Peninsula is not as clearly defined as in the north. Lazko (1962) describes a Kola-like White Sea Series of biotite gneisses, biotite-garnet gneisses, and sillimanite gneisses, containing many amphibolites and pyroxenic slates. These may be correlative with some part of the Katarchean of the northern Kola region according to Maciyevskii and Nikolayeva (1962). In the Kandalash Bay shore region, a belt of granitized gneisses contains amphibolites which include carbonate lenses. A *K-Ar* date of 3.25 b.y. on a granite from the southern part of the Kola Peninsula is given by Polkanov and Gerling (1960). Although positive correlations have not yet been made, this date supports the view that the rocks of the southern part of the Kola Peninsula are at least equivalent to some part of the Katarchean of the northern portion of the Peninsula.

North America

One of the older parts of the Canadian shield is found along the Winnipeg River in southeastern Manitoba. The oldest rocks, which are known as the Rice Lake Series are divided into three phases. (Wright, 1932). The lowest, Mangotagan Phase, is composed of quartzite, graywacke, slate, chert and conglomerate in part recrystallized into granite schist and gneiss.

Numerous pebbles and boulders of granite are found in the Mangotagan conglomerate. The granitic source from which this material has been derived must have been exposed during the time of sedimentation of many of the Mangotagan sedimentary units although no outcrops of the original rock have so far been found. It must be noted that Davies (1956) offered some inconclusive reasons to question the conglomeratic nature of the Mangotagan conglomerate, suggesting the rounded and elongated granite fragments were intrusions into an original schist, but conglomeratic origin proposed by Wright (1932), Derry (1930) and Springer (1949) seems a stronger argument.

The Rice Lake series has been invaded by pegmatites that postdate the metamorphism of the series. *Rb-Sr* age determinations (reviewed by Gastil, 1960) on lepidolite from these pegmatites give an age of about 2.65 b.y. which must be considered a terminal date for the Rice Lake cycle.

The tightly folded Rice Lake formations thus appear to be a correlative of the Bulawayan series of Southern Rhodesia and of the upper Katarchean of the Kola Peninsula. The granite boulders in the lower Rice Lake conglomerate must have been

derived from either a still older orogenic belt or a primordial granite source. If the former, the rocks involved in the orogenic belt might well be correlative with the Sebakwian System, the Swaziland System and the Kola Series.

Rocks of very great antiquity seem to crop out in a belt extending from the southern part of the Canadian shield into the northern Rocky Mountains. To the south of the Rice Lake area, very ancient granite gneiss is exposed in localities along the Minnesota River Valley in southwestern Minnesota. Catanzaro (1963) obtained *U-Pb* isotopic ages that are recognized as discordant and low, of 2.64 to 3.28 b.y. from two zircon samples from the Morton granite gneiss, and one from the Montivideo granite gneiss. Assuming the discordance to be the result of lead loss from continuous diffusion, a minimum age of zircon crystallization of 3.3 b.y. is interpreted. Assuming episodic lead loss during distinct metamorphic stages, the minimum age is 3.55 b.y. These oberrvations again lead to the inference that a substantial sialic crust was present during one of the earliest recognizable stages of geologic history.

Another study of the *U-Pb* system in zircons from ancient rocks was made by Catanzaro and Kulp (1964) for the Beartooth Mountains of Montana. The major rock units in the area of Quad Creek, where many of the samples were taken, occur in a tightly folded syncline. This group consists of Precambrian granite gneiss, migmatite, metasedimentary rocks (primarily paraamphibolites and quartzites derived from arenites, lutites, and possible calcilutites) and metamorphosed mafic dikes and stocks (Eckelmann and Poldervaart, 1957). The ages of 2.6 to 3.1 b.y. are based upon $Pb^{207}-Pb^{206}$ in zircons and are recognized as discordant and low. Assuming episodic lead loss during known metamorphic stages, a minimum age of 3.12 b.y. is obtained. On the basis of continuous diffusion of lead, a minimum age of 3.5 b.y. is interpreted.

It must again be noted that these ancient Beartooth metasediments contain zircon grains of much greater antiquity than the matrix. Although zircon occurs in most igneous rocks, it preponderates in granites which seems the likely source of the sedimentary zircon. This, together with the presence of metamorphosed quartz sandstones, the source of which has not yet been discovered, indicate the early erosion of a sialic parent rock.

The Beartooth metasediments in the Stillwater River Area, about 25 miles northwest of the Quad Creek region have also been the subject of recent reinvestigation. The Stillwater igneous complex, described in detail by Hess (1960) is intruded into the Beartooth group which have been described by A. Poldervaart (personal communication) to consist of some quartzite and ironstone members within a main mass of metamorphosed potassic shales. A *K-Ar* date on biotite from the metamorphic aureole bordering the intrusion was found to be 3.2 b.y. (J. L. Kulp, personal communication). In addition to dating the intrusion, this date establishes at least a minimum age for the deposition of the original Beartooth sediments whose composition here again implies an earlier sialic source.

In the investigation of ancient rock ages in northwestern Wyoming, to the south of the Beartooth Range, Giletti and Gast (1961) and Bassett and Giletti (1963) report *Rb-Sr* ages up to 2.7 b.y. on muscovite from pegmatites that cut highly metamorphosed rocks. These authors conclude that these ages, which represent terminal metamorphism of a sedimentary sequence, are extensive in western and central Wyoming. The metasediments of Quad Creek, northwest Wyoming and the Stillwater area appear to be a uniform geologic realm. The first two seem to be correlative with the Rice Lake, Bulawayan and upper Katarchean. The Stillwater may be older, possibly corresponding to the still older rocks in Africa.

Australia

According to Wilson, Compston, Jeffery and Riley (1960) the oldest rocks in Australia show essential unity in structure and age in a region of about 500,000 square miles in western Australia. Although the most accurately dated area is at Boya, near Perth, where the total rock method on granite gives an age of 2.7 b.y. for the emplacement of the granite, *Rb-Sr* and *K-Ar* age determinations over the vast area referred to above also give 2.7 b.y. (within a 4 percent spread) at many localities. This latter age was determined on lithium-bearing pegmatites which inject the Kalgoorlie greenstones and thus indicate the date of

metamorphism of this widespread rock system. The greenstones were derived from basaltic and andesitic lavas, pyroclastics, basic and ultrabasic intrusions, banded iron formations and ill-sorted classic sediments.

The greenstones occur in long narrow belts surrounded on all sides by granite or granite gneiss which are regionally concordant with the greenstones. The gneisses, which are in part at least, "remetamorphosed" into a complex containing charnockitic rocks of all types, are considered by Wilson (1958) to be the oldest rocks present, and to represent the more acid crust on which the original greenstone source rocks were laid down. No dates on the ancient gneisses have been given although age determinations are planned. At present, therefore, any tentative correlation with rock groups of great antiquity on other continents is difficult.

One of the clastic units included in the greenstones from Southern Cross has more recently been described by Marshall, May and Perret (1964) to be a fine-grained quartzite consisting of angular quartz grains.

Direct Conclusions from the Geologic Data

The data on the oldest rock sequences provide the basis for establishing certain limits regarding conditions and events in the earliest geologic phase of earth history. In the interpretation of these data certain assumptions or generalizations are involved. Although further study may alter or refute some of the conclusions, the method of analysis seems valid.

The terminal date for the earliest known geosyncline cycles is about 3.5 b.y. In all cases it is necessary to assume a preceding time interval sufficiently great to accumulate geosynclinal piles and raise them to a high degree of metamorphism. It seems possible to estimate the length of this interval by comparison with the interval of later cycles whose time limits are better known.

As noted earlier, the next major diastrophic events in Southern Rhodesia and the northern Kola Peninsula occurred at about 2.6 to 2.8 b.y. In view of the long time interval between successive terminal events it seems reasonable to conclude that the earliest cycle culminating at 3.5 b.y. began at least 4 billion years ago. This argument is given strong support in a later section which emphasizes the probable slow rate of bedrock erosion in an early geologic time as a result of low temperature, low atmospheric water vapor content and absence of organic acids.

This line of reasoning leads to the conclusion that subaerial erosion, involving running water, was in operation at least 4 billion years ago. Because these estimates are based in part on the *K-Ar* method of isotropic dating, they may be considered minimal owing to the possible loss of argon as noted earlier. It may also be concluded that at least a portion of the continental crust undergoing erosion about 4 billion years ago consisted of quartz-bearing sialic rock material.

The timing of events for North America is not as definitive but nevertheless leads to conclusions of great antiquity regarding the age of the first erosional cycle. The terminal event of the first clear-cut sedimentary cycle in Manitoba is about 2.6 to 2.7 b.y. If the granite inclusions in the Rice Lake conglomerates are considered to be primordial crust, the first cycle would have begun about 3.5 billion years ago. If, however, this granite represents the culmination of a preceding geosynclinal cycle, then the first cycle could also be extrapolated back to at least 4 billion years. In the Beartooth region, if the estimated minimum age of 3.5 b.y. for the Beartooth zircons is valid, then a similar history for this region can be argued, particularly in view of the quartz-bearing sediments which must certainly have been derived from sialic material. The Stillwater date certainly establishes the existence of a sedimentary pile at 3.2 b.y. which may have begun about 3.5 b.y. As the Stillwater sediments indicate a sialic source they may represent the erosion of a primordial granite crust or an earlier cycle granitized mass.

Evidence for an initial acid crust over at least a large part of Australia is also strong in view of the granite and granite gneiss basement in the western part of the continent, and the quartzite composed of angular particles in the overlying greenstones. The latter must be greater in age than 2.7 b.y., the time of pegmatite injection and metamorphism. Because the greenstones are mostly igneous in origin, it is difficult to estimate the time of initiation of this system as is possible with geosynclinal deposits. Certainly the remetamorphosed gneisses, described as the basement, are

considerably older. If first cycle, they probably represent primeval acid crust; if second cycle, they must represent geological activity involving erosion of preexisting acid rocks at least 3.5 billion years ago, and possibly 4.

A summary of our direct conclusions from the geologic data are (1) at least parts of three major continents were present about 4 billion years ago; (2) quartz-bearing sialic material existed in these continents which must have been larger than the areas in which rocks of this antiquity have been found fortuitously, (3) subaerial erosion by water transport was accumulating sediments in geosynclinal troughs about 4 billion years ago; and (4) above-freezing temperatures must have prevailed over at least part of the earth's surface 4 billion years ago.

GEOLOGICAL AND ASTRONOMICAL IMPLICATIONS OF THE OLDEST ROCKS

Solar Luminosity

The data and interpretations given above lead to the conclusion that running water required for rock erosion and sediment transport existed at a number of places having a wide distribution over the earth's surface about 4 billion years ago. Consequently the temperature of much of the surface atmosphere must have been above freezing for much of the year. This leads directly to a consideration of the source and maintenance of the atmospheric heat energy and an evaluation of astrophysical theory of solar evolution.

The ultimate source of this energy is the sun and the level of this energy is fundamentally determined by solar luminosity, which has been slowly increasing according to astrophysical theory. If we consider the present atmosphere and neglect that portion of the solar constant lost through reflection by the earth and atmosphere (about 43 percent), only one-fourth of the isolation is directly absorbed by the atmosphere and the remaining three-fourths by the earth's surface. This results from the transparency of the atmosphere to the visible spectrum which composes about 90 percent of the solar constant. For the most part, the lower atmosphere is directly heated from below in order of importance by (1) absorption of long wave (infrared) reradiation from the earth's surface, (2) release of latent heat of condensation from water originally evaporated through the energy of the short radiation by the oceans, and (3) direct conduction from the surface followed by turbulent (convective) transfer.

Infrared absorption in the atmosphere is accomplished chiefly by water vapor and carbon dioxide with the former being much more important at present in view of its greater atmospheric abundance and its much more effective absorption over a broad spectral range. The composition of the atmosphere thus plays a very important role in its heat budget. This composition must be considered for the primitive atmosphere before solar luminosity can be directly related to early surface temperature conditions.

We can make a good estimate of the present greenhouse effect of the atmosphere from the infrared absorption just described. The radiation temperature of the earth's surface is 246°K for a terrestrial albedo of 40 percent. As the average surface temperature of the atmosphere is 288°K (International Standard Atmosphere), the present greenhouse effect of 42°K is quite significant.

To consider water vapor first, a mutual interaction exists between air temperature and water vapor content. The temperature of the air determines the vapor density of water whose greenhouse effect further increases the heat budget of the atmosphere. With a fairly unlimited oceanic supply and a given lumi osity, equilibrium air temperature and water content will tend to form as at the present time. Unless the air temperature were higher than that at present, from the effect of radiation absorption by other constituents, the water vapor density could never have exceeded the present. Even the present atmosphere, with an available water supply covering 70 percent of the earth's surface is not saturated.

If the oceans accumulated slowly through exhalation from the earth's interior, as discussed by Rubey (1951) and supported by Kulp (1951), the their early volume and surface area must have been quite limited. Hutchinson (1954) also estimates the prebiological ocean volume to be no more than 5 to 10 percent of the present oceans. On the basis of a probable slow accumulation of the hydrosphere together with the dependence of temperature on solar luminosity, we must regard water vapor as unimportant in affecting the temperature in the primitive atmosphere.

At present these is also incomplete agreement about the composition of the early atmosphere as regards the other volatiles, including those which might strongly affect the atmospheric heat budget. Urey (1952, 1959) considers an early atmosphere as consisting of water vapor, hydrogen, methane and ammonia with the quantities of each being uncertain. Although methane has some absorption bands in the infrared, the intensity and spread of this absorption is very small compared to water vapor, and even to carbon dioxide. Further, as discussed by Hutchinson (1954), oxygen, produced by photo-decomposition, would have oxidized methane to carbon dioxide, which is suggested as an important source of the latter. If exhaled slowly during geologic time, as suggested for other volatiles, the concentration of methane may never have been great. Urey (1952) also notes that in the presence of water carbon dioxide pressure must remain low because of reactions with silicates. (Apparently carbon dioxide would have had a low concentration in the early atmosphere and therefore would have been of little importance in the atmospheric heat budget.)

Rubey (1955) rejects Urey's estimate of the composition of the early atmosphere on the basis of geochemical observations and reasoning, proposing instead that nitrogen and carbon dioxide were the dominant gases present. Rubey examines two primitive atmospheric models having high and low densities of original volatiles. He concludes from the geologic record that the carbon dioxide concentration could not have been excessive having accumulated gradually during geologic time.

Ammonia is the only other volatile suggested for a primitive atmosphere that is effective in the absorption of infrared in the band important in the greenhouse effect. However, the high solubility of ammonia in water would prevent the accumulation of an ammonia atmosphere significantly denser than the aqueous content of the present atmosphere. Presumably, the production of large quantities of ammonia would involve the simultaneous generation of large quantities of water into which the ammonia would dissolve indicating that it could not cause a large greenhouse effect in the primitive atmosphere.

In his three-stage model for the evolution of the atmosphere Holland (1962) proposes that methane and possibly hydrogen were the major constituents in the first stage, lasting about 0.5 b.y. (or not less than 4 b.y. ago). In his second stage, which lasted about 3 b.y. during which the pressure probably never exceeded one atmosphere, nitrogen became the dominant component, with water vapor and carbon dioxide important minor components, and with methane and ammonia being trace constituents. Holland's first and early second stages would embrace the interval important to this investigation.

The most reasonable conclusions we can make at this time from the observations and deductions of geochemistry are that the early atmosphere was probably no more, and very likely less, absorbent in the infrared region than the present atmosphere. Surface air temperatures must therefore have been controlled primarily by the direct influence of solar luminosity rather than the greenhouse effect. As such, the time of earliest detectable liquid water able to cause erosion and transportation of sediments (about 4 billion years ago) establishes the earliest time at which solar luminosity reached a value necessary to sustain above-freezing temperatures.

In applying this conclusion to astrophysical theory of solar evolution, we can utilize the luminosity calculations of Haselgrove and Hoyle (1959) which lead to a solar age of 5.3 b.y. In view of the uncertainty of nuclear data and theory and initial solar composition, Hoyle (1960) has estimated the uncertainty in solar luminosity calculations to be about 25 percent. The geologic results may thus be of considerable value in establishing permissible values for these calculations.

Table 1, which has been developed from the age calculations of Haselgrove and Hoyle, shows a time scale of solar evolution and resulting terrestrial surface temperatures according to a choice of initial conditions. Time before the present is shown in the first column, the ratio of past to present solar luminosities in the second column, the ratio of past to present black body temperatures in the third column (based on the fourth root of the luminosity) and three sets of values of resulting surface air temperatures in the last three columns.

In the column headed T_1, the ratio $T-T_0$ is applied to the present average surface air temperature of 288°K. Because studies of Cenozoic

TABLE 1.—*Variation in solar luminosity and related terrestrial surface air temperatures*

Time b.y.	$\frac{L}{Lo}$	$\frac{T}{To}$	Surface Air Temperature °K		
			T_1	T_2	T_3
0	1	1	288	295	246
−1.5	.89	.97	279	286	239
−3.0	.80	.95	272	279	233
−4.0	.74	.93	267	273	288
−4.5	.72	.92	265	271	226

climate indicate a slow cooling which culminated in the Pleistocene Ice Age, preceded by a warmer and more equable global climate, it may be more correct to use an estimated pre-late Cenozoic average of 295°K. Column T_2 is based on this value. As noted earlier, the present radiative equilibrium temperature of the earth (which omits the present greenhouse effect) is 246°K. Column T_3 indicates past temperature based on this value.

If terrestrial surface temperatures crossed the freezing threshold 4 billion years ago, the values in Column T_2 seem to give the most compatible fit between obsevration and theory. We must emphasize that the assumptions involved in determining T_2 are the most favorable for producing an early crossing of the freezing threshold. As a consequence of the flatness of the luminosity curve in Figure 2, the change in a billion years is only 3 degrees so that the theoretical resolution of the time—temperature points is rather coarse. Small changes in the theoretical $L-L_0$ can shift the freezing threshold over a long time interval. According to Column T_1 (based on present temperatures) the freezing point would not have been crossed until about 2 billion years ago. Although the data in Column T_3 do not apply directly to recent air temperatures, which, as noted earlier, are controlled by a strong greenhouse effect, they might very well apply to the very early atmosphere. The resulting greenhouse effect could have been considerably less than that of the present. It may therefore be reasonable to trace the surface air temperature curve from the lower part of Column T_3 to the upper

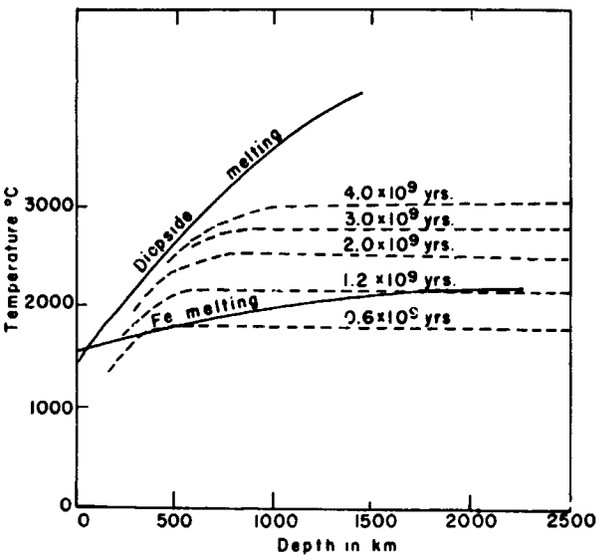


FIGURE 2.—Temperature variation in a homogeneous earth of chondritic composition having an initial temperature of 1300°C. (After MacDonald, 1959)

part of Column T_2 as the greenhouse effect became more important at which time a steepening of the temperature curve would occur.

Before evaluating these temperature relations further, another important consideration must be examined. It was noted previously that the current radiation temperature of 246°K is based on an albedo of 40 percent. However, if the early temperatures were sub-freezing as indicated by present astrophysical theory, then a considerable terrestrial frost coating must have existed, comparable to the probable frost coating responsible for the Martian polar cap. Such a coating requires a very minute amount of water and would have raised the albedo above 40 percent resulting in lower early temperatures in each of the last three columns.

The importance of a highly reflecting primitive frost coating is strongly emphasized from recent theoretical and empirical meteorological observations. For an atmosphere transparent to incident radiation, Budyko (1962) has calculated that a snow and ice cover over the entire earth for a short period would decrease the mean temperature by about 100°C. His work suggests that the present strong contrast in zonal temperatures between low and high latitudes is the result of reflection from the ice cover in both polar regions.

On a more empirical basis, Namias (1963) showed that temperatures over central United

States were as much as 6°C below expected values as a consequence of the additional reflection from somewhat extended snow cover in this region from mid-February to mid-March 1960. In a very similar investigation, Adem (1964) considered the effects of the snow cover of December 1962 which extended several degrees farther south of the normal coverage for northern United States and large parts of Europe and Asia. Observed departures of temperature from normal for January 1963 reached 6°C and 10°C below expected values for North America and Eurasia, respectively.

Although it seems certain to us that a primitive frost coating existed, the time would depend upon the rate of formation of the atmosphere and hydrosphere as well as on solar luminosity. If the atmosphere grew slowly, a higher than predicted luminosity would have been required to produce above-freezing temperatures on the primitive earth. We cannot do more than touch on the problem of the frost coating at present. Temperatures would certainly rise discontinuously at the time the frost disappeared.

Although there is, of course, much uncertainty in the geologic interpretations of the age and composition of the earliest rocks and the implied geologic precesses described earlier, it appears that somewhat warmer conditions probably existed 4 billion years ago than current theory indicates. If the interpretations are valid, alternative hypotheses affecting early temperature must be considered. 1. The composition of the primitive atmosphere may have been such as to produce a much greater greenhouse effect than the present, used in Table 1. As discussed above, we have rejected this possibility based on current information. 2. The zero age of the sun may be greater than that used in current astrophysical calculations, as suggested earlier, in which case the earth may also be older than 4.5 b.y. 3. The model of solar composition may require some modification to permit higher luminosities than the values in Table 1.

Rates of Erosion

The data and deductions given above all lead to the conclusion that rates of weathering and erosion during the early period of earth history must have been distinctly slower than at present. According to the current astrophysical theory described, terrestrial surface temperatures were significantly lower about 4 billion years ago than they are at present. It is also likely that for a long time, running water may only have existed during the summer seasons. Hence, both chemical and mechanical weathering in addition to water erosion and transportation would have operated relatively slowly.

In addition to the temperature effect, the composition of the atmosphere, which was so different from that of the present, would have also acted to slow the rates of weathering and erosion. Owing to the probably absence or low atmospheric density of water vapor, oxygen, and carbon dioxide referred to earlier, rock decomposition must certainly have been very retarded in comparison to the current activity of the hydration, oxidation, and carbonation processes. The importance of the latter has been stressed in the important equilibrium reaction described by Urey (1952) as follows:

$$Ca\ SiO_3 + CO_2 \quad CaCO_3 + SiO_2$$

$$Mg\ SiO_3 + CO_2 \quad MgCO_3 + SiO_2$$

Further, Rubey (1951, 1955) and Turekian (1959) have emphasized the importance of the presence of organic acids in the weathering of silicates. They point out the slowness of the decomposition processes referred to above if such minerals are in contact with water lacking these impurities. Thus when all of these factors—the probable low temperature, the effects of atmospheric composition on rock weathering and the lack of organic acids early in earth history—are taken together, it seems reasonable that early weathering and erosional processes must have been relatively slow, and must have increased in tempo during geologic time. This reasoning has led in part to the estimate of 0.5 b.y. for the duration of the earliest geosynclinal cycles, referred to in an earlier section, and also leads to the conclusion that the duration of individual geosynclinal cycles should have decreased with time.

Formation of the Continents

All of the direct conclusions from the oldest rocks lead directly to a consideration of the mechanism of formation of the earth's continental crust. It would seem that the presence of

early continental platforms must have been the product of a low density crust approximating the composition of granite just as is true of the present continents. This leads directly to the age old question of how these early granitic continental masses originated. Also, how different were they from the present continents, which have a volume of about 6×10^9km^3 and a mass of about 1.7×10^{22}Kg, 95 percent of which are composed, exclusive of sediments, of granites and grandiorites or their metamorphic derivatives.

It is commonly held (e.g., Barth, 1962, p. 352) that "most continental rocks are simply recycled, modified sediments." And we now see evidence for the existence of granitic continental material prior to the earliest recognizable sedimentary cycle. The major problem in this connection is whether the granite was secondary, having been derived by differentiation or granitization, or whether this granite was in a sense an original part of the earth's crust.

Closely involved with this problem is the time scale available for primeval terrestrial processes. The most fundamental information related to the time scale comes from studies on the thermal history of the earth. MacDonald's comprehensive study (1959) considers a large number of models which include a spread of initial temperatures and a different distribution of heat sources, all for an earth of chondritic composition. Of particular interest here is his model 14 using an initially homogeneous earth with a uniform temperature distribution of 1300°C after accumulation. The discussion given earlier on the origin of the earth suggests a model similar to this but probably at a lower temperature. In any case, as we discussed, the surface must have been well below the melting point.

MacDonald finds, even for the high initial temperature of model 14, that 10^9 years would be required to reach the melting point of the metallic phase. The mantle temperature would never exceed the melting point of silicates during this time. According to these results, the separation of the earth into core and mantle would hardly have begun by a billion years after formation. Elsasser (1963, p. 2) notes that the separation of the iron would not occur until the silicate matrix becomes sufficiently soft as its temperature approaches the melting point. MacDonald's calculations (Fig 2) show that below the 1000 km the temperature after a billion years is about 1500°C below the silicate melting point. Although the upper 500 km remained at about a constant temperature, that of the deeper portion are shown to have been warming nearly uniformly.

The thermal history of the earth developed by Lubimova (1958) is essentially in agreement with the somewhat more comprehensive treatment of MacDonald. The separation of the earth into a metallic core and silicate mantle must be concluded to have been a slow process based upon the arguments and conclusions given here. Segregation in the upper mantle would also be delayed until about a billion years after accumulation.

In the light of the data and conclusions given so far, we will now examine mechanisms which have been proposed for the development of the continents.

1. The generation of secondary granite by the extreme differentiation of a basaltic magma presents several problems. First, such a basaltic magma can develop only about five percent of granite. We do not find the rocks that would represent the remaining 95 percent in the regions where granites are prevalent. Second, the residual differentiates of most rocks are retained within their masses rather than being excluded to form a granite body. Barth (1962, p. 121–125) discusses this problem further and expresses the difficulty involved in the derivation of the great bulk of granite by differentiation of basaltic magma.

Metasomatic granitization has been called upon to explain the origin of granites by many authors (summarized by Barth, 1962, p. 349–357). The mechanism proposed involves the upward migration of K, Na, Al, and Si and the downward migration of Ca, Mg, and Fe. We do not see how there could have been enough time for this process to have generated the primordial continental masses between the times of aggregation of the earth and of the earliest demonstrably geosynclinal cycle especially in view of the time scale developed from the thermal history. This is supported by the geologic data from Southeastern Africa which indicates the presence of a relatively large amount of granitic material during the earliest sedimentary cycle. It therefore seems hard to escape the conclusion that a substantial primordial continental crust of granitic composition

existed in at least the regions described and that this crust provided much of the source material for earliest erosion.

Further, the conclusions of Rubey (1951) and Kulp (1951), and Holland (1962) that the hydrosphere and part or all of the atmosphere accumulated slowly and continuously during geologic time, tend to refute the possibility of a fairly rapid separation of sialic components from the mantle to form the early continents. If the volatiles were released from the interior slowly, the rapid rise of the mineral constituents required for continental formation would hardly have been possible.

Engel (1963) has reworked somewhat the concept of continental growth by accretion through the development of granitic material in orogenic processes within island arcs. It seems to us that the instability related to geosynclines and island arcs is the result of the presence of continents, rather than the cause of them. Further, sialic intrusive and extrusive masses within modern and ancient arc complexes may represent in large part at least, the recycling of material already present in the margins of continental blocks. The andesite zone may simply represent the region of mixed continental and oceanic material. If we try to visualize the primitive earth before the development of sialic material, it is difficult for us to imagine the development of geosynclinal piles without the presence of blocks of significant relief resulting from original differences in density.

(2) The convectional process involves the cellular motion of the mantle resulting in the accumulation of a granitic differentiate into localized masses at the earth's surface. This is often visualized as occurring from an originally uniform surface scum of sialic composition and seems more readily explainable under the condition of an original molten earth than under the conditions of cold accumulation stressed in this report.

Convection within the hot mantle has been proposed as an explanation of the continents by Vening-Meinez (1962) and has been elaborated by himself and others. Urey (1952b) suggests that continuous segregation of metal and silicate phases may supply energy for convection through the gravitational energy released in building the metallic core. However, the existence of viscous flow in the mantle basic to any convective process has been questioned by Jeffreys (1959), and more recently by MacDonald (1963) and Knopoff (1964).

On the basis of the known thermal gradient the rate of surface creep from convection has been estimated by Scheidegger (1958) and Dietz (1961) to spread over an order of magnitude from about 2.5 mm to 2 cm per year. In the application of recent experimental and theoretical results from upper mantle processes of the present and relatively recent past, Orowan (1964) has concluded that a convectional velocity of the order of 1 cm per year is possible. In the Pacific Ocean this would mean, for a value of 2 cm, that the crust could have been displaced from the center of the ocean to the shores of the Americas and Asia (about one-quarter cycle of the convection cell) since early Mesozoic. Four to five complete cycles could have occurred since 4.5 b.y. ago, the estimated age of the earth's crust, if the rate has been fairly constant.

Differentiation in the earth was certainly an inefficient process. If we assume only one cycle was sufficient, the accumulation of a continental crust would have taken about a billion years at the above rate. To this must be added the time to heat the earth sufficiently for convection to occur. According to the results of both MacDonald and Lubimova given above, the thermal gradient of the mantle has been steepening with time, so that present convection, if occurring, could hardly have begun until about a billion years after the earth accumulated. Convection does not seem therefore to be a valid mechanism for the accumulation of the continents. Evidence presented in the section on solar luminosity leads to a possible interpretation of a much greater age for the sun, and consequently of the earth. If true, a longer time interval would become available for the initial warming of the interior.

In addition to the arguments above, we therefore do not see how differentiation or metasomatic granitization or convection could have produced discrete continental masses within the required time limits despite recent proposals by Engel (1963) and MacDonald (1963). To accomplish this within the currently accepted age of the earth of 4.5 b.y. we are led into the problem of the accumulation of this granitic crust by some mechanism other than by geologic processes. An

alternative mechanism that seems worth considering at this time is the impact of large-size sialic bodies.

3. The origin of the continents from the impact of one or more continent-size bodies of low density would avoid many of the problems arising from their origin by differentiation of original terrestrial material. If the volume of the sialic continents is taken as 6×10^9 km^3 (30 km in thickness; 208×10^6 km^2 in area), the radius of the equivalent single body is about 1.13×10^3 km. The present distribution of the continents suggests that if such an origin occurred, it was more likely from several large bodies, with the larger continents resulting from an overlapping effect. Australia and Antarctica would thus require the infall of bodies with dimensions of about the largest of the asteroids, Ceres, which has a radius of 370 km.

In considering this mechanism of continental origin, Howell (1959) advanced as objections the lack of known sialic meteorites and the probability that complete vaporization would occur from the energy if impact.

Although the first objection raises a real problem, two ways around it exist. The moon has more than the required radius to produce the largest continents and has very nearly the right density. Why the moon's density is so low we do not know and indeed Urey (1962) has struggled with this difficulty for a long time. But if we accept the moon as forming separately from the earth then we have demonstrable evidence that continent-forming objects did exist.

Further, one should not equate the composition of the present crust with that of the impacting objects. The latter would have been modified by mixing with surface material during the collision as well as by subsequent geologic processes particularly involving invasions from the mantle.

The second objection offered by Howell to the impact theory does not seem very strong, especially if the impact was not higher than the escape velocity. If complete fusion occurs, the process would be aided by greater ease of both mixture with the mantle and differentiation of petrologic components. At the present time it cannot be said with great certainty what proportion of the energy of impact will be involved in the processes of melting, vaporization and shattering.

Some fraction of the material must have spread over the surrounding region to a considerable distance. Geologic activity over 4 billion years would have erased any initial structural details. The question to be answered is whether sufficient material would remain to serve as primeval continents. In view of the difficulties with gradual formation of the continents raised by this study and the fact that continents do exist, an affirmative answer appears to be a reasonable working hypothesis.

If formed in this manner, geological processes within the crust and upper mantle would certainly have modified the original shapes, dimensions and composition of the continents. If convectional processes have operated in the mantle as postulated by so many, then continental deformation would have been restricted to the outer rather than the inner portions of continents. The recycling of sediments in the border region while rock from the interior was carried outward almost continuously might give the impression of continental growth by accretion at the margins. At first glance this accretion picture looks much better for North America with its central shield than it does for the other continents. And even in North America, one of the oldest rock occurrences so far recognized occurs in the Northern Rocky Mountains which show recycling and new orogenic activity as recently as late Cretaceous, while in Africa and Eurasia, the oldest rocks occur near continental margins.

The impact hypothesis does not provide immediate explanations for the apparent differences in the nature of the continental and oceanic mantle to a depth of several hundred kilometers. The most cogent differences are those related to differences in seismic velocities. MacDonald (1963) summarizes the seismic results and offers other evidences for differences between continental and oceanic mantle based on conclusions from gravity and thermal observations. Although we cannot at present evaluate the effects of impact on the mantle to a depth of some hundreds of kilometers, it may be expected that some differences would develop during several billions of years.

The small differences in heat flow between the continental and oceanic crust, the data for which are also summarized by MacDonald, cannot be readily explained by the impact hypothesis except

to note that original differences may have been minimized by geological processes involving interactions between the crust and mantle. It should be noted however that the many locations of high heat flow on the continents are considered anomalous and excluded from comparisons.

Unless the age of the earth is substantially higher than 4.5 b.y. a terrestrial mechanism of origin seems unlikely in view of the presence of continents and normal geological processes as remote as 4 billion years ago. Without a substantial revision of the age of the earth we submit that the impact hypothesis of origin is as cogent as any other so far offered. The only alternative seems to be an accumulation temperature closer to the silicate melting point which is contrary to the evidence for cold accumulation discussed previously.

CONCLUSIONS

A number of geological and astrophysical inferences can be made from the nature and ages of the oldest rocks, especially when considered in the light of conclusions regarding the age and accumulation of the earth and the evolution of the sun. We wish to emphasize that many of these conclusions must be considered tentative pending further investigation, but we believe the conclusions to be of value in setting up guide lines and certain limiting conditions.

1. Radioisotope dates lead to the inference that at least parts of three major continents were present at least 4 billion years ago.

2. Since the finding of the oldest rock record is in part fortuitous and in part the result of subsequent geologic history, the primitive continents must have been considerably larger than the regions of oldest rock occurrence.

3. Quartz-bearing sialic rocks must have existed in the early continents in at least sufficient quantities necessary to provide the observed sediments and granite gneiss complexes.

4. Subaerial erosion by running water probably existed 4 billion years ago.

5. The surface temperature must have been above freezing over much of the continents about 4 billion years ago, with solar luminosity being the major factor in controlling terrestrial surface temperatures. According to current models of solar evolution the surface temperatures would probably have been sub-freezing at this time.

6. Aqueous erosion was probably present 4 billion years ago. To account for this, it may be in order either to push back the ages of the earth and sun or to modify the models used.

7. Rates of erosion of bed rock are proportional to temperature, humidity and organic acids. As these all increased during geologic time, it seems that sedimentation rates should also have increased with consequent decrease in the length of geosynclinal cycles. (N.B. This argument refers to the weathering of bed rock rather than mantle rock which may become more coherent through the binding action of plants.)

8. From published calculations on the thermal history of the earth it can be concluded that the melting of silicates in the mantle would not have occurred during the first billion years of earth history. This seems to obviate the possibility that continents formed by any segregation process from the mantle unless the earth were considerably older than 4.5 billion years.

9. The alternative proposal is offered that the continents originated from the impact of a swarm or swarms of low density meteoritic material. Fusion and admixture with mantle material by geologic processes would have modified the original material into the present continents, which for the most part seem to consist of recycled sedimentary rocks. The unstable continental margins would tend to suffer fairly continual orogeny creating the impression of accretion around a central nucleus or craton.

REFERENCES

ADEM, J., 1964, On the physical basis for the numerical prediction of monthly and seasonal temperatures in the troposphere-ocean-continent system; Monthly Weather Review, v. 92, p. 91–103.

ALLER, L. H. 1958, The abundances of the elements in the sun and stars: *Encyclopedia of Physics* Ed: S. Flügge, Springer-Verlag, Berlin, Göttingen, Heidelberg, v. 51, p. 324–352.

ALLSOPP, H., H. ROBERTS, G. SCHREINER and D. HUNTER, 1962, Rb-Sr age measurements on various Swaziland granites: Jour. of Geophys. Res., v. 67, p. 5307–5314.

ANDERS, E. and G. C. GOLES, 1961, Theories on the origin of meteorites, Jour. Chem. Ed., v. 38, p. 58–66.

ARP, H. C., 1958, The Russel-Hertzsprung Diagram, *Encyclopedia of Physics* Ed: S. Flügge, Springer-Verlag, Berlin, Göttingen, Heidelberg, v. 51, p. 75–133.

BARTH, T. F., 1962, *Theoretical Petrology 2nd Edition:* New York, John Wiley and Sons, Inc., 416 p.

BASSETT, W. and B. GILETTI, 1963, Precambrian ages in the Wind River Mountains, Wyoming: Geol. Soc. America Bull., v. 74, p. 209–212.

BENFIELD, A. E., 1950, The temperature in an accreting earth: Trans. Amer. Geophys. Un., v. 31, p. 53–57.

BROWN, H., 1952, Rare Gases and the formation of the earth's atmosphere: *The Atmospheres of the Earth and the Planets 3rd* Ed. G. P. Kuiper University of Chicago Press, p. 258.

BURBRIDGE, E. M. and G. R. BURBRIDGE, 1958, Stellar evolution: *Encyclopedia of Physics* Ed: S. Flügge, Berlin, Göttingen, Heidelberg, Springer-Verlag, v. 51. pp. 134–295.

BUDYIC, W., 1962, The thermal conductivity of the earth, Meteorology and Hydrology: No. 11, p. 1–18 (Translated by Office of Technical Services, U. S. Department of Commerce).

CAMERON, A. G. W., 1962, The formation of the sun and the planets: v. 1, p. 13–69.

CATANZARO, E., 1963, Zircon ages in southwestern Minnesota, J. Geophys, Res., v. 68, p. 2045–2048.

CATANZARO, E. and P. GAST, 1960, Isotopic composition of lead in pegmatitic feldspars: Geochimistry et Cosmochim. Acta 19, p. 113–126.

CANTANZARO, E., and J. L. KULP, 1964, Discordant zircons from the little Belt (Arizona), Beartooth (Montana) and Santa Catalina (Arizona) Mts., Geochimica et Cosmichimica Acta, v. 28, p. 87–124.

CHOUBERT, B., 1964, Ages absolus du Précambrienguganais: C. R. Acad. Sc. Paris, t. 258, p. 631–634.

DAVIES, J. F., 1956, Geology of the Booster Lake Area: Manitoba Dept. of Mines and Natural Resources, Pub. 55–1, p. 1–15.

DAVIS, L. and L. T. Aldrich, 1953, Determination of the age of lepidolites by the method of isotopic dilution: Geol. Soc. Am. Bull., v. 64, p. 379–380.

DERRY, D. R., 1930, Geology of the area from Minaki to Sydney Lake: Dist. of Kenorai Ont. Dept. of Mines, Ann. Rept. v. 39, p. 24–41.

DIETZ, R., 1962, Ocean basin evolution by sea floor spreading: *The Crust of the Pacific Basin*, Ed. G. A. Macdonald and H. Kuno, Amer. Geophys, Un., Pub. 1035, p. 11–22.

DONN, B. and G. W. SEARS, 1963, Planets and comets: role of crystal growth in their formation: Science, v. 140, p 1208–1211.

ECKELMANN, F. D. and A. POLDERVAART, 1957, Geologic evolution of the Beartooth Montana and Wyoming: Bull. Geol. Soc. Amer., v. 68, p. 1225–1262.

ELSASSER, W., 1963, in *Earth Science and Meteoritics*, Ed. by J. Geiss and E. D. Goldberg, Amsterdam, North Holland Pub. Co., 312 p. pp 1–30.

ENGEL, A., 1963, Geologic evolution of North America: Science, v. 140, p. 143.

EUCKEN, A., 1944, Physikalish-chemiche betrachtungen über die fruheste Entwicklungsgeschichte der Erde, Nachrichten der Akad. d. wissenscheften, Göttingen, math-physik, Klasse 1, p. 1–25.

FISH, R. A., G. C. GOLES and E. ANDERS, 1960, The record in the meteorites III: on the development of meteorites in asteroidal bodies: Astron. Jour., v. 132, p. 243–258.

GASTIL, G., 1960, The distribution of mineral dates in time and space: Am. Jour. Sci., v. 258, p. 1–35.

GILETTI, B. and P. GAST, 1961, Absolute age of Pre-Cambrian rocks in Wyoming and Montana: Ann. N. Y. Acad. Sci., v. 91, p. 454 458.

HASELGROVE, C. P. and F. HOYLE, 1959, Main sequence stars: Monthly Notices Roy. Ast. Soc., v. 119, p. 112–123.

HERBIG, G. M., 1962, The properties and problems of Tauri Stars and related objects: Advances in Astron. and Astrophys., v. 1, p. 47–104 Academic Press N. Y. pp 366 Z-Kopal, ed.

Hess, H., 1960, *Stillwater Igneous Complex, Montana*, Geol. Soc. Am. Mem. 80, 230 p.

HILLS, G. F., 1947, *The Formation of the Continents by Convection*. E. Arnold London.

HOLLAND, H. D., 1962, Model for the evolution of the earth's atmosphere; in *Petrologic Studies: A Volume in Honor of A. E. Buddington*, Ed. A. E. Engel, H. L. James and B. F. Leonard Geol. Soc. Am. New York, p. 447–477.

HOLMES, A. and L. CAHEN, 1957, Géochronologie africaine 1956: Acad. Roy. des Sci. Colonials (Brussels), Classe des sciences naturelles et medicales, Memp. Collection in octavo, fasc. 1, tome 5, 169 p.

HOWELL, B. F. Jr., 1959, *Introduction to Geophysics:* New York McGraw Hill Bk. Co., Inc., 399 p.

HOYLE, F., 1959, The ages of type I and type II subgiants: Monthly Notices, Roy. Ast. Soc., v. 119, p. 124–133.

HOYLE, F., 1960, On the origin of the solar nebula: Quart. Jour. Roy. Astnom. Soc., v. 1, p. 28–55.

HUNTER, DR. R., 1957, Petrology and classification of the Swaziland granites: Trans. Geol. Soc. S. Africa, v. 57, p. 85–126.

HUTCHINSON, G., 1959, The biochemistry of the terrestrial atmosphere in *The Earth as a Planet:* Ed. G. Kuiper, Univ. Chicago Press, p. 371–433.

IRVING, E., 1959, Paleomagnetic pole positions: a survey and analysis: Geophys. Jour., v. 2, p. 68.

JEFFREYS, H., 1959, *The Earth*, Fourth Ed., Cambridge Univ. Press, London. pp 420.

KHARATONOV, L., N. OZHINSKII, YE. VOLODIN and D. POLFYEROV, 1958, *Geology of the USSR: Murmansk Region, Part I. Regional Geology:* Moscow, Ministry of Geology and Mineral Conservation.

KNOPOFF, L., 1964, The convection current hypothesis; Reveiw of Geophys., v. 2, p. 89–122.

KOUVO, O. and J. L. Kulp, 1959, Isotopic composition of Finnish galenas: N. Y. Acad. Sci., Ann., v. 91, p. 477.

KULP, J. L., 1951, Origin of the hydrosphere: Bull. Geol. Soc. America, v. 62, p. 326–330.

KULP, J. L. and J. ENGELS, 1963, Discordances in *K-Ar* and *Rb-Sr* Isotopic Ages, Radioactive Dating, Internat. Atomic Energy Agency, Vienna, p. 219–238.

LATIMER, W. M., 1950, Astrochemical problems in the formation of the earth: Science, v. 112, p. 101–104.

LEVIN, B. Y., 1957, On the character and causes of the separation of molecules during planet formation: Mem. Soc. Roy. Sci. Liege, 4th Ser., v. 15, p. 186–197.

LEVIN, B. Y., 1958, The origin of the earth and the planets: Moscow Foreign Languages Publishing House, 84 pp.

LEVIN, B. Y., 1959, The formation of the earth from cold materials and the problem of formation of the simplest organic compounds, *The Origin of Life on the Earth:* Eng. Tr. Ed: F. Clark and R. C. M. Synge, New York Pergamon Press. pp 691. p. 67–75.

LUBIMOVA, H., 1958, Thermal history of the earth with consideration of the variable conductivity of its mantle: Geophys. Jour. v. 1, p. 115–134.

MACDONALD, G., 1959, Calculations on the thermal history of the earth: Jour. Geophys. Res., v. 64, p. 1967–2000.

MACDONALD, G., 1963, The deep structure of continents: Reviews of Geophys., v. 1, p. 587–665.

MACGREGOR, A. M., 1941, A Pre-Cambrian algal limestone in Southern Rhodesia: S. African Geol. Soc., Trans. and Proc., v. 43, p. 9.

MACGREGOR, A. M., 1951, Some milestones in the Precambrian of Southern Rhodesia: Anniversary Address, South African Geological Society, Trans. and Proc., v. 54.

MARSHALL, C. J. MAY and C. PERRET, 1964, Fossil microorganisms: possible presence in the Precambrian shield of Western Australia: Science, v. 144, p. 290–292.

NAMIAS, J., 1963, Surface-atmosphere interactions as fundamental causes of drought and other climatic fluctuations: *Changes of Climate*, Paris, UNESCO, p. 345–358.

ÖPIK, E. J., 1962, Jupiter: chemical composition, structure and origin of a giant planet: Icarus, v. 1, p. 200–257.

OROWAN, E., 1964, Continental drift and the origin of mountains. Science, v. 146, p. 1003–1010.

PATTERSON, C., 1956, Age of meteorites and the earth, Geochimica et Cosmichimica Acta, v. 10, p. 230–237.

RINGWOOD, A. E., 1961, Changes in solar luminosity and some possible terrestrial consequences: Geochimica et Cosmochimica Acta, v. 21, p. 295.

RUBEY, W. W., Geological history of sea water: Bull. Geol. Soc. Amer., v. 62, p. 1111–1148.

RUBEY, W., 1955, Development of hydrosphere and atmosphere, the crust of the earth: Geol. Soc. Amer., Sp. Pap. 62, p. 631–650.

SCHEIDEGGER, A., 1958, *Principles of Geodynamics*, Springer-Verlag, Berlin, Göttingen, Heidelberg, 280 p.

SCHMIDT, O. J., 1955, Role des particules solides dans a cosmogonie planetaire: Mem. Soc. Roy. de Liege, 4th Ser., v. 15, p. 638–649.

SCHMIDT, O., 1958, *A Theory of Earth's Origin:* Moscow, Foreign Languages Publishing House. 139 pp.

SCHWARZSCHILD, M., 1958, *Structure and Evolution of Stars*, Princeton University Press, Princeton, N. J. p. 296.

S , R. L., 1959, An evolutionary sequences of solar models: Astron. Phys. Jour., v. 129, p. 489–495.

SPITZER, L. JR., 1963, Star Formation, *The Origin of the Solar System*, Ed: R. Jastrow and A. Cameron, Academic Press, p. 39–53.

SPRINGER, G. D., 1949, Geology of the Cat Lake-Winnipeg River Area, Manitoba: Manitoba Mines Br., Prelim. Rept. 48–7, p. 6.

STRUVE, O., 1950, *Stellar Evolution*, Princeton University Press, Princeton, N. J. p. 264.

SUESS, H. and H. C. UREY, 1956, Abundances of the Elements: Rev. Mod. Phys., v. 28, p. 53–74.

TER HAAR, D. and H. WERGELAND, 1948, On the temperature of the earth's crust: Kon Norske Videnskab. Sels., v. 20, p. 52–54.

TUREKIAN, K., 1959, Probable aquatic environment of the Precambrian: International Oceanographic Cong. Reprints, Amer. Assoc. Adv. Sci., p. 81–82.

UREY, H. C., 1952a, Chemical Fractionation in the meteorites and the abundances of the elements: Geochim. et. Cosmochim. Acta, v. 2, p. 269–282.

UREY, H. C., 1952b, *The Planets, Their Origin and Development:* Yale University Press, New Haven. p. 242.

UREY, H. C., 1953, On the concentration of certain elements in the earth's surface: Proc. Roy. Soc., Ser. A, v. 219, p. 281–292.

UREY, H. C., 1956, Diamonds, meteorites and the origin of the solar system: Astrophys. Jour., v. 124, p. 623–637.

UREY, H. C., 1957a, Boundary conditions for theories of the origin of the solar system: *Phys. and Chem. of Earth V. 2*, London Pergamon Press, p. 46–76.

UREY, H. C., 1957b, Meteorites and the origin of the solar system: London Year Book of Physical Society, v. 14, p. 14–29.

UREY, H. C., 1958, The early history of the solar system as indicated by the meteorites: Proc. Chem. Soc. London, p. 67–78.

UREY, H. C., 1959, The atmospheres of the planets, Handbuch der physik, v. 52, p. 363–418.

VAN DE KAMP, P., 1953, The nearest stars: Amer. Scientits, v. 42, p. 573–588.

VINOGRADOV, A. P. and A. I. TUGARINOV, 1961, The geological age of Pre-Cambrian rocks of the Ukranian and Baltic shields: N. Y. Acad. Sci., Ann., v. 91, p. 504–506.

WILSON, A. F., 1958, Advances in the knowledge of the structure and petrology of the Pre-Cambrian rocks of southwestern Australia: Jour. Roy. Soc. Western Australia, v. 41, p. 57–83.

WILSON, A., W. COMPTON, P. JEFFERY and G. RILEY, 1960, Radioactive ages from the Precambrian rocks of western Australia: Jour. Geol. Soc. Western Australia, v. 6, p. 179–185.

WOOD, J. A., 1963, On the origin of chondrules and chondrites: Icarus, v. 2, p. 152–180.

WRIGHT, J. F., 1932, Geology and mineral deposits of a part southeastern Manitoba: Geol. Surv. Can., Mem. 169, p. 1–50.

VII. PLANETARY ATMOSPHERES

PRECEDING PAGES BLANK NOT FILMED.

N65-33720

LOW INTENSITY DECAMETER EMISSIONS FROM JUPITER*

J. K. ALEXANDER AND R G. STONE

Observations of Jupiter with the 26.3 Mc/s Christiansen-type array of the Clark Lake Radio Observatory have been analyzed with a view towards the determination of how the noise storm morphology might be altered when very weak activity (flux $< 10^{-23}$ w/m²/cps) is included. Analysis of the distribution of activity with system III central meridian longitude shows the same general shape as reported by other workers. Although the occurrence-probability histogram peaks are somewhat broadened in comparison to other less sensitive surveys in this frequency range, the region from 330° to 45° longitude still appears completely void of activity. Analysis of the intensity distribution at 26.3 Mc/s suggests two components of the emission. One component is comprised of a large amount of weak activity of flux density between 10^{-23} and 10^{-22} w/m²/cps. The other, secondary component is comprised of strong activity of flux density greater than 7×10^{-22} w/m²/cps and may correspond to the radiation commonly observed by workers at lower frequencies.

During September and October, 1963, and February and March, 1964, the 26.3 Mc/s radiation from Jupiter was systematically observed with the University of Maryland decametric antenna at Clark Lake, California. In view of the particularly high sensitivity of this instrument ($<5\text{x}10^{-24}$ w/m²/cps) special attention was given to the determination of how the apparent properties of the Jovian decameter radiation might be modified by the ability to detect emission considerably weaker than that observed by most other workers. A report of the results of this study follows.

The antenna is a Christiansen-type array consisting of 8 elements spaced at intervals along a two-mile E–W base line. Each element is an N–S colinear array of 91 full-wave dipoles. The response pattern of the array is multilobed—the individual lobes having half-power widths of 10′E–W by 3°N–S and E–W spacing of 1.5°. The declination of the response pattern can be set between 3°N and 63°N by adjusting the phase difference between the individual dipoles of the array. Therefore as a discrete radio source moves across the sky at sidereal rate it passes through one of the lobes of the grating response every six minutes. This feature simplifies identification of the emission since the times at which activity from the planet is detected must correspond to the passage of the source across one of the lobes of the response pattern. We note in this regard that the average difference between the observed and calculated lobe crossing times was about ±18 sec. which corresponds to apparent angular shifts in right ascension of about 4′. Since ionospheric refraction normally causes position shifts of this order in the case of "radio star" observations, we can say that within the limits of our system the noise storm did indeed originate at the planet. A typical record of a Jovian noise storm is shown in Figure 1.

The receiving system is a Ryle-Vonberg radiometer having an integration time of 2 sec. The receiver bandwidth was 35 kc/s during the 1963 observations and 800 kc/s during the 1964 observing period.

The 1963 observations were made nightly from September 19 to October 11, inclusive, with an average observing time of 3.3^h per night. Since meridian transit occurred between 00^h and 02^h local time, the observations were free from interference due to solar noise, and man-made interference was at a minimum. The 1964 data were collected between Feburary 13 and March 14, inclusive, with an average daily observing time of 2.5^h. Jupiter was near conjunction during this period, and hence, meridian transit occurred in mid-afternoon. As a result, interference from

*Published as *Goddard Space Flight Center Document X-615-64-149*, June 1964.

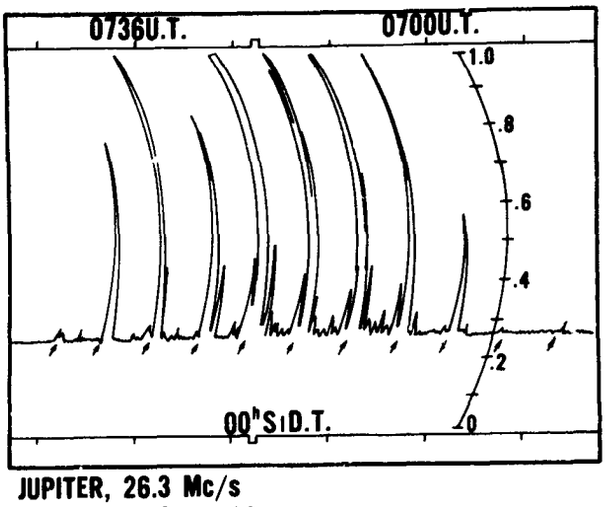


FIGURE 1.—Tracing of a typical Jovian noise storm as observed with the Clark Lake array. The short arrows indicate the calculated lobe-crossing times for the optical planet.

communications traffic and occasional solar noise reduced the total number of daily observations acceptable for analysis to 17.

The variation of the 26.3 Mc/s noise storm activity with system III central meridian longitude is shown in Figure 2. The histogram gives the longitudinal distribution of occurrence probability which is found by taking the ratio of the number of times Jupiter activity was observed to the number of times the planet passed one of the lobes of the grating response for each 10° longitude interval. Although the observations may not be sufficient to insure that the data are statistically complete one can still distinguish the three regions discussed by Douglas (1960): Region 1, 70° to 190°, Region 2, 190° to 265°, and Region 3, 265° to 330°. The distribution of total observations with longitude plotted in Figure 3 serves to show that at least all longitudes were nearly equally sampled during the study. One notices that the probability histogram peaks of Figure 2 are somewhat broadened in comparison to the results of other workers who have found the peaks grow narrower toward higher frequencies. In confirmation of a suggestion by Barrow (1962), it appears that detection of activity at all longitudes within the three main regions only requires sufficient sensitivity at the higher frequencies. On the other hand, this result does not lend support to a theoretical prediction by Ellis and McCulloch (1963) that the 26 Mc/s histogram peaks should be confined to narrow longitude bands. The significance of weak emission will be discussed in more detail later.

It is significant that the quadrant from 330° to 40° is completely void of activity. Although some workers have reported observing a few weak bursts in this region, especially below 18 Mc/s, Douglas and Smith (1963) report that the Yale workers have never detected storms at 22.2 Mc/s having flux greater than 10^{-21} w/m²/cps between longitudes 350° and 70°. In the case of our observations, no activity of flux greater than 5×10^{-24} w/m²/cps was detected. (On only one occasion did we observe any activity below 70° longitude.) All flux measurements were made relative to the discrete source Hercules A (3C 348)

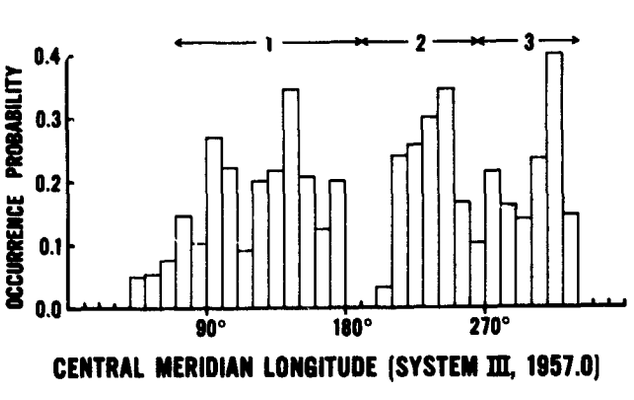


FIGURE 2.—Variation of 26.3 Mc/s activity occurrence with system III central meridian longitude. Arrows indicate location of the major regions at 22.2 Mc/s according to Douglas (1960) for comparison.

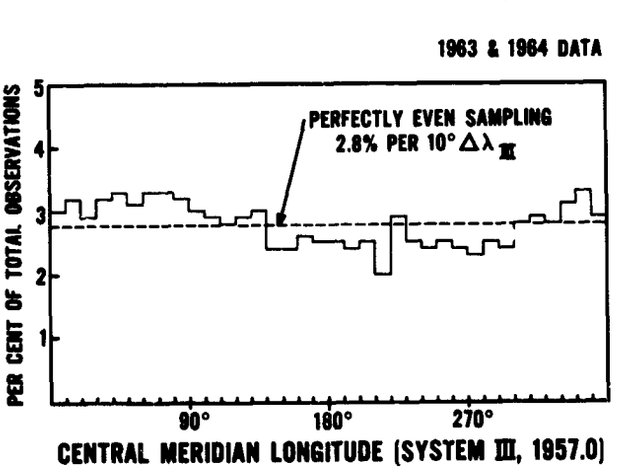


FIGURE 3.—Longitude distribution of total observations.

which has nearly the same declination as did Jupiter during these observations. The 26.3 Mc/s flux of Hercules A was taken to be 2×10^{-23} w/m²/cps (Conway, Kellermann, and Long, 1963).

One of the primary objectives of this study was to determine the number-intensity distribution of emission and to find, within instrumental limits, if there was a threshold below which activity did not occur. We must emphasize that here "activity" is defined as an abrupt increase in antenna temperature occurring at the time Jupiter transits one of the lobes of the antenna response pattern. No attempt has been made to further analyze any fine structure within the 40 seconds required for a source to drift between the half-power points of a single lobe. Furthermore, only activity detected when the planet was between the half-power points of the N–S envelope of the antenna response pattern were considered for this analysis. This condition was satisfied for an average of 1.8^h per night in 1963 and 2.5^h per day in 1964.

The histogram in the upper half of Figure 4 shows the intensity distribution for all storm activity observed in the 1963 series. One division on the abscissa corresponds to approximately 10^{-22} w/m²/cps. The most striking features of this diagram are the large concentration of activity at low intensities and the secondary peak at the high intensity end of the figure. This result would suggest that at 26.3 Mc/s the Jupiter storms are composed primarily of "weak" activity (flux less than 10^{-22} w/m²/cps) and a smaller amount of "strong" activity (flux greater than 7×10^{-22} w/m²/cps), with only rare activity at intermediate intensities.

We examine the intensity distribution of the weak activity in more detail in the lower half of Figure 4. Now the major portion of the activity is clustered at intensities just above 10^{-23} w/m²/cps

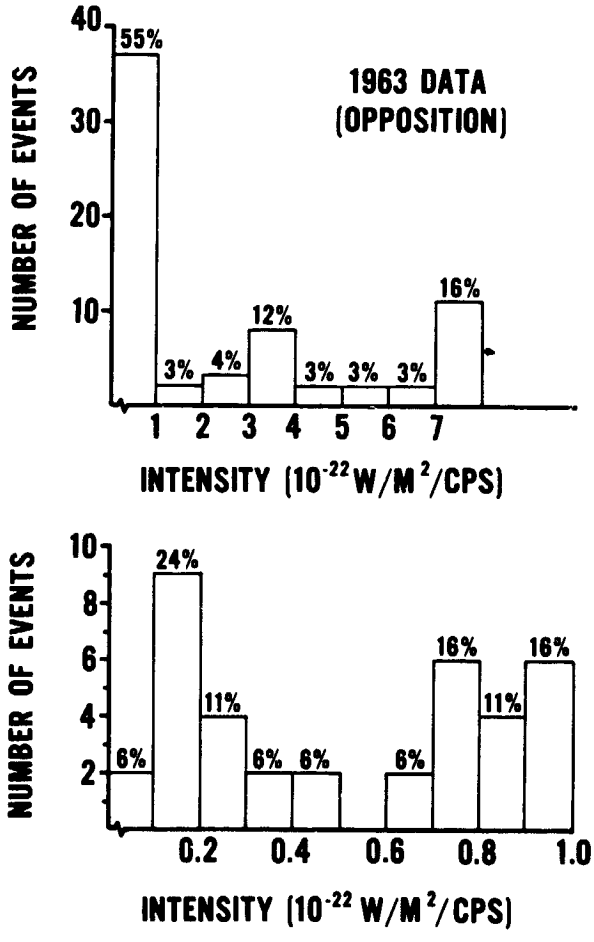


FIGURE 4.—(Upper half) Number-intensity distribution of 1963 events. (Lower half) Distribution of activity falling in first interval of upper histogram.

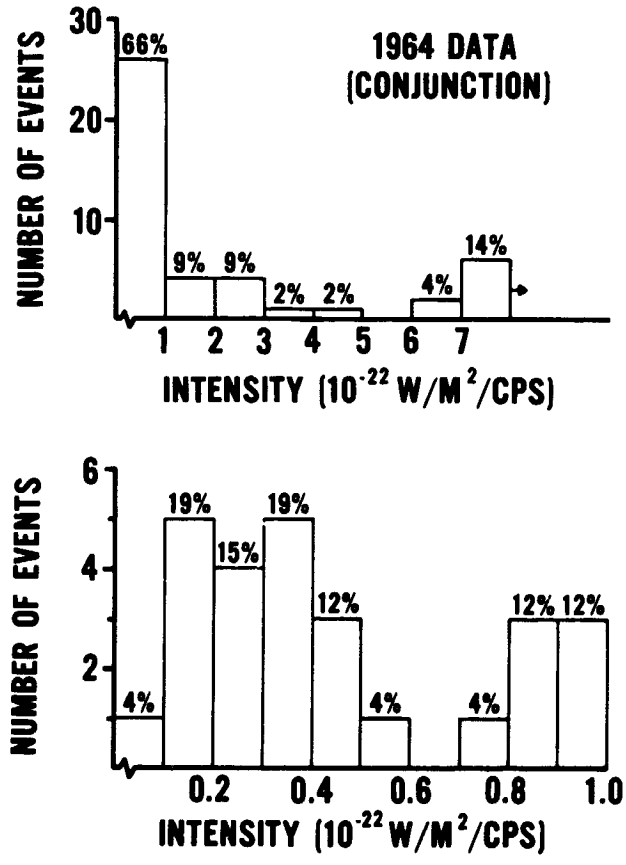


FIGURE 5.—(Upper half) Number-intensity distribution of 1964 events. (Lower half) Distribution of activity falling in first interval of upper histogram.

or at least two times the sensitivity of our antenna. A secondary peak appears around 9×10^{-23} w/m^2/cps.

The results of the same analysis for the 1964 data are shown in Figure 5. Again the intensity distribution of all activity observed is plotted in the upper half of the figure and the detailed distribution of activity falling in the first interval of the upper plot is shown in the lower half of the figure. As before, the data show a concentration of activity between 10^{-23} and 10^{-22} w/m^2/cps.

The combined results of the 1963 and 1964 measurements are given in Figure 6. The intensities of the individual events from the 1964 series were first doubled to correct for the fact that the Earth–Jupiter distance was greater by a factor of $\sqrt{2}$ than during the 1963 observations. Using the intensity distribution data of Figure 6 we may now make some general assumptions regarding the flux spectra of "typical" noise storms. Since

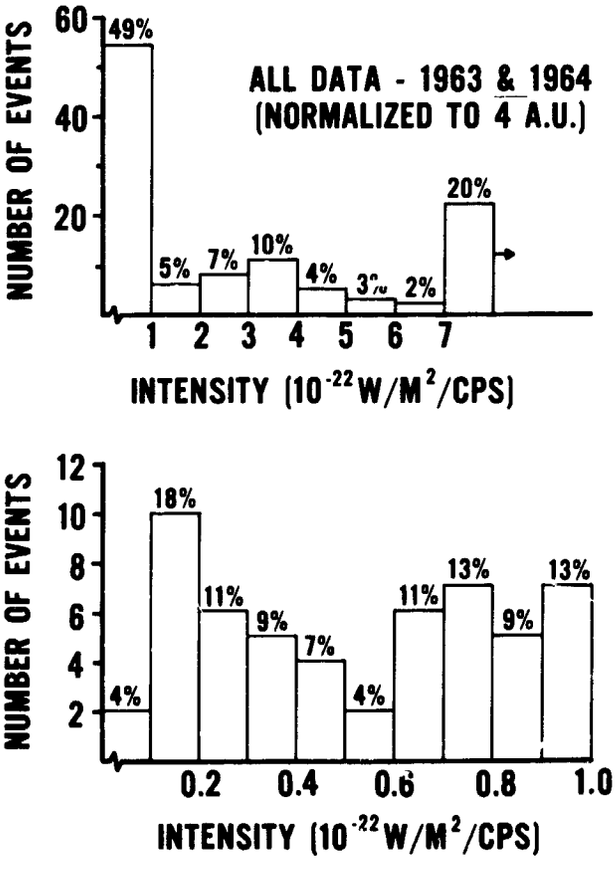


FIGURE 6.—(Upper half) Combined number-inte··· dis-tribution for 1963 and 1964 observations norma· ·o 4 A.U. (Lower half) Distribution of activity falli. first interval of upper histogram.

the Florida workers (Smith, et al.; 1963) have suggested that the flux density falls faster than f^{-5} as frequency, f, increases above 17 Mc/s, we shall assume a spectral index of -5.5. If the flux spectra obtained by extrapolation of the 26.3 Mc/s data to lower frequencies are indeed typical, then they offer an explanation of the fact that several observers report that the noise storms are generally well in excess of the sensitivity limits of their equipment. Taking the flux density for strong activity to be 8×10^{-22} w/m^2/cps, we find that at 18 Mc/s a flux density of about 6×10^{-21} w/m^2/cps in rough agreement with other reports (e.g., Smith, et al; 1963). The weak activity having a 26.3 Mc/s flux density less than 10^{-22} w/m^2/cps, however, would not exceed 9×10^{-22} w/m^2/cps at 18 Mc/s, and, therefore, would not be detected by many of the instruments presently used for Jupiter studies.

When the intensity distribution for all 26.3 Mc/s activity is determined separately for each of the three major longitude regions as shown in Figure 7, the same general shape is found with one

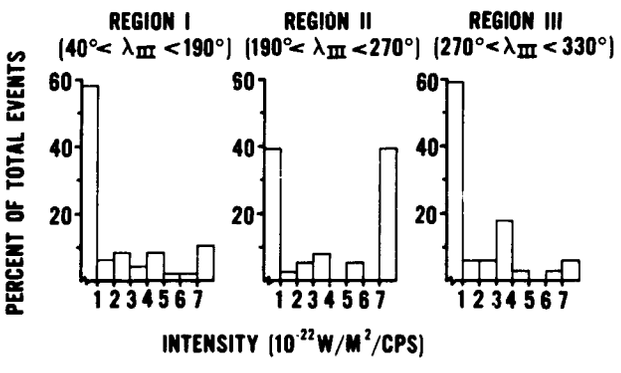


FIGURE 7.—Number-intensity distribution as a function of major longitude region. (1963 and 1964 data combined.)

interesting exception. In regions 1 and 3 the concentration of weak activity is pronounced. In the middle region, however, the weak and strong events are equally divided—each comprising 39.4 percent of the total activity observed. Such a result might have been expected from the fact that the greatest probability of occurrence for activity at frequencies nearer 20 Mc/s is found in region 2.

SUMMARY

When the 26.3 Mc/s radiation from Jupiter is observed with sufficient sensitivity to detect emission as weak as 5×10^{-24} w/m^2/cps, the variation of occurrence with system III longitude is found to have the same general shape as reported by other workers. Although the occurrence probability histogram peaks are somewhat broadened in comparison to other less sensitive surveys in this frequency range, the region from 330° to 40° longitude still appears completely void of activity. Analysis of the intensity distribution at 26.3 Mc/s suggests two components of the emission. One component is comprised of a large amount of weak activity of flux densities between 10^{-23} and 10^{-22} w/m^2/cps. The other, secondary component is comprised of strong activity of flux density greater than 7×10^{-22} w/m^2/cps and may correspond to the radiation commonly observed by other workers at low frequencies. It now remains to supplement these data with further observations at 26.3 Mc/s and new observations with comparable sensitivity at lower frequencies. If the suggestion of two components to the decameter emission is supported, then one must consider the possibility of radiation in two modes, from two radiation belts, or even by two different mechanisms.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the continued guidance and assistance of Professor W. C. Erickson who first suggested parts of this work and offered the facilities of the Clark Lake Radio Observatory for its realization. They also express their appreciation to Miss Lynn Jones who assisted in reduction of the observations. The Clark Lake Radio Observatory is operated by the University of Maryland through the support of the National Science Foundation.

REFERENCES

BARROW, C. H., 1962, *Ap. J.*, **135**, 847.

CONWAY, R. G., KELLERMANN, K. I. and LONG, R. F., 1963, MNRAS, **125**, 261.

DOUGLAS, J. N., 1960, *Dissertation*, Yale University.

DOUGLAS, J. N. and SMITH, H. J. 1963, *A. J.*, **68**, 163.

ELLIS, G. R. A. and MCCULLOCH, P. M., 1963 *Australian J. Phys.*, **16**, 380.

SMITH, A. G., SIX, N. F., CARR, T. D. and BROWN, G. W., 1963, *Nature*, **199**, 267.

N64-15970

AN ANALYSIS OF TIROS II RADIATION DATA RECORDED OVER NEW ZEALAND AT NIGHT*

LEWIS J. ALLISON

Infrared radiation measurements in the 8–12μ, 8–30μ, and 6.0–6.5μ wavelength regions taken by Tiros II on its initial orbit over New Zealand are examined in detail. A subjective analysis of the synoptic situation confirmed the two broad-scale frontal systems clearly outlined by the Tiros II radiation data recorded over the Tasman Sea at night. The equivalent blackbody temperatures measured in the 8–12μ region were found to be 6° to 12°K colder than the sea water temperatures in warm radiation source areas. In the 6.0–6.5μ region the equivalent blackbody temperatures, when converted to "effective radiation height," averaged 8705 feet below the tropopause at five upper air stations.

INTRODUCTION

The Tiros II meteorological satellite (1960 β I) was launched from Cape Canaveral, Florida, on November 23, 1960. It was spin-stabilized and space-oriented in a nearly circular orbit (apogee 454.4 statute miles, perigee 385.6 statute miles) with a period of 98.3 minutes. Tiros II was the first satellite to be equipped with a five-channel scanning radiometer to measure reflected solar radiation and infrared emission from the earth and its atmosphere, in addition to the two-television-camera system.

The radiation data selected for this analysis were obtained during Orbit Zero as the satellite initially viewed the New Zealand area. This study is a preliminary attempt to analyze the radiation data to delineate nighttime cloud cover over a remote oceanic area where meteorological information is at a minimum.

A brief description of the scanning radiometer and its regions of spectral sensitivity precedes the subjective analysis of the storm area. A comparison of frontal clouds with satellite radiation patterns and a preliminary tropopause height relationship with Channel 1 radiation data are presented.

THE SCANNING RADIOMETER

The five-channel scanning radiometer (Figure 1) measures the terrestrial and reflected solar radiation in selected regions of the infrared spectrum.

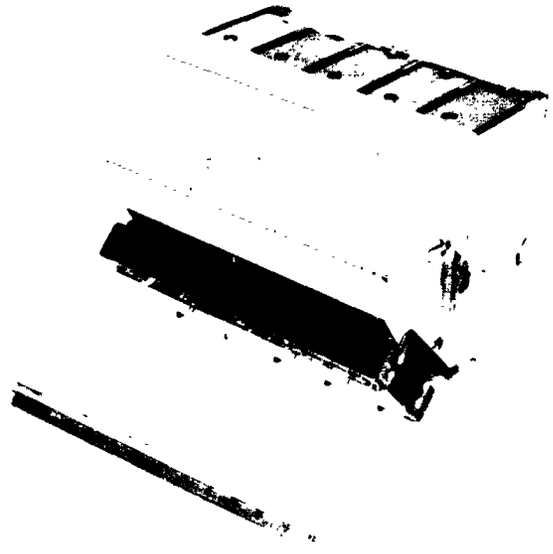


FIGURE 1.—External view of the 5-channel scanning radiometer showing the view apertures in one direction.

The five spectral regions and their characteristic functions[1] are:

- Channel 1: 6.0-6.5μ—water vapor absorption band
- Channel 2: 8–12μ—atmospheric window
- Channel 3: 0.2–6μ—reflected solar radiation band
- Channel 4: 8–30μ—thermal radiation from the earth and atmosphere
- Channel 5: 0.55–0.75μ—visible reference and daytime cloud cover

*Published as *NASA Technical Note* D-1910, March, 1964.

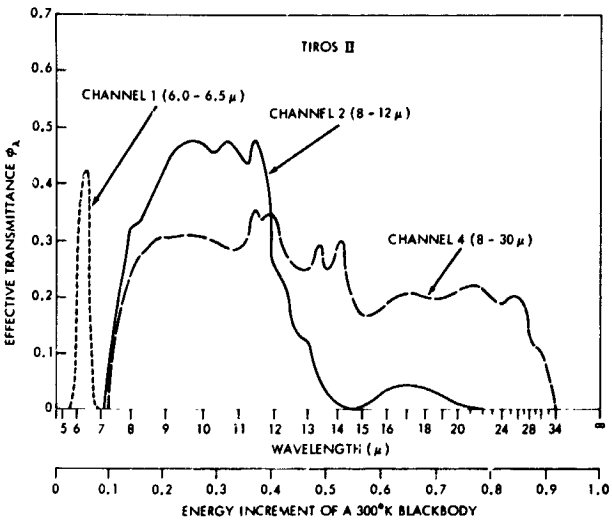


FIGURE 2.—Spectral response characteristics of channels 1, 2, and 4 of the 5-channel scanning radiometer.

The approximate transmission characteristics of the three long-wave channels (1, 2, and 4) are shown in Figure 2.

Channel 2 primarily measures the radiation (8–12μ) from the earth's surface if there are no clouds, or from the tops of clouds if they exist. The relationship between the cloud layers and the emitted infrared radiation has been expressed[2] as follows: "An overcast of dense middle and high clouds whose tops are cold, emits relatively little energy. By contrast, nearby areas of lower clouds or broken clouds or cloudless areas emit more energy. If the radiating surface is opaque, the energy emitted by the surface and measured by the satellite is related to the temperature of the surface; otherwise, the radiation is a complicated function of cloud amount and structure". A preliminary estimate of overcast cloud top heights can be made by relating the Channel 2 equivalent blackbody temperatures to nearby radiosonde temperature-height data.

Channel 4, which covers the 8–12μ window region as well as the 8–30μ region, is characterized by strong absorption bands of carbon dioxide and water vapor.

Channel 1 responds to radiative contributions, depending upon the water vapor, temperature, and pressure profiles, from a broad region in the middle and upper troposphere.

The radiometer detects the radiation energy flux viewed during a scan by means of an optical system designed to chop the radiation and apply it to five thermistor detectors. A filter system separates the five spectral regions.

The radiometer's 5 degree view angle results in a field of view encompassing approximately 40 × 40 statute miles on the earth when the sensor's nadir angle is zero. An increase in the nadir angle elongates the field in the direction viewed. The optical axes of the five channels are all parallel and inclined to the satellite spin axis by 45 degrees. Three earth-scanning modes or patterns of the radiometer (Figure 3) may be described[3] as follows:

Closed Mode: All scan spots throughout a number of spin cycles of the satellite are earth-viewed, either through the "wall" (side) sensor or the "floor" (baseplate) sensor.

Single Open Mode: Some scan spots of a spin cycle are space-viwed and the remainder are earth-viewed through the wall sensor only or through the floor sensor only.

Alternating Open Mode: The scan spots of a spin cycle are a combination of space- and earth-viewed, alternately through the wall sensor and the floor sensor.

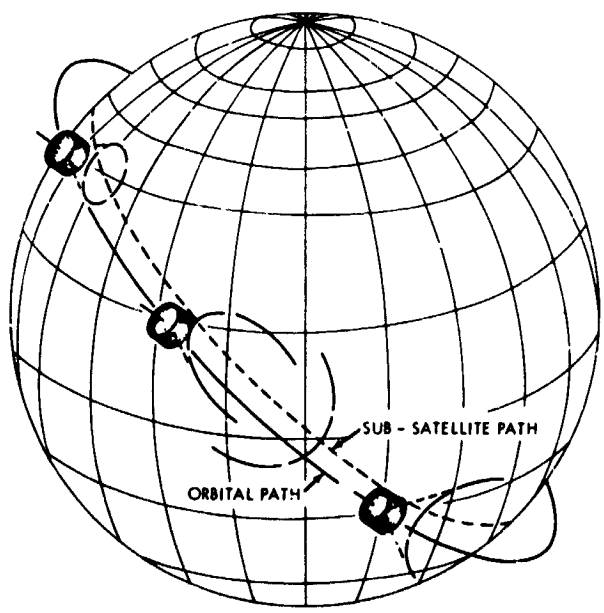


FIGURE 3.—The scan patterns of the Tiros II 5-channel medium resolution scanning radiometer. (The closed, single-open, and alternating-open modes are illustrated from left to right.)

The satellite's 8–12 rpm rotation provides the radiometer with a scan sweep that alternately views outer space (near zero radiation) and the earth's surface. The sensors measure the difference in energy flux over the scanned field of view. The sensor's output is recorded and transmitted to the ground stations where it is converted electronically into usable form. A detailed description of the data handling and conversion process has been published elsewhere.[4]

The output of the radiometers[5] can be described in terms either of equivalent blackbody temperatures (°K) or of radiant emittance (w/m²) within the spectral response curve of each channel (Figure 4). The relative accuracies of Channels 1

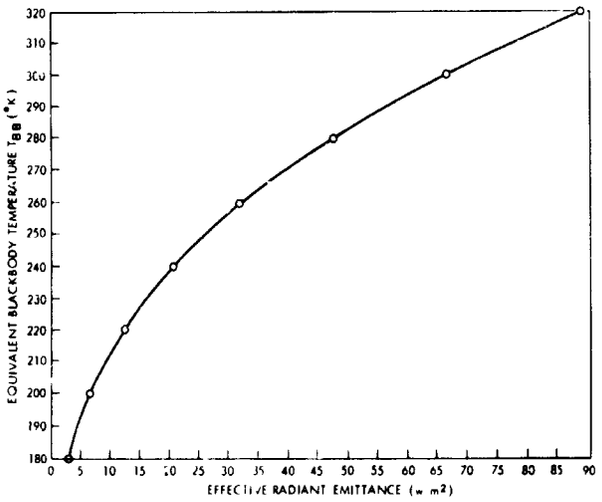


FIGURE 4.—Equivalent blackbody temperature versus effective radiant emmittance for channel 2 of the Tiros II scanning radiometer.

and 2 were estimated to be ±2°C while the absolute values may vary by ±5°C as a result of second-order effects in the calibration procedure.[6]

The radiation data for Channels 2, 4 and 1 were recorded during a 7 minute period at approximately local midnight and were reduced to usable form[5] with the aid of an IBM 7090 computer. The satellite was spinning at 8.0 rpm at swath 1 (1223 GMT), 700 miles southeast of Australia. The camera nadir angle was 127 degrees at swath 1 and slowly changed to 159 degrees at swath 74 (1232 GMT). The swath (radiation viewing path) changed from a parabolic arc to a nearly circular sweep pattern during this period.

ANALYSIS OF THE TIROS II RADIATION DATA

Four independently analyzed surface charts for 1200 GMT, November 23, 1960, were requested and received from the New Zealand and Australian Meteorological Services, the Commander of the U. S. Naval Support Force in Antarctica, and the International Antarctic Analysis Center at Melbourne, Australia, by courtesy of the U. S. Weather Bureau representative. A final composite synoptic chart and nephanalysis (Figures 5 and 6) were made from the above charts with the assistance of Lt. Comdr. John D. Ploetz, Wintering Over Officer 1960, U. S. Naval Weather Service, Washington, D. C.

The two cold-type surface meridional occlusions shown in Figures 5 and 6 had moved east and southeast off the east coast of Australia, and had intensified in the Tasman Sea, within the preceding 24 hours (November 22–23, 1960). A long north-south band of low equivalent blackbody temperatures (cold radiation sources) in the Tasman Sea (Figures 7, 8 and 9) 240° to 258°K,

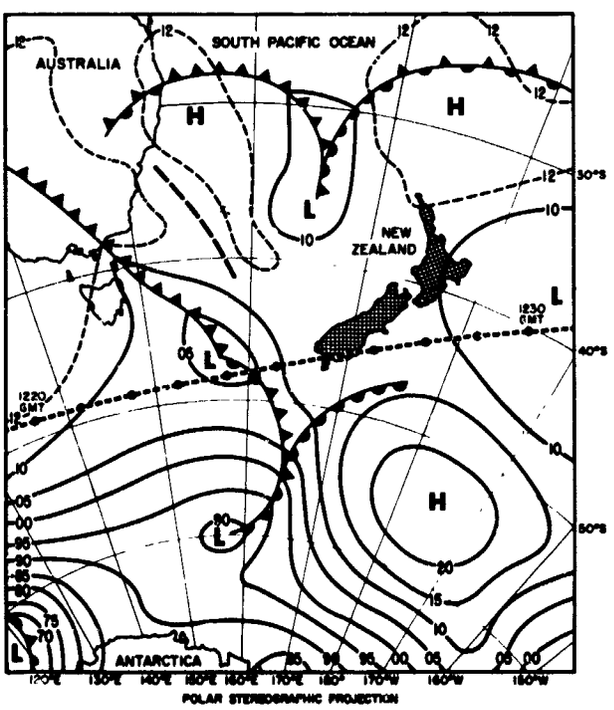


FIGURE 5.—Surface synoptic chart for the New Zealand area, showing sea-level isobars and fronts at, 1200 GMT, November 23, 1960. The satellite track during orbit 0 is shown by a heavy dashed line, with a dot indicating every minute of time.

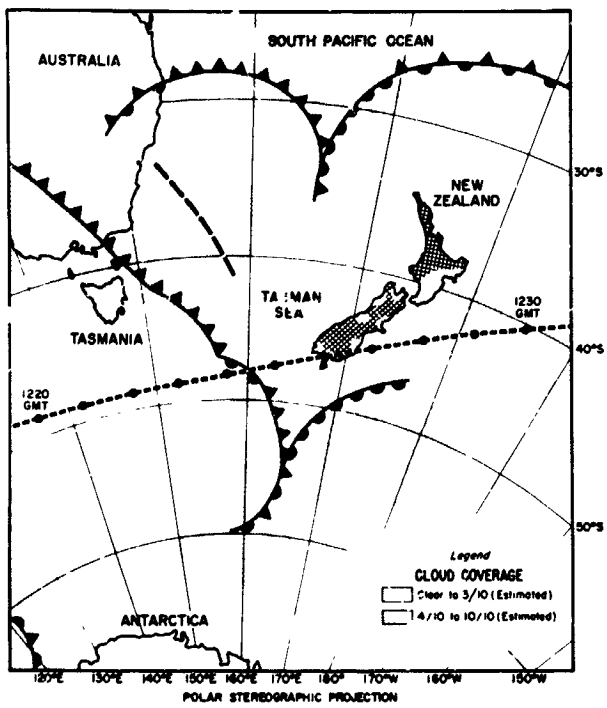


FIGURE 6.—Nephanalysis of the surface synoptic chart for 1200 GMT, November 23, 1960, on orbit 0 or Tiros II.

233° to 255°K, and 225° and 237°K, for channels 2, 4, and 1 respectively—clearly outline these well-developed occluded fronts. The shading indicates areas of cooler equivalent blackbody temperatures. By relating the above mentioned Channel 2 temperatures to nearby radiosonde data, it was estimated that the dense high overcast cloud layers in the meridional occlusions extended to heights of 5-7 kilometers.

Fritz and Winston[2] estimated the heights of the cloud tops over the United States during Orbit 4 (daytime) by use of upper air constant pressure charts, radiosonde reports, and the conversion to height of Channel 2 equivalent blackbody temperatures. It was noted that the Channel 2 blackbody temperatures recorded over thin cirriform clouds over surface fronts related to the infrared emission of the less opaque cirriform clouds and the denser (warmer) altiform and/or stratiform cloud layers below. Thus, the presence of thin high cloud layers can only be deduced by an intelligent analysis of the radiation data, synoptic situation and from climatological data.

Smaller cold radiation areas over the rugged mountain ranges of New Zealand indicate orographic-type clouds while larger cold radiation sources at 173°E longitude and 50°S latitude probably indicate alto- or cirriform-type clouds on the west side of a maritime 1022-mb surface high. Broad areas of warm equivalent blackbody temperatures 265° to 282°K, 261° to 273°K, and 237° to 246°K for channels 2, 4 and 1—occur over two surface high pressure regions. Warm radiation sources indicate areas that are: (1) cloudless, or (2) covered by warm low clouds, or (3) covered by scattered or broken clouds over a warm surface.[2] The first and third weather condition best describes the synoptic cloud cover over the weak surface maritime highs.

An air temperature and air-sea temperature difference chart for 1200 GMT, November 23, 1960 was drawn and analyzed. The waters within 10° to 15° latitude of Christchurch showed a negligible ($\pm 1°-2°$K) air-sea temperature difference both for the 1200 GMT synoptic period and for a Marsden Square oceanographic summary for November[7] (Figures 10 and 11). On air

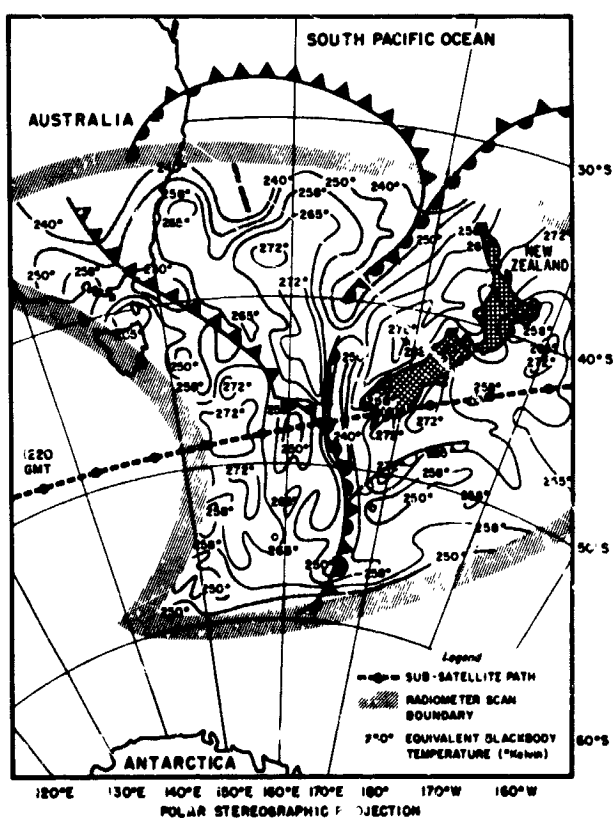


FIGURE 7.—Radiation maps from the Tiros II 5-channel scanning radiometer: Channel 2, orbit 0, 1220–1230 GMT, November 23, 1960. Equivalent blackbody temperatures of 582°K or less are shaded.

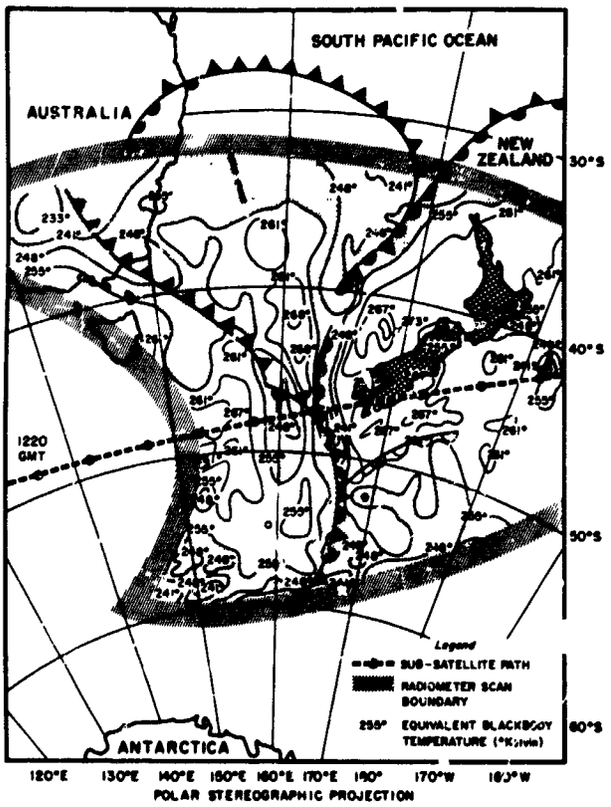


FIGURE 8.—Radiation maps from the Tiros II 5-channel scanning radiometer: Channel 4, orbit 0, 1220–1230 GMT, November 23, 1960. Equivalent blackbody temperatures of 255°K or less are shaded.

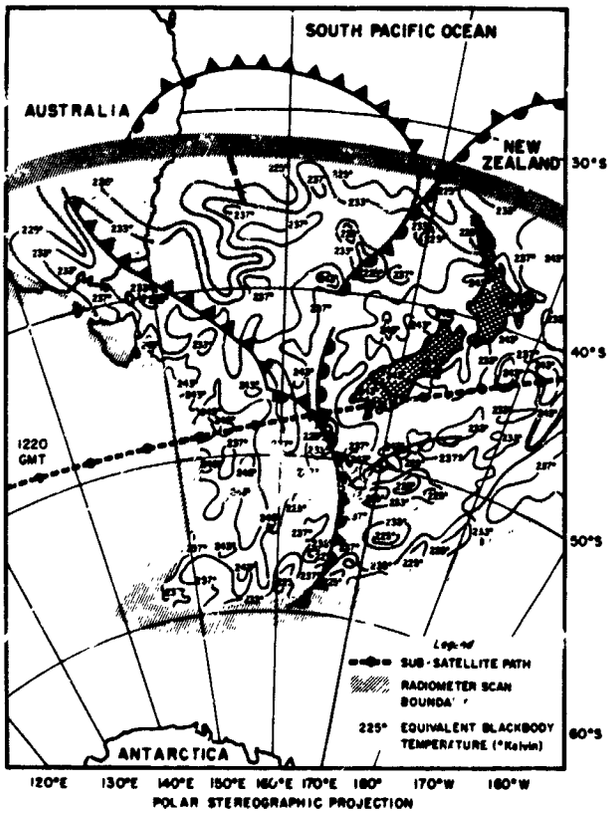


FIGURE 9.—Radiation map from the Tiros II 5-channel scanning radiometer: Channel 1, orbit 0, 1220–1230 GMT, November 23, 1960. Equivalent blackbody temperatures of 237°K or less are shaded.

temperature and air-sea temperature comparison was made where the warmest Channel 2 equivalent temperatures (282° and 278°K) had occurred. At 168°E, 42°S, the Channel 2 equivalent temperatures were noted to be 6° to 12°K colder than the sea water temperatures (Figure 10). This temperature difference was probably caused by the presence of low clouds and/or by the fact that some energy is still received in Channel 2 from atmospheric water vapor and ozone; i.e., this channel is not an entirely transparent "window."[8]

Aircraft weather reports[9] plotted from 0013 GMT, November 23, 1960 (Figure 12) helped to verify the 1200 GMT surface synoptic analysis (Figure 5). Multiple layers of low clouds (stratiform and cumuliform, with bases at 3000–5000 and tops at 6000–8000 feet), middle clouds (altostratus-altocumulus, with bases at 8000–11,000 and tops at 18,000 feet), and high clouds (cirrostratus with bases at 25,000 feet), and linse of thunderstorms (cumulonimbus—0351 GMT) were reported in the area of the meridional occlusion between 160° and 170°E and 30° and 40°S. Broken altostratus and cirrostratus cloud layers were reported by aircraft at 0612 and 0627 GMT in the same area. The presence of these dense, opaque cloud decks with tops at 18,000–25,000 feet is confirmed by the 240°–258°K equivalent blackbody temperatures recorded 6 hours later by Channel 2 (Figure 7) in the Tasman Sea.

Pilot weather reports, aided by aircraft radar, were considered to be among the most valuable sources of cloud information at night over this remote maritime area.

The most detailed upper air information for the 850 mb up through the 200 mb constant pressure level was available at 0000 GMT, November 23 and 24, 1960. The 300 mb chart (Figure 13) for 0000 GMT, November 23 was selected to show the strong circulation pattern which persisted from the 700 mb to the 200 mb level for the 24

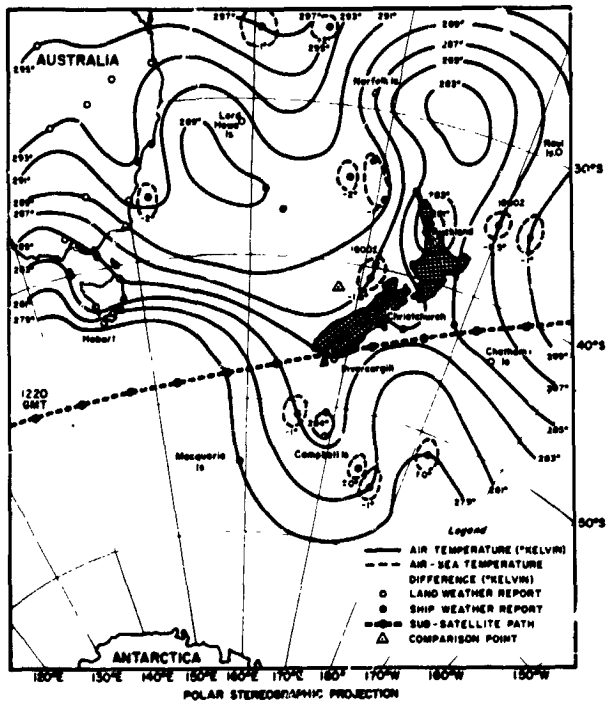


FIGURE 10.—Surface air temperature and air-sea temperature difference chart for the New Zealand area at, 1200 GMT, November 23, 1960.

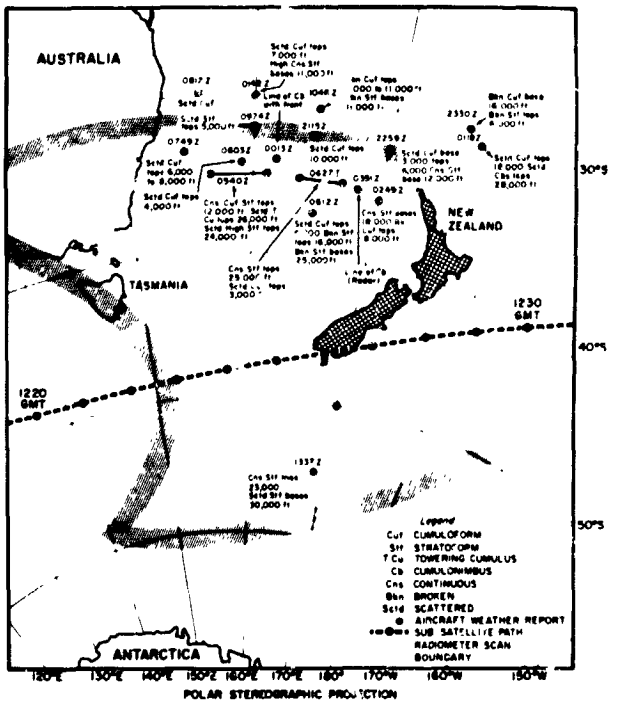


FIGURE 12.—Aircraft weather reports for the New Zealand area, 0013-2330 GMT, November 23, 1960.

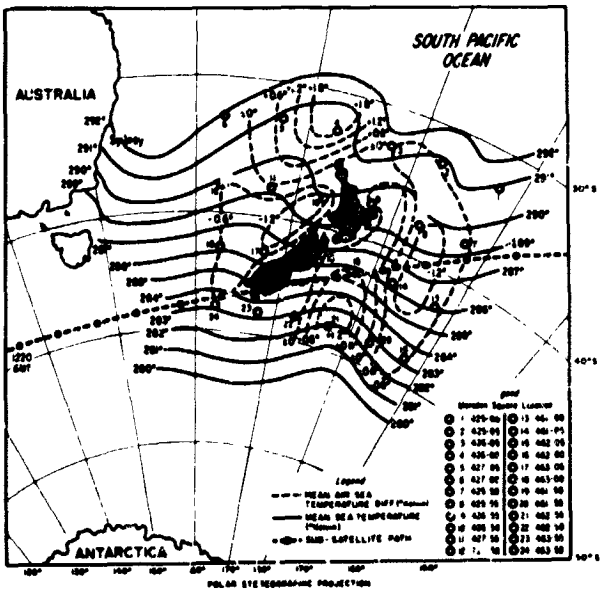


FIGURE 11.—Mean air-sea temperature difference and sea temperature chart for November (1000–1956) for the New Zealand area.

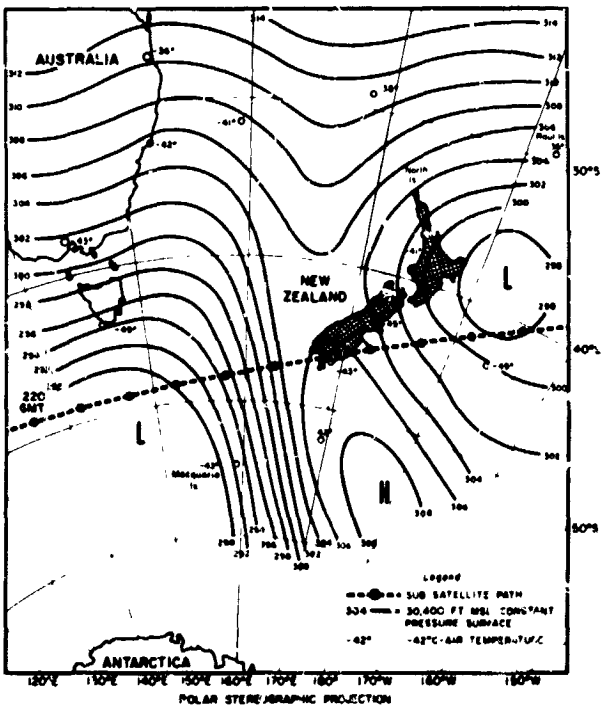


FIGURE 13.—300 mb constant pressure chart for the New Zealand area at, 0000 GMT, November 23, 1960. Tiros II, orbit 0.

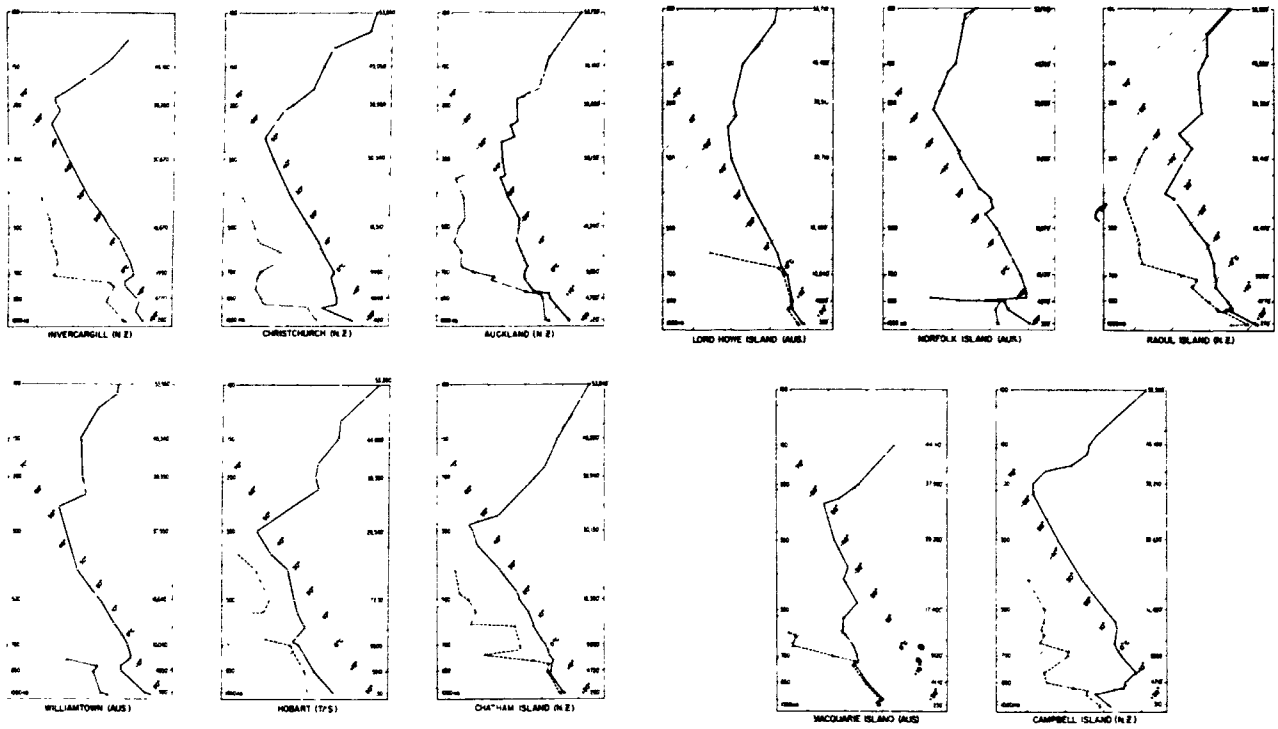


FIGURE 14.—Radiosonde soundings, at 0000 GMT, 23 November, 1960. The solid line shows temperature distribution; the dashed line, dewpoint distribution.

hour period. A strong jet stream (140 knots maximum) was oriented SSE/NNW over Macquarie Island, approximately 5 degrees west of the surface front. An 80 knot isotach maximum was located over the northeast tip of North Island, New Zealand. The 300 mb chart was included in order to complete the upper air analysis for the area.

On the basis of the subjective interpretation of the three well correlated sets of radiation data (Channels 2, 4, and 1) and all synoptic charts, the warm front originally drawn south of New Zealand was redrawn as an upper warm front and the surface portion was placed meridionally to the west of South Island, New Zealand (Figures 7, 8, and 9).

Thus, the meterorologist, through the judicial combination of Tiros II radiation data, aircraft weather reports, and all available surface charts and upper air information, can obtain a reasonably detailed nocturnal description of the horizontal and vertical cloud distribution in the stormy Tasman Sea.

TROPOPAUSE HEIGHT ESTIMATION

Widger and Touart[10] and House and Blankenship[11] have indicated that the Channel 1, 6.0–6.5μ radiation data may be useful in estimating tropopause heights. Hanel and Wark[6] have shown that Channel 1 responds to water vapor emission

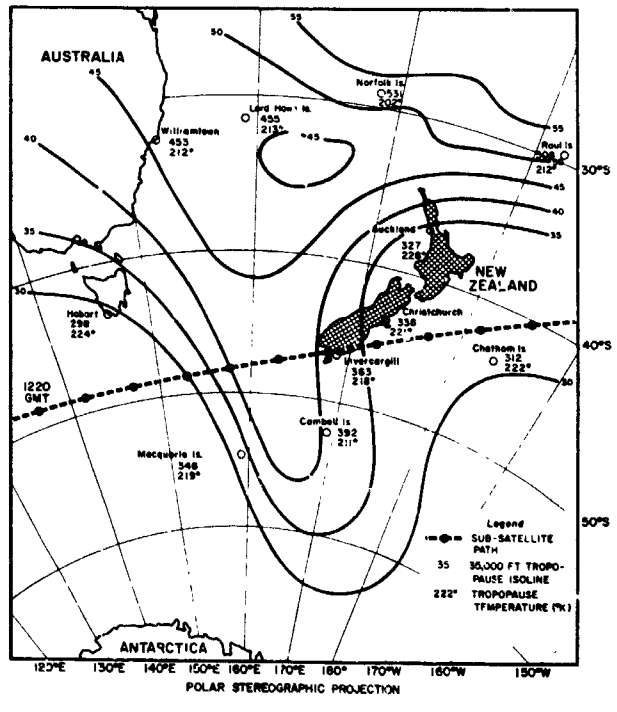


FIGURE 15—Tropopause height chart, for the New Zealand area, 0000 GMT, November 23, 1960.

in a broad region of the middle and upper troposphere, the maximum contributions emanating from the vicinity of 400 mb.

In order to verify this height estimate, the tropopause heights for the New Zealand area were determined in accordance with World Meteorological Organization definitions[12] from Figure 14 and drawn in Figure 15. Channel 1 equivalent blackbody temperatures over 5 radiosonde stations were converted to "effective radiation heights" by means of Figure 14. The effective "heights" were subtracted from the actual tropopause sounding heights at 0000Z on November 23, 1960 and averaged (Table 1). The base of a lower inversion at 36,350 feet at Invercargill was used in Table 1, because it closely approximated the WMO definition for a conventional tropopause. The lower polar tropopause height (32,700 ft) was used at Auckland, which exhibited a multiple tropopause structure.

Table 1 indicates that the Channel 1 (6.0–6.5μ) maximum integrated radiation level averaged 8705 feet below the tropopause; this qualitatively confirms the work of Hanel and Wark. The T_{BB} for Channel 1 averaged 18°K warmer than the tropopause temperatures, in this study.

W. R. Bandeen, (Private communication, September 1962) suggested an empirical equation to obtain a closer tropopause height estimate from the Tiros Channel 1 data:

$$\text{Tropopause Height} = 55{,}000 - 530\,(T_{BB}\ \text{Chan. 1-200}). \quad (1)$$

A comparison of the tropopause height from the same radiosonde soundings with the tropopause heights estimated by use of Equation 1 is shown in Table 2. The difference between the radiosonde and estimated tropopause heights ranged from +1010 to −2690 feet at radiosonde stations with single and multiple tropopauses.

Defant and Taba,[13] in their worldwide hemispheric investigation of tropopause heights and temperatures, stated: "With the exception of the very northern latitudes (75° to 90°N), the close relationship between tropopause height and tropopause temperature seem to be almost perfectly fulfilled in the sense that an increase in height goes parallel with a decrease in temperature or vice versa (compensation principle)." One can presume that the same relationship holds for the southern hemisphere.

Since tropopause temperatures and heights are operational requirements for the stratosphere flights of military and civilian aircraft,[14] more comprehensive studies of these parameters with Tiros radiation data will be made in the future.

Table 1

Comparison of Heights and Temperatures at the Tropopause as Determined from Radiosonde and Tiros Channel 1 Data.

Station	Radiosonde Tropopause Data (Figure 13)			Channel 1 "Effective" Height			Height Diff. (ft)	Temp. Diff. (°K)
	Height (ft)	Pressure (mb)	T_{BB} (°K)	Height (ft)	Pressure (mb)	T_{BB} (°K)		
Auckland	32,700	266	228	27,280	340	237	5,420	9
Invercargill	36,350	230	218	26,930	358	240	9,420	22
Campbell Is.	39,200	200	211	30,980	295	230	8,220	19
Macquarie Is.	34,800	230	219	22,350	405	242	12,450	23
Christchurch	33,850	257	221	25,840	365	238	8,010	17
							AVG: 8,705	18

Table 2

Comparison of Tropopause Heights Estimated by Use of Equation 1 with Those Obtained from Radiosonde Data.

Station	Tropopause Height (feet)			Tropopause Characteristics
	Radiosonde	Eq. 1 Est.	Difference	
Auckland	32,700	35,390	-2,690	multiple
Invercargill	36,350	33,800	-2,550	single
Campbell Is.	39,200	39,100	- 100	single
Macquarie Is.	34,800	32,740	-2,060	single
Christchurch	33,850	34,860	+1,010	single

CONCLUSION

The results described in this study point to the important meteorological potential of satellite infrared radiation data in synoptic observation of frontal and air mass movement, interpretation of large scale upper air flow patterns, and investigation of the atmospheric heat budget.[15,16,17]

ACKNOWLEDGMENTS

The author wishes to express his thanks to Dr. R. A. Hanel, Mr. W. R. Bandeen, and Dr. W. Nordberg of Goddard Space Flight Center; Lt. Comdr. John D. Ploetz of the U. S. Navy; and Mr. Thomas Gray of the U. S. Weather Bureau for their instructive comments and encouragement during the course of this service-interrupted study.

REFERENCES

1. Bandeen, W. R., Hanel, R. A., Licht, J., Stampfl, R. A., and Stroud, W. G., "Infrared and Reflected Solar Radiation Measurements from the Tiros II Meteorological Satellite," NASA Technical Note D-1096, November 1961; also published in *J. Geophys. Res.* 66(10):3169-3185, October 1961.
2. Fritz, S., and Winston, J. S., "Synoptic Use of Radiation Measurements from Satellite TIROS II," *Monthly Weather Rev.* 90(1):1-9, January 1962.
3. Tiros II Radiation Data Users' Manual Supplement, Goddard Space Flight Center, May 1962.
4. Davis, J. F., Hanel, R. A., Stampfl, R. A., Strange, M. G., and Townsend, M. R., "Telemetering Infrared Data from the Tiros Meteorological Satellites," NASA Technical Note D-1293, August 1962; also presented at the IRE Professional Group on Space Electronics and Telemetry, Albuquerque, N. M., September 1961.
5. "TIROS II Radiation Data Users' Manual," Goddard Space Flight Center, August 1961.
6. Hanel, R. A., and Wark, D. Q., "Physical Significance of the Tiros II Radiation Experiments," NASA Technical Note D-701, December 1961; also *J. Opt. Soc. Amer.* 51(12):1394-1399, December 1961.
7. U. S. Navy Oceanographic Office, Washington, D. C., Marine Decks: 110, 115, 116, 192, 193, 194, 195, 196, 281.
8. Wark, D. Q., Yamamcto, G., and Lienesch, J., "Infrared Flux and Surface Temperature Determinations from TIROS Radiometer Measurements," Meteorological Satellite Laboratory Report No. 10, U. S. Weather Bureau, Washington, D. C., August 1962.
9. Daily Weather Bulletins, New Zealand Meteorological Service, December 4, 1960.
10. Widger, W. K., Jr., and Touart, C. N., "Utilization of Satellite Observations in Weather Analysis and Forecasting," *Bull. Amer. Meteorological Soc.* 38(9):521-533, November 1957.
11. House, F. B., and Blankenship, J. R., "Applications of Infrared Measurements in Meteorology," Air Weather Service Technical Report 157, June 1961.
12. "Radiosonde Analysis," Air Weather Service Manual 105-124, Vol. 1, March 15, 1961.
13. Defant, F. R., and Taba, H., "The Threefold Structure of the Atmosphere and the Characteristics of the Tropopause," *Tellus* 9(3):259-274, August 1957.
14. Jones, D. L., Cooley, D. S., and Carey, T. A., "A Survey of Observational Requirements from Meteorological Satellites 1963-1966," Travelers Research Center, Inc., July 1962.
15. Winston, J. S., and Rao, K. P., "Preliminary Study of Planetary Scale Outgoing Long Wave Radiation as Derived from TIROS II Measurements," *Monthly Weather Review*, 90(8):307-310, August 1962.

16. WEXLER, R., "Interpretation of TIROS II Radiation Measurements," Allied Research Associates, Inc., Boston, Mass., Final Report, Contract AF 19(604):-5968, May 31, 1962.

17. BOLDYREV, V. G., "Ob ispol' zovanii radiatsionnykh izmerenii so sputnikov v sinopticheskon analize (Use of Radiation Measurements from Satellites in Synoptic Analysis); *Meteorologiia i Gidrologiia Moscow*, No. 10:28–34, October 1962.

See N65-35230 +
Serv # 17971

LATITUDINAL DISTRIBUTION OF CLOUD COVER FROM TIROS III PHOTOGRAPHS*

ALBERT ARKING†

Television pictures from the Tiros III satellite have been analyzed on a computer to give the latitudinal distribution of cloud cover during the summer of 1961. The results, which will be useful in studying the heat balance of the atmosphere and in the determination of vertical motion, show good agreement with the long-term average cloudiness derived from data accumulated during 50 years of ground observations.

The global distribution of cloud cover is now being investigated by the analysis of photographs from Tiros meteorological satellites. The aim of this cloud-cover study is to derive basic information on the radiation energy balance of the atmosphere, and on the vertical atmospheric motions which are revealed by the existence of clouds The type of information yielded by this analysis is important for studies of climate and for investigations of the general circulation of the atmosphere.

The energy input which sets the atmosphere in motion is given by the difference between the incoming solar energy, consisting of radiation primarily in the visible part of the spectrum, and the outgoing energy, consisting of radiation from the earth and the atmosphere in the far infrared part of the spectrum. The main control over the incoming solar radiation is provided by clouds, which can reflect up to 80 percent of the incident visible radiation, depending upon their thickness and type. Reflection by the atmosphere and the underlying surface is much less; it is about 8 to 10 percent for the atmosphere, from 3 to 20 percent for most terrains, and, in general, only a few percent for large bodies of water. Calculations of the available solar energy are thus strongly dependent on knowledge of the cloud-cover distribution.

The most important factors which govern the distribution of the outgoing infrared radiation are the ground temperatures, the amount of water vapor in the atmosphere, and particularly the extent and height of the clouds.

It is thus seen that the distribution of cloudiness plays an important role in the determination of both the inflow and the outflow of energy through the earth's atmosphere. Thus far, the distribution of clouds—amount, types, and approximate heights—have all been taken from ground-based observations. Satellite observations enable us to obtain extensive cloud-cover data on a global scale in a relatively short period of time.

Each Tiros satellite contains two television camera systems which photograph the cloud cover. Some satellites (Tiros II, III, IV, and VII) contain, in addition, radiometers to measure emitted infrared radiation and reflected visible radiation from the earth, atmosphere, and clouds. The data from the visible and infrared radiometers would ordinarily be sufficient for deriving the energy balance, but only on Tiros VII did the visible radiometer channel work successfully. From the cloud-cover distribution, however, one can estimate the reflected radiation in the visible spectrum and hence, the incoming solar radiation.

Use of the radiation data in determining the energy balance is discussed in a simultaneous publication by Rasool (*1*) in which references are made to other studies of the radiation data. Although the cloud-cover pictures have been used extensively for their synoptic value in the detection and tracking of storms and in present methods of weather forecasting, the analysis of the pictures

*Published in *Science*, 143(3606):569–572, February 7, 1964.
†Goddard Space Flight Center and New York University, New York.

to obtain statistical data on cloud cover, for application to energy-balance studies, has not been undertaken previously.

The analysis of the cloud-cover photographs was performed on an I.B.M. 7090 computer. Photographs are converted into digital form for insertion into the computer by dividing the photographic image into 250,000 picture elements to form a matrix of numbers, 500 by 500, each number representing the brightness of the corresponding picture element.

The main problem in the computer analysis of cloud pictures is the choice of a criterion, suitable for machine execution, which will distinguish clouds from clear areas. The brightness of the picture element should be a suitable criterion, since clouds will have higher reflectivities than clear areas except in snow-covered regions. The analysis is complicated by the fact that a cloud of given reflectivity can have a range of brightness depending on the relative angles of the sun and camera with respect to the surface appearing in the photograph, the structure and thickness of clouds, and the nature of the underlying terrain. In addition, there may be variations in the characteristics of the vidicon tube and associated

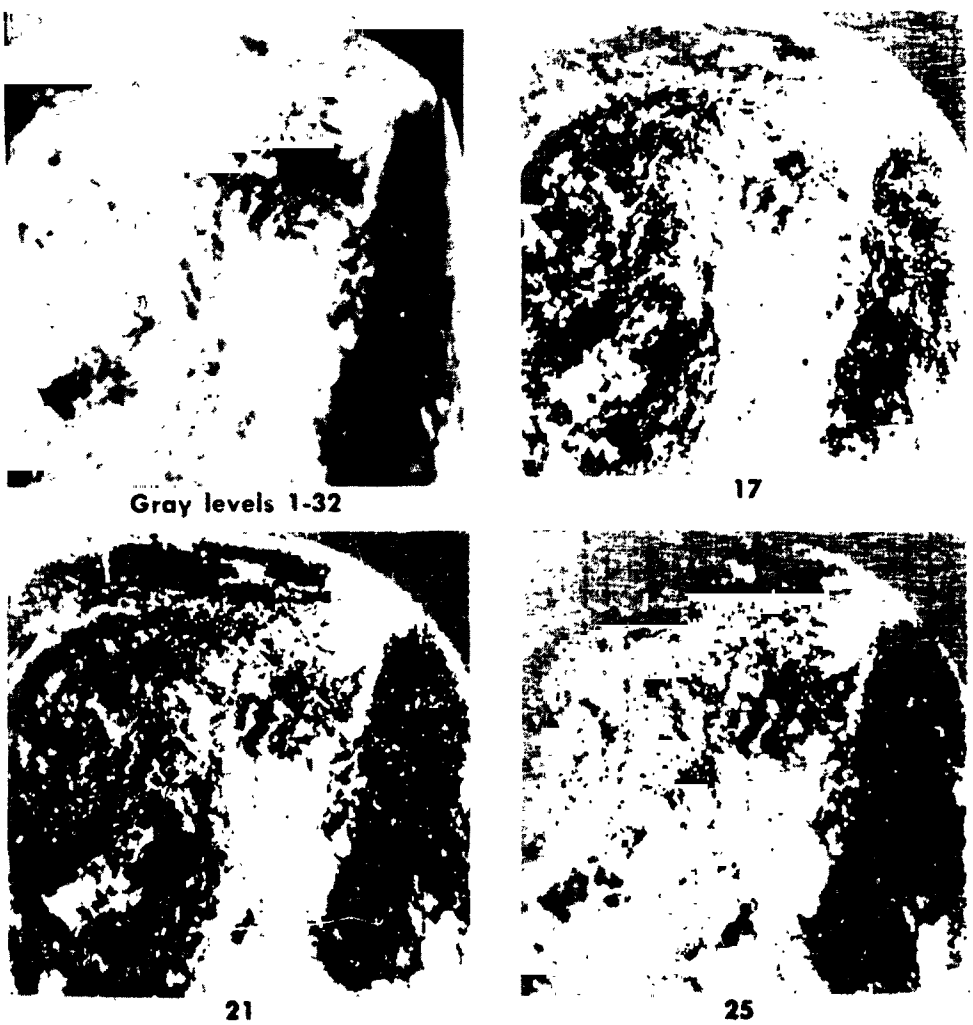


FIGURE 1.—Determination of brightness threshold of clouds (criteria for discriminating clouds from clear areas). A series of two-level images, in which intensities greater than a given threshold appear white and all others black, are produced for a number of thresholds, three of which are shown here. The number underneath each two-level image is the brightness level used as the threshold for the discrimination. The original full-tone image (gray levels 1–32) is compared with each two-level image to find the threshold which most closely reproduces the cloud boundaries reported by the human observer in the original image, in this case 21.

electronics after the satellite is put into orbit. Nonetheless, it was found, by analysis of sample pictures, that the differncee between the reflectivities of clouds and the underlying terrain is great enough so that it is possible to choose a brightness threshold which defines the cloud boundaries reliably (Fig. 1). It was further found that the threshold may vary considerably for pictures taken during different orbits, but it varies only slightly from picture to picture within an orbit. It was therefore necessary to determine a new threshold for each orbital sequence but not for each picture. This method does not distinguish between clouds and snow-covered terrain, and, furthermore, it may result in an overestimation of the cloudiness over regions of bright sand. However, only a very small fraction of the earth is covered by bright sand, and the season and regions covered by Tiros III rule out extensive snow-cover.

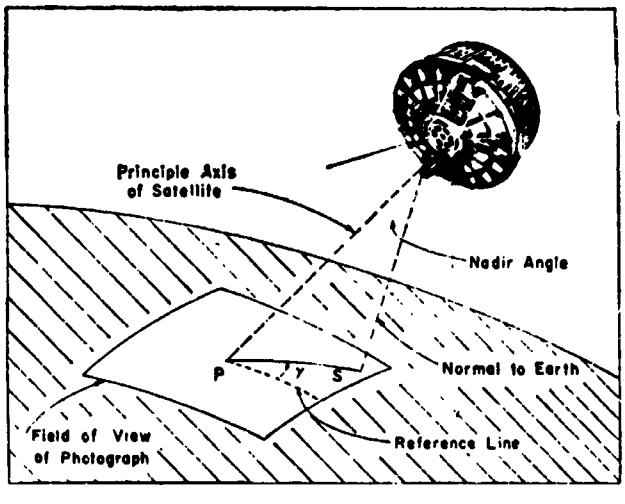


FIGURE 2.—Geometry of Tiros photography. The two television cameras point along the principle or spin axis of the satellite, which is fixed in space. The angle between the principle axis and the normal to the earth, the nadir angle, changes as the satellite orbits the earth. The intersection of the principal axis with the earth is call the principal point, denoted by P, and the point immediately below the satellite, called the subsatellite point, is denoted by S. As the satellite rotates, the image on the vidicon tube appears to rotate about the point P. The angle between PS and a fixed reference line on the vidicon tube defines the roll angle, γ, the value of which is required, in addition to the geographic positions of P and S, to transform from picture coordinates to geographic coordinates.

The location of cloudy and clear areas on a world map requires precise knowledge of the satellite's position and angular orientation at the time the photograph was taken (see Fig. 2). These data were supplied by the Meteorological Satellite Laboratory of the Weather Bureau and the Aeronomy and Meteorology Division of the Goddard Space Flight Center. The transformation from the image plane of the photograph into geographical coordinates is described by Frankel and Bristor (*2*) and by Mach (*3*).

There were 1447 Tiros III photographs available on video tape which were free from excessive noise and for which we were able to obtain all the data required for the transformation to geographical coordinates. All the pictures were processed on the 7090 computer and form the basis of the results presented here.

The end product of the computer analysis is a magnetic tape which contains a summary of every picture analyzed. A computer program interrogates this tape to provide the mean percentage of cloud cover over any specified geographic region during any interval of time. For the purposes of this introductory note, however, we will present only the latitudinal distribution of cloud cover for the summer of 1961.

The geographical distribution of the pictures is shown in Fig. 3; each asterisk on the map represents a single photograph. All were taken during daytime between 12 July and 30 September 1961, between latitudes 60°S and 60°N. The number of pictures in the Southern Hemisphere is appreciably less than in the Northern Hemisphere and, consequently, errors due to statistical sampling will be greater for southern latitudes. There may also be some statistical bias in northern latitudes due to the sparseness of photographs over Asia.

The average latitudinal distribution of cloud cover is shown in Fig. 4 for the period 12 July to 30 September 1961. The solid horizontal bars give the mean percentage of earth area covered by clouds in intervals of 10° of latitude as determined from the Tiros III photographs. The vertical lines passing through the bars indicate the uncertainty estimated to arise in the threshold determination. Possible errors due to statistical sampling are not shown.

The results in Fig. 4 show that the cloud cover in middle latitudes is the same in the Northern

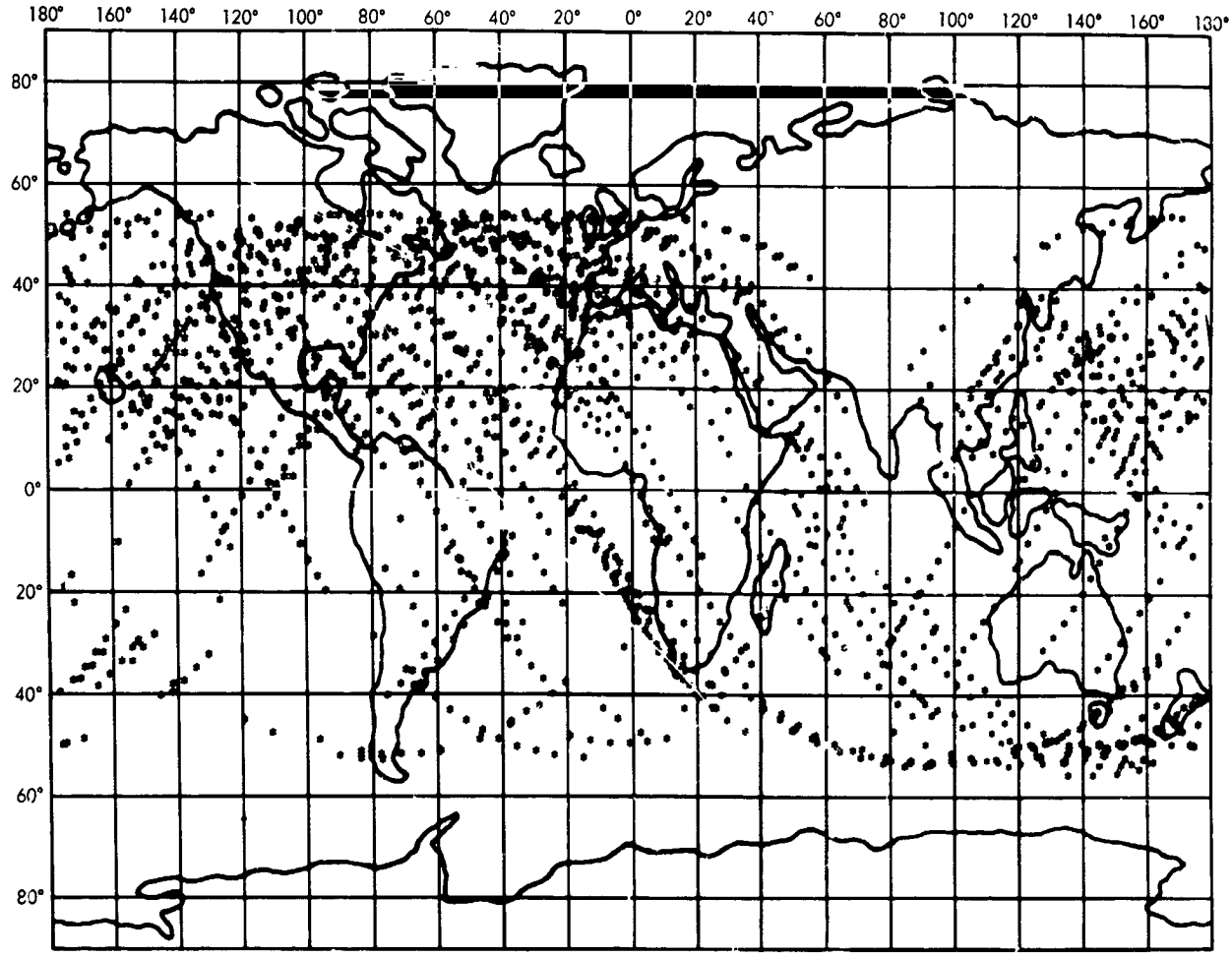


FIGURE 3.—The geographical distribution of the 1447 Tiros III pictures taken between 12 July and 30 September, 1961, which were used in the analysis.

and Southern Hemispheres. However, in tropical latitudes there is an asymmetry, with a local maximum of the cloud cover in the tropics centered at 10°N latitude. This is the average position of the "thermal equator" during the period 12 July to 30 September.

The broad features of the latitudinal distribution of cloud cover obtained from the Tiros photographs are consistent with the known pattern of the general circulation (*4*). Air rising at the thermal equator produces condensation and a relative maximum in the cloud cover, while on the average there is downward motion of cool, dry air at 30°C, which explains the relative minimum of cloudiness. The relationship between cloud-cover distribution and vertical air currents suggests that the Tiros cloud-cover statistics may have an important application in the determination of vertical motions in the atmosphere.

It is of considerable interest to compare the Tiros observations for the summer of 1961 with the climatological distributions of cloudiness found in the literature for the same season. Such distributions have been published by Haurwitz and Austin (*4*), Landsberg (*5*), and others for the globe; and, based on more extensive data, by Telegadas and London (*6*) for the Northern Hemisphere. The climatological results for the period July through September are shown by the dashed histogram in Fig. 4. The Northern Hemisphere data are taken from Telegadas and London (*6*); the Southern Hemisphere data are taken from Landsberg (*5*).

The Tiros results for 1961 are seen to be in good agreement with the long-term mean of cloud-cover distribution obtained from ground observations. The degree of correspondence between our results and the ground-based data gives us confidence in

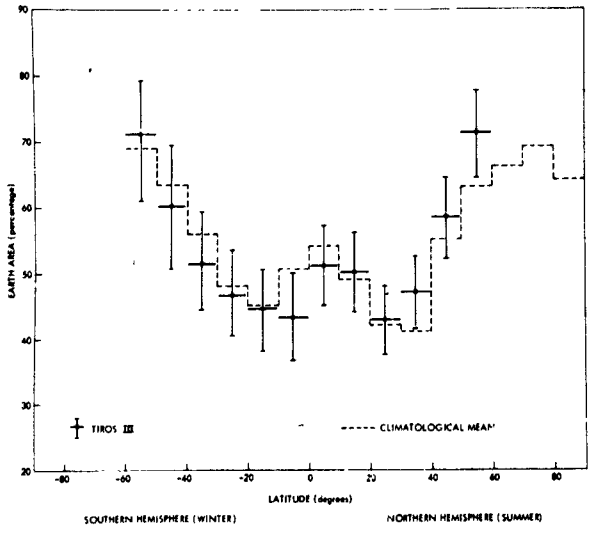


FIGURE 4.—The latitudinal distribution of cloud cover. The solid horizontal bars are the results derived from Tiros III photographs from 12 July to 30 September, 1961. The vertical lines show the estimated uncertainty due to threshold determination. The dashed histogram represents the climatological mean cloud cover based upon ground observations, taken from Telegadas and London (6) for the Northern Hemisphere and from Landsberg (5) for the Southern Hemisphere.

this method of analysis of satellite photographs. The availability of more data from subsequent satellites will permit the determination of the geographical distribution of cloud cover over short intervals of time, for which no reliable information is now available. An increase in the density of observations, coupled with improvements in the techniques for picture analysis now being developed, should eventually lead to the use of cloud-cover pictures, automatically analyzed on a computer to provide the cloud-cover distribution as a function of time for use in studies of atmospheric dynamics and long-term changes in climate.

REFERENCES AND NOTES

1. S. I. RASOOL, *Science*, this issue.
2. M. FRANKEL and C. L. BRISTOR, *Meteorological Satellite Laboratory Report No.* **11**, Weather Bureau, U.S. Department of Commerce, 1962.
3. R. E. MACH, "Research on the processing of satellite photography by digital techniques," *Final Report, Contract AF* **19(604)8432**, International Business Machines Corp., 1962.
4. B. HAURWITZ and J. M. AUSTIN, *Climatology* (McGraw-Hill, New York, 1944).
5. H. LANDSBERG, "Climatology," *Handbook of Meteorology*, F. A. BERRY, JR., E. BOLLAY, N. R. BEERS, Eds. (McGraw-Hill, New York, 1945), pp. 927–97.
6. K. TELEGADAS and J. LONDON, "A physical model for the Northern Hemisphere for winter and summer," *Scientific Report No.* **1**, Contract AF 19(122)-165, Research Division, College of Engineering, New York University, 1954.

N64-33637

A RADIATION CLIMATOLOGY IN THE VISIBLE AND INFRARED FROM THE TIROS METEOROLOGICAL SATELLITES*

W. R. BANDEEN, M. HALEV, AND I. STRANGE

Radiometric measurements in the visible and infrared have been made intermittently since 1960 by the meteorological satellites, TIROS II, III, IV and VII. Measurements of reflected solar radiation have been acquired within filtered passbands of approximately 0.2−6μ and 0.55−0.75μ and, of emitted long wave radiation, within passbands of approximately 8−12μ, 8−30μ, 6−6.5μ, and 14.8−15.5μ. The TIROS VII satellite has acquired continuous radiometric data over the earth for more than one year, increasing significantly the potential of its data for climatological studies, compared to the earlier experiments which lasted for shorter periods of time.

A limited climatology of the spatial and temporal variations of the total outgoing long wave radiation and of the planetary albedo is derived from the essentially beam measurements within the 5° field of view of a five-channel, scanning radiometer. The results show essential agreement with previous studies of the planetary heat balance, based upon conventional data. However, several interesting differences appear. The satellite data indicate that a disproportionately high amount of radiation is emitted from the subtropical anticyclones, while the meridional gradient of the mean zonal albedo is larger than indicated in several recent studies.

In addition to the heat balance results, the relative planetary distribution of atmospheric moisture is inferred from measurements made within the 6.3μ band of water vapor, and the seasonal changes of mean temperatures in the stratosphere between about 15 and 35 km are shown from measurements made within the 15μ band of carbon dioxide.

I. INTRODUCTION

Climatological investigations by means of radiation data from the TIROS meteorological satellites have been hindered, at least in part, by the relatively short lifetimes of the early experiments flown on TIROS II, III, and IV. The medium-resolution scanning radiometer flown on TIROS VII, however, has acquired continuous data over the earth for more than one year, increasing significantly the potential of the radiometric measurements for climatological studies.

The design of the radiometers, their calibration, the information flow, and the data reduction techniques have been described previously (Bandeen et al., 1961; Nordberg et al., 1962; Kondratiev, 1963), as have the physical significance of the experiment and its synoptic use potential (Hanel and Wark, 1961; Fritz and Winston, 1962; Bandeen et al., 1964). In this paper we shall discuss three climatological applications of radiation data from TIROS VII, and from TIROS IV for one type of measurement (in the 6.3μ H_2O band) that was not made by TIROS VII.

The radiometer measures essentially beam radiation within a 5° field-of-view from five channels, each made sensitive to a different part of the spectrum by means of filters and other optical elements. The nominal wavelength limits of the sensors flown on the various satellites are shown in Table I.

II. CALIBRATION

The calibration of the infrared channels is carried out in the laboratory by viewing a blackbody cavity, and the primary calibrated quantity is the "equivalent blackbody temperature, T_{BB}(°K)". It follows that the "effective radiant emittance, W (watts m^{-2})" and the "effective radiance N

*Published as NASA Technical Note D-2534; also presented at the International Radiation Symposium, International Association of Meteorology and Atmospheric Physics (I.U.G.G.), Leningrad, August 1964.

TABLE I.—*Nominal Wavelength Intervals of TIROS Radiometers*

Nominal Wavelength Interval		TIROS II	TIROS III	TIROS IV*	TIROS VII**
6–6.5μ	(H_2O absorption)	X(1)	X(1)	X(5)	
8–12μ	(atmospheric window)	X(5)	X(2)	X(5)	X
8–30μ	(long-wave radiation)	X(1)	X(1)		X
14.8–15.5μ	(CO_2 absorption)				X
0.2–6μ	(reflected solar radiation)	X(0)	X(2)	X(5)	X
0.55–0.75μ	(reflected solar radiation in the visible)	X(0)	X(1)	X(5)	X

*Only four channels were used on TIROS IV.
The parenthetical numbers indicate the approximate number of months of usable data for each channel.
**The radiometer on TIROS VII, launched 19 June 1963, is still operating as of this writing.
Detailed characteristics of all experiments are found in their respective Users' Manual (Staff members, 1961, 1962a, 1963, 1964).

(watts m^{-2} ster^{-1})" to which a sensor responds is given by

$$\bar{W} = \pi \bar{N} = \int_{\lambda=0}^{\infty} W_\lambda(T_{BB})\phi_\lambda d\lambda \qquad (1)$$

where $W\lambda$ (T_{BB}) is the spectral radiant emittance from a blackbody at temperature T_{BB} (i.e., Planck radiation) and ϕ_λ is the effective spectral response function for the particular channel. The constant π transforms Planck radiance (intensity) to blackbody radiant emittance (flux).

The calibration of the short-wave channels is carried out in the laboratory by viewing a diffuse sheet of white paper of known spectral reflectivity, ρ_λ, illuminated at normal incidence by a standard lamp of known spectral radiant intensity, J_λ (watts ster^{-1} micron^{-1}), at a known distance R (meters). The reflectance of the white paper is *assumed* to be ideally diffuse (i.e. Lambertian). The effective radiant emittance and radiance to which a sensor responds, then, are

$$\bar{W} = \pi \bar{N} = \frac{1}{R^2} \int_{\lambda=0}^{\infty} J_\lambda \rho_\lambda \phi_\lambda d\lambda \qquad (2)$$

In order to interpret the short-wave measurements in terms of reflectance of solar radiation from the earth and atmosphere, one must know the "effective solar constant, $\bar{W}^*$ (watts m^{-2})", i.e., the measurement which would result if the radiometer field-of-view were completely filled by an ideally diffuse reflector of unit reflectivity when illuminated by one solar constant at normal incidence, expressed by

$$\bar{W}^* = \int_0^{\infty} S_\lambda \phi_\lambda d\lambda \qquad (3)$$

where S_λ is the spectral irradiance of the earth by the sun at unit distance. The solar constant

$\int_0^{\infty} S_\lambda d\lambda$ equals 1395 watts m^{-2}

$= 2.0\ \varsigma$ cal cm^{-2} min^{-1}.

The reflectance of solar radiation, r_o, assuming that the reflectance is *independent of wavelength* and that the radiation is *reflected isotropically* is given by

$$r_0 = \frac{\bar{W}}{\bar{W}^* \cos \zeta} (R_s)^2 \qquad (4)$$

where ζ is the solar zenith angle and R_s is the distance of the earth from the sun in astronomical units. Equation (4) was used to calculate reflectance (and, assuming isotropy, albedos) in this paper, although there is evidence that appreciable errors may result if the angular dependence of the backscattered radiation is not included (Arking, 1964). The maximum allowable solar zenith angle, to insure adequate illumination, was $\zeta = 70°$.

III. THE PLANETARY HEAT BALANCE

We shall only investigate the heat balance at the top of the atmosphere, and not consider the more complicated study of the heat budget of the atmosphere alone at this time. In our study we shall use data obtained over a full year from TIROS VII. A number of problems have arisen concerning the interpretation and treatment of the satellite data. We shall discuss the most important of these below.

a. Coverage

The orbit of TIROS VII is nearly circular, at a mean height of 635 km, and inclined to the equator by 58°. For this reason, the polar regions cannot be observed (Figure 1). The maximum allowable angle of the sensor optical axis from the nadir (the "nadir angle") was 45° for data to be accepted. Therefore, the coverage was extended beyond the subsatellite track to latitudes 63.5°N. and S. For our purposes this broad zone of coverage, which includes 89.5 percent of the world's surface area, shall be termed the "quasi-globe." The two "V-shaped" gaps near Asia and South America (Figure 1) result from the combined constraints of the location of the three Command and Data Acquisition Stations in the United States and the limit of the storage capacity of the satellite tape recorder to one orbit of data. In addition, there is a temporal coverage problem. Because of the orbital regression and the motion of the sun, a period of 76 days (the "orbital synodic cycle") is required for the orbital nodes to move through one complete cycle relative to the sun and, hence, for all latitudes to be sampled at all local times. Figure 2 shows the heliocentric views of the TIROS VII orbit at launch and at four quarters throughout a 76-day cycle.

b. Selection of Spectral Intervals

Ideally the best inference of total long-wave radiation should derive from measurements by the 8–30 micron channel and the best inference of total reflected solar radiation should derive from

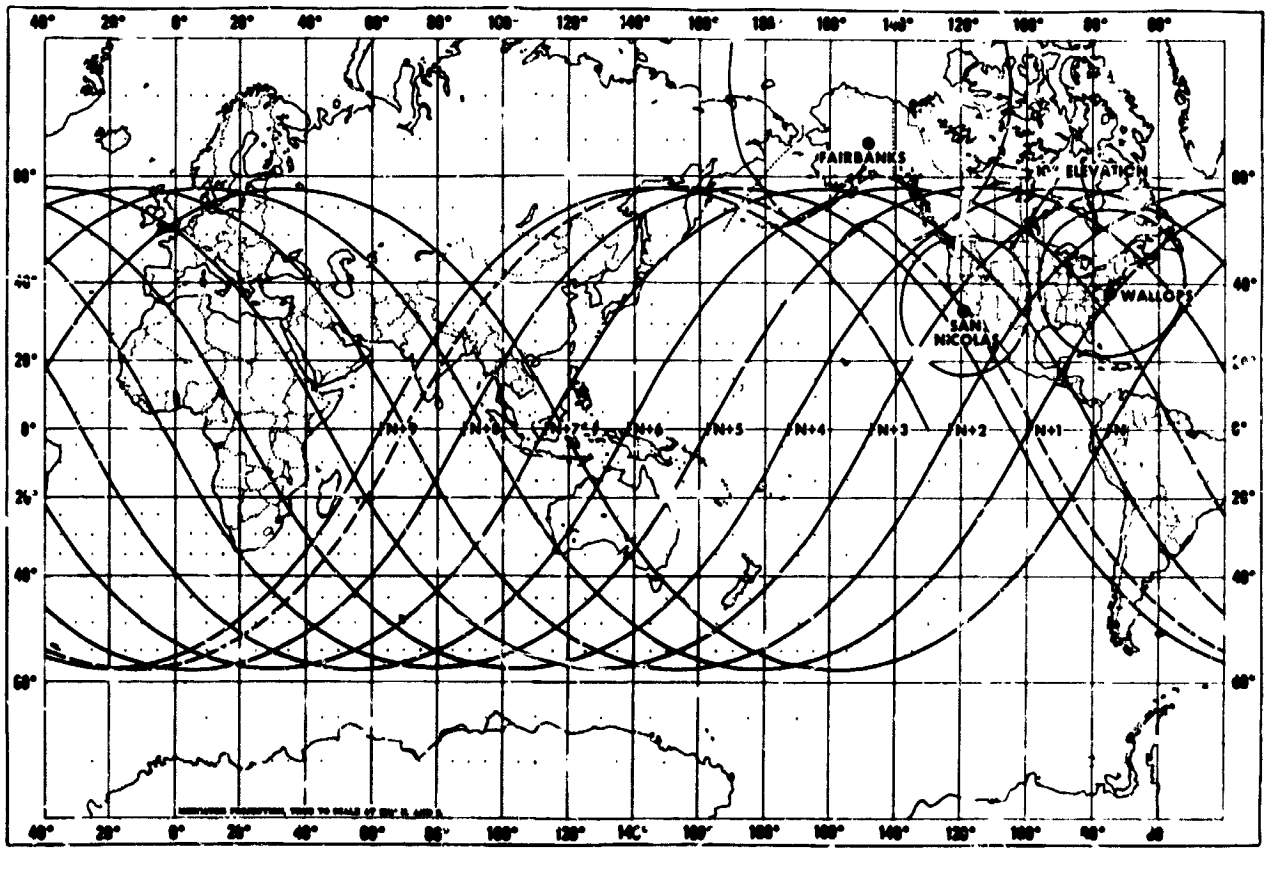


FIGURE 1.—Nominal world-wide radiation data coverage for the Tiros VII satellite.

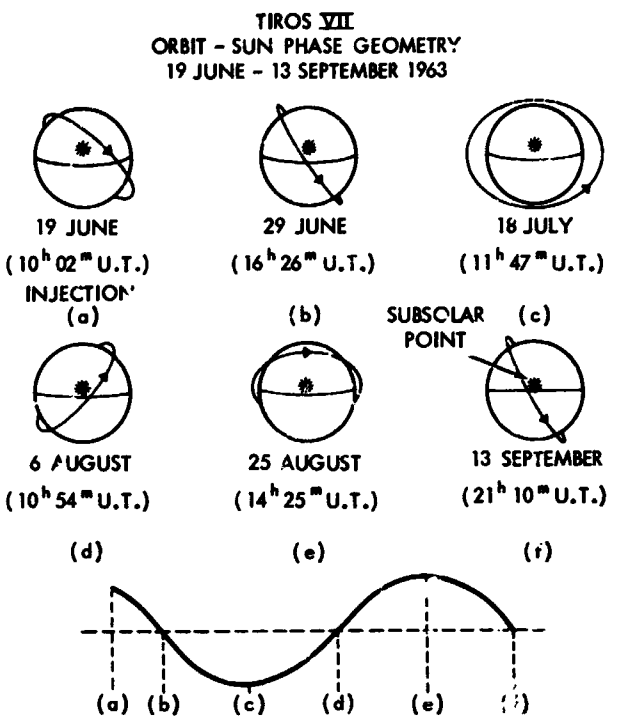


FIGURE 2.—Heliocentric views of the Tiros VII orbit, illustrating the 76-day orbital synodic cycle. The sinusoid shows the convention for indicating the orbit-sun phase geometry in Figures 5 and 6.

measurement by the 0.2–6 micron channel. However, because of engineering problems and the long-term stability of the sensors, it was decided to use the 8–12 micron and 0.55–0.75 micron channels instead (cf. Staff members, 1964).

c. Method of Converting T_{BB} Measurements to Total Outgoing Long-Wave Flux, W

The method of Wark, Yamamoto, and Lienesch (1962) was adopted for converting the essentially beam measurements of the 8–12 micron channel to total outgoing flux. Plots of W vs. T_{BB} (and $\bar{W}$) calculated for three zenith angles are shown in Figure 3. The linear fit

$$W(\text{ly min}^{-1}) = 0.141 + 0.00907\ \bar{W}\ (\text{watts } m^{-2}) \quad (5)$$

was adopted for manually applying second-order corrections for observed degradation in the instrumental response to the data printed out by an electronic computer (cf. IIIe. below).

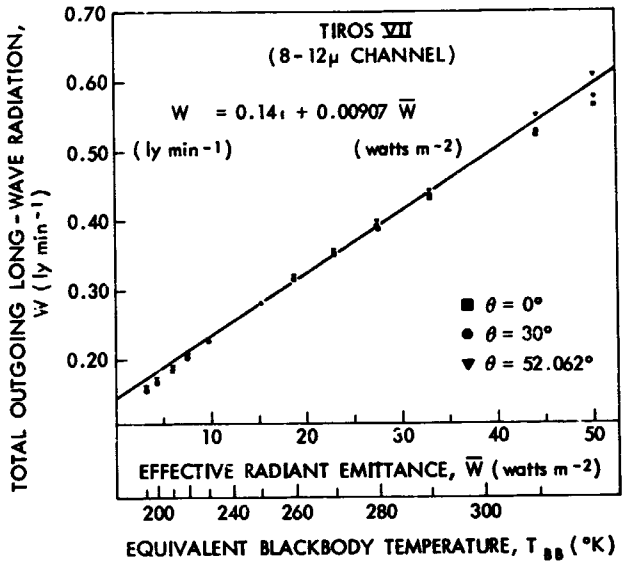


FIGURE 3.—Total outgoing long-wave radiation as a function of equivalent blackbody temperature (and effective radiant emittance) for the 8–12μ channel of Tiros VII for three zenith angles, θ. The straight line approximation was adopted for applying corrections for observed instrumental degradation.

d. Modification Factor for Limb Darkening

The theoretical limb darkening contained in the work of Wark et al. (1962) was determined under ideal conditions of either completely clear or completely overcast skies. The actual limb darkening was investigated using a two-dimensional distribution matrix containing 30 increments of radiance and 10 increments of zenith angle. The curve shown in Figure 4 was developed from the 50th percentile value of radiance in each of the zenith angle increments (after correcting for instrumental degradation; cf. TIROS IV Catalog-Manual: Staff members, 1963), calculated by a computer from nearly 1.3×10^6 measurements by the 8–12 micron channel of TIROS IV. The corresponding theoretical limb darkening, calculated for the same vertical outgoing radiance, i.e., $\bar{N}(T_{BB}=268.18°K)$, is noticeably less in Figure 4. The ratio of the integral incorporating the empirical curve to that incorporating the theoretical curve, calculating the total fluxes from an arbitrary radiance measurement, N', at an average zenith angle, $\theta_{ave}=35.5°$, yielded the value

$$\frac{W_{\text{EMPIRICAL}}}{W_{\text{THEORETICAL}}} = \frac{N'/E(\theta=3.35°)\int_0^{\pi/2}E(\theta)\sin\theta\cos\theta\,d\theta}{N'/T(\theta=35.5°)\int_0^{\pi/2}T(\theta)\sin\theta\cos\theta\,d\theta} = 0.986 \quad (6)$$

where $E(\theta)=[N(\theta)/N(0)]_{\text{EMPIRICAL}}$ and $T(\theta)=[N(\theta)/N(0)]_{\text{THEORETICAL}}$.

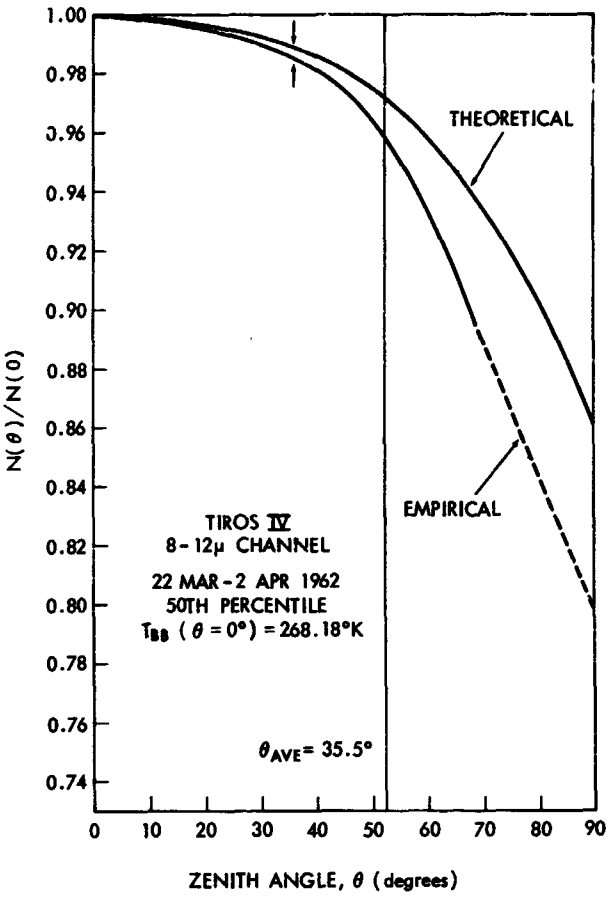


FIGURE 4.—Theoretical vs. empirical limb darkening for the 8–12μ channel of Tiros IV.

Values of total flux calculated by the computer, then, were reduced by the factor 0.986 as a first approximation to correct for the greater observed limb effects compared to the theoretically determined limb darkening.

e. Data Processing and Degradation of the Instrumental Response

An electronic computer was employed to process the data contained on magnetic tape (Final Meteorological Radiation Tapes, FMRT; cf. TIROS VII Catalog-Manual: Staff members, 1964) to a grid print map format (cf. Staff members, 1962b). Briefly, a square-mesh grid containing 2376 elements was superimposed on a Mercator map base of the quasi-globe, and measurements were accumulated in appropriate grid squares by the computer over a period of about seven days. Then averages of all measurements in each grid square were taken resulting in a weekly map of data. In producing the weekly map the computer also calculated and printed out the average of all individual measurements processed (the "map average").

In Figure 5, the effective radiant emittances, $\bar{W}$, calculated by means of Equation (5) from the map average values of total outgoing long-wave flux, are shown for each of 19 weekly periods during the first year of TIROS VII (dots with bars). The 19 periods are numbered along the upper abscissa, and the sinusoid indicates the orbit-sun phase geometry (cf. Figure 2). The Roman numerals at the top of the figure indicate the calendar months making up the four seasons, each of which is represented by averages of data from four weekly periods, equally spaced every 19 days over a 76-day synodic cycle. The lengths of the bars indicate the number of days making up the data period, usually seven. A least-squares quadratic curve is shown for the map averages. The decrease with time is interpreted as indicating a degradation in the response of the instrument while in orbit. There are unavoidable gaps in the quasi-global coverage, caused by such factors as noise and engineering considerations preventing the successful acquisition and reduction of every possible orbit, which might bias the quasi-global map averages. Upon close examination of the data, such biases were ascertained in

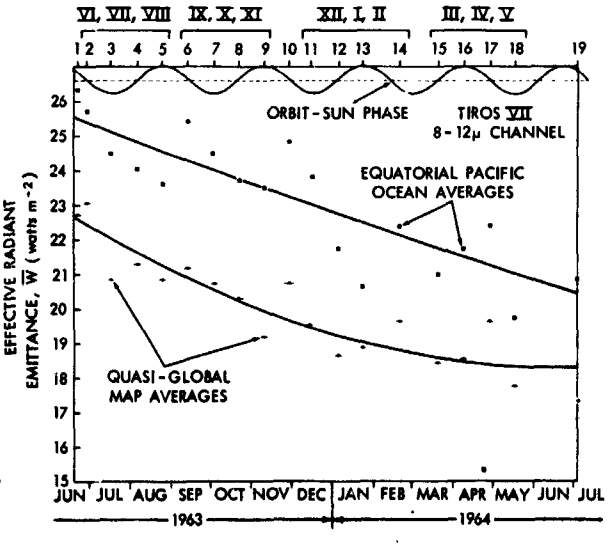


FIGURE 5.—Effective radiant emittances, $\bar{W}$, vs. time from "quasi-global" map averages and equatorial Pacific Ocean averages, indicating a degradation of the instrumental response. Tiros VII, 8–12μ channel.

the map averages of Figure 5. Therefore, values of W were calculated by means of Equation (5) for averages of the data over an equatorial region of the Pacific Ocean which was uniformly covered by all 19 weekly maps. The squares in Figure 5 refer to these values. A least-squares quadratic [illegible] for the Pacific Ocean averages is shown in Figure 5 (top curve), and was adopted as representing the true instrumental degradation. Based upon this curve, then, correction factors, $h_2(t)$, were calculated to correct for degradation in the instrumental response. Because of the linear response characteristic of the radiometer, a degradation-corrected value, $\bar{W}'$, can be expressed by

$$\bar{W}' = h_2(t)\ \bar{W} \tag{7}$$

where $\bar{W}$ is the uncorrected value. Since no degradation in the instrumental response of the 8–12 micron channel is presumed to exist at the time of launch ($t=0$), $h_2(0)=1$. The value of $h_2(t)$ increases throughout the year, reaching 1.250 at the midpoint of period 19 on 5 July 1964.

Finally the values of total outgoing flux calculated by the computer, W, were corrected manually for (1) instrumental degradation and (2) an empirically determined increase in limb darkening. From Equations (5) and (7) we have for the degradation-corrected value of flux, W'

$$W' = 0.141 + 0.00907\ h_2(t)\ \bar{W} \tag{8}$$

Combining Equations (5) and (8) to eliminate $\bar{W}$, we have

$$W' = h_2(t)\ W - 0.141\ [h_2(t) - 1] \tag{9}$$

Finally from Equations (6) and (9) we have for the outgoing flux, W'', corrected for both instrumental degradation and an empirically determined increase in limb darkening

$$W'' = 0.986\ \{h_2(t)W - 0.141[h_2(t) - 1]\} \tag{10}$$

where W is the value of flux calculated by the computer.

Quasi-global averages of reflectances calculated by the computer using Equation (4) from measurements made by the 0.55–0.75 micron channel for each of 17 weekly periods processed during the first year of TIROS VII are shown in Figure 6. Because of an unfavorable orbit-sun geometry,

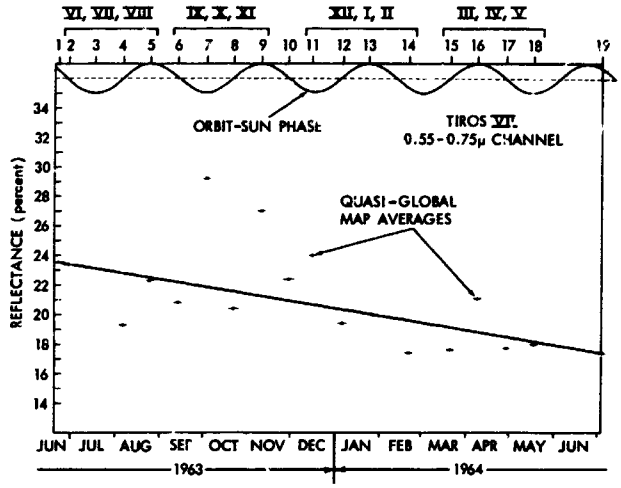


FIGURE 6.—"Quasi-global" map average reflectances vs. time, indicating a degradation of the instrumental response. Tiros VII, 0.55–0.75μ channel.

only very small amounts of data were acquired for periods 3 and 13, and these were rejected as being unrepresentative. A least-squares linear curve was calculated for the quasi-global map averages, showing again, a decrease with time. After some comparisons of reflectances calculated over "constant" targets (e.g., the Sahara Desert) for several periods throughout the year, the curve of Figure 6 was adopted as representing the *apparent* degradation of the instrumental response, and was used in determining the factor $h_5(t)$ for correcting the reflectances printed out by the computer. A reflectance value, corrected for apparent instrumental degradation, r_0', becomes

$$r_0' = h_5(t)\ r_0 \tag{11}$$

where r_0 is the uncorrected value.

In initially treating the data using Equations (10) and (11), it was found that the annual planetary albedo required to balance the annual long-wave flux determined from the 8–12 micron channel was 32.2 percent (cf. IIIf. below). On the other hand the annual planetary albedo inferred from values of reflectance calculated by Equation (11) was only 20.1 percent. If planetary radiative equilibrium exists over one year (and we presume that it very nearly does), there is a glaring inconsistency between the two types of data. There has been a history in the previous TIROS radiometers of abnormally low reflectances and of observed decreases in instrumental

response *even before launch* (Bandeen, et al., 1963). On the other hand there have been no instances where the long-wave channels were observed to degrade before launch or yield "abnormal" results immediately after launch. Therefore, we shall accept the long-wave fluxes as modified by Equation (10), and multiply the reflectance values, r'_0, calculated by Equation (11) by a factor required to satisfy annual radiative equilibrium. This factor is

$$\frac{32.2}{20.1}=1.60 \qquad (12)$$

Finally, from Equations (11) and (12) we have for the reflectances (equivalent to albedo under our assumption of isotropy), corrected for an apparent instrumental degradation and further increased by factor 1.60 to effect annual planetary radiative equilibrium,

$$r_0''=\text{albedo}=1.60\ h_5(t)r_0 \qquad (13)$$

The additional 1.60 factor is presumed to be due to (1) uncertainties in the backscattered fluxes resulting from not including possible angular dependencies, (2) uncertainties in the original calibration, (3) unobserved deterioration of the instrument between the last check-of-calibration on 5 June 1963 and the first orbit of data on 19 June 1963, (4) the assumption that the reflectance of radiation is independent of wavelength, or (5) some combination of the foregoing. This problem is very puzzling and is under continuing study, but to date no satisfactory explanation has been found.

Zonal averages of the weekly maps were "corrected" in accordance with Eqs. (10) and (13), and averages from the applicable set of four weekly maps constituted the data for each season. Meridional integrations were accomplished graphically by means of a planimeter, using "energy diagrams" like those of Figs. 7–11 for long-wave and Fig. 12 for short-wave radiation.

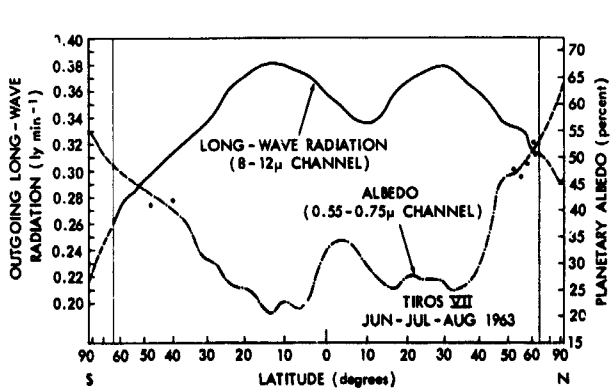


FIGURE 7.—Outgoing long-wave radiation and albedo for June–July–August 1963. The abscissa is scaled in proportion to zonal areas (sine of the latitude).

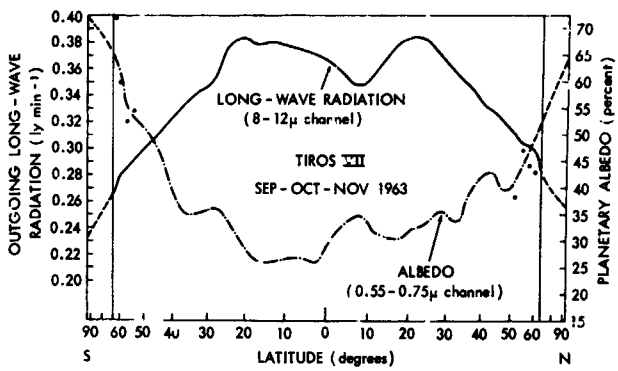


FIGURE 8.—Outgoing long-wave radiation and albedo for September–October–November 1963. The abscissa is scaled in proportion to zonal areas (sine of the latitude).

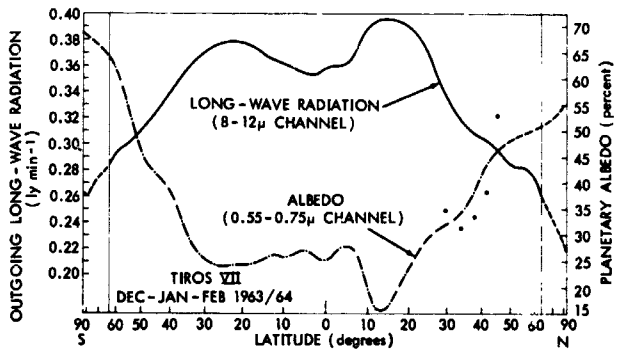


FIGURE 9.—Outgoing long-wave radiation and albedo for December–January–February 1963/64. The abscissa is scaled in proportion to zonal areas (sine of the latitude).

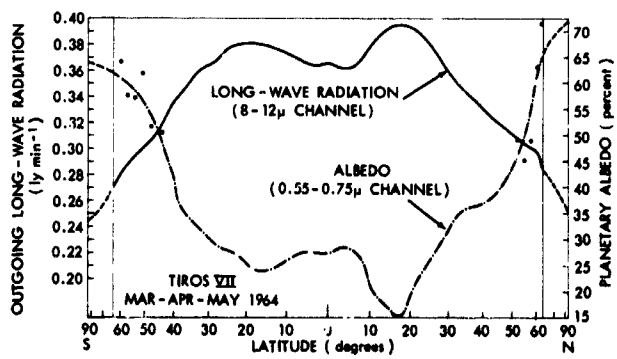


FIGURE 10.—Outgoing long-wave radiation and albedo for March–April–May 1964. The abscissa is scaled in proportion to zonal areas (sine of the latitude).

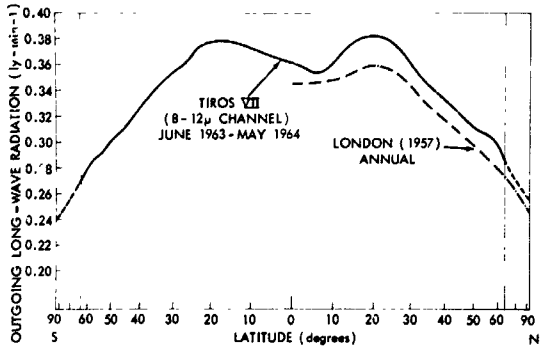


FIGURE 11.—Annual mean outgoing long-wave radiation. The corresponding values of London (1957) for the northern hemisphere are shown for comparison. The abscissa is scaled in proportion to zonal areas (sine of the latitude).

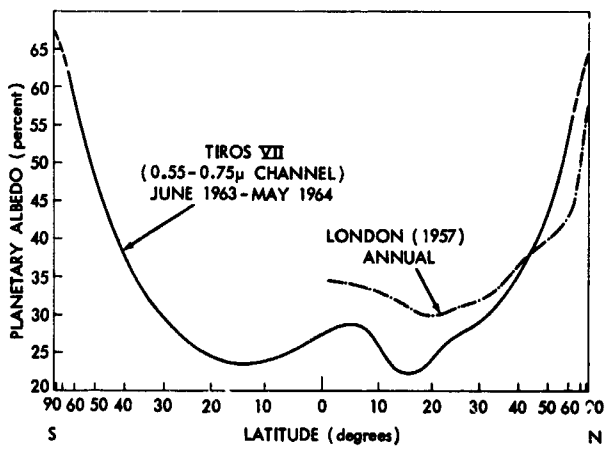


FIGURE 12.—Annual mean planetary albedo. The corresponding values of London (1957) for the northern hemisphere are shown for comparison. The abscissa is scaled in proportion to the annual zonal isolation (top of the atmosphere).

f. Results

The final values of outgoing long-wave radiation and albedo are summarized in Table II.

The seasonal values of long-wave radiation and albedo are shown in Figures 7, 8, 9 and 10 and the annual values are shown in Figures 11 and 12. Vertical lines at latitudes 63.5°N. and S. mark the boundaries of the quasi-globe. Dashed lines indicate an extrapolation of the curves beyond (or near) these boundaries where satellite data were not acquired. Dots show the scatter of the short-wave seasonal data at high latitudes where the population of the data samples becomes increasingly smaller. Data from London (1957) are shown in Figures 11 and 12 for comparison.

Two maxima in the long-wave radiation, associated with the subtropical high pressure regions, show up clearly in all figures as does a tropical minimum, associated with the cloudiness of the Intertropical Convergence Zone (I T C). The location of the long-wave tropical minimum remains in the northern hemisphere in all seasons except in December-January-February, clearly pointing up the asymmetry between the hemispheres. Similarly, the long-wave maximum occurs over the subtropical high of the northern hemisphere in all seasons except June-July-August, when the long-wave radiation over the subtropical high in the southern hemisphere becomes only slightly larger. Generally more long-wave radiation is emitted from the extratropical latitudes of the northern hemisphere than from the same latitudes of the southern hemisphere in

TABLE II.—*Final Seasonal and Annual Values of Long and Short-Wave Radiation*

Season (Months)	Planetary Long-Wave Radiant Power (10^{15} cal min^{-1})	Solar Radiation		Planetary Albedo (%)
		Reflected (10^{15} cal min^{-1})	Incident on Planet (10^{15} cal min^{-1})	
J-J-A	1725.2	699.0	2457.8	28.4
S-O-N	1725.2	922.2	2574.8	35.8
D-J-F	1712.2	859.7	2609.8	32.9
M-A-M	1746.5	796.0	2544.2	31.3
Annual	1727.3	819.4	2546.7	32.2

corresponding seasons. However, in the tropics the situation is reversed where less radiation is emitted from the northern hemisphere than from the southern hemisphere in corresponding seasons. This characteristic tends to equalize the total radiation emitted from the two hemispheres annually. The albedo minimums occur in the tropics of the winter and spring hemispheres but are lowest when they occur in the northern hemisphere.

The tropical annual minimum in the long-wave radiation (and, hence, the approximate mean annual position of the ITC) is located at latitude 5°N, whereas the maximum long-wave radiation is located at latitude 20°N (Figure 11). The emission is smaller in the northern hemisphere between the equator and latitude 15° but larger between latitude 15° and the pole than in the southern hemisphere. The overall long-wave radiation is slightly larger than London's (1957) as would be expected from a comparison of the TIROS-inferred albedo of 32.2 percent with London's 35.2 percent.

The tropical maximum in the albedo of Figure 12 occurs at latitude 5°N, consistent with the long-wave minimum, but the minimum albedo occurs at latitude 16°N, slightly south of the position of the long-wave maximum. The albedo in the northern hemisphere generally is larger between the equator and latitude 12° but smaller at higher latitudes than in the southern hemisphere. It is interesting to note the greater meridional gradient of the TIROS-inferred albedo compared to London (1957) in Figure 12. The consequences of this characteristic is a larger required poleward flux of heat across latitude circles.

The maximum poleward flux of heat (Figure 13) is 9.8×10^{16} cal min^{-1} across latitude 36°S, whereas in the northern hemisphere the maximum is only slightly smaller at 9.6×10^{16} cal min^{-1} across latitude 36°N. A somewhat surprising result is that the poleward flux curve crosses zero near the equator at latitude 0.4°S whereas previous studies (e.g., Simpson, 1929) placed the crossover at about latitude 4.0°N. The annual curve given by London (1957) is shown for comparison.

The seasonal variations of albedos inferred from TIROS VII are shown in Figure 14. Monthly averages of the stellar magnitudes of measurements of earthlight reflected from the moon made

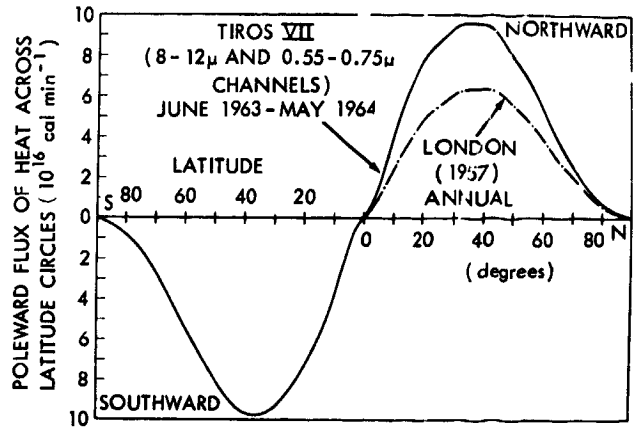


FIGURE 13.—Annual required poleward flux of heat across latitude circles. The corresponding values of London (1957) for the northern hemisphere are shown for comparison.

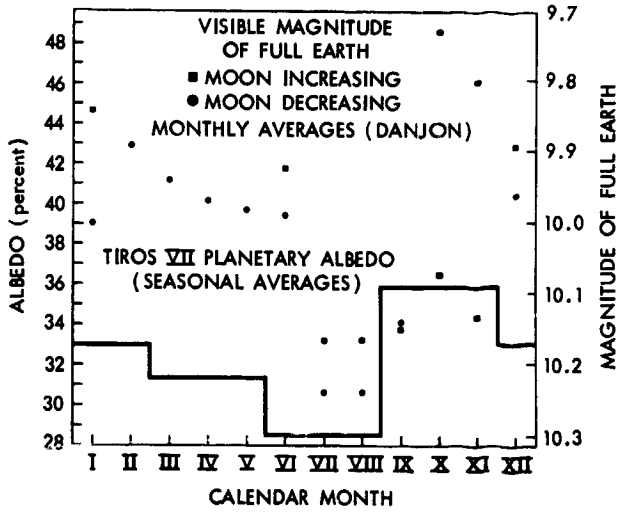


FIGURE 14.—Tiros VII seasonal albedos compared with monthly averages of the magnitudes of earthlight reflected from the moon (after Danjon).

during the years 1926–1935 by Danjon (1936, 1954) are also plotted. There is an approximate correspondence between the albedo and magnitude ordinates. Danjon found a seasonal minimum in the visual albedo of 32 percent in July and a maximum of 52 percent in October. Fritz (1949) pointed out that Danjon's measurements did not include the Pacific Ocean. More recent measurements of this type made by Dzhasybekova, Kazachevskii, and Khari. ov (1960) in the Kazakh SSR during the years 1957-1958 showed considerable variations in the visible albedo (between 31 percent and 48 percent) but did not

show any clear seasonal fluctuations. However, the data of Figure 14 do seem to give further evidence of a minimum around July and a maximum around October.

IV. DETERMINATION OF THE MEAN RELATIVE HUMIDITY AND WATER VAPOR MASS OF THE UPPER TROPOSPHERE

Möller (1961 and 1962) developed a method for the determination of the mean relative humidity of the troposphere and of the surface temperature of clouds or of the ground from radiation data of the TIROS meteorological satellites, and Möller and Raschke (1963/64) have further developed the method. We shall apply the method to data from the 6–6.5 micron and 8–12 micron channels of TIROS IV during the week 11–18 April 1962, and we shall extend the method to include the inference of the mass of water vapor above 500 mb.

The principle of the method is illustrated in Figure 15. Since approximately the same lapse rate exists in all atmospheres ($\gamma \sim 6.5°C/km$), and

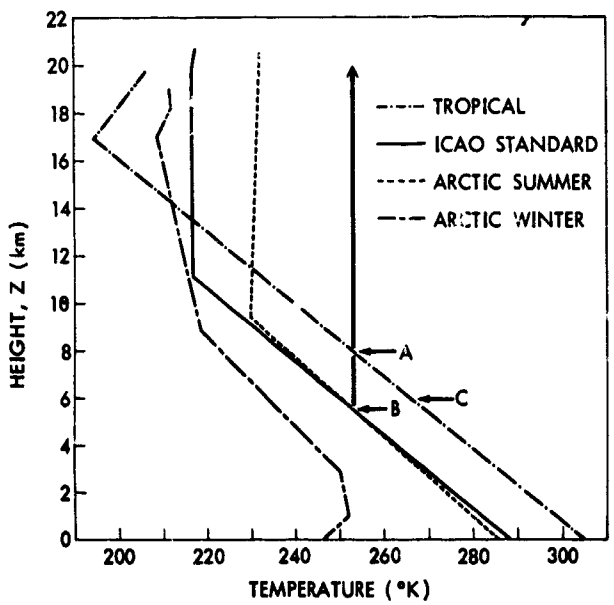


FIGURE 15.—Schematic illustrating that radiation of a given temperature is emitted from approximately the same optical depth in all atmospheres with the same relative humidity (after the method of Möller).

the saturation water vapor pressure is a function only of temperature, it follows that the same water vapor mass exists above the same temperature in all atmospheres having the same relative humidity. The water vapor mass, u, above height, z, can be expressed by

$$u = \int_z^\infty \frac{e_s(T(z))r(z)}{T(z)R^*/m_w} dz \qquad (14)$$

[(a) in Fig. 15]

where e_s is the saturation water vapor pressure, T is temperature, r is relative humidity, R^* is the universal gas constant, and m_w is the molecular weight of water vapor. Thus in Figure 15 Equation (14) would calculate approximately the same water vapor mass above height A in the tropical atmosphere as it would above height B in the ICAO standard atmosphere for the same relative humidities. The reduced mass of water vapor

$$u^* = \int_0^u \left(\frac{P}{P_0}\right)^n du \qquad (15)$$

[(b) in Fig. 15]

which is required for the radiative transfer problem, would be slightly different above the two heights of the previous illustration because of the term $(P/P_0)^n$ which takes into account the pressure broadening of the spectral lines in the absorption band. However, Möller (1961 and 1962) has shown that for the same relative humidity and a lower boundary (ground or cloud) at the same temperature, the outgoing radiance in the interval 6–6.5 microns is nearly the same for all atmospheres. Therefore, to a good approximation one can apply the calculations for one atmosphere to all atmospheres. The evaluation diagram in Figure 16 is designed to determine the mean relative humidity and the surface temperature (of the ground or of clouds) from simultaneous measurements of the 6–6.5 micron channel ("channel 1") and the 8–12 micron channel ("channel 2") (see Möller, 1962 for a detailed discussion of the construction of the diagram). The evaluation diagram in Figure 16 was constructed for the tropical atmosphere of Figure 15 for TIROS III, using the radiation tables of Elsasser (1960) with a pressure dependence exponent of $n = 1$ in Equation (15). No attempt was made to "correct" the 6–6.5 micron data, but the 8–12 micron data were corrected for instrumental degradation in accordance with the nomogram given in the TIROS IV Catalog-Manual (Staff members, 1963). The maximum allowable nadir angle was set at

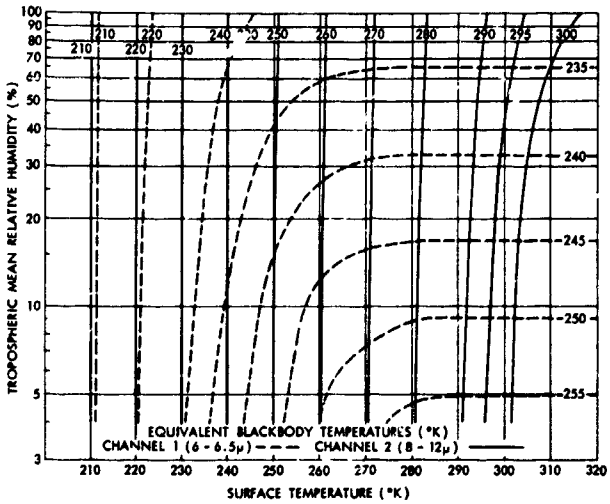


FIGURE 16.—Evaluation diagram for determining tropospheric mean relative humidities and surface temperatures of the ground or of clouds.

40° to eliminate excessive limb effects. We merely state that the spectral response functions of the 6–6.5 micron and 8–12 micron channels of TIROS III and TIROS IV are very nearly the same as a justification for using the TIROS III diagram with TIROS IV data. Also, the pressure dependence exponent, n, should probably be somewhat less than 1. These items notwithstanding, maps of the 6–6.5 micron and 8–12 micron data of TIROS IV were produced by a computer (similar to the maps used in the heat balance study) for the week 11–18 April 1962, and the conversion to relative humidities was carried out by hand using Figure 16. The resultant "quasi-global" relative humidities are shown in Figure 17.

However, if one is interested in the actual water vapor mass, relative humidities can be misleading (i.e., a high relative humidity over the winter arctic may involve much less water vapor than a low relative humidity over the Sahara Desert). To determine the water vapor mass above a given level of a particular atmosphere, having already determined the mean relative humidity, we merely computed the integral of Equation (14) above the temperature in the tropical atmosphere corresponding to the temperature at the given level of the atmosphere in question. For example, suppose that "B" and "C" in Figure 15 are both at the 500 mb levels of their respective atmospheres. Then the water vapor mass above 500 mb in the tropical atmosphere would be determined by insert-

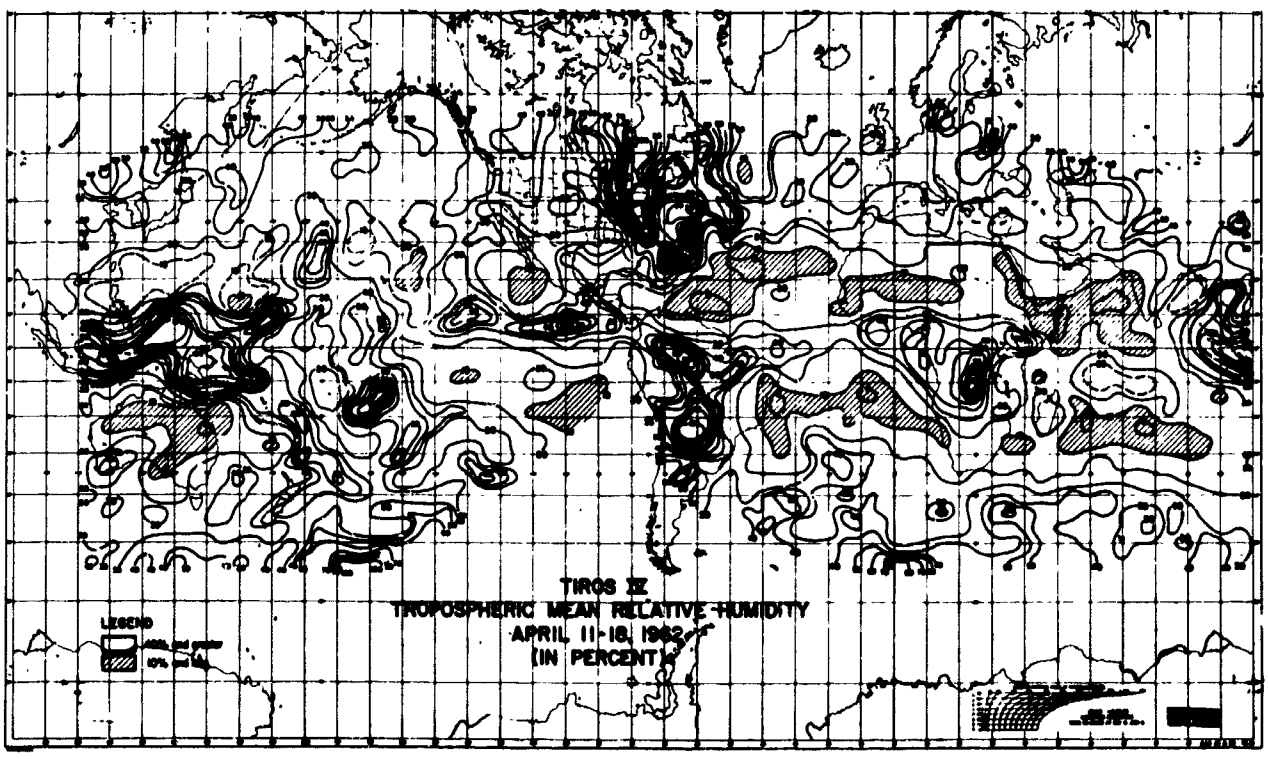


FIGURE 17.—Mean tropospheric relative humidities determined from simultaneous 6.0–6.5 micron and 8–12 micron measurements by Tiros IV during the period 11–18 April 1962, in percent.

ing the measured relative humidity in Equation (14) and integrating over the tropical atmosphere from level "C" to infinity. The water vapor mass above 500 mb in the ICAO standard atmosphere would be determined by inserting the measured relative humidity in Equation (14) and integrating over the tropical atmosphere again, but from level "A" (at the same temperature as "B") to infinity. It was decided to make the determination of water vapor mass above 500 mb because the greater part of the radiation sensed by the 6–6.5 micron channel originates above 500 mb (Hanel and Wark, 1961), and it was convenient to compare the results with a previous study using conventional data by Bannon and Steele (1960).

A world map of 500 mb temperatures was constructed from radiosonde and climatological data for the week of 11–18 April 1962. Using this map and the relative humidities of Figure 17, calculations of water vapor mass were carried out using the technique described above.

The results of the calculations are shown in Figure 18. The agreement with the results of Bannon and Steele (1960) shown in Figure 19 is generally good. Three major concentrations of water vapor are located over Indonesia, South America, and Africa, comparing favorably both in position and magnitude with Figure 19, although there are smaller included areas of much higher concentrations indicating near-saturation. Two additional indicated moist regions over 6°N, 100°W and 15°S, 60°W (which are not seen in Figure 19) are located near the position of the ITC in April and were found to persist in an analysis of additional data two weeks later. Hence, they too may be semi-permanent in nature, but conclusions cannot be drawn without more extensive analyses. The marked dry tongue over the Sahara Desert extending westward to the Caribbean Sea and the relative dryness everywhere near latitude 20°N, with the consequent decrease in atmospheric absorption and emission, corroborates the long-wave peak near this latitude in Figure 10. Considering the many approximations used, it is not possible at this time to fully assess the validity of the results, but the patterns are encouraging for additional work in this area.

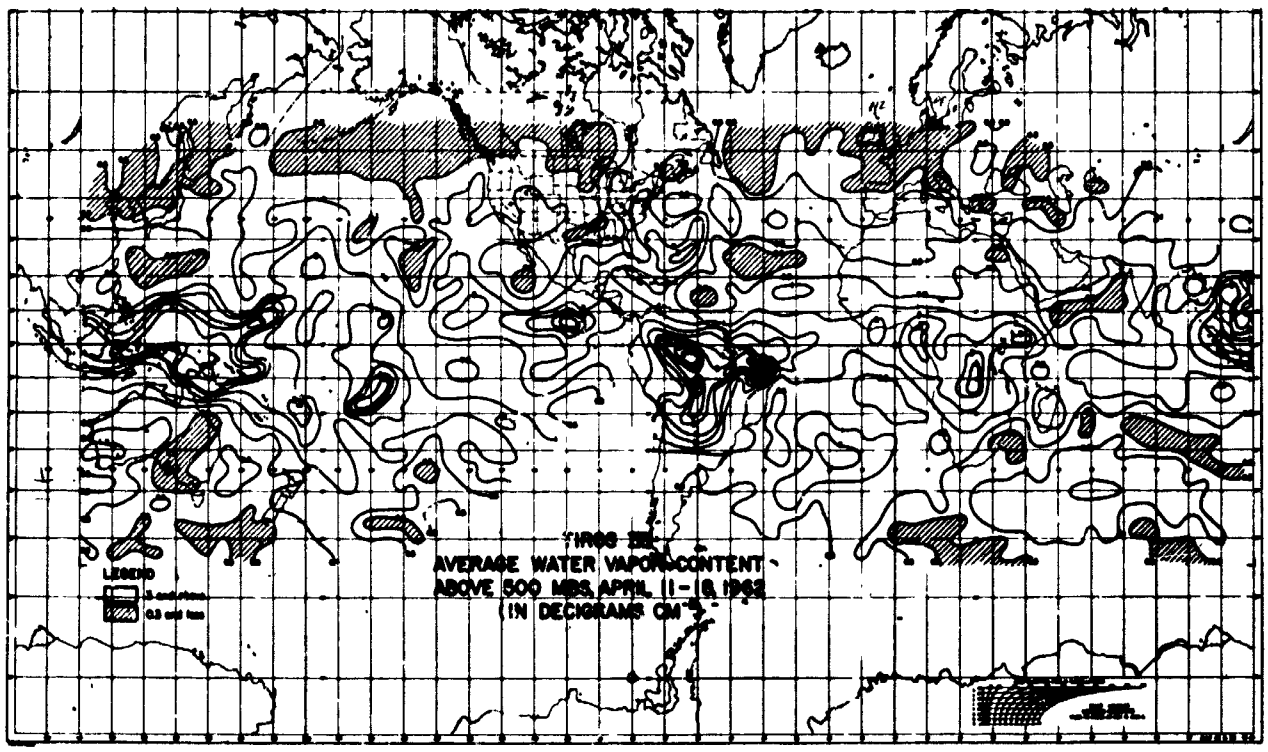


FIGURE 18.—Average water vapor content above 500 mb determined from simultaneous 6.0–6.5 micron and 8–12 micron measurements by Tiros IV and 500 mb temperatures from radiosonde data during the period 11–18 April 1962, in decigrams cm^{-2}.

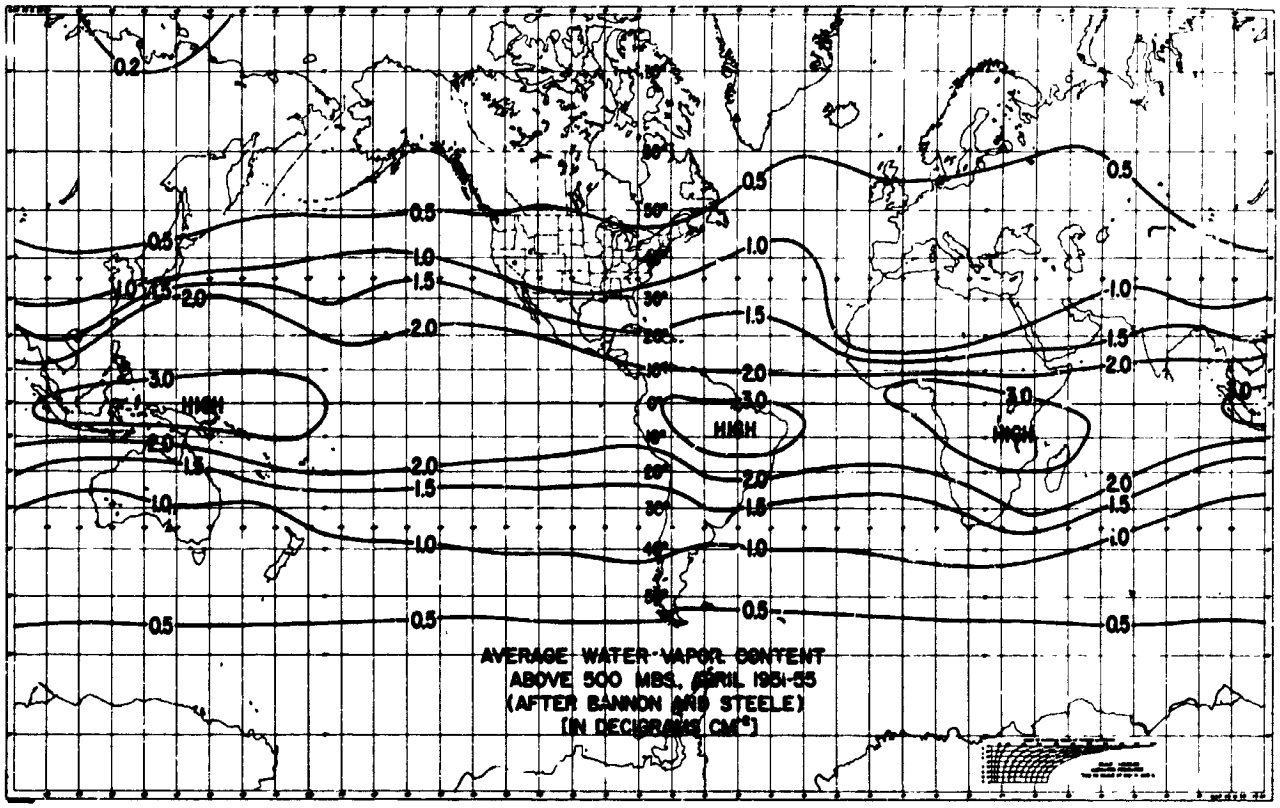


FIGURE 19.—Average water vapor content above 500 mb, April 1951–55, in decigrams cm^{-2}. (After Bannon and Steele).

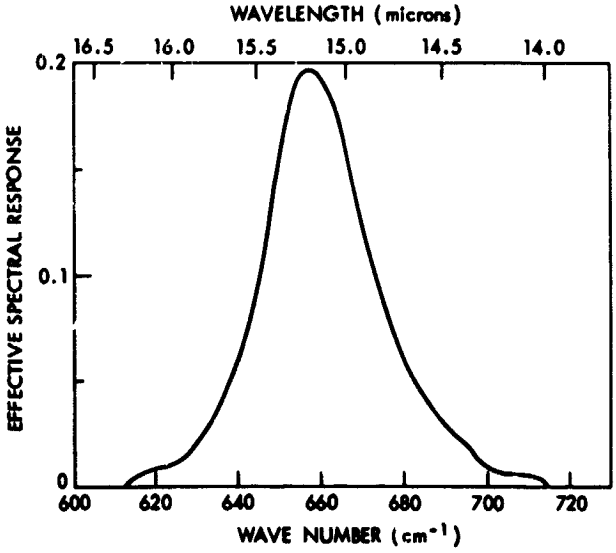


FIGURE 20.—Effective spectral response of the Tiros VII 15μ channel as a function of wavelength.

V. MEASUREMENT OF STRATOSPHERIC TEMPERATURES

One of the optical channels of the TIROS VII radiometer measured infrared radiation emitted mainly by carbon dioxide in the earth's atmosphere at wavelengths ranging from 14.8 to 15.5 microns. A detailed discussion of the physical significance of the measurements in this spectral interval and of the temperature patterns mapped during the first nine months of TIROS VII has been given by Nordberg et al. (1964). We shall only demonstrate by a few examples the ability of this channel to map the mean temperatures of the lower stratosphere on a global scale, and we refer the reader to the original paper for a complete discussion.

The outgoing effective radiance, $\bar{N}$, in the direction of the satellite can be expressed by

$$\bar{N} = \frac{1}{\pi} \int_{\lambda=0}^{\infty} \int_{\lambda=0}^{\infty} B_\lambda(T(h)) \phi_\lambda \frac{\partial \tau_\lambda(h)}{\partial h} dh \, d\lambda$$

$$\bar{N} = \frac{1}{\pi} \int_{\lambda=0}^{\infty} \Psi(h) \, dh \tag{16}$$

$$\Psi(h) = \int_{\lambda=0}^{\infty} B_\lambda(T(h)) \phi_\lambda \frac{\partial \tau_\lambda(h)}{\partial h} d\lambda$$

where

B = Planck function (flux)

ϕ_λ = spectral response function of channel

τ = atmospheric beam transmittance in the direction of the satellite

h = height above the surface

The spectral response function of the channel is shown in Figure 20. Four model atmospheres are shown in Figure 21(a). A homogeneous distribution of CO_2 of 0.00031 fractional parts by volume was assumed in the atmosphere. Using the radiation tables of Elsasser (1960) and a pressure-dependence exponent of $n=1$, the $\Psi(h)$ functions for the four model atmospheres were computed for nadir angles of 0° and 58° (equivalent to a zenith angle of 69°) and plotted in Figures 21(b) and (c). From Figure 21 it may be seen that about 65 percent of the integrated radiance reaching the satellite originates in the region from 15 to 35 km for a nadir angle of 0° and that the weighting curves are shifted upward about 2.5 km for a nadir angle of 58°. Thus even though there are contributions from higher and lower layers, it

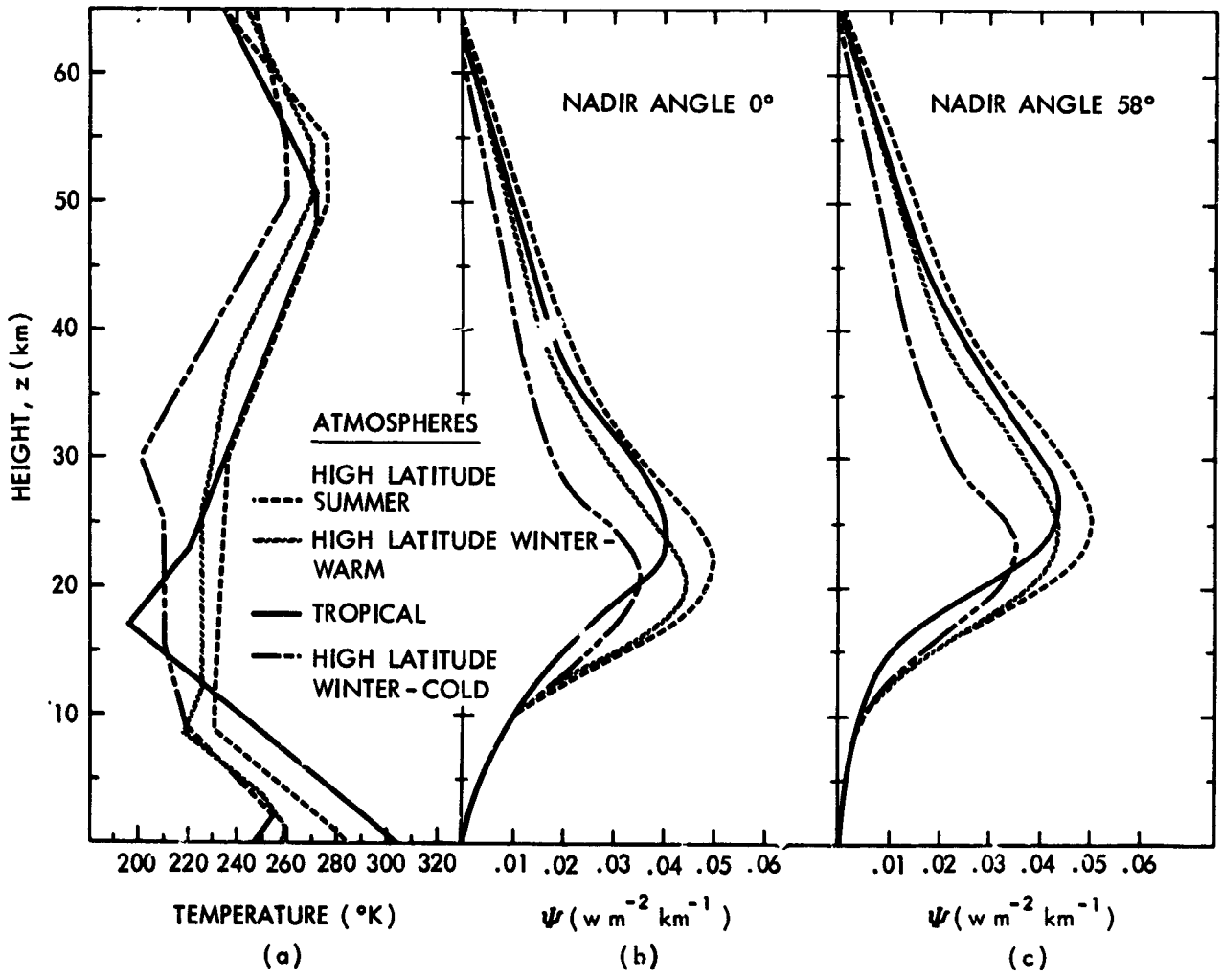


FIGURE 21(a).—Typical temperature profiles based on proposed supplements to the "U. S. Standard Atmosphere 1962" for 60° North summer, 60° North winter (warm and cold) and 15° North. The "warm" and "cold" temperature profiles for 60° N ("high latitude winter") can be considered typical at these latitudes depending on the state of the stratosphere in these regions.

FIGURE 21(b).— Weighting functions Ψ (h), applying to the measured outgoing radiance N; nadir angle = 0°.

FIGURE 21(c).—Weighting functions Ψ (h); applying to the measured outgoing radiance N; nadir angle = 58°.

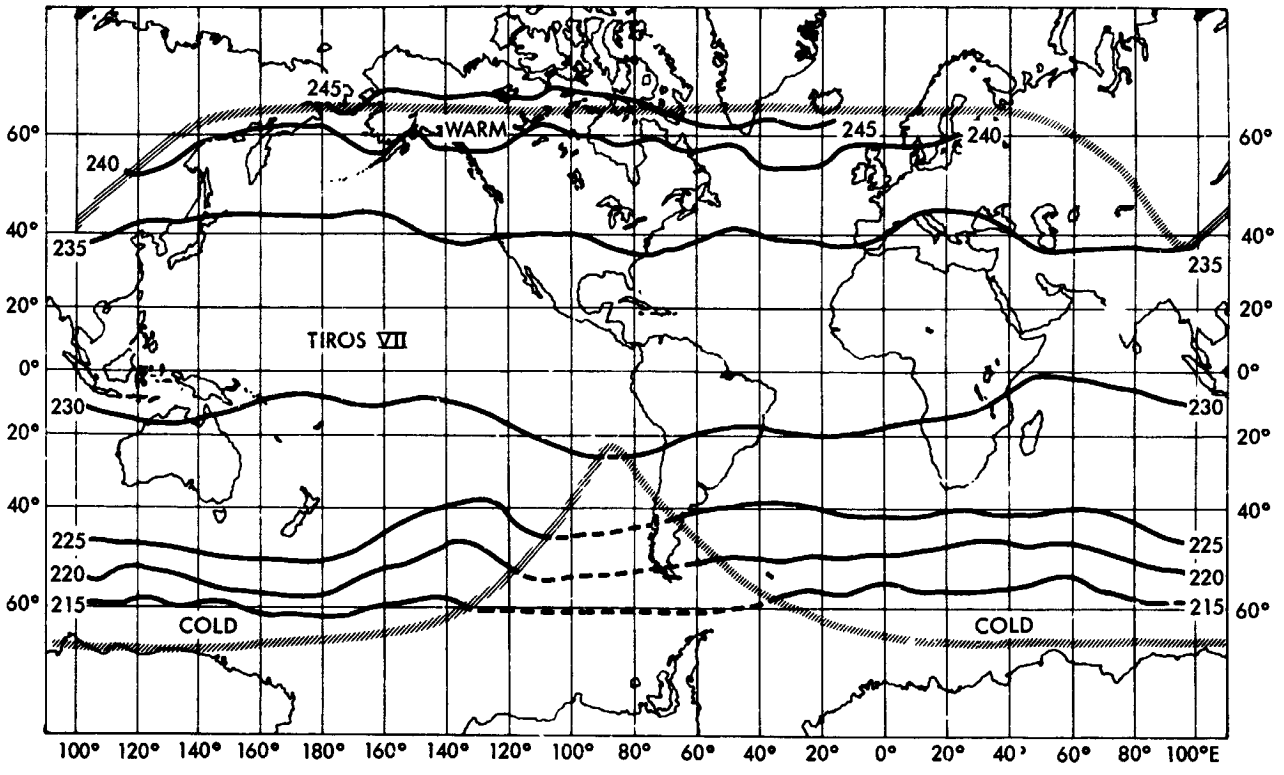


FIGURE 22.—Quasi-global map of 15μ equivalent blackbody temperatures averaged over the week 19–25 June 1963; nadir angle 0°–40°.

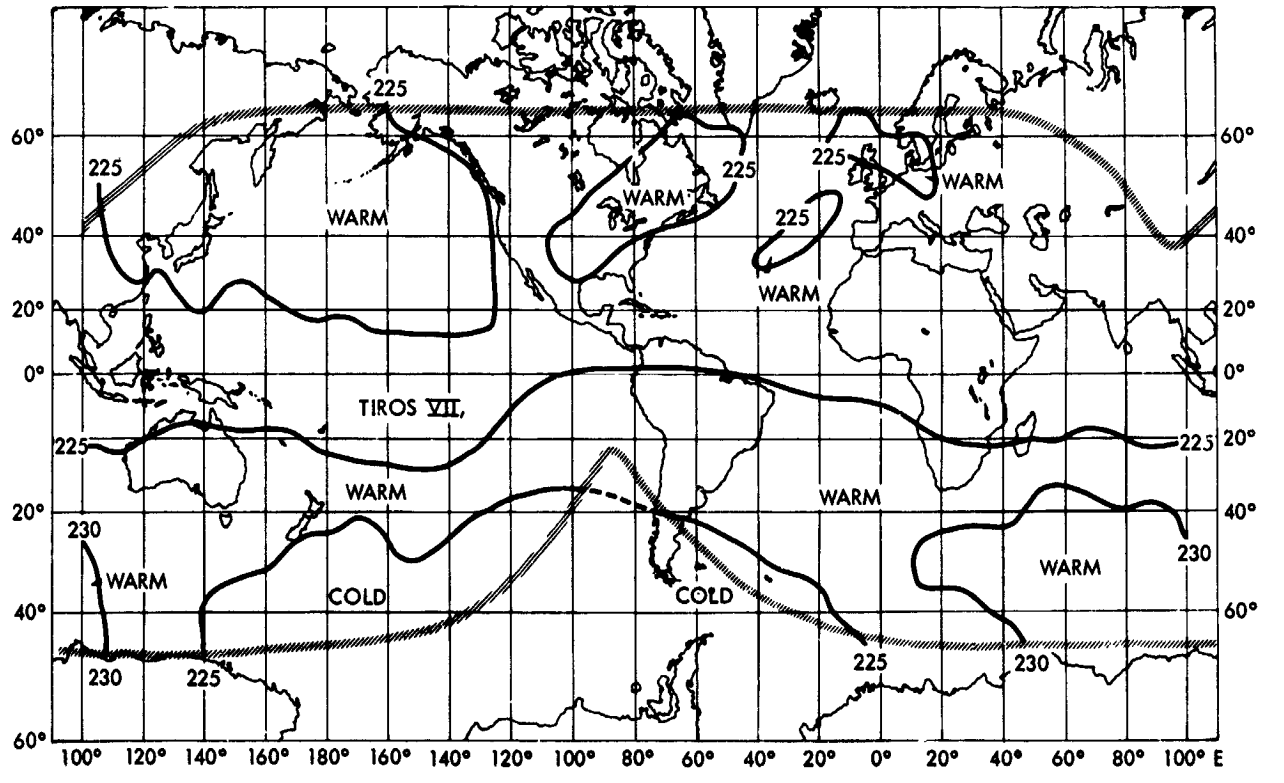


FIGURE 23.—Quasi-global map of 15μ equivalent blackbody temperatures averaged over the week 25 September-1 October 1963; nadir angle 0°–40°.

is seen that the measurements may primarily be interpreted in terms of temperatures of the lower stratosphere.

A typical temperature pattern at the start of the northern hemisphere summer and southern hemisphere winter is shown in Figure 22. Stratospheric mean temperatures averaged over the week immediately after the launch of TIROS VII, 19–25 June 1963, range from 240°K near latitude 60°N to 215°K near 60°S. The generally zonal structure of the temperature pattern confirms the established knowledge of the lower stratosphere, namely, the existence of a very strong cyclonic circulation system around the cold low pressure core centered near the winter pole and an anticyclonic circulation system around the warm high pressure system centered over the summer pole.

The map for the period 25 September–1 October 1963 illustrates the transition period. It is important to point out that due to instrumental degradation, all temperatures shown in Figure 23 and all succeeding maps are about 7°K too low, inasmuch as the degradation seemingly leveled off in July and August after an initial rather rapid decline. In Figure 23 the southern winter vortex has broken down and the temperature gradients between the two hemispheres have reached thier minimum during the transition period.

The map for the period 15–22 January 1964 (Figure 24) illustrates the state of the stratosphere in opposite seasons to those depicted in Figure 22. The southern hemisphere summer pattern is very much like its northern hemisphere counterpart in Figure 22. However, a striking difference is seen in the northern hemisphere winter. The "Aleutian Anticyclone", a regular climatological feature, displaces the cold stratospheric cyclone from the position of symmetry that one might expect it to have. The strong Aleutian Anticyclone established itself in the first week of December and remained throughout the entire winter.

During the last week of January 1964 a stratospheric warming of apparently moderate intensity occurred over the region of the Caspian Sea. The situation is summarized in Figure 25 which shows that on 27 January the radiation temperatures had increased by 15°K over the averages for the week of 9–15 January near the center. During the

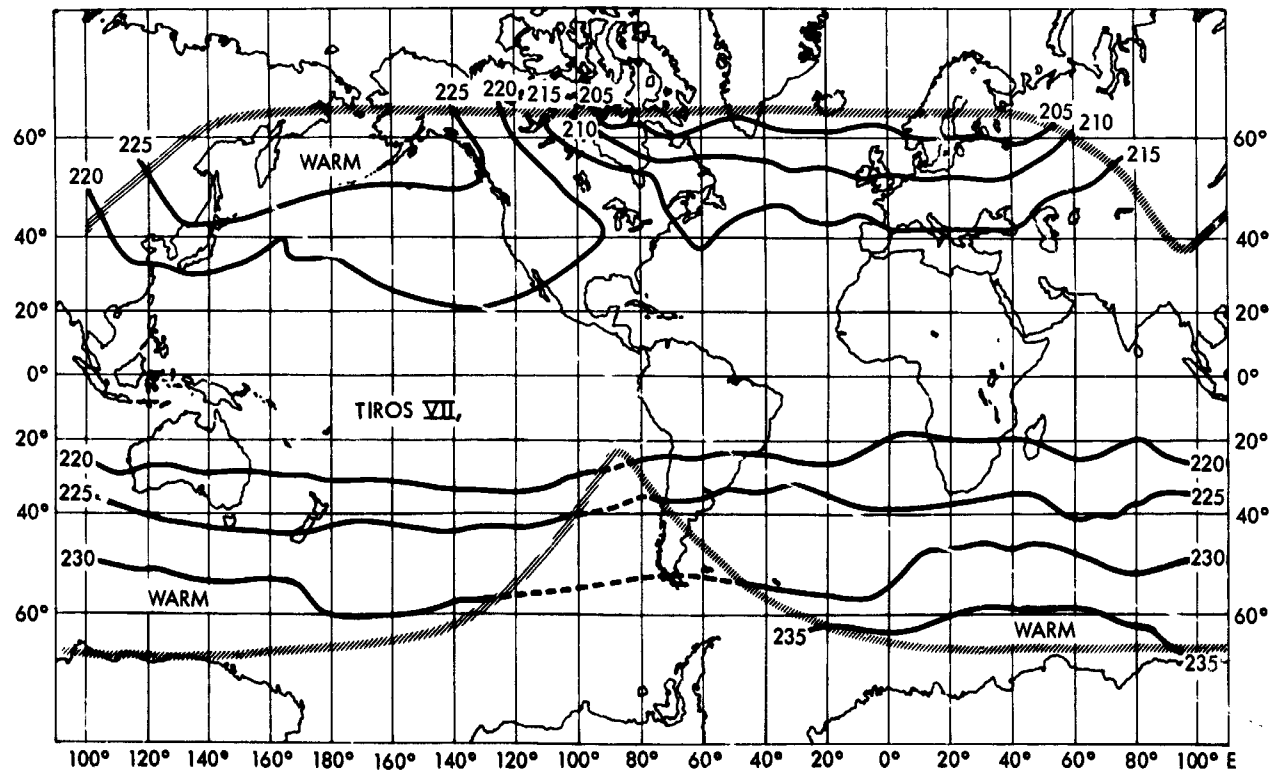


FIGURE 24.—Quasi-global map of 15μ equivalent blackbody temperatures averaged over the week of 15–22 January 1964; nadir angle 0°–40°.

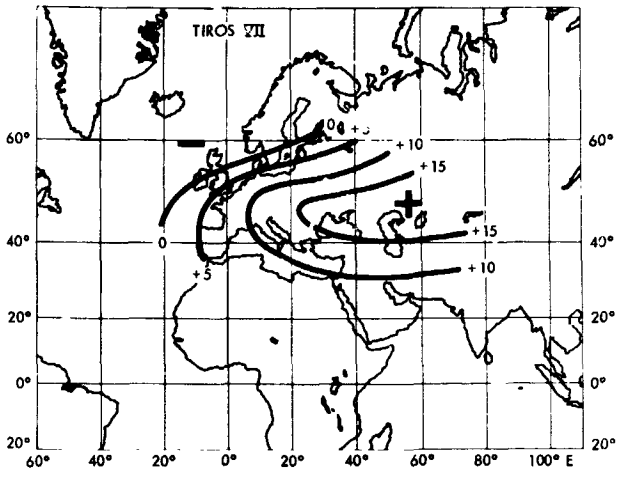


FIGURE 25.—Regional map of the differences between the 15μ equivalent blackbody temperatures measured on 27 January 1964 and those averaged over the week 9–15 January; nadir angle 0°–40°. The effects of a stratospheric warming over Eastern Europe are illustrated.

following week, 31 January–8 February, the warming passed its climax and dissipated soon thereafter.

The global temperature structure undergoes a major change during the period from early February to the period around the Vernal Equinox in Figure 26. The strong zonal winter temperature patterns, particularly over the eastern hemisphere, as well as the zonal summer pattern in the southern hemisphere have disappeared. The southern hemisphere has cooled considerably, resulting again, as in September, in an almost constant temperature, 223°+7°K, prevailing over the entire globe.

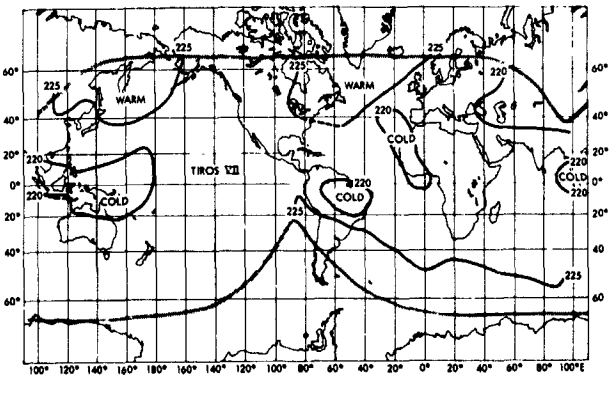


FIGURE 26.—Quasi-global map of 15μ equivalent blackbody temperatures averaged over the week 18–25 March 1964; nadir angle 0°–40°.

VI. SUMMARY

Studies involving satellite radiometric data have been carried out in three areas of climatological interest—the planetary heat balance, the upper tropospheric content and distribution of water vapor, and the temperature and associated circulation patterns of the lower stratosphere. Even though difficulties exist in certain areas, such as instrumental stability, and interpretations of the essentially-beam measurements, the results give a new insight into the large-scale, long-term radiative characteristics of our planet and the physical processes which can be derived from them. The comparative relationships indicating hemispheric asymmetries and seasonal fluctuations in the various parameters appear to be valid.

Finally, the potentialities of climatological investigations from future satellites—in polar orbits and with improved instrumentation—have been clearly demonstrated, we believe, by the studies reported in this paper.

VII. ACKNOWLEDGEMENTS

We gratefully acknowledge the contributions of the many persons, too numerous to mention individually, who have contributed to the success of the TIROS series of meteorological satellites. We are particularly indebted to Mr. Robert Hite and his group for their computational efforts in reducing the vast amount of radiation measurements and to Mr. Lewis Allison for his significant contributions in regard to the water vapor analyses.

REFERENCES

1. ARKING, ALBERT, 1964: "Angular Distribution of Scattered Radiation as Measured from TIROS." Paper to be presented at the International Radiation Symposium, IAMAP (IUGG), Leningrad, August 1964.
2. BANDEEN, W. R., R. A. HANEL, JOHN LICHT, R. A. STAMPFL, and W. G. STROUD, 1961: "Infrared and Reflected Solar Radiation Measurements from the TIROS II Meteorological Satellite." *J. Geophys. Res.*, **65**, 3169–3185.
3. BANDEEN, W. R., R. E. SAMUELSON, and I. STRANGE, 1963: *TIROS III Radiation Data Users' Manual Supplement—Correction Models for Instrumental Response Degradation.* NASA Goddard Space Flight Center, Greenbelt, Maryland, 59 pp.
4. BANDEEN, W. R., V. KUNDE, W. NORDBERG, and H. P. THOMPSON, 1964: "TIROS III Meteorological Satellite Radiation Observations of a Tropical Hurricane". TELLUS (to be published).

5. BANNON, J. K. and L. P. STEELE, 1960: "Average Water-Vapor Content of the Air." *Geophys. Mem. Roy. Meteor. Office*, No. 102, 38 pp.

6. DANJON, A., 1936: "Nouvelles Recherches sur la Photometrie de la Lumière Cendrée et l'Albedo de la Terre." *Ann. l'Obs. Strasbourg*, **3**, 139–181.

7. DANJON, A., 1954: "Albedo, Color, and Polarization of the Earth," Chapter 15 in *The Earth as a Planet* ed. by Gerard P. Kuiper, University of Chicago Press, 751 pp.

8. DZHASYBEKOVA, E. K., V. M. KAZACHEVSKII, and A. V. KHARITONOV, 1960: "Determination of the Earth's Albedo." *Astronomicheskii Zhurnal*, **37**, no. 1, 131–134. (English translation in *Soviet Astronomy*, Vol. 4, no. 1, pp. 125–128.)

9. ELSASSER, W. M., 1960: "Atmospheric Radiation Tables." *Meteorological Monographs*, Vol. 4, No. 23.

10. FRITZ, S., 1949: "The Albedo of the Planet Earth and of Clouds." *J. Meteor.* **6**, 277–282.

11. FRITZ, SIGMUND and JAY S. WINSTON, 1962: "Synoptic Use of Radiation Measurements from Satellite TIROS II." *Mo. Weather Review*, **90**, 1–9.

12. HANEL, R. A., and D. Q. WARK, 1961: "TIROS II Radiation Experiment and its Physical Significance." *J. Opt. Soc. Amer.*, **51**, 1394–1399.

13. KONDRATIEV, K. YA., 1963: *Meteorologicheskiye Sputniki*. Hydrometeorological Publishing House, Leningrad, 311 pp.

14. LONDON, J., 1957: "A Study of the Atmospheric Heat Balance." Final Report Contract No. AF 19(122)-165, Dept. of Meteorology and Oceanography, New York University, 99 pp.

15. MÖLLER, F., 1961: "Atmospheric Water Vapor Measurements at 6–7 Microns from a Satellite." *Planet. Space Sci.*, **5**, 202–206.

16. MÖLLER, F., 1962: "Einige Vorläufige Auswertungen der Strahlungsmessungen von TIROS II." *Arch f. Met. Geophys. und Biokl.*, Ser. B., **12**, 79–94.

17. MÖLLER, F., and E. RASCHKE, 1963/64: "Evaluation of TIROS III Radiation Data." Interim Report No. 1 (1963) and Final Report (1964). NASA Research NsG-305, Meteorolog. Institut der Universität München.

18. NORDBERG, W., W. R. BANDEEN, B. J. CONRATH, V. KUNDE, and I. PERSANO, 1962: "Preliminary Results of Radiation Measurements from the TIROS III Meteorological Satellite." *J. Atmos. Sci.*, **19**, 20–30.

19. NORDBERG, W., W. R. BANDEEN, G. WARNECKE, and V. KUNDE, 1964: "Stratospheric Temperature Patterns Based on Radiometric Measurements from the TIROS VII Satellite." (Paper presented at the Fifth International Space Science Symposium, COSPAR, Florence, May 1964.) (To be published in *Space Research V*, North-Holland Publishing Co.)

20. SIMPSON, G. C., 1929: "The Distribution of Terrestrial Radiation." *Mem. Royal Meteor. Soc.* **3**, No. 23, 53–78.

21. Staff members, Aeronomy and Meteorology Division, GSFC, and Meteorological Satellite Laboratory, USWB, 1961: *"TIROS II Radiation Data Users' Manual.* NASA Goddard Space Flight Center, Greenbelt, Maryland, 57 pp.

22. Staff members, Aeronomy and Meteorology Division, GSFC, 1962a: *TIROS III Radiation Data Users' Manual*, NASA Goddard Space Flight Center, Greenbelt, Maryland, 71 pp.

23. Staff members, Aeronomy and Meteorology Division, GSFC, and Meteorological Satellite Laboratory, USWB, 1962b: *TIROS III Radiation Data Catalog.* NASA Goddard Space Flight Center, Greenbelt, Maryland, 388 pp.

24. Staff members, Aeronomy and Meteorology Division, 1963: *TIROS IV Radiation Data Catalog and Users' Manual.* NASA Goddard Space Flight Center, Greenbelt, Maryland, 250 pp.

25. Staff members, Aeronomy and Meteorology Division, 1964: *TIROS VII Radiation Data Catalog and Users' Manual, Volume 1.* NASA Goddard Space Flight Center, Greenbelt, Maryland (to be published).

26. WARK, D. Q., G. YAMAMOTO and J. H. LIENESCH, 1962: "Methods of Estimating Infrared Flux and Surface Temperature from Meteorological Satellites." *J. Atmos. Sci.*, **19**, 369–384. (Also, "Infrared flux and surface temperature determinations from TIROS radiometer measurements." Meteorological Satellite Laboratory Report No. 10(1962) and Supplement thereto (1963), U.S. Weather Bureau, Washington, D.C.)

Lee - N65-27455

TIME CORRELATION OF EXTREME ULTRAVIOLET RADIATION AND THERMOSPHERIC TEMPERATURE

R. E. BOURDEAU, S. CHANDRA† AND W. M. NEUPERT

Thermospheric temperatures deduced from satellite drag observations are compared with the intensity of extreme ultraviolet radiation measured on the first Orbiting Solar Observatory (OSO–1). The comparison leads to a conclusion that, for the two complete solar rotation periods during which the EUV data are available, the 27-day periodic variation in upper atmospheric density was due to corresponding changes principally of ultraviolet rather than corpuscular radiation.

INTRODUCTION

It generally has been accepted that the absorption of solar extreme ultraviolet (EUV) radiation constitutes an important mechanism for atmospheric heating. The possibility of an additional heat source of importance comparable to the EUV effect has been introduced by many investigators to explain semi-annual, diurnal and 27-day variations in atmospheric densities deduced from satellite drag observations. However, the need for a heat source as important as EUV radiation in order to explain temporal density changes does not appear to be conclusive.

Paetzold and Zschörner (1960) first suggested that semi-annual variations in atmospheric density could be associated with corresponding changes in geomagnetic activity and thus with possible slow changes in a second heat source associated with the solar wind. However, they depended on correlating satellite deceleration data for the period 1958–1961 with geomagnetic indices for the corresponding months averaged over the previous solar cycle. According to Bartels (1963), a positive correlation is not obtainable if the satellite deceleration data are compared with simultaneously-observed geomagnetic indices. This sheds doubt on the need for a solar wind effect to explain semi-annual density variations.

Harris and Priester (1962), by solving the heat balance equation have proposed that the EUV effect should result in a diurnal maximum in atmospheric density at 17^h local time. The observed maximum deduced from satellite drag studies occurs at about 14^h local time (Jacchia, 1963). To account for the difference between the observed and their theoretically-computed time for the diurnal maximum, Harris and Priester empirically introduce a second heat source equivalent to the EUV effect and which they identify with the solar wind. However, an alternative explanation for the phase shift is that the heat source is entirely EUV but a transport mechanism exists which shifts the absorbed energy toward the west. (Priester, private communication).

MacDonald (1963) has suggested that there is a systematic 56-hour phase delay between the 27-day cycles in observed decimetric solar flux and satellite-drag inferred densities, a delay corresponding to a straight line velocity from the sun of 750 km sec^{-1}. Since 750 km sec^{-1} is characteristic of solar wind velocities, MacDonald concludes that the 27-day period in atmospheric density is entirely due to corpuscular rather than EUV radiation.

In this paper daily averages of 170–400Å radiation measured by use of the first Orbiting Solar Observatory during March–May 1962 are time-correlated with satellite drag observations and with geomagnetic activity. The comparison leads to a conclusion that for the two complete solar rotation periods during which the EUV data are available, the 27-day periodic and the day-to-day

†Goddard Space Flight Center. Mr. Chandra is a NAS-NASA Postdoctoral Resident Research Associate.

variations in upper atmospheric density are principally due to EUV radiation, a conclusion which is not in harmony with MacDonald's.

MEASURED ULTRAVIOLET SPECTRUM

The solar radiation data used in our analyses were obtained by the use of a grazing incidence spectrometer pointed at the center of the solar disk within several minutes of arc. In this orientation the spectrometer analyzed radiation from the entire solar disk and inner corona, covering the nominal spectral range from 10Å to 400Å with a resolution of 0.85Å once every 8.5 minutes. Variations in several of the more intense lines in the observed spectrum have already been associated with solar flares (Behring, et al., 1963) and the slowly varying component of solar activity (Neupert, et al., 1964). In this paper, we present time variations of the EUV flux representing the sum of the intensities of the twenty-two most prominent lines in the range 170–400Å. These lines are identified by asterisks in the two spectral scans shown in Figure 1.

Since an absolute calibration of the OSO–1 spectrometer has not yet been completed, an indirect method of calibration has been used to derive flux from measured counting rates. This method consists of comparing the counting rates obtained from the OSO–1 instrument with fluxes

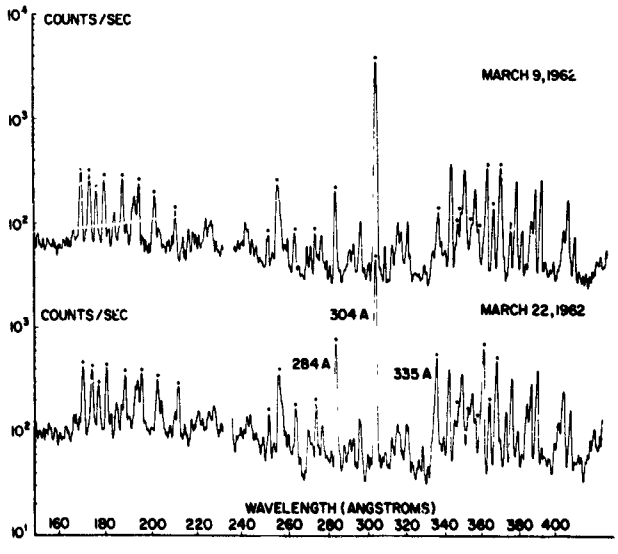


FIGURE 1.—Spectra obtained by OSO–1, grating spectrometer for two days of different solar activity. The lines marked by asterisk were used in the present analysis. Unmarked lines above 340A represent second order images of lines observed at shorter wavelengths.

measured by Hall et al., (1963) who used a calibrated rocketborne spectrometer. The comparison could only be made in the spectral region for which the two instruments overlapped, 250Å to 400Å. Furthermore, the comparison is meaningful only if the solar radiation on the day of the rocketborne measurement was indeed the same as on the days of OSO–1 observations chosen for comparison. The mean daily solar radio flux at 2800 Mc, measured by the National Research Council, Ottawa, Canada, was used as a measure of solar activity to obtain equivalent days of OSO–1 observations. Calibration of the instrument was extrapolated below 250Å using the measured reflectivity of the grating and an estimate of the detector sensitivity.

Each of the spectral lines used in the intensity estimate was observed 75 to 100 times during each day of observation, thereby yielding an accuracy of about one percent in the determination of the counting rate for each line, assuming a constant line intensity throughout the averaging period. The estimated error of the sum of intensities of the twenty-two brightest lines is therefore significantly smaller than the observed day-to-day variations. It has been found that the details of the time variations are relatively insensitive to likely changes in the assumed calibration of the instrument and also to the inclusion of numerous fainter lines in the sum of intensities. The daily averages of the sum of the intensities of the twenty-two discrete lines are illustrated in Figure 2a. Because of the uncertainties in calibration, the absolute value of the overall curve should not be taken as reliable but the time variations should be valid.

PHASE COMPARISON BETWEEN ATMOSPHERIC TEMPERATURE AND MEASURED ULTRAVIOLET RADIATION

Altitude variations in EUV spectra obtained by use of rocket-borne nomochromators (Hinteregger and Watanabe, 1962) establish that the portion of the spectrum which contributes most heavily to thermospheric heating lies between 1026 and 170Å. A summary of these and other EUV measurements (Watanabe and Hinteregger, 1962) indicates that the twenty-two discrete lines identified in Figure 1 represent approximately one-third of the pertinent ultraviolet radiation. Thus

conclusions derived from the EUV data presented here will depend on the reasonable assumption that the entire spectrum between 1026–170Å exhibits the same temporal variation as the sum of the intensities of these twenty-two discrete lines. Conclusions drawn from the EUV data additionally will depend on the assumption that the contribution of the twenty-two individual lines to thermospheric heating is proportional to their respective intensities. One weakness in such an assumption arises from uncertainties in our knowledge of the absorption and photoionization cross sections of the atmospheric constituents as a function of wavelength. However, a review (Watanabe and Hinteregger, 1962) does suggest that the cross sections of the important atmospheric constituents (O, O_2 and N_2) individually are reasonably independent of wavelength for the spectral region considered here.

Plotted in Figure 2b are the daily values of the temperature (T) representative of diurnal maximum conditions in the isothermal altitude region. These values of T were computed by Jacchia and Slowey (1963) from their observations of the deceleration of the Explorer 9 satellite and a theoretical model (Nicolet, 1961) of atmospheric density profiles as a function of temperature. It should be emphasized that, although the absolute value of T and the EUV intensity may be somewhat uncertain, we are interested here only in the validity of the time variation.

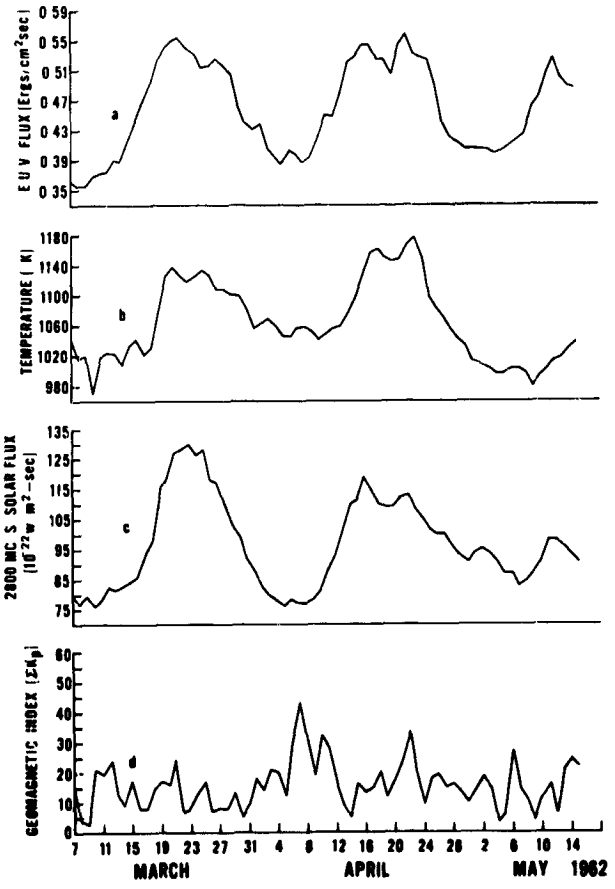


FIGURE 2.—Comparison of the EUV flux with the atmospheric temperature, the 2800 Mc/s solar flux, and the geomagnetic index ΣK_p.

By comparing Figures 2a and 2b, we conclude both from visual inspection and from a least-square fit performed by a computer analysis that the phase difference between T and the measured EUV radiation for the March and April solar rotations is of the order of one day. A one day delay can be explained either by the uncertainty of the satellite deceleration data or by the relaxation time of the atmosphere. Thus, it can be considered that there is a positive correlation between T and the EUV flux, leading to the conclusion that the day-to-day and the 27-day periodic variation of T are due principally to the EUV effect, at least for the months of March and April, 1962. It should be noted that despite the uncertainty in the individual T points, the short term variations in the atmospheric temperature correlate quite well with corresponding EUV radiation changes. Note, for example, the existence of double peaks in both the T and the EUV data during the maxima of the March and April solar rotations.

The 10.7 cm decimetric radiation has been plotted in Figure 2c. Comparison with Figure 2a shows that decimetric flux is an approximate but not precise index of the EUV effect. MacDonald (1963) has suggested that there is a systematic 56-hour delay between the emission of solar decimetric radiation and the response of the atmosphere. Since this corresponds to typical velocities of the solar wind, this delay led him to conclude that the 27-day density fluctuations are entirely due to interaction of the upper atmosphere with the solar wind. It should be noted that Jacchia (1963) reports that there was no appreciable lag between 10.7 cm solar flux and atmospheric density for the same period analyzed by MacDonald.

It also is possible to conclude that the two 27-day cycles in T considered here are not significantly associated with the solar wind from

simultaneously-measured indices of geomagnetic activity. Snyder et al., (1963) have measured the velocity of the solar wind and have quantitatively related this parameter to the geomagnetic index, Σk_p. Consequently, we have plotted the Σk_p index for each day of the pertinent period in Figure 2d. Using the relationship between solar wind velocity and Σk_p established by Snyder and et al., we estimate that if the solar wind is entirely responsible for the 27-day variation of the upper atmosphere temperature there should be an average phase delay of 4 days between the EUV flux and T, a delay much in excess of the average value derived from Figures 2a and 2b. Thus, we conclude that the solar wind was not significantly effective in heating the upper atmosphere during March and April, 1962. This does not contradict the findings of Jacchia (1963) who has associated significant transient increases in atmospheric density with geomagnetic activity generally much more intense than what was characteristic of the time interval studied here.

There is a suggestion of a significant enough phase delay between T and the EUV radiation during the May rotation to support the need for another heat source. However, it would be dangerous to draw any firm conclusion from this cycle without data throughout the complete solar rotation.

COMPARISON BETWEEN DIURNAL AMPLITUDE OF ATMOSPHERIC TEMPERATURE AND MEASURED ULTRAVIOLET RADIATION

It is important to compare the respective amplitudes as well as the phases of the 27-day temperature and EUV variations. In Figure 2, even though the amplitudes of the two complete EUV cycles are approximately the same, the amplitudes of the corresponding T cycles differ significantly from each other. Due to the motion of the perigee of the satellite under study with respect to the sun-earth line, the diurnal temperature variation shows up in the data of Jacchia and Slowey in addition to the day-to-day variation. For our purposes, we should have temperatures reduced to the same local time. For this reason we will now compare the differences between diurnal temperature maximum (T) and minimum (T_0) with heat source variations. It will be shown that the amplitudes of the 27-day variation of ($T-T_0$) and of the EUV cycles present a consistent picture. Jacchia (1963) has suggested the following empirical formula which relates day and night temperatures deduced from drag observations:

$$T = T_0\left(1+0.33\cos^n\frac{\Psi}{2}\right), \qquad (1)$$

where T_0 corresponds to the nighttime temperature minimum, Ψ is the angular distance between satellite perigee and a point 30 degrees east in longitude of the subsolar point and where $n=4$.

In Figure 3, we have set $n=1, 2, 4$, and compare the three resultant $T-T_0$ curves with the EUV data, using observed values of Ψ applicable to the Explorer 9 satellite. It is seen that by the use of $n=4$, one tends to overcorrect for the change in Ψ

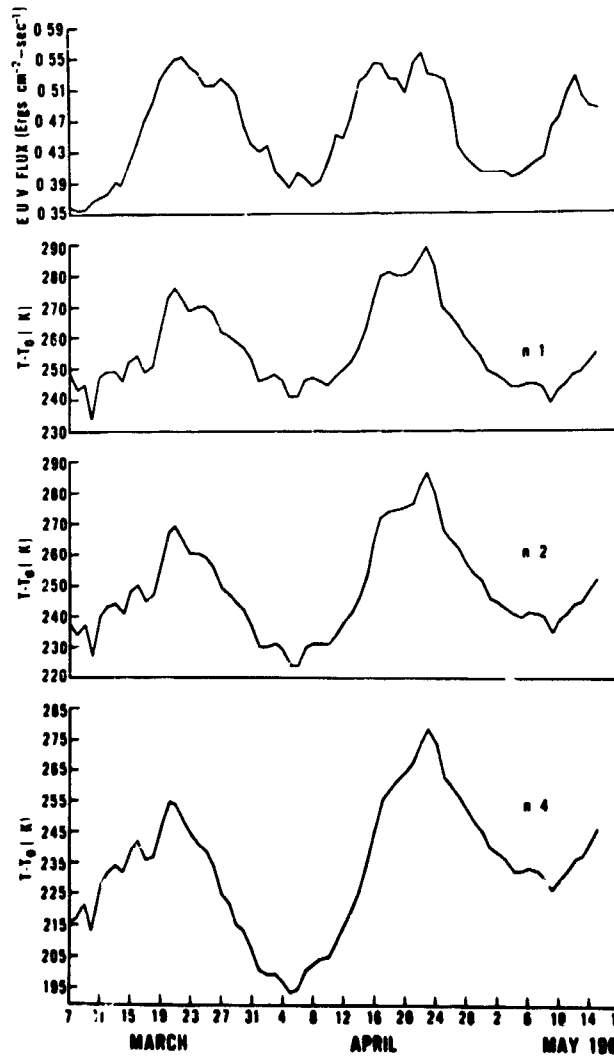


FIGURE 3.—Comparison of the EUV flux with the diurnal amplitude as computed from Equation 1 for $n=1$, 2 and 4.

with time in that now the amplitude of the second temperature cycle is larger than the first. An appropriate value for n seems to be between 1 and 2, as demonstrated by the very favorable comparison between the relative amplitudes of the two EUV cycles with the relative amplitudes of the two temperature cycles for $n=1$ and 2. Jacchia (private communication) independently concludes that a much lower value of n than was previously used is more consistent with satellite drag observations.

CONCLUSIONS

Daily values of atmospheric temperature have been compared with EUV radiation data and also with indices of geomagnetic activity. The analysis which includes both a phase and amplitude study favors the EUV rather than the solar wind effect as principally responsible for the 27-day variations in atmospheric temperature at least for the months of March and April 1962. The question as to whether this conclusion applies at all times and whether a second source of heating is required to explain semi-annual and diurnal temperature variations remains open.

ACKNOWLEDGEMENTS

The discussions with Dr. L. G. Jacchia and Dr. W. Priester during the preparation of this manuscript are gratefully acknowledged.

REFERENCES

BARTELS, J., Discussion on Time Variation of Geomagnetic Activity. Indices K_p and A_p, 1932–61. Annals de Geophysique **19**, 1–20, 1963.

BEHRING, W. E., W. M. NEUPERT and J. C. LINDSAY, Preliminary Solar Flare Observations with a Soft X-Ray Spectrometer on the Orbiting Solar Observatory. Space Research III, ed. W. PRIESTER (Amsterdam, North-Holland Publishing Co., 1963).

HALL, L. A., K. R. DAMON and H. E. HINTEREGGER, Solar Extreme Ultraviolet Photon Flux Measurement. Space Research III, ed. W. PRIESTER (Amsterdam, North-Holland Publishing Co., 1963).

HARRIS, I. and W. PRIESTER, Time-Dependent Structure of the Upper Atmosphere. J. Atmospheric Sci., **19**, 286–301, 1962.

HINTEREGGER, H. E. and K. WATANABE, Photoionization Rates in the E and F Regions. J. Geophys. Res., **67**, 3373–3392, 1962.

JACCHIA, L. G., Influence of Solar Activity on the Earth's Upper Atmosphere. Review prepared for the Inter-Union Commission on Solar and Terrestrial Relationships. Presented at the Tokyo, Japan meetings in September 1963.

JACCHIA, L. G. and J. SLOWEY, An Analysis of the Atmospheric Drag of the Explorer IX Satellite from Precisely Reduced Photographic Observations. Smithsonian Astrophys. Obs., special report No. 125, 1963.

MACDONALD, G. J. F., The Escape of Helium from Earth's Atmosphere. Review of Geophysics **1**, 305–349, 1963.

NEUPERT, W. M., W. E. BEHRING and J. C. LINDSAY, The Solar Spectrum from 50Å to 400Å. Space Research IV, ed. P. MULLER (Amsterdam, North-Holland Publishing Co., 1964).

PAETZOLD, H. K. and H. ZSCHÖRNER, The Structure of the Upper Atmosphere and its Variations after Satellite Observations. Space Research II, ed. H. C. VAN DE HULST, C. DE JAGER and A. F. MOORE (Amsterdam, North-Holland Publishing Co., 1961).

SNYDER, C. W., M. NEUGEBAUER and U. R. RAO, The Solar Wind Velocity and its Correlation with Cosmic Ray Variations and with Solar and Geomagnetic Activity. J. Geophys. Res. **68**, 6361–6370, 1963.

N65-18192

DETAILED BEHAVIOR OF THE MIDLATITUDE IONOSPHERE FROM THE EXPLORER XVII SATELLITE*

L. H. BRACE, N. W. SPENCER AND A. DALGARNO†

Measurements of electron temperature and ion density by electrostatic probes on the Explorer XVII satellite have revealed the detailed diurnal and latitudinal behavior of the summer ionosphere near the altitude of the F_2 maximum over the eastern United States at the time of solar minimum. The electron temperature at 40° North magnetic latitude is observed to rise rapidly from a nighttime value of about 1100°K to a mid-morning maximum of 2700°K followed by an afternoon plateau of 2000°K. The electron temperature always exceeds accepted values of neutral gas temperature and thus reflects the existence of heat sources in both the daytime and nocturnal ionosphere. The ion density displays a maximum value about three hours after local noon. A strong degree of latitude control, evident near the F_2 maximum, causes the temperature to increase and the density to decrease with increasing latitude. The electron temperature is relatively independent of altitude between 260 and 450 kilometers in the forenoon but displays a slight increase with altitude at night. The diurnal variations found over stations at 10° and 60° North magnetic latitude display general characteristics similar to those found at 40° N and also reveal the inverse gradients of temperature and density with latitude. Calculations based on the measurements show that the ion temperature begins to exceed the neutral temperature significantly above 300 kilometers and approaches the electron temperature near 600 km. The ratio of electron to ion temperature at 300 km is between 2 and 3 in the daytime and about 1.5, but highly variable, at night. The measurements are also employed to calculate the amount of electron heating in the F region. It appears that heating by solar ultraviolet radiation is adequate to explain the daytime results below 300 kilometers, and that the available flux of ionizing photoelectrons is adequate to account for the heating at higher altitudes. The heating at night is consistent with a corpuscular flux of soft electrons with a total energy of about 1×10^{-2} ergs cm^{-2} sec^{-1}.

INTRODUCTION

Between 1959 and 1961 a series of measurements of electron temperature using rocket-borne electrostatic probes at Ft. Churchill, Manitoba and at Wallops Island, Virginia, established that there are marked differences between the electron temperature and the heavy particle temperature in both daytime and nighttime ionospheres (Boggess, Brace and Spencer, 1959),[1] (Spencer Brace and Carignan, 1962),[2] (Brace, Spencer and Carignan, 1963).[3] The data show also that the magnitude of the temperature difference increases with latitude, a conclusion in harmony with later observations from the Ariel satellite (Willmore, Henderson, Bowen and Boyd, 1964).[4]

It was clear that further measurements were desirable to determine more precisely the extent of non-equilibrium for a wide variety of ionospheric conditions, especially since the degree of thermal non-equilibrium is a sensitive indicator of atmospheric energy sources. Theoretical studies (Hanson and Johnson, 1961),[5] (Dalgarno, McElroy and Moffett, 1963),[6] (Hanson, 1963),[7] (Dalgarno, 1963)[8] had demonstrated the close relationship of the charged particle temperature and densities to the composition and structure of the neutral particle atmosphere, and thus had emphasized the need for simultaneous measurements of ionospheric and neutral particle parameters. Accordingly, cylindrical electrostatic probe experiments, similar to that employed in the rocket flights, were included with the neutral partical mass spectrometer and total density gauge experiments in the Explorer XVII satellite, (Reber, Nicolet, 1965)[9] (Newton, Horowitz, Priester, 1965).[10] A brief description of some of the satellite probe measurements has been given

*Published as *Goddard Space Flight Center Document X-651-64-368*, November 1964.

†The Queen's University of Belfast, North Ireland.

(Brace and Spencer, 1964).[11] In this paper, we shall report the probe results in more detail and discuss their significance in the context of the heat balance in the ionosphere.

THE ELECTROSTATIC PROBE EXPERIMENTS

The cylindrical probes employed for the measurements of electron temperature T_e and ion density N_i were dimensioned to operate in an orbital-motion-limited mode (Mott-Smith and Langmuir, 1926).[12] Each collector was 23 cm in length and 0.056 cm in diameter. A guard electrode, 10 cm in length, was employed to prevent the collection of charged particles within the ion sheath surrounding the satellite. Both the collector and guard were made of stainless steel. The location of the sensors is discussed in an accompanying paper (Spencer, 1965).[13]

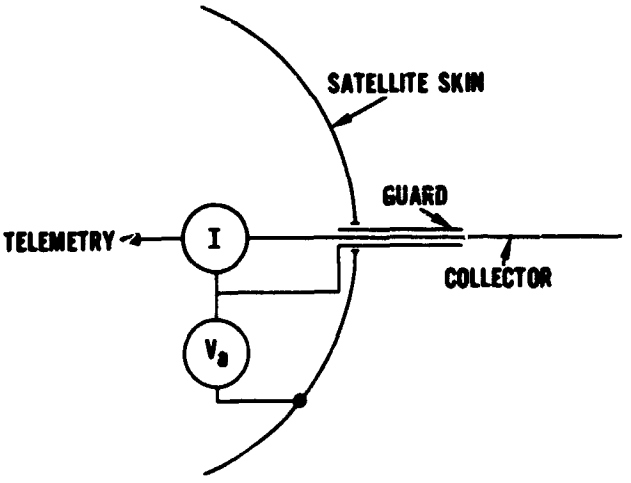


FIGURE 1.—Functional diagram of either the electron temperature (T_e) or the positive ion density (N_i) probe experiment.

Figure 1 is a functional block diagram which applies for either the T_e or the N_i experiment. The electronic circuitry associated with each sensor provided suitable sawtooth sweep voltages and current measuring sensitivities to optimize one experiment for the determination of N_i and the other for the measurement of T_e. The experiments operated simultaneously throughout each 4-minute satellite pass and provided simultaneous point by point measurements of both T_e and N_i along the satellite path.

The Ion Density Measurement

The voltage applied between the N_i probe and the satellite was swept both negative and positive (−3 v to +2 v) so that both the ion and electron saturation regions of the volt-ampere characteristics were observed. Since the satellite potential was normally about −0.5 v, the probe potential V relative to the plasma was −3.5 to +1.5 v. The ion saturation region of a cylindrical collector (when employed on a satellite) is particularly useful for determining the local charged particle density because the flux of ions intercepted by the probe is determined largely by the satellite velocity and the collector orientation, which are known factors. Because of the high satellite velocity the ion current I_i is relatively insensitive to the thermal velocity of the ions and the accelerating potential of the collector, factors which are less well known.

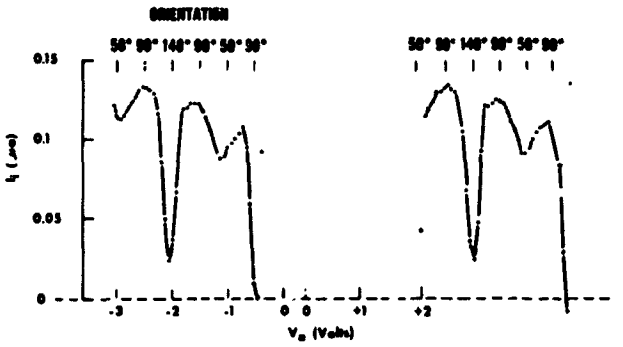


FIGURE 2.—Raw telemetry data (points) for two consecutive volt-ampere characteristics of N_i probe. The ion saturation regions, modulated by the changing probe orientation, display maximum ion current when the probe is perpendicular to velocity vector (90°) and minimum current when the probe looks forward (50° in this case) and aft (140°).

Figure 2 shows a machine-plot of raw telemetry data points representing two consecutive volt-ampere characteristics from the N_i probe. At positive applied voltages the measured current is off-scale negative due to the predominance of electron current, and only the ion saturation regions are visible. The sweep period (2 seconds) was made several times the spin period (0.7 seconds) so that the effects of changes in orientation upon collection area and wake effects could be studied through the manner in which they modulate the

ion saturation current. Maxima in the ion current (I_{max}) occur twice per satellite spin, corresponding to those times when the probe axis is perpendicular to the velocity vector. At this orientation the collector presents its maximum projected area to the plasma stream. In addition, a major minimum of ion current is observed each time the probe is rotated into the wake region behind the satellite. A secondary minimum occurs 180° later when the sensor is ahead of the satellite and presents its minimum projected area to the stream.

The relatively small effect of the applied voltage is evident in the slight decrease of I_{max} between successive maxima. This decrease corresponds to a voltage change of about 1 v. Curves such as those shown in Figure 2 permit one to derive empirically the functional dependence of I_{max} upon V, a relationship which is needed for an accurate evaluation of N_i from the measured ion currents. Using the slope of I_{max} versus V derived from a representative selection of probe data, I_{max} was found to be about 20 percent ± 3 percent higher than would be observed at $V=0$, where no acceleration of the ions occurs. Neglecting a very small enhancement of the current due to the thermal motion of the ions (<4 percent), the current at $V=0$ is simply the charge per unit time swept out by the frontal area of the collector. Therefore the ambient positive ion density is given by:

$$N_i=\frac{0.8I_{max}}{\ell d e v_s}=V=0 \qquad (1)$$

where ℓ is the length and d is the diameter of the collector, e is the unit charge and v_s is the satellite velocity.

Because I_i is always measured at the same orientation, the relative accuracy of the N_i measurements depends primarily upon the current detection error, which does not exceed 5 percent for $N_i>5\times10^4$ cm^{-3}. The absolute accuracy of N_i is better than 10 percent when 0^+ ions predominate (below 600 km).

The Electron Temperature Measurement

Between the extremes of ion and electron saturation, the probe current characteristic exhibits an exponential shape whose curvature permits a direct interpretation of the curves in terms of the electron temperature (Mott-Smith and Langmuir, 1926).[12] The analysis of individual curves is carried out by the log-plot technique in which T_e is given by the slope of $\log_e I_e$ with respect to V, as given by (2)

$$\frac{d\log_e I_e}{dV}=-\frac{e}{kT_e} \qquad (2)$$

Based on the experience gained with identical probes on ejectable rocket-borne instruments (Spencer, Brace and Carignan, 1962),[2] the T_e experiment was designed to permit high resolution of the exponential portion of the volt-ampere characteristics, with primary emphasis on thermal electrons in the $0-1$ eV range. To permit ample voltage resolution through a wide range of T_e, two saw-tooth sweeps were employed (0 to +0.75 v and 0 to +1.5 v). The telemetry sampling rate assigned to the experiment (180 samples/sec) was sufficient to permit a sweep rate of 10 curves per second. The necessary current resolution was provided by two linear current detectors of 1 μa and 5 μa full scale sensitivity which time-shared the assigned telemetry channels on a 10-second-on, 10-second-off basis. Each detector recorded 100 consecutive volt-ampere curves during its on-period.

Figure 3 is a computer plot of 1 μa and 5 μa raw telemetry data at the beginning of two consecutive 10 second on-times recorded during a nighttime Blossom Point pass. The temperature and density at this time were moderately low ($T_e\simeq1200°$ K, $N_i\simeq1\times10^5$ cm^{-3}). The inflight current calibrations at the left were obtained by substituting appropriate known resistances for the probe sensor at the beginning of each 10 second series of curves. The 5 μa data taken with a 1.5 v sweep permit a "coarse" measurement of the characteristics. A "high resolution" measurement is provided by the 1 μa detector with a 0.75 volt saw-tooth applied. This combination of sensitivities and voltages permitted high resolution measurements of T_e between 500° and 5000° K, when N_i exceeded 3×10^4 cm^{-3}. The satellite spin period is apparent in Figure 3 in the modulation envelope of the electron current curves (not particularly marked in this example). This modulation stems from both the reduction of

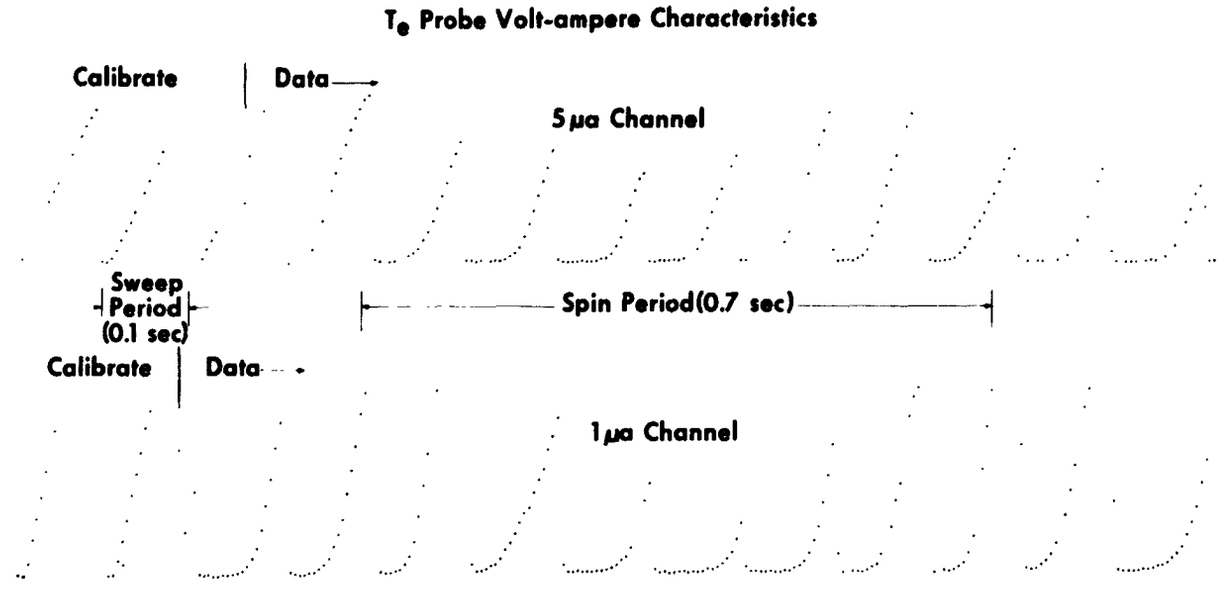


FIGURE 3.—Raw telemetry data showing a series of 5μa (0 to 1.5 v applied) and 1μa (1 to 0.75 v) current characteristics, each preceded by an inflight current calibration. The 1μa channel provides the high resolution necessary to permit accurate determination of the low electron temperature in nighttime passes such as this one.

charged particle density in the wake region immediately behind the satellite and the modulation of probe voltage by the induced $\vec{B}\times\vec{v}_s$ voltage caused by motion of the satellite through the Earth's magnetic field, $\vec{B}$. Measurement error due to these easily recognized effects has been avoided by selecting for analysis only those non-wake curves which correspond to the minimum rate of change of induced voltage. Following this criterion for curve selection, the relative accuracy of the T_e measurements within a given satellite pass is better than 5 percent. The absolute accuracy depends upon the maintenance of the linearity and amplitude of the saw-tooth sweep voltage throughout the life of the satellite. Frequent checking of inflight calibrations has shown no detectable change and the absolute accuracy is believed to be better than 5 percent on most passes.

Figure 4a and 4b help to demonstrate the computer methods employed in the T_e analysis. Figure 4a is a computer printout arranged to display a single volt-ampere characteristic as it appeared on the 1 μa channel, and 4b is a log-plot of the same characteristic. Because of a limited plotting interval imposed by the printer, the plotting accuracy shown is only 1/10th the accuracy of the actual data. The computer is programmed to make a best straight-line fit to the linear part of the log plot and then print out the corresponding T_e according to equation (2). Plots of this type are made by the computer for each curve of a group of 15 consecutive curves. Several groups are chosen from each satellite pass to permit hand verification of the computer values of T_e. All temperature data presented in this paper have been verified in this manner.

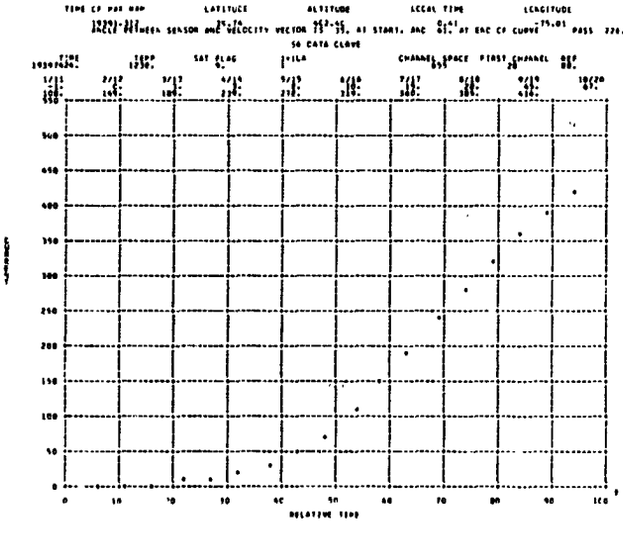


FIGURE 4a.—Computer plot of single current characteristic demonstrating the computer analysis procedure. Telemetered points (current) and plotted versus time (voltage, 0→0.75 v) to permit visual identification of exponential portion of curve (0 to 60 milliseconds).

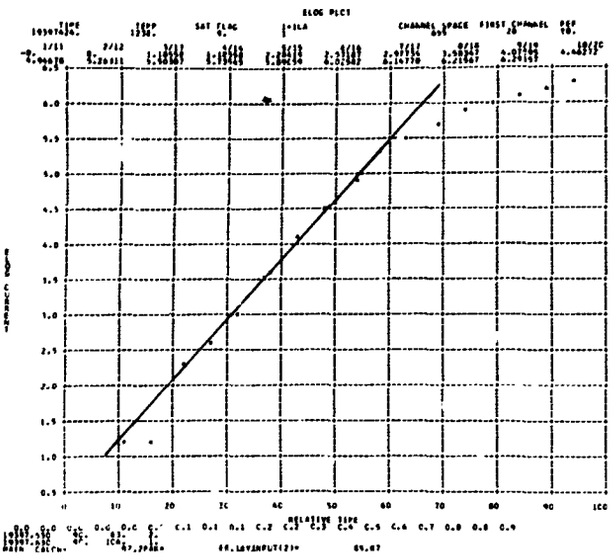


FIGURE 4b.—Computer plot of natural log of currents in Figure 4a showing the linear portion which is related to T_e. Line drawn by hand permits manual check of computer value of T_e printed automatically above.

RESULTS

Motion of the Satellite

The inclination of a satellite orbit simultaneously determines the precession rate of the orbit plane and the movement of perigee, so that a given satellite may sample the atmosphere at only unique combinations of local times and altitudes over each telemetry station. Thus, the orbital inclination determines the aeronomic coverage which a particular satellite can provide. For example, in its 100 day lifetime, Explorer XVII's inclination of 58°, combined with apogee and perigee altitudes of 927 and 258 km, provided the coverage over Blossom Point shown in Figure 5. Each pair of points connected by a line represents the beginning and end of a particular satellite pass. Although the pattern of coverage was similar at other latitudes, perigee and apogee occurred at other local times.

Figure 5 aids in selecting those Blossom Point passes in which diurnal and altitude effects can be most effectively separated. For example, the southbound set of passes provides an excellent opportunity to study the diurnal variation of T_e and N_i at a fixed altitude (perigee) through the daylight hours.

In processing the data from each telemetry station we derived a single value of T_e and of N_i

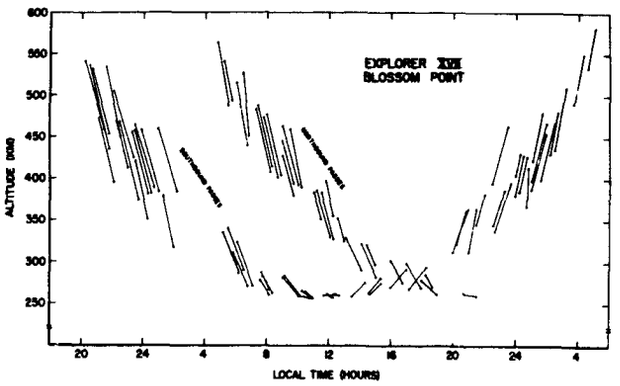


FIGURE 5.—Altitude and local solar time coverage over blossom point produced by the precession of the orbit plane and the movement of perigee during the 100-day active lifetime of Explorer XVII. The points joined by lines represent the positions of the satellite at the beginning and end of each pass.

from groups of volt-ampere curves taken at one-minute intervals through each satellite pass and plotted the beginning-of-pass and end-of-pass values of T_e and N_i against local solar time. Although the gross variations evident in the resulting local time plots for each station may reflect some altitude and seasonal variations in the atmosphere, the rapid precession of the orbit plane ensures that local time effects will predominate. Variations of T_e and N_i within a particular pass, however, are controlled by latitude or altitude changes in that pass because the local time change within a pass is small.

Blossom Point Data

Table 1 contains the values of T_e and N_i measured at the beginning and end of each Blossom Point pass, with the corresponding positions of the satellite and concurrent values of the solar index ($F_{10.7}$) and of the magnetic index (a_p).

To permit the local time, altitude, latitude and season of each measurement to be identified, data from southbound and northbound passes have been plotted separately in Figures 6 and 7, respectively. The points joined by lines are the beginning and end-of-pass values and the figures show the average altitude and the month of the pass.

Latitude and Local Time Effects

In order to show the effects of latitude, smoothed curves have been drawn in Figure 6 through the values measured at magnetic latitudes

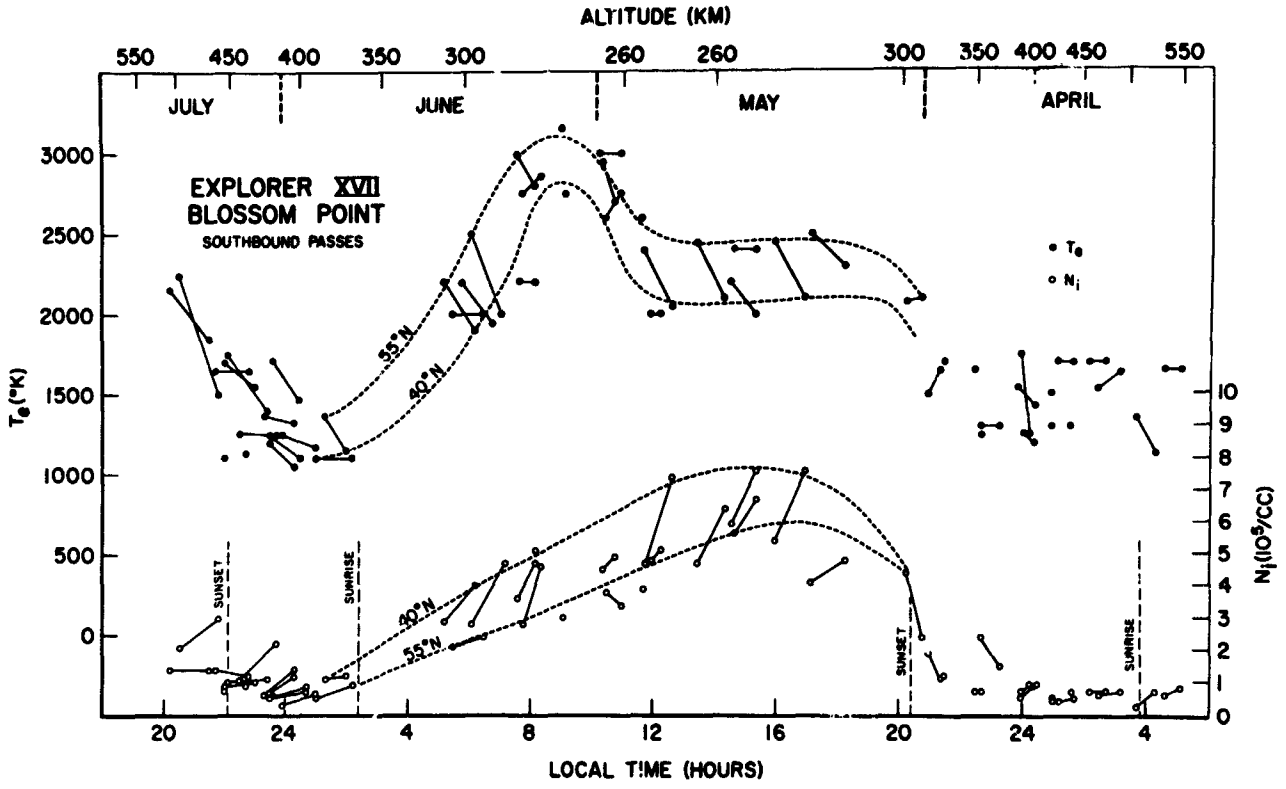


FIGURE 6.—T_e and N_i data from southbound Blossom Point passes identified with the altitude, month, and local times. The points joined by lines represent beginning-of-pass and end-of-pass values. The latitude gradient in T_e and N_i, evident in the daytime, made possible latitude identification of the data as well.

of 40° N and 55° N, the extremes of latitude observed from Blossom Point. The smoothed curves therefore represent an averaged diurnal variation of T_e and N_i near F_2 maximum at the indicated magnetic latitudes. The major characteristics of the diurnal variation of electron temperature are a nighttime minimum, a steep rise to a mid-morning maximum, and an afternoon plateau, followed by a decline to a lower, more variable, value at night. This behavior is very different from the diurnal variation of the neutral gas temperature (T_g), which exhibits a maximum in the afternoon rather than the morning. The values of T_e are clearly much in excess of T_g, even at night. The diurnal variation in N_i is characterized by a gradual morning rise, an afternoon maximum and a sharp decrease at sunset.

It is particularly significant that in all southbound daytime passes the satellite was near perigee and therefore was changing altitude very little. Thus the changes in T_e and N_i observed within each of these passes must reflect the latitude control which was first evident in the smoothed diurnal variation curves for 40° and 55° N magnetic latitude. The average temperature gradient is about 25° K/° latitude, at this latitude and altitude.

Seasonal Effects

A major seasonal effect which may be seen in Figure 6 is made evident by the longer time T_e remained low in the April night passes as compared to the June-July night passes when nights were shorter. The less rapid decay of T_e at sunset in July than at sunset in April may also be a seasonal effect, or it may simply reflect the greater influence of heat conduction in the electron gas at the higher altitude of the July sunset passes.

Altitude Effects

The night-time passes in April (Figure 6), where altitude is the most important variable, suggest that there is a slight positive gradient of T_e with altitude at night of the order of 1° K/km between 350 and 550 km.

As noted earlier, a comparison of northbound passes (Figure 7) and southbound passes (Figure

Tabulation of T_e and N_i from Blossom Point Passes (Explorer XVII)

PASS	DATE	GMT	GEOGRAPHIC LATITUDE (degrees)		GEOGRAPHIC LONGITUDE (degrees)		ALTITUDE (kilometers)		LOCAL TIME (hours)		T_e (degrees Kelvin)		N_i (part./cc)		$F_{10.7}$	a_p
		(hr:min)	begin	end	begin	end	begin	end	begin	end	begin	end	begin	end		
15	4/4/63	2:10	33	43	−82	−70	263	260	20.7	21.5	1100	1150	2.6+5	2.4+5	70	3
50	4/6/63	10:47	35	27	−93	−86	534	592	4.6	5.1	1650	1650	5.5+4	7.8+4	78	22
79	4/8/63	9:20	39	30	−84	−77	492	550	3.7	4.3	1350	1125	2.2+4	7.2+4	81	9
118	4/10/63	23:35	30	36	−80	−68	285	270	18.3	18.7	1600	-------	1.0+6	-------	82	2
123	4/11/63	7:59	41	32	−82	−73	446	510	2.5	3.2	1525	1625	6.4+4	6.8+4	88	3
138	4/12/63	8:40	41	35	−88	−82	440	481	2.2	2.7	1700	1790	6.5+4	6.7+4	93	7
148	4/12/63	23:47	35	46	−86	−73	279	262	18.0	19.0	1800	1800	7.7+5	6.7+5	73	7
152	4/13/63	6:32	46	36	−77	−64	400	468	1.6	2.5	1300	-------	6.9+4	-------	89	4
167	4/14/63	6:35	46	36	−84	−70	388	456	1.0	2.0	1500	-------	4.5+4	-------	87	4
182	4/15/63	6:43	43	33	−87	−74	398	467	1.0	2.0	1300	-------	3.9+4	-------	88	27
192	4/15/63	22:25	32	43	−85	−72	299	269	17.0	18.0	2600	-------	3.6+5	-------	88	5
197	4/16/63	6:49	38	30	−85	−78	427	479	1.2	1.7	1700	1700	4.4+4	5.4+4	88	4
211	4/17/63	5:15	46	39	−78	−68	368	414	0.1	0.8	1250	1200	6.3+4	9.2+4	87	7
226	4/18/63	5:21	43	36	−79	−71	385	427	0.2	0.7	1400	1250	7.7+4	7.6+4	88	3
236	4/18/63	21:04	34	44	−78	−68	303	275	16.0	16.8	1850	1700	7.5+5	8.2+5	88	27
241	4/19/63	5:25	42	34	−84	−75	381	429	23.9	0.5	1550	1425	5.4+4	9.0+4	84	18
256	4/20/63	5:34	37	33	−84	−80	405	431	24.0	0.3	1750	1250	6.7+4	9.0+4	78	6
270	4/21/63	3:55	47	37	−81	−66	338	396	22.7	23.7	1250	-------	6.7+4	-------	74	3
285	4/22/63	4:40	45	38	−82	−73	345	387	22.6	23.3	1300	1300	2.4+5	1.5+5	72	3
310	4/23/63	19:51	36	42	−81	−75	321	298	14.5	15.0	2300	2000	2.4+5	2.4+5	71	3
325	4/24/63	19:57	37	47	−86	−72	322	282	14.2	15.2	2050	-------	3.2+5	-------	73	4
330	4/25/63	4:21	32	22	−89	−79	396	464	22.5	23.5	1650	-------	6.7+4	-------	72	6
359	4/27/63	2:52	39	32	−82	−70	346	382	21.5	22.0	1700	-------	1.2+5	-------	75	6
384	4/28/63	18:39	40	49	−82	−69	329	290	13.2	14.2	2000	-------	3.0+5	-------	78	4
388	4/29/63	1:20	44	32	−75	−61	313	365	21.0	21.4	1500	1650	1.9+5	1.1+5	78	3
398	4/29/63	17:04	36	41	−68	−62	353	325	12.6	13.0	2250	-------	2.5+5	2.8+5	78	4
413	4/30/63	17:07	29	36	−83	−73	398	356	11.8	12.3	2800	3200	2.9+5	2.9+5	80	39
418	5/1/63	1:32	39	31	−80	−72	333	362	20.3	20.8	2080	2100	4.4+5	2.4+5	82	80
443	5/2/63	17:18	34	44	−88	−76	384	328	11.5	12.3	2500	-------	2.1+5	-------	82	32
458	5/3/63	17:22	35	45	−95	−81	386	330	11.2	12.1	2200	2400	2.4+5	2.2+5	81	18
472	5/4/63	15:49	36	42	−73	−67	384	352	11.0	11.5	2250	2500	2.1+5	2.6+5	82	18
506	5/6/63	22:41	49	39	−83	−67	268	294	17.2	18.3	2500	2300	4.1+5	4.8+5	87	6
516	5/7/63	14:23	27	39	−75	−64	460	390	9.5	10.2	2900	2750	2.2+5	2.2+5	88	6
531	5/8/63	14:27	28	39	−80	−69	464	393	9.0	10.0	3000	3000	-------	-------	86	7
546	5/9/63	15:32	32	40	−82	−73	446	395	9.0	9.7	3000	3000	1.7+5	1.8+5	88	9
561	5/10/63	14:37	36	44	−85	−70	428	380	9.0	9.7	2800	2800	1.9+5	1.8+5	87	9
590	5/12/63	13:06	31	42	−77	−65	477	405	8.0	8.9	2700	-------	1.2+5	-------	87	9
595	5/12/63	21:28	42	32	−83	−70	270	292	16.0	17.0	2450	2100	5.4+5	7.6+5	87	7
605	5/13/63	13:08	33	43	−81	−68	474	402	7.8	8.7	2400	2450	1.7+5	1.7+5	89	23

620	5/14/63	13:13	32	43	−88	−75	488	415	7.4	8.3	2600	2600	1.3+5	1.3+5	95	6
635	5/15/63	13:17	34	45	−91	−78	482	409	7.3	8.2	2600	2600	1.5+5	1.6+5	98	5
654	5/16/63	20:06	43	34	−82	−69	262	275	14.7	15.4	2400	2400	5.6+5	6.7+5	100	2
669	5/17/63	20:10	40	30	−84	−72	263	281	14.6	15.4	2200	2000	5.9+5	7.6+5	100	12
679	5/18/63	11:49	36	41	−87	−75	527	452	6.4	6.8	2100	2100	1.2+5	1.2+5	98	3
694	5/19/63	11:53	33	44	−88	−79	515	440	6.0	6.7	2500	2450	1.2+5	1.3+5	99	3
698	5/19/63	18:98	44	36	−76	−66	259	265	13.5	14.4	2450	2100	4.7+5	6.4+5	99	6
723	5/21/63	10:21	33	39	−78	−71	542	494	5.2	5.7	2700	2700	5.7+4	6.6+4	88	2
738	5/22/63	10:24	31	41	−85	−74	562	488	4.8	5.5	2500	----	6.3+4	----	89	2
772	5/24/63	17:15	44	41	−80	−76	261	258	12.0	12.3	2000	2000	4.8+5	5.1+5	89	3
787	5/25/63	17:19	41	32	−82	−73	261	260	11.8	12.7	2400	2050	4.7+5	7.4+5	83	7
802	5/26/63	17:22	41	29	−84	−76	260	262	11.7	12.5	2600	----	3.9+5	----	76	4
846	5/29/63	15:52	43	37	−81	−74	265	258	10.5	11.0	2600	2750	3.8+5	3.4+5	80	4
861	5/30/63	15:55	42	33	−86	−76	266	257	10.3	11.0	3000	3000	----	----	83	5
876	5/31/63	15:58	37	30	−85	−79	259	257	10.4	10.8	2950	2700	4.5+5	4.9+5	89	6
890	6/1/63	14:22	47	36	−80	−65	282	259	9.1	10.1	3150	----	3.0+5	----	84	11
905	6/2/63	14:24	46	34	−86	−71	284	259	9.2	10.2	2750	----	----	3.2+5	81	15
964	6/6/63	12:57	43	36	−77	−68	288	268	7.7	8.2	2200	2200	----	5.1+5	77	5
979	6/7/63	12:59	41	32	−79	−70	282	263	7.8	8.4	2750	2850	2.8+5	4.6+5	84	22
994	6/8/63	13:02	38	29	−80	−74	278	261	7.6	8.2	3000	2800	3.5+5	4.7+5	90	6
1023	6/10/63	11:27	48	34	−78	−63	324	271	6.1	7.1	2500	2000	2.8+5	4.7+5	99	5
1038	6/11/63	11:27	50	40	−90	−74	341	290	5.5	6.5	2000	2000	2.2+5	2.4+5	103	4
1053	6/12/63	11:31	43	31	−84	−71	311	271	5.8	6.8	2200	1950	2.8+5	4.4+5	109	6
1068	6/13/63	11:31	47	36	−95	−81	336	287	5.2	6.2	2200	1900	2.9+5	4.0+5	107	6
1260	6/26/63	6:57	52	42	−100	−83	460	385	1.0	2.2	1100	1100	5.4+4	9.0+4	74	48
1275	6/27/63	7:02	40	28	−86	−75	379	318	1.3	2.0	1375	1150	1.1+5	1.2+5	72	18
1289	6/28/63	5:24	50	40	−83	−67	459	384	23.9	1.0	1250	1175	3.2+4	1.1+5	74	6
1304	6/29/63	5:25	50	40	−88	−72	465	389	23.5	0.7	1700	1425	4.6+4	6.9+4	73	4
1319	6/30/63	5:26	48	37	−89	−74	458	382	23.5	0.5	1250	1100	4.5+4	7.7+4	76	12
1334	7/1/63	5:28	47	36	−92	−78	456	381	23.3	0.3	1375	1325	5.5+4	1.4+5	76	7
1349	7/2/63	5:31	41	29	−89	−77	421	351	23.5	0.3	1200	1050	6.3+4	1.2+5	77	3
1363	7/3/63	3:52	50	41	−86	−69	505	425	22.1	23.4	1750	1400	9.5+4	1.1+5	78	2
1378	7/4/63	3:54	45	31	−81	−65	468	374	22.5	23.7	1250	1250	9.9+4	3.0+5	78	5
1393	7/5/63	3:55	47	37	−89	−75	492	413	22.0	23.0	1700	1550	8.6+4	9.9+4	78	32
1408	7/6/63	3:57	50	41	−101	−82	534	451	21.6	22.5	1125	1125	9.0+4	1.8+5	77	67
1423	7/7/63	3:58	43	31	−93	−80	473	396	21.1	22.1	1100	1100	8.1+4	2.2+5	77	9
1437	7/8/63	2:20	49	39	−85	−69	531	454	20.7	21.8	1650	1650	1.4+5	1.2+5	77	15
1452	7/9/63	2:21	49	34	−88	−70	538	435	20.5	21.8	2250	1500	2.1+5	3.0+5	77	27
1467	7/10/63	2:22	48	35	−92	−74	541	459	20.2	21.5	2150	1850	1.4+5	1.4+5	76	6

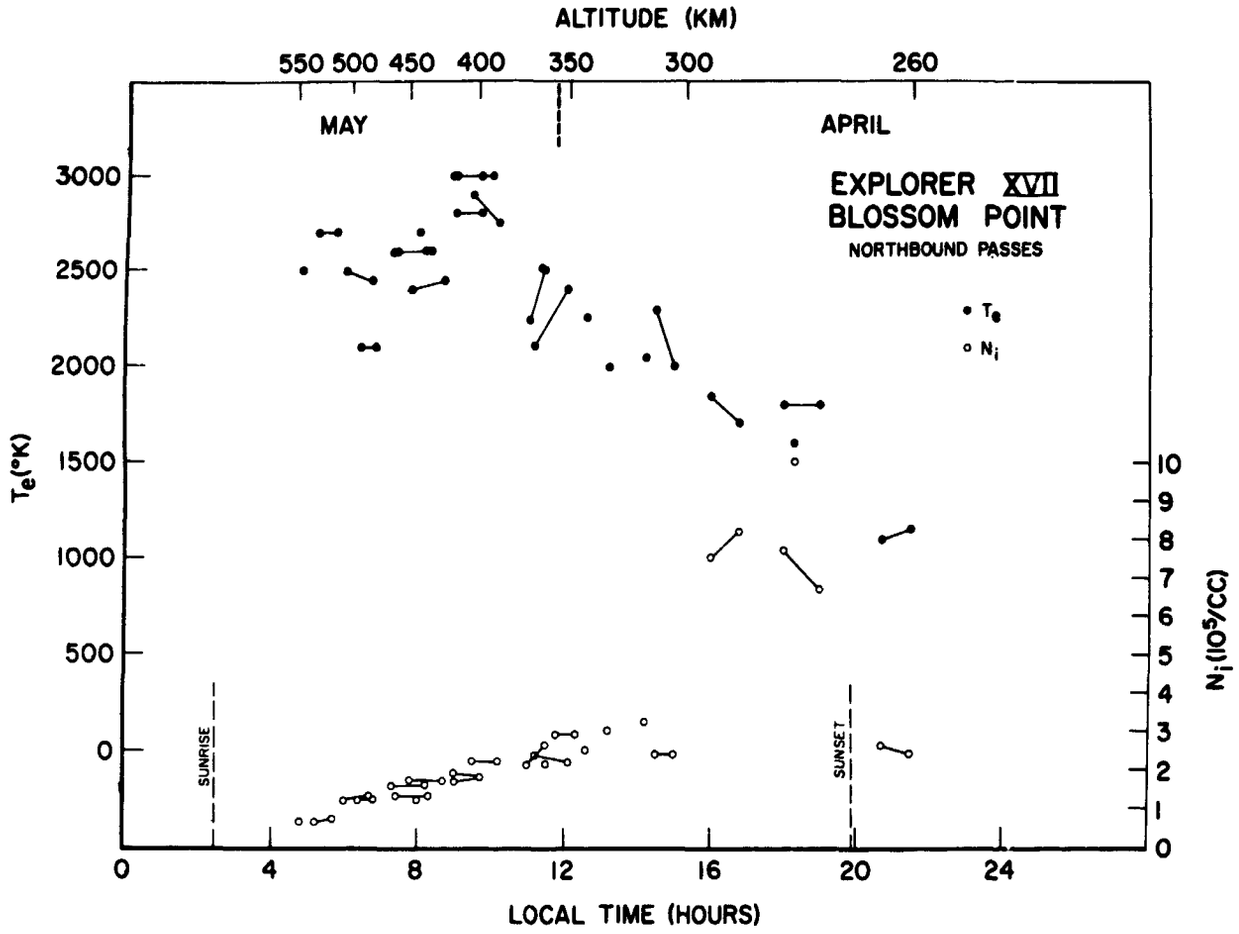


FIGURE 7.—T_e and N_i data from southbound Blossom Point passes.

6) permits the effect of altitude to be evaluated for the morning hours. Such a comparison reveals no systematic gradient of T_e with altitude between 0600 and 1200 hours.

Inverse Relationship Between T_e and N_i

In nearly all daytime passes where beginning and end values of T_e and N_i have been derived, T_e increased and N_i decreased with increasing latitude, the changes being especially marked in the southbound passes which were at lower altitudes. A few exceptions occurred in the early morning, higher altitude, passes when neither T_e nor N_i changed significantly within individual passes.

Data from Other Latitudes

Graphs of the measurements of T_e and N_i during passes at Quito, Equador (10° N magnetic) and at College, Alaska (60° N magnetic) have been prepared. The smoothed diurnal variations of T_e and N_i derived from all three stations are shown in Figure 8. These curves refer to the region between 260 and 400 km containing the F_2

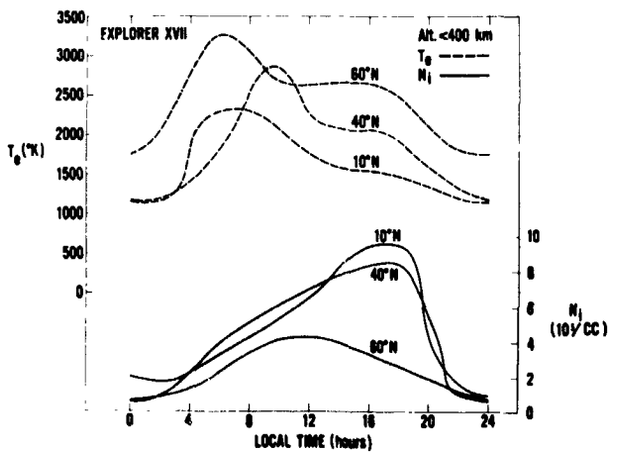


FIGURE 8.—Smoothed diurnal variations in T_e and N_i at three latitudes (10°, 40°, and 60° north magnetic) for altitudes below 400 kilometers.

maximum. Although the general behavior of the F-region at Blossom Point, Quito and College is similar, the times of the morning maxima of T_e and afternoon maxima of N_i and the magnitudes of the temperatures and densities differ considerably. Not all the differences can be ascribed to magnetic latitude alone, for the T_e maximum at College may occur earlier because of the earlier sunrise there in summer, which is a result of the higher geographical latitude. Indeed, the Blossom Point results at 40° and 55°N magnetic latitude suggest that magnetic latitude has little control over the time of occurrence of the maxima in T_e. Differences in magnetic longitude may also be of importance.

The higher nocturnal electron temperatures of the College passes also may be related to the higher geographic latitude (58° N) at which they were measured. The mid-summer F_2 region at 58° N is always illuminated, although the zenith angle of the sun exceeds 90° near midnight.

ION TEMPERATURE AND HEAT SOURCE

If we assume that the temperature of the electron gas is in equilibrium such that the rate at which it gains heat equals the rate at which it loses heat, the heat source Q necessary to maintain a temperature difference can be calculated from the cooling rates. Only elastic processes contribute significantly to cooling in collisions with the positive ions, and the rate at which the electron gas loses heat to a positive ion gas mixture of 0^+, He^+ and H^+ at a temperature T_i is given approximately by

$$Q_e=\frac{5\times10^{-7}(T_e-T_i)n_e}{T_e^{3/2}}\{n(0^+)+4n(He^+)+16n(H^+)\}eVcm^{-3}\,\mathrm{sec}^{-1} \quad (3)$$

(cf. Spitzer 1956),[14] (Hanson and Johnson, 1961).[5] Equation (3) demonstrates that simultaneous measurements of temperature and density are required for an accurate calculation of the energy source responsible for the temperature difference. The strong inter-dependence of temperature and density require that the measurements be essentially point measurements.

The Positive Ion Temperature

The heat gained by the positive-ion gas causes its temperature to rise above that of the neutral

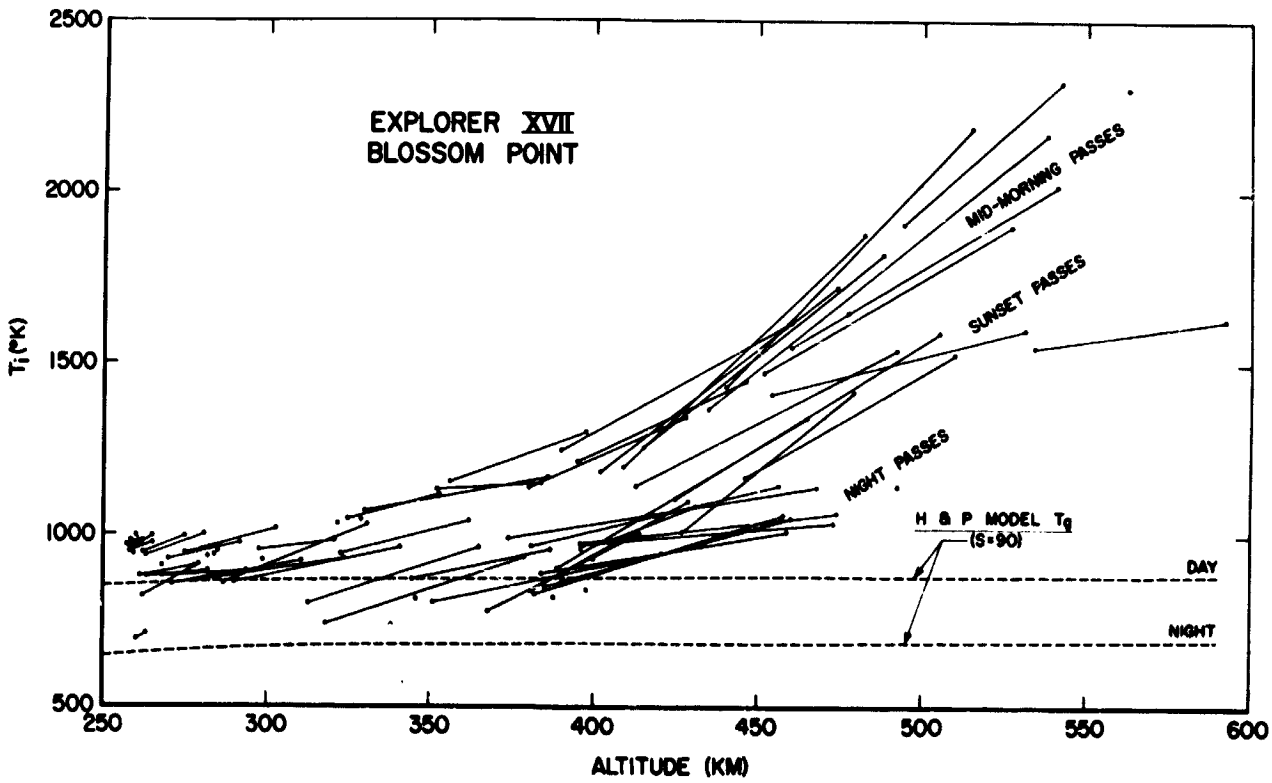


FIGURE 7.—Calculated ion temperature (T_i) for all Blossom Point passes. The Harris and Priester models, used with measured T_e and N_i for computation, are also shown.

particle gas (Hanson, 1963)[7] (Dalgarno, 1963).[8] The positive ions cool mainly by resonance charge transfer collisions with the parent atoms and by elastic collisions with unlike atoms and molecules. The rates of heat loss from the positive ions to the neutral gas at a temperature T_g may be derived from the collision frequencies (Dalgarno, 1958,[15] 1961,[16] 1964[17]) and there results

$$Q(0^+)=n(0^+)\ \{9\times10^{-14}n(0)+6\times10^{-14}n(N_2)+6\times10^{-15}n(He)+2\times10^{-15}n(H)\}(T_i-T_g)\text{eV cm}^{-3}\text{ sec}^{-1}, \quad (4)$$

$$Q(He^+)=n(He^+)\ \{6\times10^{-14}n(0)+4\times10^{-14}n(N_2)+2\times10^{-13}n(He)+1\times10^{-14}n(H)\}(T_i-T_g)\text{eV cm}^{-3}\text{ sec}^{-1} \quad (5)$$

and

$$Q(H^+)=n(H^+)\ \{3\times10^{-14}n(0)+3\times10^{-14}n(N_2)+5\times10^{-14}n(He)+4\times10^{-13}n(H)\}(T_i-T_g)\text{eV cm}^{-3}\text{ sec}^{-1}. \quad (6)$$

In equilibrium,

$$Q_e=Q(0^+)+Q(He^+)+Q(H^+), \quad (7)$$

an equation which determines T_i in terms of T_e, n_e and the neutral particle and positive-ion concentrations.

The altitudes of the Blossom Point measurements were usually such that 0^+ was the major ion and thus (7) can be simplified to yield

$$T_i-T_g=\frac{6\times10^6 n_e(T_e-T_i)}{n(0)T_e^{3/2}}. \quad (8)$$

We have substituted the satellite measurement of $N_i=n_e$ and of T_e into (S) together with the number densities of atomic oxygen given by the model of Harris and Priester (1962)[18] for the local time of the measurement. The resulting values of T_i and of T_e/T_i for Blossom Point are shown in Figures 9 and 10 as a function of altitude. In Figure 9 the passes divide naturally into those occurring in mid-morning (when T_e reaches its

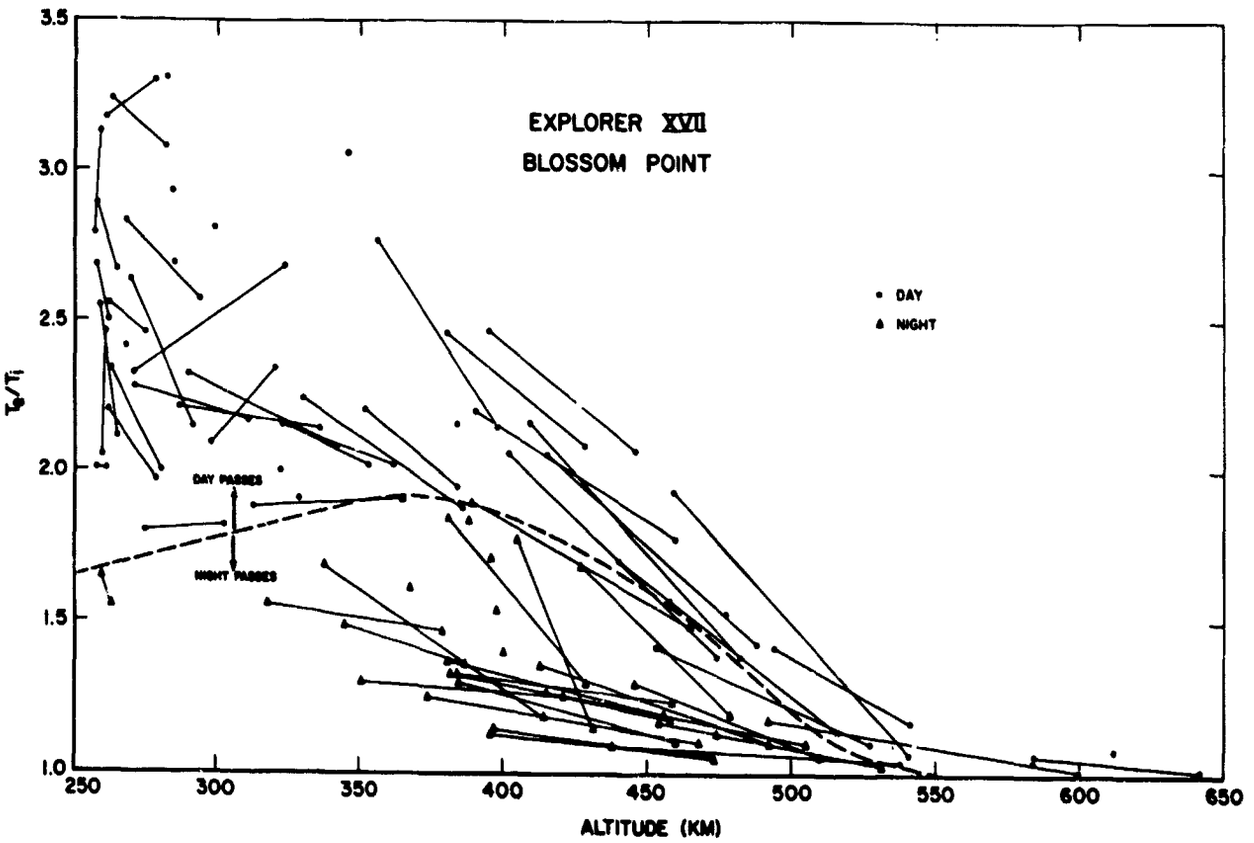


FIGURE 10.—T_e/T_i at Blossom Point.

maximum), near sunset and at night. Figure 9 reveals a positive-ion temperature increasing with altitude, a difference between T_i and T_e first emerging in the region of 300 km. The positive-ion temperature is larger during the day than at night but even at night it appears to exceed T_e at high altitudes. Local time effects are distinguished in Figure 10 only between day and night but the data are probably sufficiently numerous to establish that during the day T_e/T_i passes through a maximum value of between 2 and 3 at an altitude near 300 km and approaches unity at high altitudes. Similar descriptions of the average T_e variations with altitude and of T_e/T_i have been given by Evans (1964)[19] and by Evans and Lowenthal, (1964)[20] in analyses of ground-based radar observations made at the latitude of the Blossom Point passes.

The Heat Source

Having determined T_i, the rate of heat loss of the heated electrons to the positive ion gas may be computed. The heated electrons also cool in collisions with the neutral particle gas, the most significant processes being excitation of atomic oxygen to the 'D level at high altitudes and rotational and vibrational excitation of molecular nitrogen at low altitudes. Cooling by excitation of oxygen to the 'D level is of special importance since it may be responsible for the variability in the intensity of the red arc in the dayglow (Noxon 1964),[21] (Dalgarno and Walker, 1964).[22] The cooling rates have been tabulated by Dalgarno and McElroy (1964)[23] and a comparison of the contributions of the various processes at an altitude of 250 km is given in Figure 11.

The accuracy of the cooling rates is not yet adequate to warrant the use of the neutral particle densities and temperatures measured by the other Explorer XVII experiments (Newton, et al, 1964),[10] (Reber, et al, 1964).[9] To calculate the heat necessary to produce the measured electron temperature we have instead adopted the model distributions of Harris and Priester (1962)[18] appropriate to the local time of the measurement. The results for the southbound passes over Blossom Point are shown in Figure 12 as a function of local time and altitude. Local time effects dominate during the daytime southbound passes and it appears that the heat input is roughly symmetrical about local noon, a variation in marked contrast to the asymmetries in T_e and N_i. The heat input at 260 km has a maximum value of 4×10^3 eV cm^{-3} sec^{-1}. The nighttime values, presented in Figure 13, depend more strongly on altitude. They show clearly a non-vanishing heat source decreasing with increasing altitude with a value of not less than 20 eV cm^{-3} sec^{-1} at an altitude of 400 km. The importance of using calculated ion temperatures is shown by a comparison of Figures 13 and 14. No clear pattern for Q emerges from Figure 14, which is based upon the assumption that $T_i = T_e$.

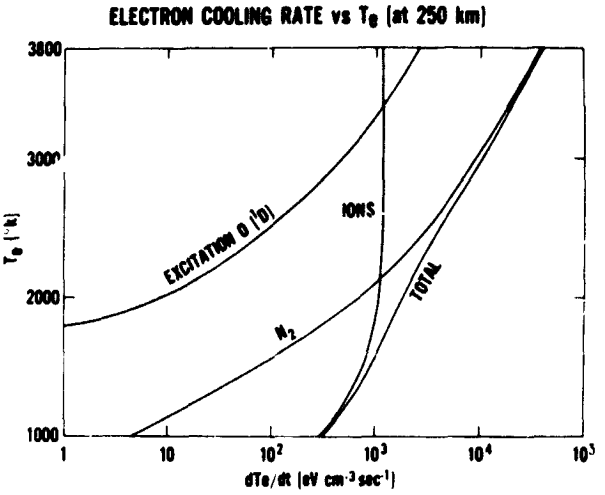


FIGURE 11.—Electron cooling rates at 250 kilometers demonstrating their sensitivity to the electron temperature.

The calculated values of Q for the northbound passes are shown in Figure 15. Local time and altitude effects are mixed and we have attempted to remove the dependence on local time by adopting for the northbound passes the diurnal variation suggested by the southbound passes. The corrected values of Q, normalized to noon, are given in Figure 16. They show a heat input decreasing with altitude with an apparent change in the gradient in the region of 350 km.

DISCUSSION

The probe measurements of T_e from the Explorer XVII satellite confirm the conclusions of the rocket flights (Boggess, Brace and Spencer 1959),[1] (Spencer, Brace and Carignan, 1962),[2] (Brace, Spencer and Carignan, 1963),[3] (Nagy, Brace, Carignan and Kanal, 1963)[24] that the difference between the temperature of the electrons

in the F region of the ionosphere and the neutral particle temperature is large during the daytime and small but non-vanishing at night. These conclusions are in harmony with measurements on the Ariel satellite (Willmore, Henderson, Bowen and Boyd, 1964)[4] and on the Discoverer satellite (Sagalyn and Smiddy, 1964),[25] with observations of incoherent backscatter (Bowles, Ochs and Green, 1962),[26] (Evans, 1962),[27] (Pineo and Hynek, 1962),[28] (Evans, 1964),[19] (Evans and Lowenthal, 1964),[20] (Greenhow, Sutcliffe, Watkins)[21] and with a recent analysis of density profiles (Bauer, and Blumle, 1964).[30]

There are significant differences between the various data. In particular, Willmore, et al do not find the morning maximum in T_e which is a

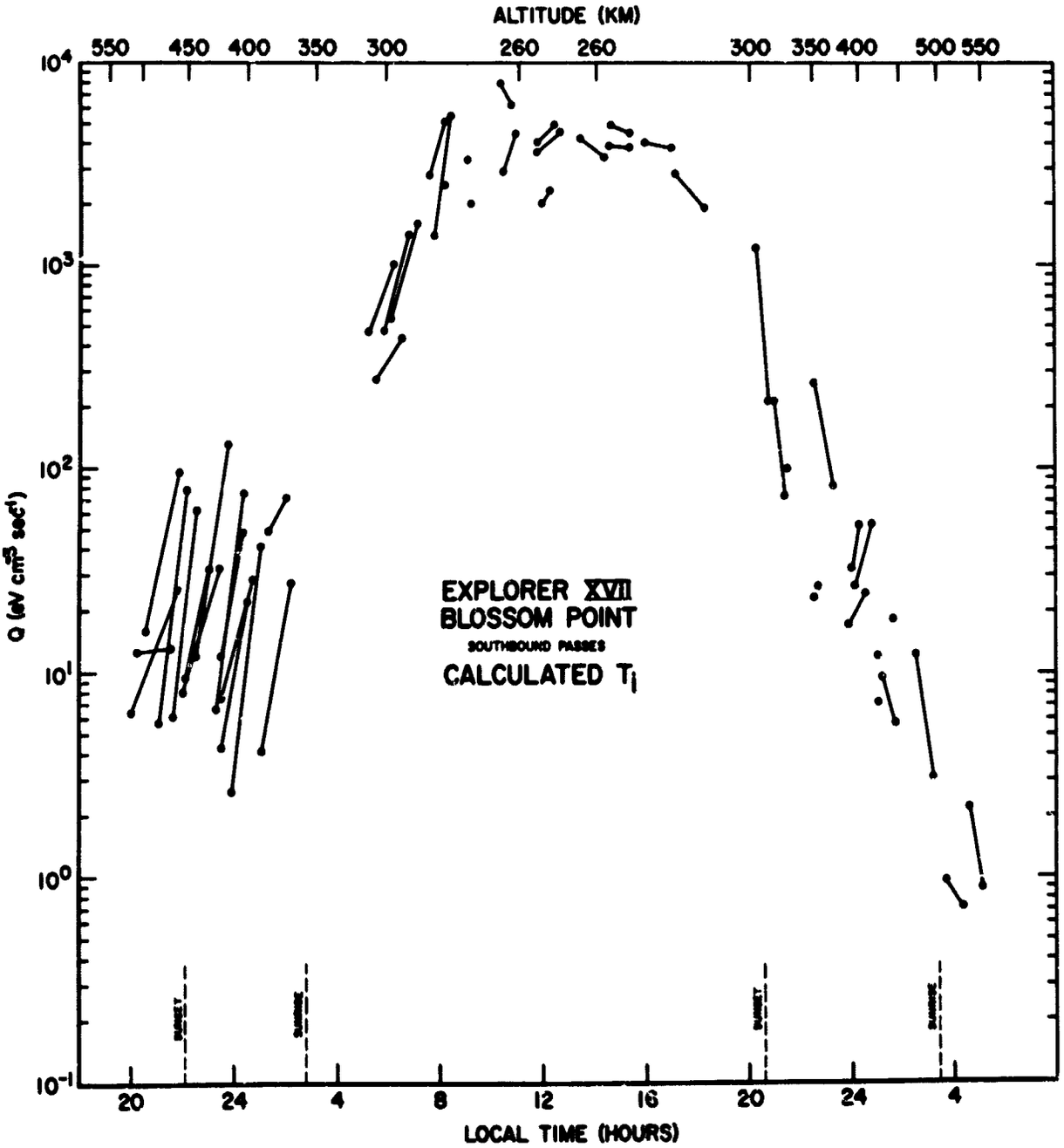


FIGURE 12.—Locally absorbed energy (Q) calculated for southbound passes. Local time effects predominate in the daylight passes and altitude effects control the variations in the nighttime values of Q.

feature of the Explorer XVII data and which also appears in the backscatter observations of Evans and Loewenthal, the satellite data of Sagalyn and Smiddy and the density profile analysis of Bauer and Blumle. Bowles, et al, observing at the Equator, find a maximum at sunrise. As discussed earlier, Figure 8 shows that the time of the maximum depends at least partly upon geographic latitude.

Again in harmony with the rocket flight data, the Explorer XVII data show that T_e increases with latitude, a phenomenon also observed by Willmore, Henderson, Bowen and Boyd (1964).[4] The inverse relationship between T_e and n_e described in Equation (3) strongly suggests that the increase in T_e is caused by the decrease in n_e, a possibility noted by Dalgarno (1963).[8] The temperature may not continue to increase at high latitudes and a preliminary analysis of passes over Grand Forks at a latitude of approximately 65°N magnetic indicates that T_e had passed through a maximum (and n_e through a minimum).

Strong evidence that the positive ion temperature is also higher than the neutral particle temperature at altitudes above 300 km has been advanced by Evans (1964)[19] and by Evans and Loewenthal (1964)[20] and the averaged profiles of T_e and of T_e/T_i that they present are consistent with those given in Figures 9 and 10, which are based upon Equation (8). The magnitudes of T_i at 400 km are in harmony with those measured by Boyd and Raitt (1964)[21] and it seems that over a considerable altitude extent $T_e > T_i > T_g$ with T_e/T_i approaching unity at high altitudes, a possibility to which Hanson (1963)[7] and Dalgarno (1963)[8] drew attention.

The heating effect of absorbed solar ultraviolet radiation has been investigated by Hanson and Johnson (1961),[5] by Dalgarno, McElroy and Moffett (1963),[6] by Hanson (1963)[7] and most recently by Dalgarno and McElroy (1964)[23] whose calculations make use of the improved photoionization calculations of Dalgarno, Henry, and Stewart (1964).[22] The assumption that the heat

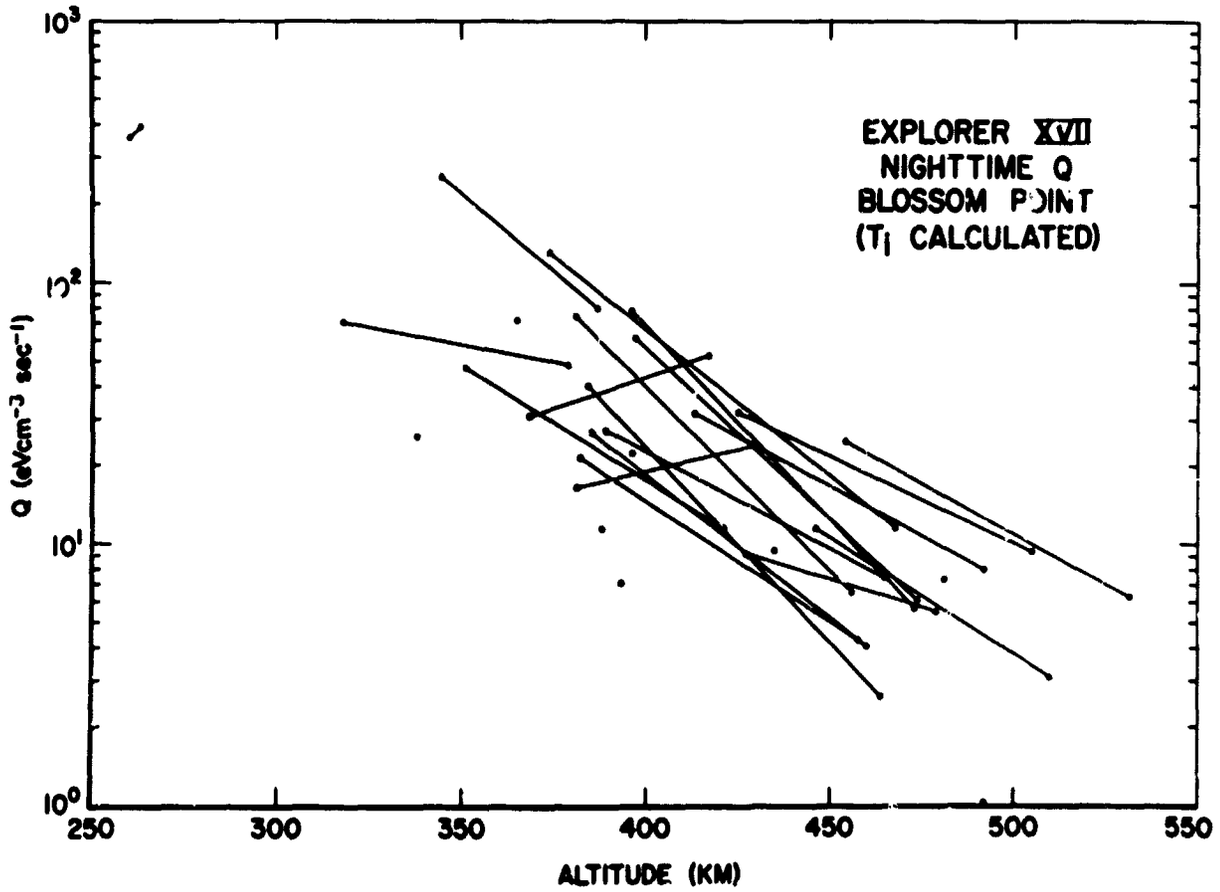


FIGURE 13.—Q vs. altitude in the nighttime ionosphere using the calculated values of T_i.

is deposited locally leads to the predicted euv heating shown in Figure 16, which is in agreement with the heating derived from the Explorer XVII measurements at altitudes below 300 km. Because of the escape of fast photoelectrons (Hanson, 1963)[7] the assumption of local heating is not valid above 300 km.

The gradient of the heat deposition from solar extreme ultraviolet radiation is equivalent to the neutral particle scale height, which is also shown in Figure 16 to emphasize the change in gradient near 350 km. This change in gradient demonstrates that absorption of energy by neutral particles is not the controlling process at great altitudes, and it suggests that the controlling process may be absorption by the ambient electrons, a typical gradient for which is also shown in Figure 16. Such a heat flux is to be expected at great altitudes from the escaping photoelectrons which lose energy by elastic collisions with the ambient electrons.

If j cm^{-2} sec^{-1} is the upward flux of photoelectrons and we adopt a mean of 10 eV for their energy, the rate of heating is given by

$$Q = 2j\, n_e \times 10^{-13} eV \text{ cm}^{-3} \text{ sec}^{-1} \tag{9}$$

If the electrons cool by collisions with the positive ions 0^+,

$$\frac{T_e - T_i}{T_e^{3/2}} = \frac{4\times10^{-7}}{n_e} j, \tag{10}$$

and if by collisions with He^+

$$\frac{T_e - T_i}{T_e^{3/2}} \quad \frac{1\times10^{-7}}{n_e} j. \tag{11}$$

It follows that an upward flux of about 5×10^8 cm^{-2} sec^{-1} may be sufficient to explain the Explorer XVII data, and it appears from the report of Hinteregger, Hall and Schmidtke (1964)[33] on the intensities of the extreme solar ultraviolet

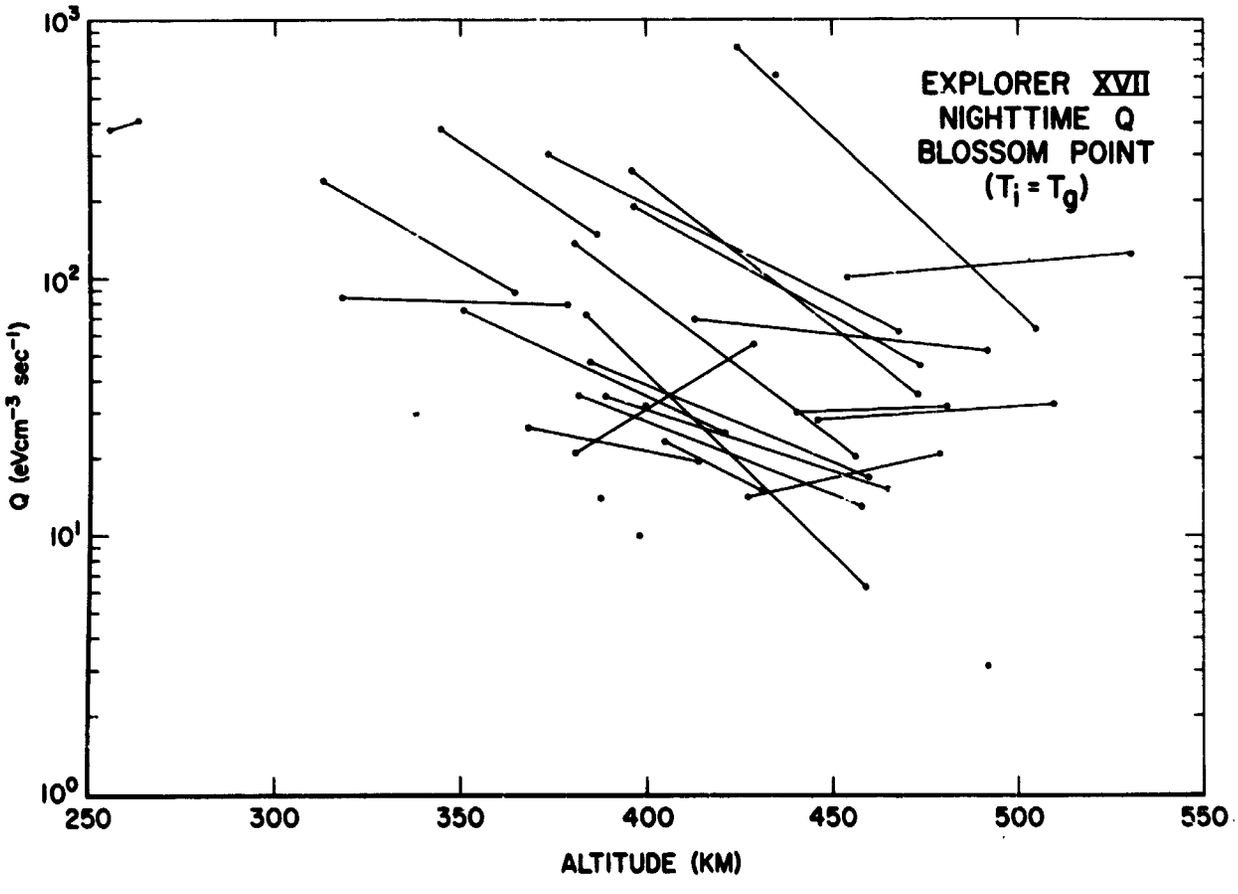


FIGURE 14.—Q vs. altitude in the nighttime ionosphere assuming $T_i = T_e$. Note the greater scatter and lack of pattern resulting from this assumption.

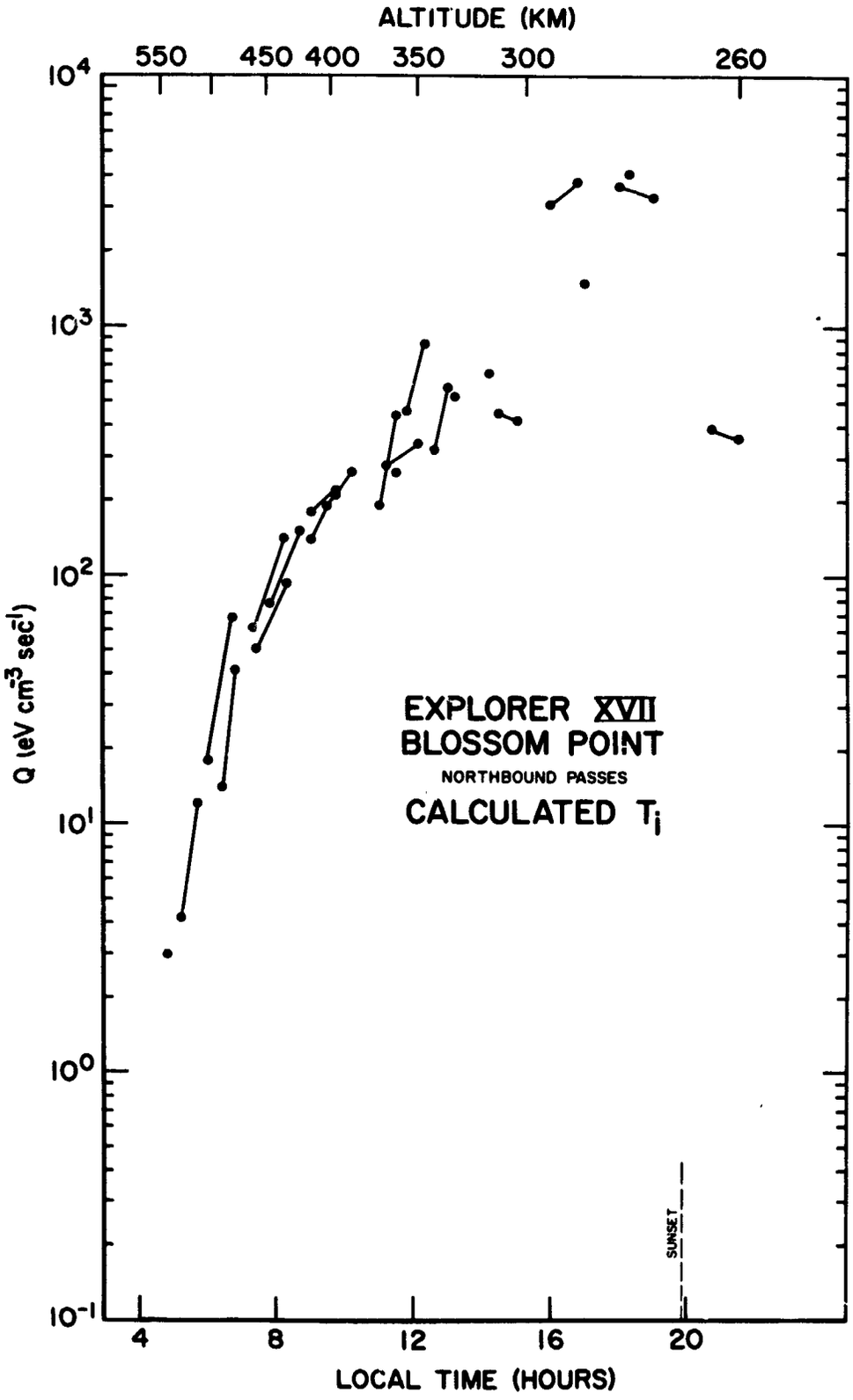


FIGURE 15.—Q calculated for northbound passes. Local time and altitude effects are strongly mixed in these passes, particularly in the early morning.

radiation, and from observations on Cosmos 3 and Cosmos 5 satellites (Mularchik and Vaisberg, 1964),[34] that such a flux is available.

Similar arguments have been advanced by Willmore (1964)[35] in a discussion of the Ariel measurements which were carried out a year earlier than Explorer XVII measurements, but which lead to daytime heat inputs smaller by a factor of two or three.

The fact that the electron temperature does not fall to the neutral particle temperature at night is of great interest, since it reveals the presence of an energy source in the nocturnal ionosphere. The heat input at an altitude of 400 km is about 20 eV cm^{-3} sec^{-1}, which is about five times larger than that derived from Ariel data. Willmore (1964)[35] has suggested that the energy source responsible for the Ariel observations can be identified with the fluxes of 2–3 kev electrons of about 8×10^{9} cm^{-2} sec^{-1} observed by Vernon, Savenko, Shavrin, Nesterov, Pisivenko (1963).[36] This is an energy flux of 30 ergs cm^{-2} sec^{-1}, which is an order of magnitude larger than that of the solar ultraviolet radiation. If these electrons were precipitated into the atmosphere, they would give rise to 6 kilorayleighs of 3914 Å emission (c.f. Dalgarno, 1964)[37] whereas O'Brien, Allum and Goldwire (1964)[38] observe 5 rayleighs or less above 85 kilometers.

The difficulties presented by the large energy flux postulated by Willmore can be avoided by arbitrarily invoking instead a flux of soft electrons of, say, 100 eV energy. Assuming a heating efficiency of 0.1, the Ariel data require about 2×10^{-3} ergs cm^{-2} sec^{-1} and the Explorer XVII data require about 1×10^{-2} ergs cm^{-2} sec^{-4}. Fluxes of this order do not contradict any of the limits that have been derived from spectroscopic observations (c.f. Dalgarno, 1964)[37] and the resulting altitude variation in the heat deposition would be similar to that indicated by the data in Figure 16.

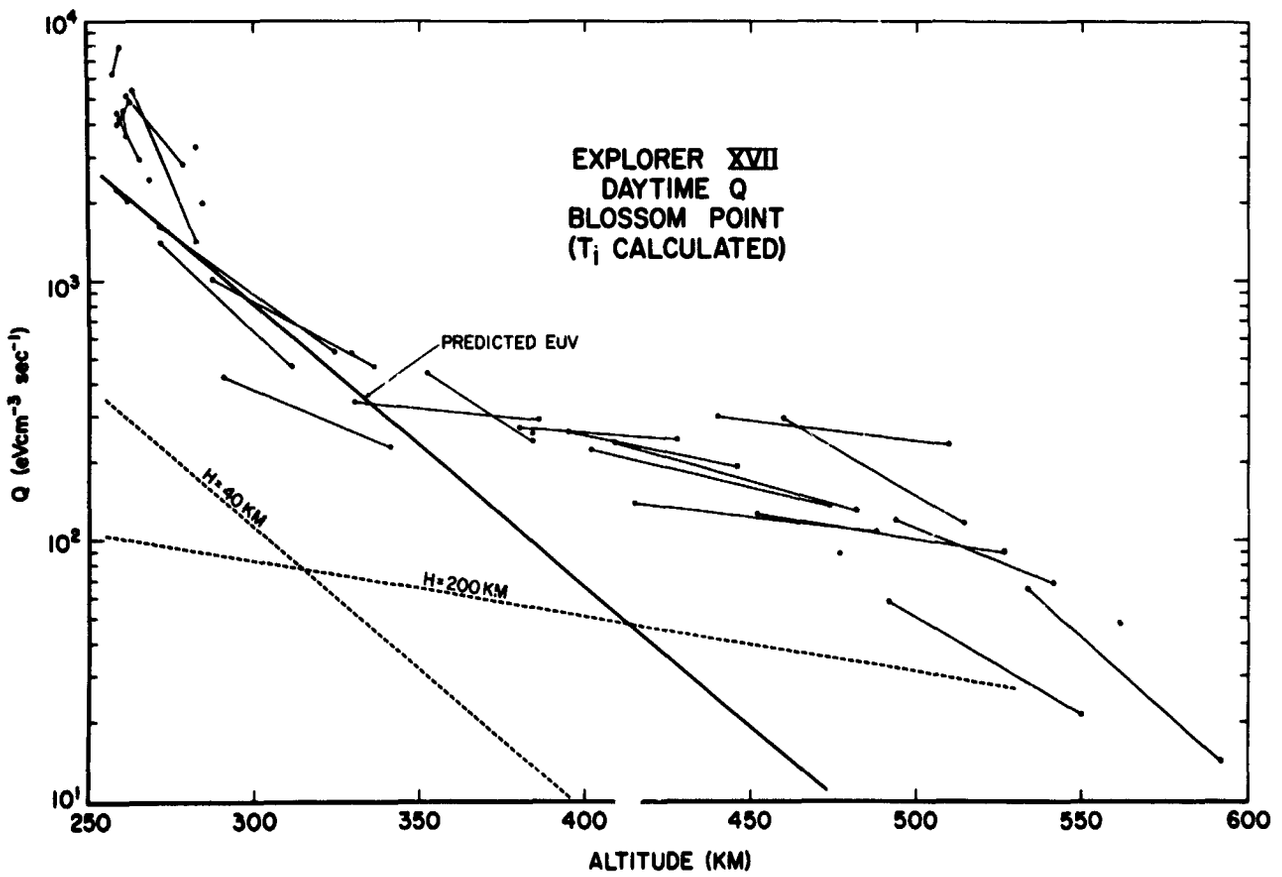


FIGURE 16.—Q vs. altitude normalized to noon. This altitude variation of Q was obtained from northbound passes by removing the diurnal variation suggested by southbound passes.

ACKNOWLEDGMENT

The authors are indebted to Mr. G. Carignan, University of Michigan, and Mr. J. Findlay, Goddard Space Flight Center for their outstanding efforts toward making the experiment successful and to G. Dunham for his aid in the analysis of data.

REFERENCES

1. Boggess, R. L., L. H. Brace, N. W. Spencer, J. Geophys. Res. **64,** 1627 (1959)
2. Spencer, N. W., L. H. Brace, G. R. Carignan, J. Geophys. Res. **67,** 157 (1962)
3. Brace, L. H., N. W. Spencer, G. R. Carignan, J. Geophys. Res. **68,** 5397 (1963)
4. Bowen, P. J., R. L. F. Boyd, C. L. Henderson, A. P. Willmore, Proc. Roy. Soc. A, **281,** 514 (1964)
5. Hanson, W. B., and F. S. Johnson, Mem. Soc. Sci. Liege, Series 5, **4,** 390 (1961)
6. Dalgarno, A., M. B. McElroy and R. J. Moffett, Planet. Space Sci. **11,** 463 (1963)
7. Hanson, W. B., Space Research III (North-Holland: Amsterdam) Edited by W. Priester (1963)
8. Dalgarno, A., National Aeronautics and Space Administration, CR-8 (1963)
9. Reber, C. A., and M. Nicolet, Planet. Space Sci. (in this issue) (1965)
10. Newton, G. P., R. Horowitz and W. Priester, Planet. Space Sci. (in this issue) (1965)
11. Brace, L. H., and N. W. Spencer, J. Geophys. Res. **69,** (1964)
12. Mott-Smith, H. M., and I. Langmuir, Phys. Rev., **28,** 727 (1926)
13. Spencer, N. W., Pllanet. Space Sci. (in this issue) (1965)
14. Spitzer, L., *Physics of Fully Ionized Gases* (Interscience, N. Y.) (1965)
15. Dalgarno, A., Phil. Trans. Roy. Soc., **250,** 426 (1958)
16. Dalgarno, A., Annales de Geophysique, **17,** 16 (1961)
17. Dalgarno, A., Journal of Atmos. and Terr. Phys., **26,** 939 (1964)
18. Harris, I., and W. Priester, J. Geophys. Res., **67,** 4585 (1962)
19. Evans, J. V., J. Geophys. Res., **69,** 1436 (1964)
20. Evans, J. V., and M. Loewenthal, Planet. and Space Sci. **10,** 905 (1964)
21. Noxon, J. F., J. Geophys. Res. **69,** 3245 (1964)
22. Dalgarno, A., and J. C. G. Walker, Journ. Atmos. Sci., **21,** 463 (1964)
23. Dalgarno, A., and M. B. McElroy, to be published (1964)
24. Nagy, A. F., L. H. Brace, G. R. Carignan and M. Kanal, J. Geophys. Res. **68,** 6401 (1963)
25. Sagalyn, R. C., M. Smiddy, Y. N. Bhargava, presented at fifth International COSPAR meeting, Florence, (1964)
26. Bowles, K. L., E. R. Ochs, and J. L. Green, J. Res. NBS, **66,** 395 (1962)
27. Evans, J. V., J. Geophys. Res., **67,** 4914 (1962)
28. Pineo, V. C., D. P. Hynek, J. Geophys. Res. **67,** 5119 (1962)
29. Greenhow, J. S., H. K. Sutcliffe and C. D. Watkins, Journ. Atmos. Terr. Phys., **25,** 197 (1963)
30. Bauer, S. J., L. Blumle, J. Geophys. Res., **69,** 3613 (1964)
31. Boyd, R. L. F., W. J. Raitt, to be published (1964)
32. Dalgarno, A., R. J. W. Henry, and A. L. Stewart, Planet. Space Sci. **12,** 235 (1964)
33. Hinteregger, H. E., L. A. Hall, G. Schmidtke, presented at fifth International COSPAR meeting, Florence, (1964)
34. Mularchik, T. M., O. L. Vaisberg, presented at fifth International COSPAR meeting, Florence (1964)
35. Willmore, P. A., Proc. Roy. Soc. A, **281,** 140 (1964)
36. Vernov, S. N., I. A. Savenko, P. I. Shavrin, V. YE Nesterov and N. F. Pisarenko, Planet. Space Sci., **11,** 567 (1963)

FIRST ELECTROSTATIC PROBE RESULTS FROM EXPLORER 17*

N66 37964

L. H. BRACE AND N. W. SPENCER

INTRODUCTION

On April 3, 1963, the Explorer 17 satellite was launched into a 58° inclination direct orbit having perigee and apogee altitudes of 258 and 927 km, respectively. In addition to the neutral particle instruments [*Spencer and Reber*, 1963, *Newton et al.*, 1963], two cylindrical electrostatic probes were used, one for the measurement of electron temperature (T_e) and the other for determining the positive ion density (N_i). The purpose of this letter is to report the variations in T_e and N_i observed above the Blossom Point, Maryland, telemetry station (latitude 38°N, longitude 77°W) and to briefly discuss the more immediate conclusions that can be drawn from the data.

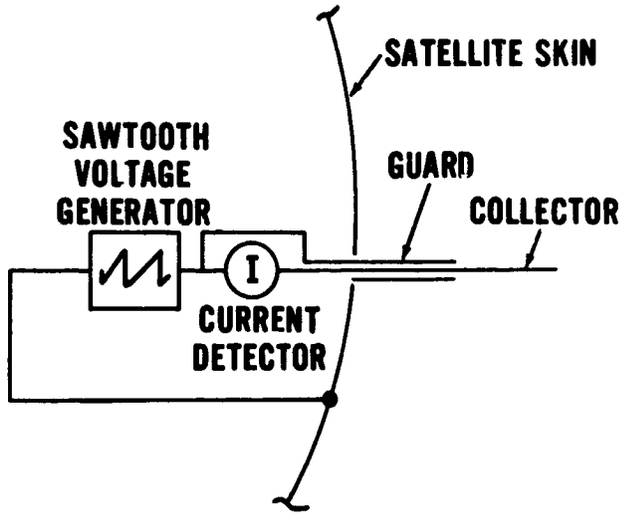


FIGURE 1.—Functional block diagram of either the T_e or N_i probe experiment. A sawtooth voltage applied to the guard and collector produces a volt-ampere characteristic from which either T_e or N_i can be derived, depending on the current sensitivity and sweep magnitude employed.

*Published in the *Journal of Geophysical Research*, 69(21):4686–4689, November 1, 1964.

THE EXPERIMENTS

Figure 1 is a block diagram illustrating either of the two independent probe instruments. Both instruments employ a two-element sensor consisting of a 10-cm-long guard electrode, concentric with a collector electrode 0.056 cm in diameter and 23 cm long. The collector and guard electrodes were insulated from each other and from the satellite, which served as the reference electrode. An appropriate sawtooth voltage was applied to both collector and guard elements, but only the current to the collector was measured and telemetered. Voltage sweep rates, current sensitivities, and telemetry sampling rates were optimized to permit a measurement of T_e in approximately 50 msec, corresponding to negligible translational and rotational motion of the satellite. The volt-ampere characteristics resulting from operation of the instruments thus permitted computation of electron-temperature and ion-density values that can be considered measurements at a point.

Figure 2 shows a series of raw telemetry points from the T_e experiments which demonstrate the measurement resolution. The computational technique employed has been reported [*Langmuir and Mott-Smith*, 1924; *Taylor et al.*, 1963; *Nagy et al.*, 1963].

To permit greater confidence in the measurement accuracy of the Explorer 17 probe technique, a rocket (NASA 6.07) carrying an ejectable instrumentation employing a probe sensor identical to that of Explorer 17 was launched concurrent with a satellite pass at Wallops Island, Virginia. An ionosonde was also operating simultaneously nearby. The ionosonde and rocket probe data agree with the satellite data within a few percent, confirming that the satellite motion did not perturb the temperature and density measurements. A more detailed documentation of the rocket-

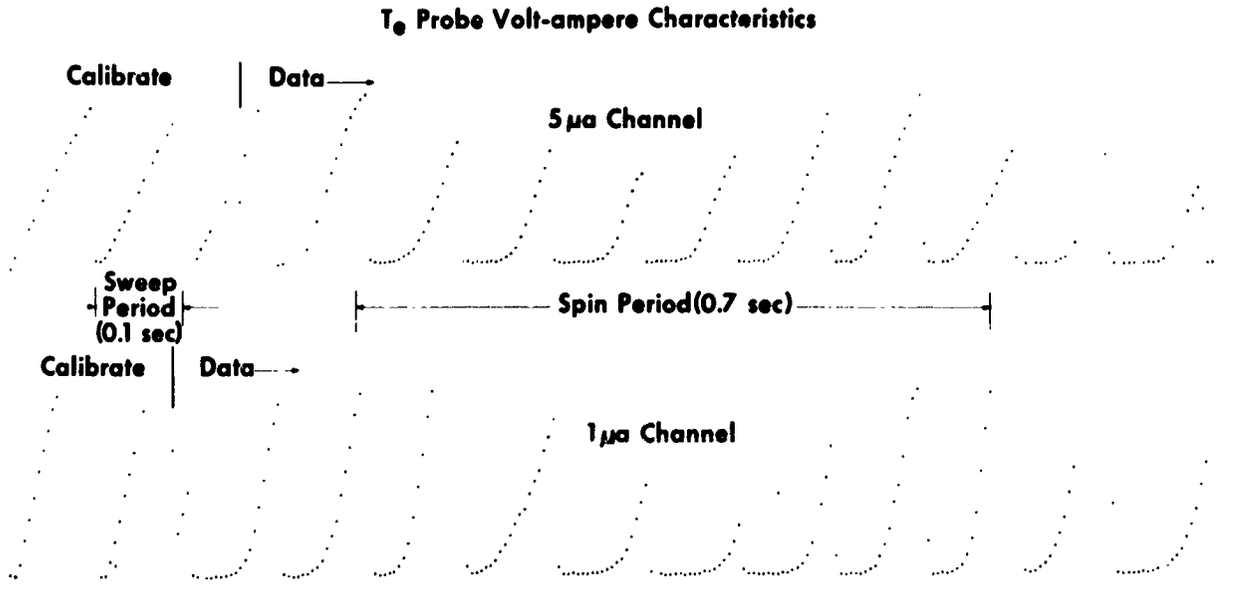


FIGURE 2.—Plot of raw telemetry data from T_e experiment during nighttime perigee pass at Blossom Point. The upper set of volt-ampere characteristics were measured by the 5-μa detector with a sweep of 0→+1.5 volts. The lower set, recorded a few seconds later, were measured by the 1-μa detector with a sweep of 0→+0.75 volts applied to the probe.

satellite experiment will be provided in a later paper.

THE BLOSSOM POINT DATA

During its 100-day operational lifetime (April 3 to July 10, 1963), Explorer 17 was interrogated successfully about 650 times, most responses providing 4 minutes of measurements along the satellite path. The Blossom Point (BP) passes, to be discussed here, provided dense coverage over the eastern United States. During the active lifetime of the satellite, the latitude of perigee, initially 38°N, moved to its maximum northward excursion of 58°, then moved southward, reaching 22°S at the time of battery exhaustion. Accordingly, most of the BP measurements were made near perigee and thus generally represent the F_2 region between 258 and 400 km. Changes in T_e and N_i during these passes reflect latitude and/or local apparent time effects (approximately 10° of latitude and 1 hour of local time) and not altitude change. These changes within passes are evident in Figure 3, which shows data measured over a three-month period. The individual points connected by solid lines are the values of T_e and N_i measured at the beginning and end of some of the individual 4-minute passes. Most of these data are from southward passes during which a consistent decrease in T_e and an increase in N_i were observed. The reverse changes are observed within the northward passes, confirming that the change observed within the pass is a latitude effect. By drawing smooth curves through the end values of T_e and N_i from the individual passes in Figure 3, the gross diurnal variation at the extremes of latitude reached in the BP passes (40° to 55° north magnetic) is made evident.

Several general characteristics of the ionosphere are also evident in Figure 3:

1. A steep morning rise in T_e and a more gradual rise in N_i.
2. A morning maximum in T_e (about 5 hours after local sunrise) and a continuing gradual rise in N_i during the day.
3. A decrease of T_e to an afternoon plateau (its value depending on latitude) and a late afternoon maximum of N_i.
4. A possible minor maximum of T_e during late afternoon.
5. A steep decline of both T_e and N_i at sunset.
6. A nighttime plateau of T_e at about 1150°K, but moderately variable, and a relatively constant value of N_i.

Although the data from other sites have not been completely analyzed, they appear to exhibit the same gross characteristics as the data from Blossom Point, the major differences being in the

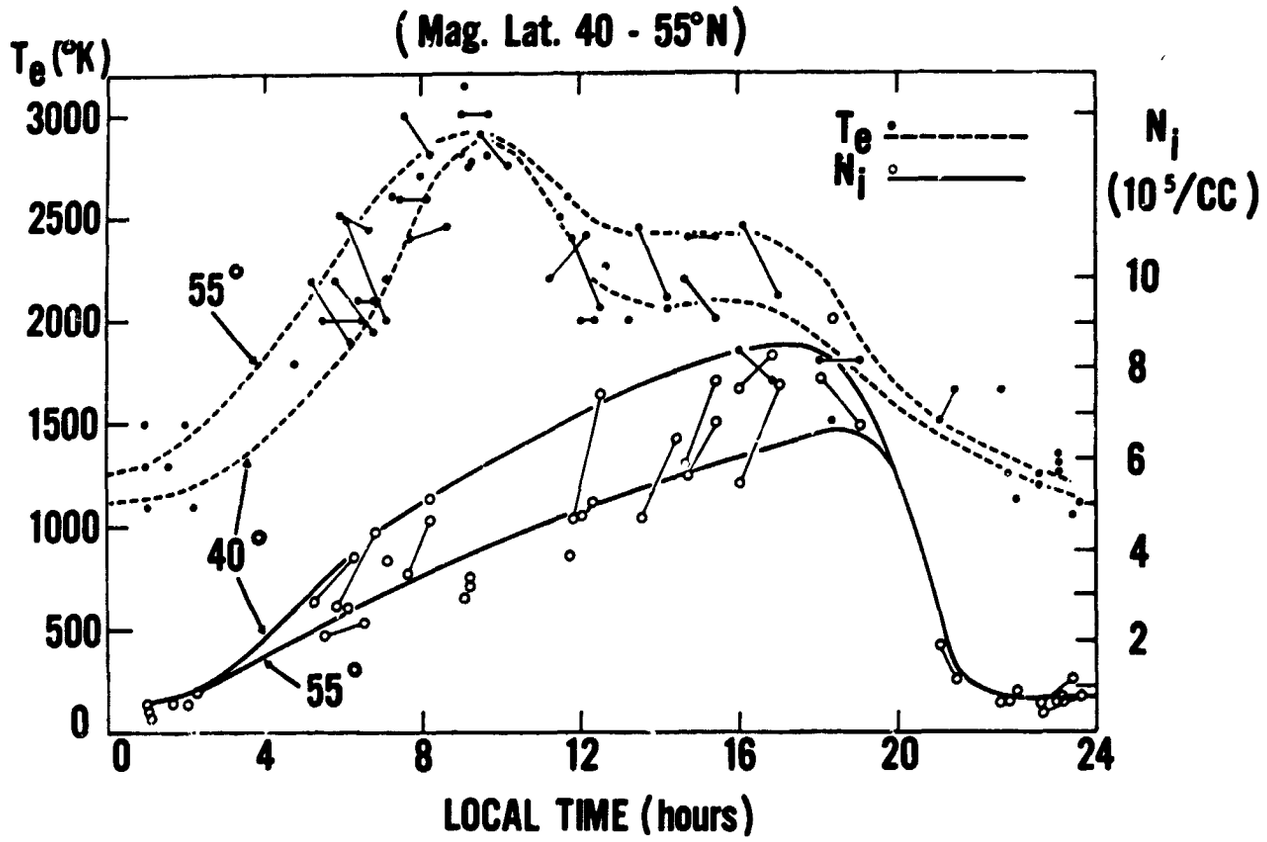


Figure 3.—Blossom Point results in the 260- to 400-km altitude range. The points connected by lines represent values recorded at the beginning and end of particular passes. The inverse variation of T_e and N_i within individual passes has been identified with the latitude change in the pass.

time of the morning T_e maximum and the value of T_e during the afternoon plateau.

GEOPHYSICAL IMPLICATIONS OF THE BLOSSOM POINT DATA

If we consider the neutral gas temperature (T_g) to be between about 700°K (at night) and 1000°K (daytime), the Figure 3 data show that thermal nonequilibrium is the normal condition near the F_2 maximum, both day and night. Thus, the Explorer 17 data confirm our earlier conclusions which were based on rocket probe data [*Boggess et al.*, 1959; *Spencer et al.*, 1962; *Brace et al.*, 1963] that (*a*) the ionosphere is not in thermal equilibrium and (*b*) the degree of nonequilibrium is greater in the auroral zone than at midlatitudes.

The major implication of the nonequilibrium is that significant sources of local heating for the electrons are present and that some of these sources exist at night as well as during the day. This is borne out by calculations of the local electron heat flux density (Q) using the equation given by *Hanson* [1963];

$$Q = -\frac{3kN_e}{2}\left[\left.\frac{dT_e}{dt}\right|_{O} + \left.\frac{dT_e}{dt}\right|_{N_2} + \left.\frac{dT_e}{dt}\right|_{O_2} + \left.\frac{dT_e}{dt}\right|_{+}\right] \quad (\text{ev cm}^{-3}\,\text{sec}^{-1}) \quad (1)$$

where the derivatives represent the cooling rates of electrons to ions and to the indicated neutral species. In a calculation using the measured values of T_e and N_i and the Harris and Priester model values ($S=90$) for the neutral particle concentrations and temperatures, the locally deposited energy (Q) at 400 km was determined to be approximately 250 ev cm^{-3} sec^{-1} at midday and 15 ev cm^{-3} sec^{-1} at night [*Brace et al.*, 1964]. The daytime value is in reasonable agreement with the predictions of both *Hanson* [1962] and *Dalgarno et al.* [1963], but differs significantly from Ariel satellite data given by *Willmore* [1964], who finds about

100 ev cm^{-3} sec^{-1} at 400 kilometers. The nighttime value of 15 ev cm^{-3} sec^{-1} indicates a significant heating source, not yet identified.

The calculations further show that both the diurnal and latitudinal variations of electron temperature in the daytime ionosphere (Figure 3) are primarily reflections of corresponding variations in the local electron density [*Willmore*, 1964]; thus there is no reason to believe that the observed latitude variation of T_e in the daytime is related to atmospheric heating by particles as a function of latitude [*Willmore et al.*, 1962; *Spencer et al.*, 1962; *Brace et al.*, 1963].

It is particularly useful to compare BP data with the radar backscatter data of *Evans* [1964], because the measurements are from essentially the same latitude, longitude, and altitude. In March 1963, a month before Explorer 17 was launched, Evans found that T_e/T_i, at 330 km reached a maximum value of 3 at midmorning, exhibited an afternoon plateau of about 2.3, and was variable between 1.3 and 2 at night. The satellite results at Blossom Point, expressed in terms of T_e/T_p, provide an essentially identical picture.

Inherent in this comparison of T_e/T_i and T_e/T_p is the assumption of equilibrium between ions and neutrals ($T_i = T_p$). In the altitude region below 400 km where these data are compared, it is reasonable to make this assumption. However, as *Hanson* [1963] has pointed out, because the neutral particle scale height is less than the ion-electron scale height, the thermal contact between ions and electrons will predominate above some altitude and the ion temperature will approach that of the electrons. *Hanson* [1963], using an assumed model atmosphere, has calculated that T_i will exceed T_p above about 600 km and will approach T_e at 900 km. Accordingly, T_i was predicted to lie about midway between T_e and T_p at 750 km. Similar calculations using Explorer 17 BP data show that this transition occurs near 500 km rather than 750 km (*Brace et al.*, 1964].

In summary, Explorer 17 measurements of T_e and N_i in the F_2 region above Blossom Point have shown that thermal nonequilibrium ($T_e > T_p$) is the normal condition both day and night and, further, that the degree of nonequilibrium is strongly latitude dependent in the daytime and moderately variable at night. The data also reveal a strong inverse relationship between the local values of T_e and N_i (or N_e), which is consistent with the equations Hanson and Dalgarno used in their studies of ionospheric heating. There is no evidence that the observed latitude dependence of T_e, seen in previous rocket flight data and apparent in Ariel satellite measurements, is related in an important way to particle fluxes at higher latitudes. Instead, the data suggest that the latitude dependence of T_e primarily reflects the global distribution of electron density and its controlling mechanism.

ACKNOWLEDGMENT

We thank Dr. A. Dalgarno of the Queen's University of Belfast for enlightening discussions of the experimental data and its geophysical implications and for his participation in the calculations of ion temperature and electron energy input.

REFERENCES

Boggess, R. L., L. H. Brace, and N. W. Spencer, Langmuir probe measurements in the ionosphere, *J. Geophys. Res.*, *64*(10), 1627–1629, October 1959.

Brace, L. H., N. W. Spencer, and G. R. Carignan, Ionosphere electron temperature measurements and their implications, *J. Geophys. Res.*, *68*(19), 5397–5412, October 1, 1963.

Brace, L. H., N. W. Spencer, and A. Dalgarno, Explorer 17 electrostatic probe data and interpretation, paper presented at Annual Meeting, Am. Geophys. Union, 45th Meeting, Washington, D. C. April 23, 1964.

Dalgarno, A., M. B. McElroy, and R. J. Moffett, Electron temperatures in the ionosphere, *Planetary Space Sci.*, *11*, 463–454, Pergamon Press, London, 1963.

Evans, J. V., Incoherent backscatter studies of the ionosphere at Millstone Hill, in *Electron Density Distributions in the Ionosphere and Exosphere*, edited by E. Thrane, pp. 256–304, North-Holland Publishing Co., Amsterdam, 1964.

Hanson, W. B., Electron temperatures in the upper atmosphere, *Space Res.*, *3*, 282–302, 1963.

Langmuir, I., and H. Mott-Smith, Studies of electric discharges in gases at low pressures, *Gen. Elec. Rev.*, part 1, *27*(7), 449–455, July 1924.

Nagy, A. F., L. H. Brace, G. R. Carignan, and M. Kanal, Direct measurements bearing on the extent of thermal nonequilibrium in the ionosphere, *J. Geophys. Res.*, *68*(24), 6401–6412, December 1963.

Newton, G. P., D. J. Pelz, G. E. Miller, and R. Horowitz, Response of modified Redhead magnetron and Bayard-Alpert vacuum gages aboard Explorer 17, *Transactions of the Tenth National Vacuum Symposium*, edited by George H. Bancroft, pp. 208–212, The Macmillan Company, New York, 1963.

NICOLET, M., Density of the heterosphere related to temperature, *Smithsonian Astrophys. Obs. Spec. Rept. 75*, Sept. 19, 1961.

SPENCER, N. W., L. H. BRACE, and G. R. CARIGNAN, Electron temperature evidence for nonthermal equilibrium in the ionosphere, *J. Geophys. Res.*, *67*(1), 157–175, January 1962.

SPENCER, N. W., and C. A. REBER, A mass spectrometer for an aeronomy satellite, *Space Res.*, *3*, 1151–1155, 1963.

TAYLOR, H. A., L. H. BRACE, H. C. BRINTON, and C. R. SMITH, Direct measurements of helium and hydrogen ion concentration and total ion density to an altitude of 940 kilometers, *J. Geophys. Res.*, *68*(19), 5339–5347, October 1963.

WILLMORE, A. P., Energy sources for the ionosphere electrons, *Nature*, *202*(4927), 41–42, April 14, 1964.

WILLMORE, A. P., R. L. F. BOYD, and P. J. BOWEN, Some preliminary results of the plasma probe experiments on the satellite ariel, *Proc. Intern. Conf. Ionosphere, Imperial College, London*, July 1962.

N66 37965

ORIGIN AND EVOLUTION OF ATMOSPHERES AND OCEANS*

A. G. W. CAMERON, P. J. BRANCAZIO, AND N. W. PANAGAKOS

In recent years it has become evident that the atmosphere and oceans have been produced by the outgassing of volatile materials from the earth's interior, principally from volcanoes. In 1950 W. W. Rubey presented geological evidence indicating that sea water has progressively accumulated in this way. At about the same time, Harrison Brown pointed out that the very low abundances of the noble gases in the earth's atmosphere constituted compelling evidence that at least the atmospheric oxygen and nitrogen were almost entirely of secondary origin, having been outgassed, possibly in different chemical form.

The main purpose of the 1963 conference was to consider new evidence for mechanisms which might add to or subtract from the contents of the atmosphere and oceans, and to consider mechanisms which could contribute to the origin of atmospheres of other planets. H. H. Hess opened the discussion with a consideration of the problems of convection currents in planetary mantles. The important factor is the time between nucleosynthesis and the formation of the solid planet. The protoplanet will contain radioactive materials which generate heat, and the rate at which it is generated will decrease as the activity of the radioisotopes decreases. A planet may or may not have convection, depending on whether enough radioactive materials remain in the mantle to cause vigorous heating at the base.

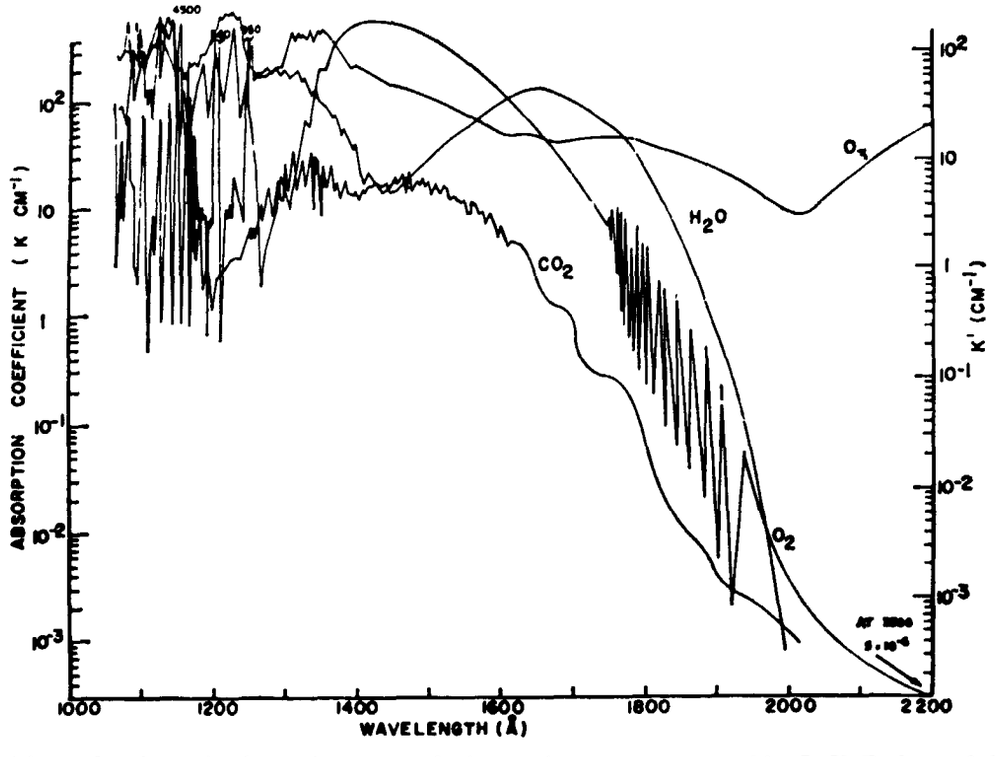
K. K. Turekian of Yale University discussed some models for the degassing of argon from the earth. He showed that at least eighty percent of the argon now in the atmosphere must have come from the mantle. He proposed that the most useful degassing model appeared to be one with continuous degassing of the earth as a whole. Turekian pointed out, however, that one should be careful in applying the results thus obtained to the escape of the other rare gases from the earth.

L. V. Berkner of the Southwest Center for Advanced Studies, Dallas, Texas, discussed the origin of oxygen in the earth's atmosphere. According to Berkner, the existence of an appreciable atmosphere of oxygen implies the presence of life. He pointed out that the production of oxygen in the primitive atmosphere from the photochemical dissociation of water would in turn lead to a layer of ozone in the atmosphere. The ozone layer is a buffer against the destruction of terrestrial life as we know it by the ultraviolet rays of the sun. If the amount of oxygen in the atmosphere were very small, the ozone layer would occur at ground level, and the sun's ultraviolet radiation would penetrate to the surface. In such a harmful environment, life would be unable to emerge from the seas onto land.

Tracing the history of the earth's atmosphere, Berkner pointed out that a sufficiently long period of time had to elapse for marine life to produce sufficient oxygen to change the composition of the atmosphere and produce high-lying ozone layers. When this took place, life was able to emerge—first in plant form, then in animal form in perhaps a few million years.

The helium problem was examined by G. J. F. MacDonald of the University of California at Los Angeles. He pointed out that if it is assumed that the escape of gases from the earth's atmosphere is due only to temperate effects, then the rate of influx of helium by outgassing is larger than the rate of escape. The helium in the earth's atmosphere should be building up at a rapid rate, but this is not the case. MacDonald noted that analyses of satellite drag data by I. Harris and W. Priester showed that an additional heat source derived from the solar wind is necessary to account for the time variations in the temperature of the

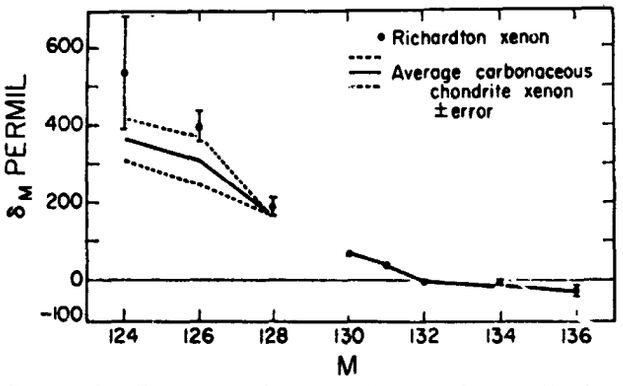
*Published in *Physics Today*, :19–22, February 1964.



SKETCH 1.—Ultraviolet absorption by various atmospheric constituents, as presented by L. V. Berkner of the Southwest Center for Advanced Studies. The absorption beyond 2000 angstroms is almost entirely due to ozone.

upper atmosphere. He also noted that if the earth's magnetic field was at one time significantly stronger, as indicated by measurements of ancient baked clay, the interaction between the field and solar plasma could increase the heating of the upper atmosphere. This could account for a greater rate of escape of helium, averaged over a long time scale.

R. O. Pepin discussed some of the recent work done by J. R. Reynolds and his group at the University of California, Berkeley, in the field of "xenology"—the study of the isotopic abundances of the xenon isotopes. These studies are providing considerable information about the early history of the solar system; for example, the anomalously high abundance of ^{129}Xe resulting from the decay of primordial ^{129}I allows us to obtain the formation ages of meteorites. He pointed out that the relative abundances of the xenon isotopes differ for different types of meteorites and, in particular, the isotopic abundances of the xenon isotopes in the earth's atmosphere differ from the abundances found in meteorites. The problem is to account for all these anomalies.



SKETCH 2.—Comparison between the general anomalies in xenon from the Richardton stone meteorite (an ordinary chrondrite) and in xenon from carbonaceous chondrites.

In this respect, Pepin cited the work of W. B. Clarke, who conducted heating experiments on irradiated uranium oxide samples. Clarke found that the xenon evolved at low temperatures was isotopically different from that evolved at high temperatures. The inference is that the isotopic anomalies may be due to temperature fractionation effects on a fissionogenic component of the xenon gas.

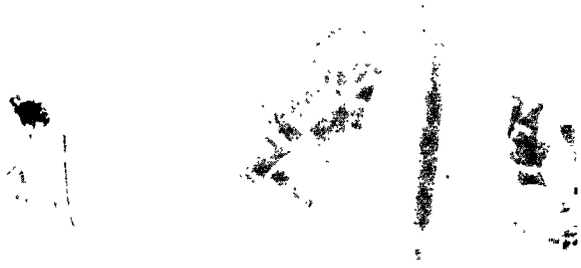
Among the studies done by the Berkeley group was an analysis of some deep-seated terrestrial rocks. They found that the xenon evolved from these rocks contained a fission component resulting from uranium spontaneous fission. This suggests the feasibility of a U-Xe dating method, to be used in conjunction with other established dating techniques.

Peter Signer of the University of Minnesota showed that the abundances of primordial gases in meteorites are generally similar to those in the atmosphere or in the sun only when there is a great amount of this gas in the meteorites. If the amount of gas is small, then a very large degree of fractionation can occur among the elements.

In order to understand the composition of the earth's atmosphere, one must be able to explain the differences between the isotopic abundances of xenon isotopes in meteorites and in the earth. A. G. W. Cameron of the Goddard Institute for Space Studies interpreted the differences in abundances of the light "shielded" xenon isotopes as resulting from neutron capture in the sun during the deuterium-burning stage of early solar history. However, this interpretation requires that the bulk of the xenon in the earth's atmosphere should have once been in the sun, and that it has since been captured by the earth from the solar wind.

The composition of Mercury's atmosphere raised considerable discussion among participants at the conference. Mercury is the smallest planet, and, because of its small size, has been generally believed to have no atmosphere, or a very tenuous one at best. George Field of Princeton University discussed measurements of the differential polarization of scattered light from Mercury made by Audouin Dollfus of the Observatoire de Paris, Meudon, France, who had interpreted them as showing the presence of a small atmosphere.

Field suggested that the atmosphere of Mercury may be composed of a considerable amount of radiogenic argon which escaped from the interior of the planet. The observed atmosphere would require that Mercury had outgassed to about the same extent as the earth. However, if the atmospheric constitution is to be nearly pure argon, the temperature of the upper atmosphere should not exceed 1400 degrees. Above this temperature, evaporative escape would take place. A problem arises in that theoretical estimates place the temperature of an upper atmosphere of pure argon at well above 1400 degrees, since argon is a very inefficient radiator of electromagnetic energy. It was agreed that much more study is needed for an understanding of the manner in which an argon atmosphere—if it exists—could be maintained on Mercury.



SKETCH 3.—Seated in the cabin of his balloon observatory, Audouin Dollfus of the Observatoire de Paris prepares for a flight in which he attempted to detect extraterrestrial water.

Important new data concerning the presence of water vapor on Mars and Venus was presented by Audouin Dollfus. Reporting on observations made with a specially designed telescope which he carried aloft in a balloon, and on observations made from an elevated mountain observatory, Dollfus said he measured 0.01 gm/cm² of water vapor for Venus above the cloud-top level.

For Mars, Dollfus found 0.02 gm/cm² of water vapor, enough to cover that planet to a depth of a fifth of a millimeter. This is five times as much as had been generally estimated. The disagreement between Dollfus' observations and those of several other scientists who give lower values for the water-vapor content of the Martian atmosphere raised sharp debate at the conference. Participants concluded that various experiments may have been subject to unforeseen sources of error, and that it is extremely important to make new measurements.

H. D. Holland of Princeton University suggested that pools of molten sulphur, rather than oceans of water, will be found on the surface of Venus. When that planet was formed, much less water was accreted than when the earth was formed, he suggested. The main substance ejected from Venusian volcanoes probably is sulphur dioxide, which would react with carbon

monoxide to form liquid sulphur on the surface.

Thomas Gold of Cornell University suggested that some of the depressions on the moon's surface may be regions of collapse caused by underground rivers which result from an "outgassing" process taking place in the moon.

He reasoned that if there is water in the interior of the moon, radioactivity should produce enough heat to vaporize it. The water vapor would then seek to rise from the interior to the surface, but it would be trapped by a permafrost or ice layer which must exist below the surface at a depth of 150 feet. However, large meteorites striking the lunar surface could form craters extending below the ice layer; the water in the area of the explosion would then flow underground towards the crater. The extensive area that is drained might then collapse, forming the rills that are a common feature in moon photographs. Gold pointed out that these rills or depressions, converging on craters in flat ground from many angles, could hardly be caused by stresses from the craters themselves. Unlike the stress pattern, the rills do not converge in a straight line, but zigzag—sometimes for hundreds of miles.

Gold also suggested that Venus may possibly be covered by water which has escaped from the interior of the planet. The consensus among scientists has been that if there is water on Venus, it is in the planet's atmosphere and does not exist in appreciable amounts. However, Gold cautioned against ruling out the possibility of water on the surface. His contention was that it is too difficult to explain what happened to the water that must have been present on Venus at some time in its past if this were not the case.

Venus resembles the earth closely in many of its properties. Hence, Gold reasoned that water should have been outgassed from Venusian volcanoes to the same extent that this outgassing had occurred on earth. If Venus' surface temperature is somewhat less than that measured by instruments on Mariner II (perhaps a little less than 600° K), then water would be in equilibrium with steam. Hence, Venus should possess a very massive lower atmosphere composed primarily of steam.

Discussing the atmosphere of Venus, [illegible] Sagan of Harvard University reviewed many of [illegible]eat uncertainties in the measured parameters [illegible]e planet. He also discussed some of the chemical equilibria that could influence the composition of gas in the Venusian atmosphere.

Sagan noted that the planet must be extremely hot at the surface, a temperature of perhaps 650° K on the dark side and about 750° K on the bright side. The pressure at the base of the Venusian atmosphere appears to be at least 30 atmospheres, which would indicate that there may be an appreciable amount of water vapor. Nevertheless, because of the high temperatures, the amount of water in the Venusian atmosphere would be enormously less than that in the earth's oceans. Hence, Sagan disagreed with Gold's suggestion that an approximation of the Venusian atmosphere could be obtained by heating the surface of the earth.

Despite the relatively large amount of observational data on Mars, there is still a considerable degree of uncertainty about the abundances of the planet's atmospheric constituents, according to Richard M. Goody of Harvard University. The amount of carbon dioxide could be reliably determined if the atmospheric pressure at ground level were known, but this is uncertain by at least a factor of two. The amount of oxygen is probably very small. One clue to the abundance of oxygen in the Martian atmosphere might be obtained by detecting that element's allotrope, ozone. This raises a problem, however. If there is oxygen on Mars, ozone should be present even at the ground level. Berkner pointed out that ozone formed near the ground may react with anything that can be oxidized, unless the entire surface of Mars is sufficiently covered by an oxide layer that further oxidation is impossible. It is possible that the continual weathering by wind on Mars would expose fresh rock surfaces and that the subsequent loss of ozone would eventually result in a very small amount of oxygen in the atmosphere.

P. J. E. Peebles of Princeton University presented new calculations showing that Jupiter and Saturn have approximately the same composition, predominantly hydrogen and helium, and that the ratio of hydrogen to helium is similar to that of the sun. Using a variety of reasonable values for different mixtures of hydrogen and helium—and with a core of heavier elements near the center of the planet—Peebles found that a helium-hydrogen ratio by number of 0.075 to 0.08 for Jupiter, and

perhaps twice this amount for Saturn, was required in order to reproduce the mechanical properties of these planets. The amount of heavier elements has not been well determined, but it is only a few percent by mass.

Rupert Wildt of Yale University Observatory summarized the results obtained by H. Spinrad of the Jet Propulsion Laboratory, who showed that at times there are clouds of ammonia in the upper atmosphere of Jupiter which have velocities as high as four kilometers per second relative to the underlying atmosphere. This has become known as the "Spinrad effect." Spinrad has also determined some preliminary values for the composition of the Jovian atmosphere. His work indicates that the ratio of carbon to hydrogen on Jupiter is greater than that in the sun, and that it is higher still in the atmosphere of Saturn.

See N64-33210

THE RED LINE OF ATOMIC OXYGEN IN THE DAY AIRGLOW*

A. DALGARNO† AND JAMES C. G. WALKER

It is argued from the observations of the red line that the rate coefficient for deactivation of the excited oxygen atoms in collisions with molecular oxygen cannot be much less than 10^{-10} cm^3 sec^{-1} and that the solar flux in the region of 1450Å cannot be as large as the reported measured values. It is further argued that photodissociation of molecular oxygen and recombination of molecular ions are inadequate to explain the observations and that an additional mechanism is operative at high altitudes. The mechanisms of fluorescent excitation and of non-thermal excitation by photoelectrons are briefly examined and it is concluded that although the latter may be a significant source of excited atoms, neither mechanism explains the observed variability. It is suggested that the variability in red line intensities is associated with the presence of hot thermal electrons and sample calculations are presented of the altitude profiles that may result from this source.

1. INTRODUCTION

From a theoretical survey, Bates and Dalgarno (1954) concluded that the red line of atomic oxygen at 6300 Å is one of the strongest features of the dayglow spectrum, and a more quantitative study of the red line by Brandt (1958) yielded an intensity of 50 kilorayleighs (kR), about half arising from photodissociation of molecular oxygen in the Schumann-Runge continuum

$$O_2+h\nu\rightarrow O(^3P)+O(^1D) \qquad (1)$$

and about half arising from dissociative recombination of the molecular ions O_2^+

$$O_2^++e\rightarrow O(^1D)+O'. \qquad (2)$$

The ground-based measurements of the red line in the dayglow by Noxon and Goody (1962) and by Jarrett and Hoey (1963) yielded intensities comparable with those predicted by Brandt, but other ground-based measurements by Noxon (1963) and rocket-based measurements by Zipf and Fastie (1963) and by Wallace and Nidey (1964) have shown that the intensity of the red line is often much less than the predicted value. This variation in intensity is significant for it suggests that the mechanisms leading to the emission of the red line have not been correctly identified and described.

2. MODEL ATMOSPHERES

To describe the structure of the neutral component of the upper atmosphere, we have employed the analytic representation of Bates (1959) which is specified by the temperature, temperature gradient and composition at a reference altitude, z_0, and by the exospheric temperature, $T(\infty)$. For the atmospheric parameters, we adopted the values

$$z_0=120 \text{ km}, \left.\frac{dT}{dz}\right|_{z_0}=20\text{K/km}, \quad T(z_0)=350\text{K}$$

and for the number densities of atomic oxygen, molecular oxygen, and molecular nitrogen at 120 km the alternative sets (i) $n(O)=5\times10^{10}$ cm^{-3}, $n(O_2)=1\times10^{11}$ cm^{-3}, $n(N_2)=5\times10^{11}$ cm^{-3}, which we shall describe as the high O_2 atmosphere, and (ii) $n(O)=5\times10^{11}$ cm^{-3}, $n(O_2)=2\times10^{10}$ cm^{-3}, $n(N_2)=5\times10^{11}$ cm^{-3}, which we shall describe as the low O_2 atmosphere. The models appear to span the possible atmospheres (Nicolet, 1961; Harris and Priester, 1962; Pokhunkov, 1963; Jursa, Nakamura and Tanaka, 1963; Schaefer, 1963; Hall, Schweizer, and Hinteregger, 1963; Nier, Hoffman, Johnson, and Holmes, 1964).

The mass density distribution corresponding to the low O_2 atmosphere with $T(\infty)=750$K and to

*Published in *Journal of the Atmospheric Sciences*, 21(6):463–474, September 1964.

†Queens University, Belfast, Northern Ireland.

high O_2 atmospheres with $T(\infty)=750$K and 1000K are compared in Fig. 1 with the density distributions derived from satellite drag data (King-Hele, 1963; Jacchia and Slowey, 1964; Bryant, 1964). The red line observations that we shall discuss were carried out in 1962 and 1963 and except when otherwise noted our results refer to the low O_2 atmosphere with $T(\infty)=750$K, an atmosphere in harmony with the drag data and an atmosphere more easily reconciled with observations of the red line than are the high O_2 atmospheres. The number density, temperature, and mass density distributions corresponding to the low O_2 atmosphere with $T(\infty)=750$K are presented in Table 1.

At altitudes below 120 km we used an atmosphere similar to the model of Jastrow and Kyle (1961).

3. PHOTODISSOCIATION IN THE SCHUMANN-RUNGE CONTINUUM

At wavelengths between 1300 Å and 1750 Å, ultraviolet radiation is strongly absorbed in the Schumann-Runge continuum of molecular oxygen

$$O_2(X^3\Sigma_g^-)+h\nu\rightarrow O_2(B^3\Sigma_u^-)\rightarrow O(^3P)+O(^1D), \quad (3)$$

the process providing a substantial source of $O(^1D)$

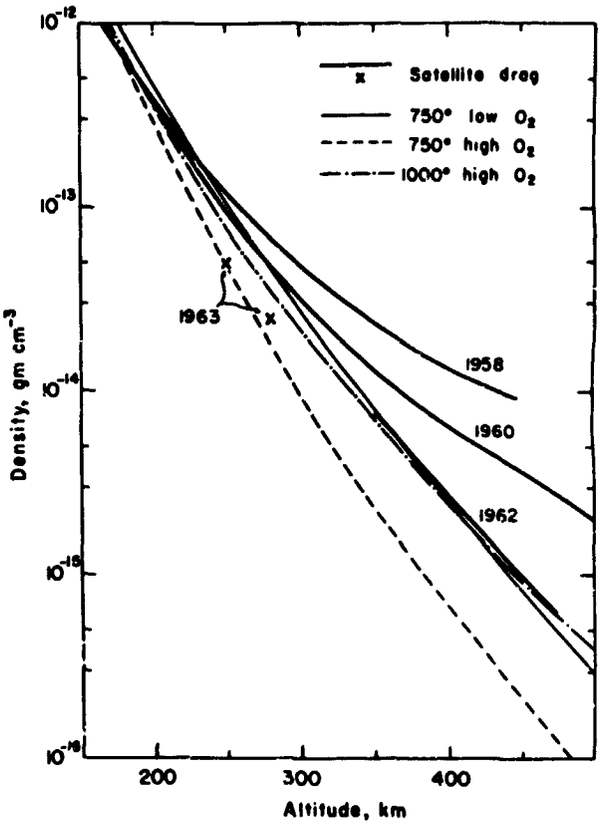


FIGURE 1.—Comparison of the model atmospheres with mid-year daytime maximum satellite drag densities (King-Hele, 1963; Jacchia and Slowey, 1964; Bryant, 1964).

TABLE 1.—*750° low O_2 model atmosphere.*

Altitude km	Temperature °K	Density gm cm^{-3}	Concentration, cm^{-3}				Emission probability, p, for $\beta=10^{-10}$ cm^3 sec^{-1}
			O	O_2	N_2	e	
120	350	2.96(−11)*	2.00(11)	2.00(10)	5.00(11)	1.25(5)	3.43(−3)
130	507	1.04(−11)	9.05(10)	5.94 (9)	1.65(11)	1.43(5)	1.14(−2)
140	602	5.26(−12)	5.51(10)	2.61 (9)	7.87(10)	1.70(5)	2.55(−2)
150	660	3.11(−12)	3.78(10)	1.35 (9)	4.37(10)	2.05(5)	4.79(−2)
160	695	1.98(−12)	2.76(10)	7.55 (8)	2.61(10)	2.53(5)	8.15(−2)
170	717	1.33(−12)	2.08(10)	4.41 (8)	1.62(10)	3.13(5)	1.30(−1)
180	730	18(−13)	1.59(10)	2.64 (8)	1.04(10)	3.80(5)	1.94(−1)
190	737	6.48(−13)	1.24(10)	1.61 (8)	6.70 (9)	4.38(5)	2.74(−1)
200	742	4.65(−13)	9.66 (9)	9.90 (7)	4.37 (9)	5.30(5)	3.63(−1)
225	748	2.14(−13)	5.30 (9)	3.01 (7)	1.54 (9)	7.65(5)	5.71(−1)
250	749	1.04(−13)	2.95 (9)	9.30 (6)	5.52 (8)	1.01(6)	6.87(−1)
275	750	5.32(−14)	1.65 (9)	2.91 (6)	2.00 (8)	1.25(6)	7.32(−1)
300	750	2.81(−14)	9.27 (8)	9.21 (5)	7.30 (7)	1.37(6)	7.50(−1)
350	750	8.36(−15)	2.97 (8)	9.46 (4)	9.95 (6)	1.09(6)	7.55(−1)
400	750	2.64(−15)	9.68 (7)	1.00 (4)	1.40 (6)	7.00(5)	7.58(−1)
450	750	8.61(−16)	3.21 (7)	1.10 (3)	2.02 (5)	4.50(5)	7.58(−1)
500	750	2.88(−16)	1.08 (7)	1.25 (2)	3.01 (4)	3.00(5)	7.58(−1)

atoms in the upper atmosphere (Bates, 1948). The rate of production of $O(^1D)$ atoms at an altitude, z, and for a solar zenith angle, θ, is given by

$$Q_1(z,\theta) = n(O_2|z)\int_{1350}^{1750} F_\infty(\lambda)\sigma(\lambda)\exp[-\mu(\lambda)]d\lambda \quad (4)$$

where $F_\infty(\lambda)$ is the incident flux, $\sigma(\lambda)$ is the absorption cross section and μ is the attenuation factor,

$$\mu(\lambda) = \sigma(\lambda)\int_{(R+z)\cos\theta}^{\infty} n(Q_2|\{y^2+(R+z)^2\sin^2\theta\}^{\frac{1}{2}})dy, \quad (5)$$

R being the radius of the earth. For the calculation of Q_1 we have used the measurements of $\sigma(\lambda)$ by Metzger and Cook (1964) and we have supposed the solar ultraviolet flux to be that of a black body at a temperature, T_s. The limited data (Purcell, Packer and Tousey, 1960; Detwiler, Garrett, Purcell, and Tousey, 1961; Tousey, 1963) are consistent with a value of 4750K for T_s. The results shown in Figs. 2 to 5 are actually appropriate to a solar temperature of 4500K, but they may all be scaled to a temperature of 4750K by multiplying by a factor of 2.9, the loss of accuracy being negligible.

Values of the production rates of $O(^1D)$ atoms are shown in Fig. 2 for several values of θ. Since the intensity of the red line at 6300 Å in the night airglow is much less than that of the green line at 5577 Å (cf. Chamberlain, 1961a) and since the red line is emitted at a much greater altitude than is the green line (Packer, 1961; Tarasova, 1963), it is clear that severe deactivation of the $O(^1D)$ atoms must occur at the lower altitudes (Bates and

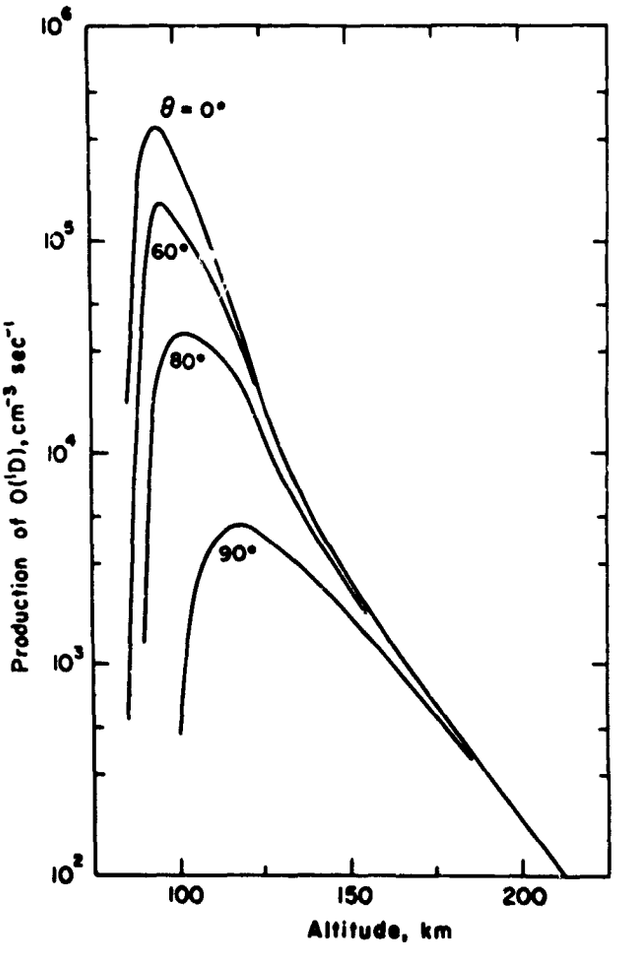


FIGURE 2.—Dependence on altitude of the rate of production of $O(^1D)$ atoms by photodissociation of O_2 for the 750° low O_2 model atmosphere with $T_s = 4500$K. Results are shown for several values of solar zenith angle, θ.

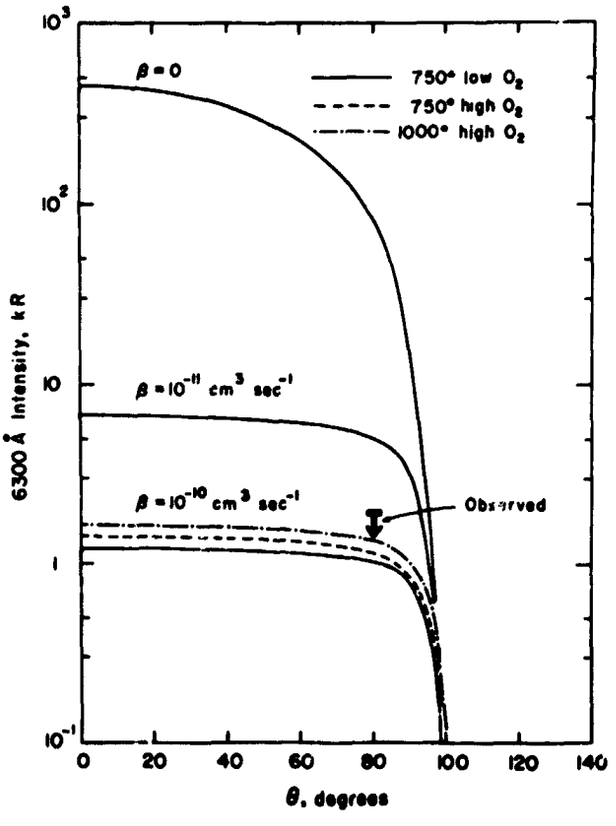


FIGURE 3.—Dependence on solar zenith angle of the intensity which results from the photodissociation of O_2 with $T_s = 4500$K. Results are shown for several atmospheric models and several values of the deactivation coefficient, β. The upper limit imposed by Noxon's (1963) observation is indicated.

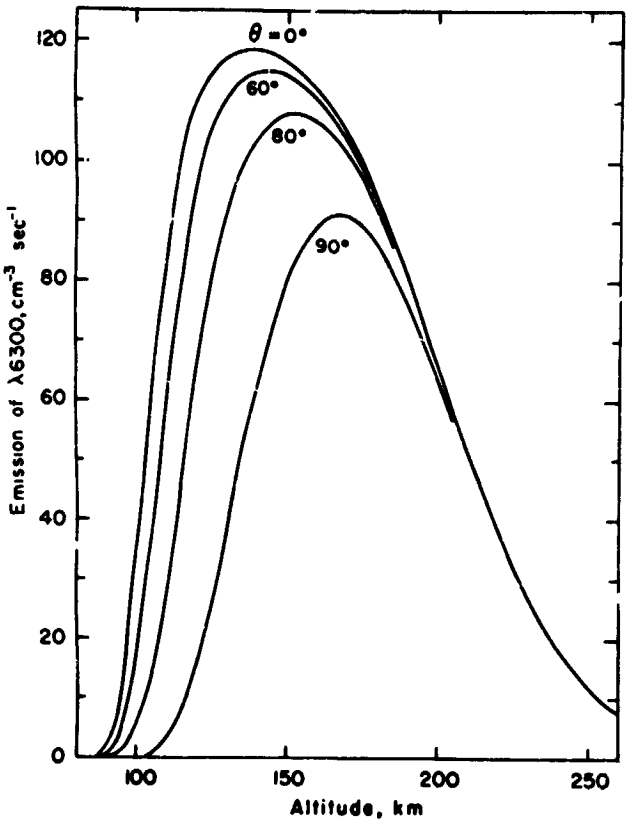


FIGURE 4.—Dependence on altitude of the emission of λ6300 which results from the photodissociation of O_2. The results are for the 750° low O_2 model atmosphere with $T_e = 4500$K and $\beta = 10^{-10}$ cm,³ sec⁻¹.

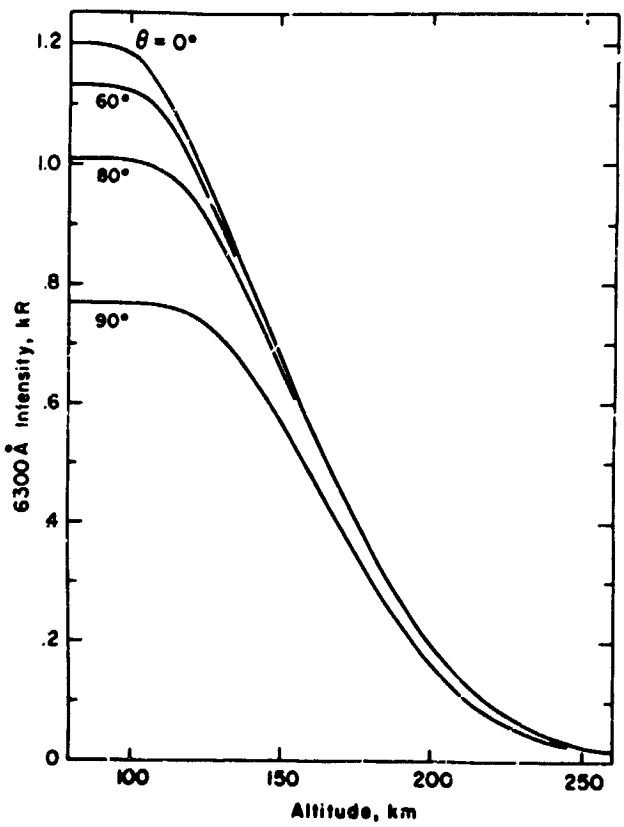


FIGURE 5.—Dependence on altitude of the zenith intensity which results from the photodissociation of O_2. The results are for the 750° low O_2 model atmosphere with $T_e = 4500$K and $\beta = 10^{-10}$ cm³ sec⁻¹.

Dalgarno, 1953; Seaton, 1954). Deactivation by electron impact is negligible (cf. Bates, 1960), and Bates and Dalgarno (1953) and Seaton (1958) have argued that the deactivation probably occurs in collisions with molecular oxygen. With this assumption, Wallace and Chamberlain (1959) have derived from an analysis of auroral observations of the O_2 atmospheric band system a rate coefficient, β, such that

$$4\times10^{-12}\ \text{cm}^3\ \text{sec}^{-1} < \beta < 10^{-10}\ \text{cm}^3\ \text{sec}^{-1}.$$

Despite the fact that the mechanism does not conserve spin, it has been suggested that deactivation of $O(^1D)$ by molecular nitrogen is comparable in efficiency (DeMore and Raper, 1964) and, although the arguments are not convincing, the possibility cannot be excluded. Because the N_2 density profile closely parallels the O_2 density profile, the inclusion of N_2 deactivation would involve no important modification of our subsequent arguments and we shall largely ignore it.

During the day, deactivation by electron impact is still negligible but there occurs the possibility that $O(^1D)$ atoms may be removed by resonance absorption of solar radiation.

There appears to be only one allowed transition, the wavelength of which coincides with a strong solar line, and that is the transition to the $O(2p^3 3s\,^1D)$ state at 1152 Å, the solar flux of λ1152 being about 10^9 photons cm^{-2} sec^{-1} (Hinteregger, 1961; Hinteregger and Watanabe, 1962). Despite the substantial flux and the wavelength coincidence, the mechanism can be discounted because the only allowed transition from the upper state returns the atom to the original ^1D state.

The probability that an $O(^1D)$ atom at altitude, z, will emit a 6300 Å photon is given by

$$p(z) = \left\{1+\frac{A_{6364}}{A_{6300}}+\frac{\beta n(O_2|z)}{A_{6300}}\right\}^{-1} = \{1.32+1.45\times10^2\beta n(O_2|z)\}^{-1}, \quad (6)$$

where A_λ is the radiative transition probability for

the line of wavelength λ (cf. Chamberlain, 1961b). Values of $p(z)$ are given in Table 1 for $\beta = 10^{-10}$ cm^3 sec^{-1}.

The intensities of red line emission for a solar temperature of 4500K and various values of the deactivation coefficient, β, are shown in Fig. 3 as a function of solar zenith angle. The intensity for an overhead sun varies from 433 kR in the absence of deactivation to 1.2 kR if β has a value of 10^{-10} cm^3 sec^{-1}, which is close to the maximum possible for a thermal collision. Fig. 3 includes also the intensities corresponding to two other model atmospheres to show the lack of sensitivity to the details of the number density distribution.

To determine β we note that Noxon (1963) has reported that the intensity of the dayglow red line at noon at Fort Churchill was less than 2 kR on 15 December, 1962, a result in harmony with an observation of Wallace and Nidey (1964). [illegible] Noxon was observing with the sun at a zenith angle of about 80°, it follows from Fig. 3 that, for the low O_2 atmosphere with $T(\infty) = 750$K, $\beta \geq 1.7 \times 10^{-10}$ cm^3 sec^{-1} if $T_e = 4750$K and that $\beta \geq 3.7 \times 10^{-11}$ cm^3 sec^{-1} if $T_e = 4500$K. If deactivation in collisions with N_2 were significant, the implied upper limit on T_e would be increased.

Bates and Dalgarno (1954) pointed out that photodissociation in the Hartley continuum of ozone is an intense source of O(^{1}D) atoms below 100 km, and Cadle (1964) has predicted an intensity of 2 kR for the resulting red line emission, his calculations following Brandt (1958) in using a value of 10^{-12} cm^3 sec^{-1} for β. With either of the two limits derived from Noxon's observations, the actual red line emission below 100 km must be negligible.

Adopting a value of 10^{-10} cm^3 sec^{-1} for β, we show in Figs. 4 and 5, respectively, the emission rate and zenith intensity as functions of altitude for a number of solar zenith angles.

4. IONIC RECOMBINATION

There is usually a good correlation between the nocturnal intensity of the red line and the critical frequency of the F region (Barbier, 1957; Barbier and Glaume, 1962; Barbier, Roach, and Steiger, 1962; Carman and Kilfoyle, 1963; Barbier, 1964), supporting the belief that the main excitation mechanism is dissociative recombination. Originally the molecular ion involved was identified as O_2^+ but it is apparently now believed that NO^+ is the major participant (cf. Wallace and Nidey, 1964). The proposed reaction

$$NO^+(X^1\Sigma^+) + e \rightarrow N(^4S) + O(^1D) \qquad (7)$$

fails to conserve spin. We therefore concentrate on O_2^+ recombination as a source of O(^{1}D) atoms. The ionospheric and airglow data presented by Barbier (1964) indicate that the rate of electron removal in the nocturnal ionosphere exceeds the rate of emission of 6300 Å photons by a factor of about 8, consistent with our suggestion that (7) is not an important mechanism for O(^{1}D) excitation.

The nocturnal intensity of the red line is between 50R and 100R (cf. Chamberlain, 1961a) and considerable daytime enhancement is to be expected. The calculations by Brandt (1958) were based upon a simple sequence of ion-removal reactions and his results were sensitive to the adopted reaction rates. It is now clear that the actual reaction paths in the ionosphere are much more complicated (Nicolet and Swider, 1963; Dalgarno, 1964a). Serious uncertainties remain, the consequences of which can be largely avoided by assuming that local equilibrium prevails between ion production and removal rates, an assumption which is valid below 300 km throughout most of the day but fails at greater altitudes because of the increasing importance of diffusion.

The ion production rates, $q(O^+)$, $q(N_2^+)$, and $q(O_2^+)$ cm^{-3} sec^{-1} have been calculated for a wide range of model atmospheres and solar zenith angles by Dalgarno and McElroy (1964a) and some of their results are reproduced in Fig. 6.

In order to estimate what fractions of the ions, O^+, N_2^+, and O_2^+, lead to (O^{1}D) atoms, it is necessary first to determine what fractions of the ions O^+ and N_2^+ are converted to O_2^+ ions or possibly to NO^+ ions. The fractions depend upon the electron density distribution and we have adopted the distribution measured by Brace, Spencer, and Carignan (1963) in a quiet ionosphere and included in Table 1. The intensity of the red line is insensitive to the details of the adopted electron density profile.

The reactions which remove N_2^+ ions have been discussed by Nicolet and Swider (1963). They are

$$N_2^+ + e \rightarrow N' + N'' \qquad (8)$$

$$N_2^+ + O_2 \rightarrow N_2 + O_2^+ \qquad (9)$$

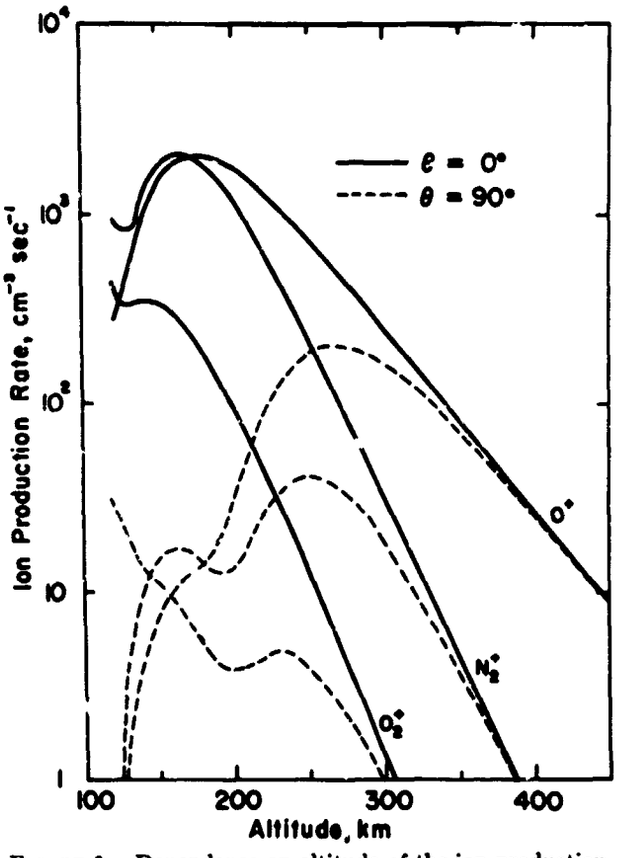


FIGURE 6.—Dependence on altitude of the ion production rates for the 750° low O_2 model atmosphere.

$$N_2^+ + O \rightarrow N_2 + O^+ \qquad (10)$$

$$N_2^+ + O \rightarrow NO^+ + N. \qquad (11)$$

The rate coefficient of (8) has been measured as 2×10^{-7} cm^3 sec^{-1} (cf. Biondi, 1964) and of (9) as 2×10^{-10} cm^3 sec^{-1} (Fite, Rutherford, Snow, and van Lint, 1962). Rate coefficients of 2×10^{-11} cm^3 sec^{-1} for (10) and (11) have been derived by Norton, Van Zandt, and Denison (1963) and by Whitten and Poppoff (1964) from ionospheric data. The analyses of ionospheric data impose arbitrary temperature variations on the rates of the possible processes and the derived coefficients have little quantitative significance. However, the measured concentrations of N_2^+ apparently require that one or both of (10) and (11) be rapid so we assume that each of them has a rate coefficient of 1×10^{-11} cm^3 sec^{-1}. Since, in the atmosphere, (10) is usually followed by

$$O^+ + N_2 \rightarrow NO^+ + N \qquad (12)$$

the distinction between (10) and (11) is unimportant for our purposes.

The O^+ ions are removed by the ion-atom interchange reactions, (12) and

$$O^+ + O_2 \rightarrow O_2^+ + O \qquad (13)$$

(Bates, 1955). For our calculations we have adopted a rate coefficient of 1×10^{-11} cm^3 sec^{-1} for both (12) and (13) (Sayers and Smith, 1964) but the red line intensity depends only upon the ratio of the rates of the two reactions.

The ion, O_2^+, disappears by dissociative recombination

$$O_2^+ + e \rightarrow O' + O'' \qquad (14)$$

throughout the altitude region in which (14) might provide a significant source of emission of the red line. Thus, if f is the probability that (14) produces an $O(^1D)$ atom, the rate of production is given by

$$Q_2 = f\left\{ q(O_2^+) + q(N_2^+)\left[\alpha_9 n(O_2) + \frac{\alpha_{10} n(O)}{1+\frac{\alpha_{12} n(N_2)}{\alpha_{13} n(O_2)}}\right] \times [\alpha_8 n(e) + \alpha_9 n(O_2) + (\alpha_{10}+\alpha_{11}) n(O)]^{-1} + q(O^+)\left[1+\frac{\alpha_{12} n(N_2)}{\alpha_{13} n(O_2)}\right]^{-1}\right\} \qquad (15)$$

The reaction is energetically capable of producing two $O(^1D)$ atoms so that f may be as large as 2. In Fig. 7 $Q_2(z)$ is shown as a function of altitude for several solar zenith angles with f taken equal to 2. The results provide an upper limit to the possible excitation rates consistent with the adopted ultraviolet flux and reaction rates.

Because $n(N_2)$ is much greater than $n(O_2)$ in the altitude region where $q(O^+)$ and $q(O_2^+)$ are comparable, the last term of (15) is important only at high altitudes where the production rates are small. Thus the intensity of the red line is insensitive to the magnitudes of α_{12} and α_{13}. In contrast is the sensitivity to the relative values of α_8 and α_9. Reaction (9) is an important source of O_2^+ ions at the lower altitudes where the production rates are large, whereas the sequence of (10) followed by (13) is important only at high altitudes where the production of $O(^1D)$ is small. Accordingly, we illustrate in Fig. 7 only the sensitivity to the ratio α_8/α_9, the dashed curves of Fig. 7 arising when the value of 2×10^{-10} cm^3 sec^{-1} for α_9 is replaced by zero so that most of the N_2^+ ions disappear by dissociative recombination.

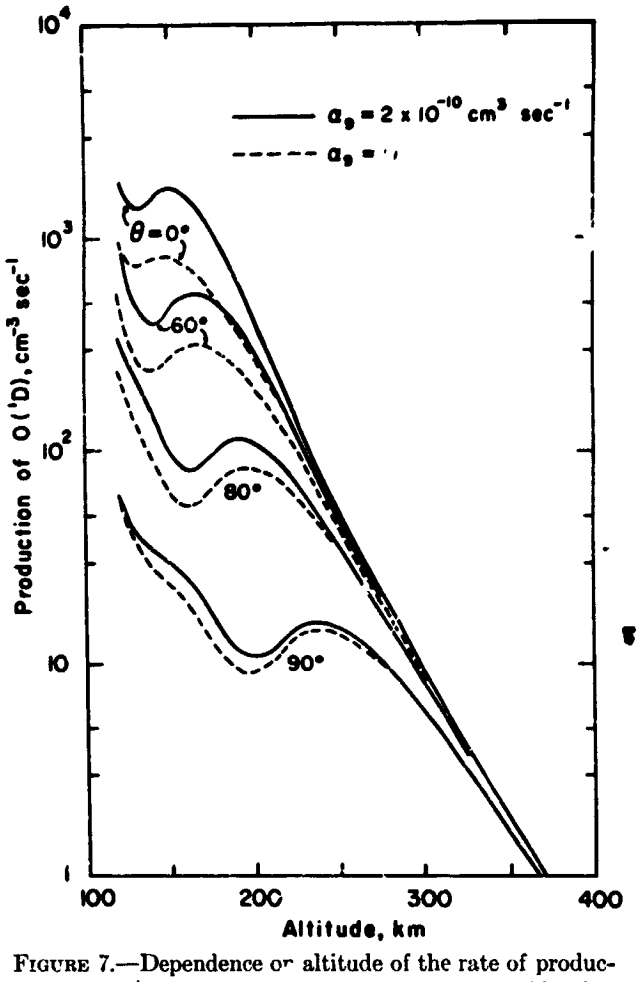


FIGURE 7.—Dependence on altitude of the rate of production of O(¹D) atoms by the dissociative recombination of O_2^+ for the 750° low O_2 model atmosphere.

The possible additional source of O(¹D) atoms provided by the dissociative recombination of NO^+ (7) is shown in Fig. 8 on the assumption that every NO^+ recombination yields an O(¹D) atom. The actual yield is presumably much less than unity.

The rate of emission of 6300 Å photons produced by dissociative recombination of O_2^+ with $\alpha_3/\alpha_9 = 10^3$, $\alpha_8/\alpha_{10} = \alpha_8/\alpha_{11} = 2 \times 10^4$, $\alpha_{11}/\alpha_{12} = 1$, $\beta = 10^{-10}$ cm^3 sec^{-1}, and $f = 2$ is shown in Fig. 9. The corresponding zenith intensities are shown in Fig. 10.

Fig. 11 shows the sensitivity of the predicted intensities to the atmospheric model.

5. PHOTODISSOCIATION AND IONIC RECOMBINATION

A comparison of Figs. 3 and 11, and 5 and 10 shows that the contributions of photodissociation and of ionic recombination to the zenith intensity of the red line may be comparable, depending upon the values of T_s and f, but there are significant differences in the variations with solar zenith angle and with altitude. The zenith intensity has been measured as a function of altitude by Zipf and Fastie (1963) at Wallops Island on 7 May 1964 with the sun at a zenith angle of 60°. We have attempted to reproduce the observational data of Zipf and Fastie by combining the contributions of photodissociation, shown in Fig. 5, and of ionic recombination, shown in Fig. 10, and regarding the solar temperature, T_s, and the excitation probability, f, as disposable parameters subject to $f \leqslant 2$.

Fig. 12 shows the best fits we have been able to obtain by this procedure for three model atmospheres. The results for the 1000° high O_2 atmosphere with $T_s = 4660$K and $f = 2$ agree closely with the observations, the total zenith intensity being 4.8 kR of which 3.1 kR is due to photodissociation and 1.7 kR to ionic recombination. The corresponding intensities appropriate to a zenith angle of 80° are, respectively, 2.7 kR and 1.0 kR,

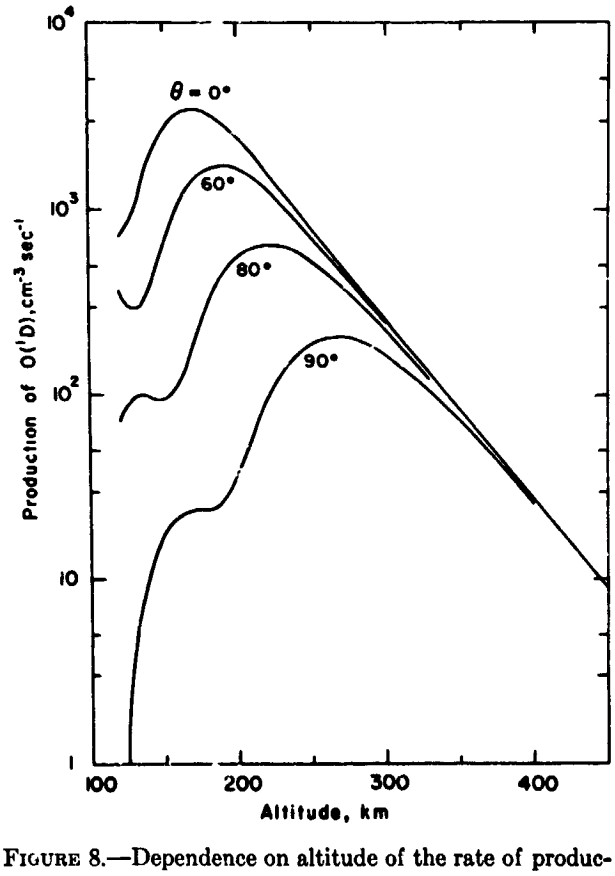


FIGURE 8.—Dependence on altitude of the rate of production of O(¹D) atoms by the dissociative recombination of NO^+ for the 750° low O_2 model atmosphere.

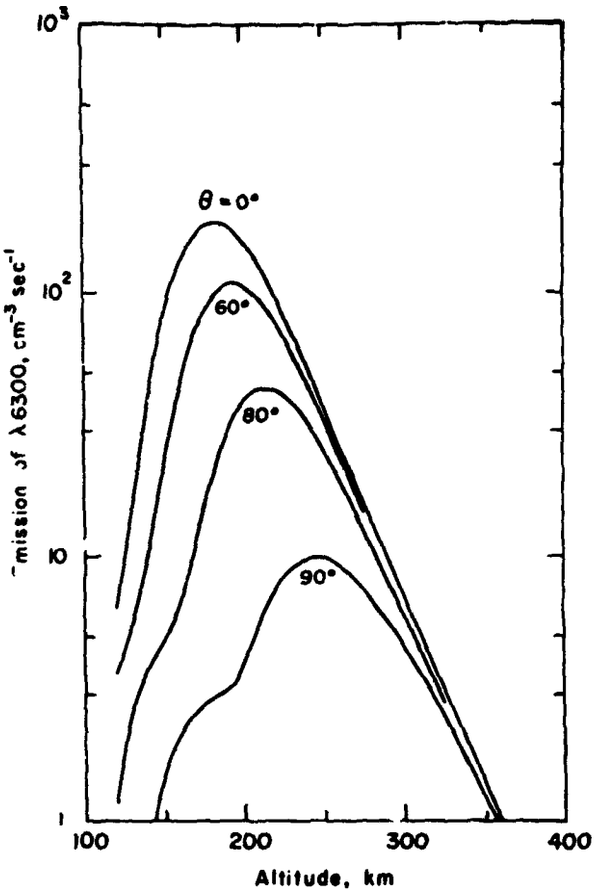


FIGURE 9.—Dependence on altitude of the emission of λ6300 which results from the dissociative recombination of O_2^+ for the 750° low O_2 model atmosphere.

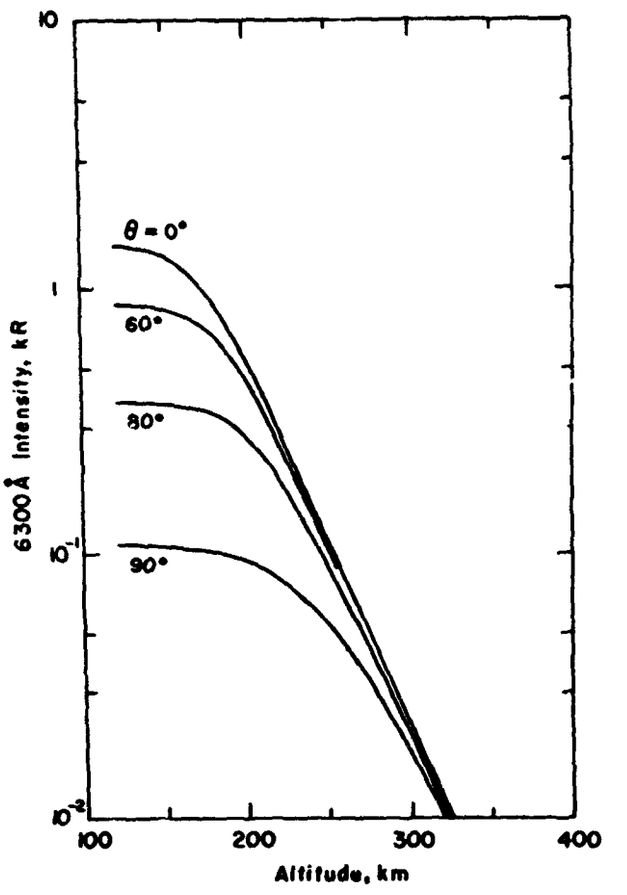


FIGURE 10.—Dependence on altitude of the zenith intensity which results from the dissociative recombination of O_2^+ for the 750° low O_2 model atmosphere.

the sum of which exceeds the upper limit of 2 kR observed by Noxon (1963) at Fort Churchill at noon on 15 December 1962. In order to be consistent with an upper limit of 2 kR at 80°, it is necessary with the 1000° high O_2 model atmosphere to adopt a solar temperature not greater than 4580K, even with $f=0$. The results for the other models shown in Fig. 12 lead to even greater conflict with Noxon's upper limit.

Temporal variations in the solar ultraviolet flux in the neighborhood of 1216 Å are small (Lindsay, 1963), and there is no evidence to suggest that large variations occur in the flux at the longer wavelengths of the Schumann-Runge continuum. Ignoring the possibility of temporal change in T_s, we have determined the values of T_s which give the largest intensities at 60° consistent with the upper limit of 2 kR at 80°. The intensities as a function of altitude for the three model atmospheres are shown in Fig. 13. All of these models lead to intensities which are markedly lower than the Zipf and Fastie results. The situation cannot be improved by using lower values of f and higher values of T_s because, with a deactivation coefficient, β, of 10^{-10} cm^3 sec^{-1}, the photodissociation contribution is nearly independent of θ as θ varies from 60° to 80°, as shown in Fig. 3.

There are also discrepancies in the detailed dependence of emission on altitude, Zipf and Fastie (1963) finding that the emission rate is 2.6×10^2 cm^{-3} sec^{-1} from 120 km to 190 km, increasing abruptly to 5.0×10^2 cm^{-3} sec^{-1} from 190 km to 220 km. The dependence of emission on altitude would be modified if the deactivation coefficient, β, were to change rapidly with increasing temperature. As Bates (1960) has pointed out, O(^{1}D) atoms produced by dissociative recombination initially possess considerable kinetic energy so that the effective temperature controlling β may be higher than the ambient temperature. The effect

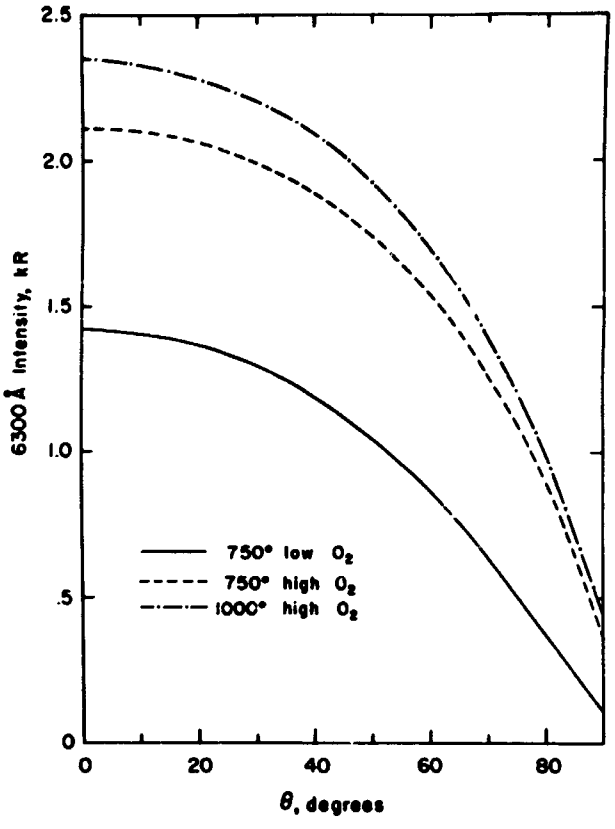


FIGURE 11.—Dependence on solar zenith angle of the intensity which results from the dissociative recombination of O_2^+ for several model atmospheres with $\beta = 10^{-10}$ cm^3 sec^{-1} and $f = 2$.

is probably small since molecular oxygen is a minor constituent at the altitudes of interest and thermalizing collisions with the major constituents, O and N_2, occur more frequently than deactivating collisions with O_2. A rapid change of β with increasing altitude will not, in any case, remove the discrepancy between the predicted intensities of Fig. 13 and the observed intensities. The discrepancy can more plausibly be attributed to an additional excitation mechanism.

6. FLUORESCENT EXCITATION

Bates (1948) and Chamberlain (1958) have shown that resonance scattering of sunlight

$$O(^3P)+\lambda 6300 \rightarrow O(^1D) \qquad (16)$$

does not contribute significantly to the twilight-glow in the red line and it follows from their calculations, but including deactivation, that the yield in the dayglow does not exceed 20R.

Fluorescence through the sequences

$$O(^3P)+\lambda 1304 \rightarrow O(^3S) \qquad (17)$$

$$O(^3S) \rightarrow O(^1D)+\lambda 1641, \qquad (18)$$

or

$$O(^3S) \rightarrow O(^1S)+\lambda 2325 \qquad (19)$$

followed by

$$O(^1S) \rightarrow O(^1D)+\lambda 5577, \qquad (20)$$

may be a more abundant source of dayglow red line emission and incidentally of dayglow green line emission also, for the atmosphere is optically very thick in the triplet centered at λ1304.

The intensity of λ1304 has been measured by Donahue and Fastie (1964) and by Fastie, Crosswhite and Heath (1964). If we assume a Doppler width corresponding to a temperature in the region of 2000K for the line and a value of 4.0×10^{-6} (Garstang, 1961) for the ratio of the transition probabilities for the emission of λ1641 and λ1304 radiation, the measured intensities of λ1304 lead to an estimate of the order of 100R for the red

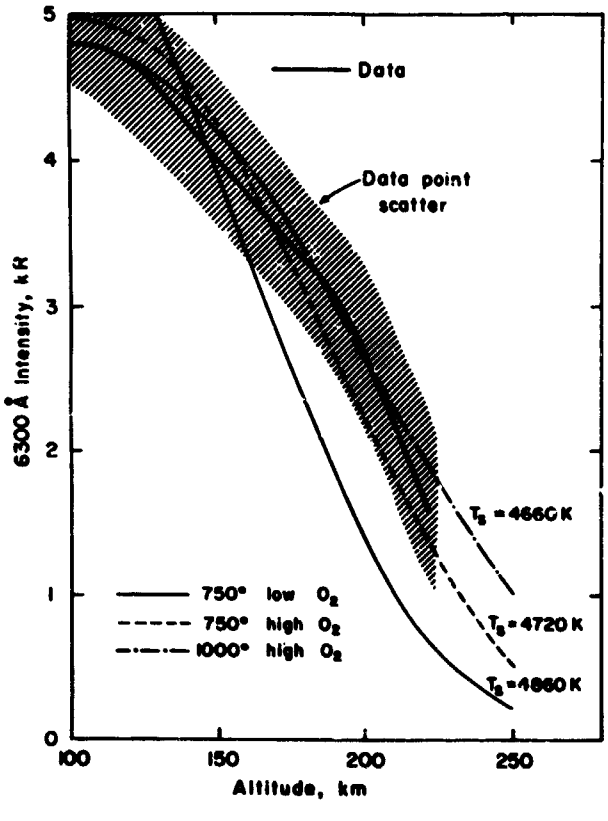


FIGURE 12.—Comparison of the data of Zipf and Fastie (1963) with the theoretical intensities at $\theta = 60°$ which result from photodissociation and ionic recombination. All three models use $f = 2$ and $\beta = 10^{-10}$ cm^3 sec^{-1}.

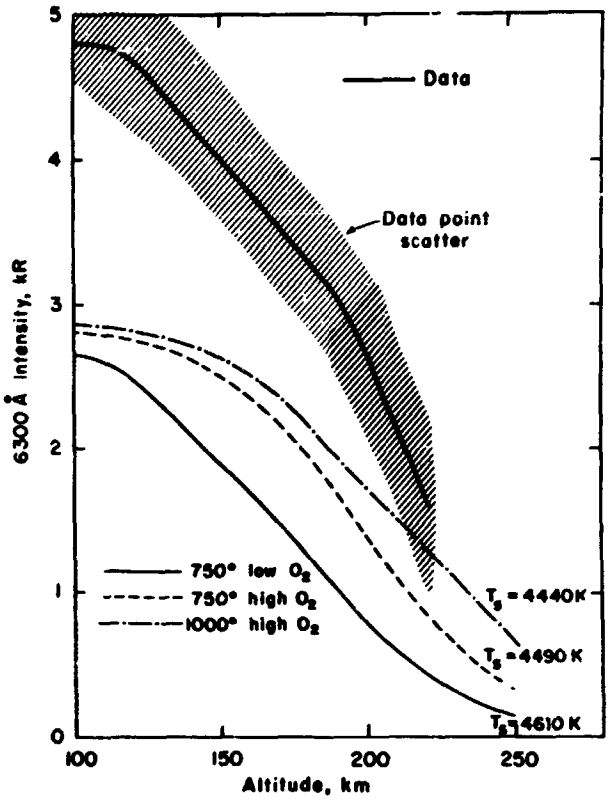


FIGURE 13.—Comparison of the data of Zipf and Fastie (1963) with theoretical intensities at $\theta = 60°$ due to photodissociation and ionic recombination for models which lead to an intensity of 2kR at $\theta = 80°$. The calculations assume $f = 2$ and $\beta = 10^{-10}$ cm^3 sec^{-1}.

line, an estimate which is consistent with the fact that λ1641 radiation has not been detected in the airglow. The contribution of (19) and (20) is much smaller because the ratio of the transition probabilities for the emission of λ2325 and λ1304 is 1.6×10^{-8} (Garstang, 1961).

None of the mechanisms discussed so far explains the great variability of the red line intensity, discovered by Noxon (1964). The electron temperature is markedly higher than the neutral particle temperature in the F region and may well be very variable (Spencer, Brace and Carignan, 1962; Brace, Spencer and Carignan, 1963; Nagy, Brace, Carignan and Kanal, 1963) and the energy source responsible may also be a source of excitation of the red line.

7. ELECTRON IMPACT EXCITATION

7.1 Thermal Excitation

It has been pointed out (Dalgarno, 1964b) that a rise in the temperature of the ambient electron gas will be accompanied by an enhancement of the red line. Spencer, Brace and Carignan (1962) and Brace, Spencer and Carignan (1963) have presented the results of measurements of electron densities and electron temperatures from several rocket flights. If we assume that there is no depletion of the high energy tail of the velocity distribution, the associated red line emission rates may be easily computed using rate coefficients derived from Seaton (1956). The emission rates and intensities as functions of altitude are shown in Figs. 14 and 15 for a deactivation coefficient $\beta = 10^{-10}$ cm^3 sec^{-1}.

In calculating the zenith intensities, it is necessary to estimate the emission at altitudes above those for which electron temperature and density data are available. We have assumed that the electron gas is isothermal and in diffusive equilibrium at these altitudes. The contribution of the high altitudes to the overhead intensity is

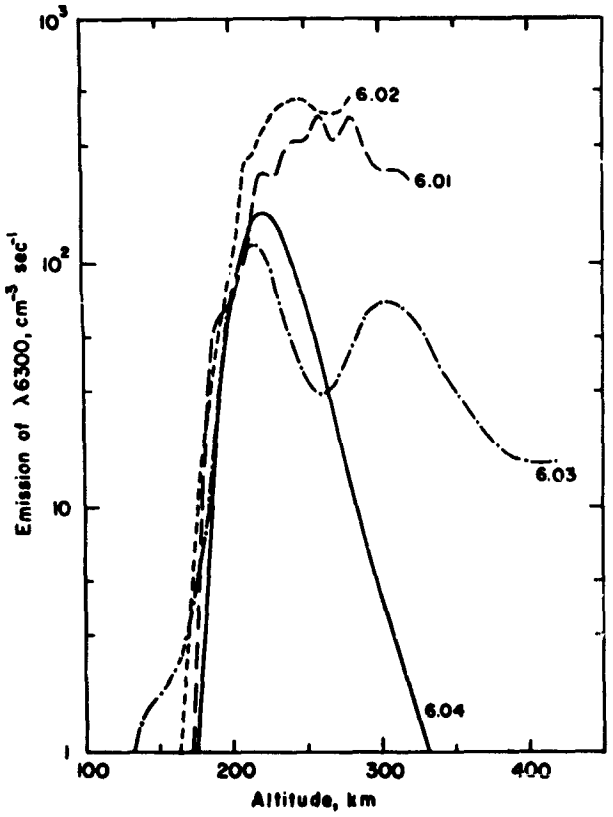


FIGURE 14.—Dependence on altitude of the emission of λ6300 which results from excitation by thermal electrons for four measured profiles of electron temperature and density (Spencer, Brace, and Carignan, 1962; Brace, Spencer, and Carignan, 1963. The results are for the 750° low O_2 model atmosphere with $\beta = 10^{-10}$ cm^3 sec^{-1}.

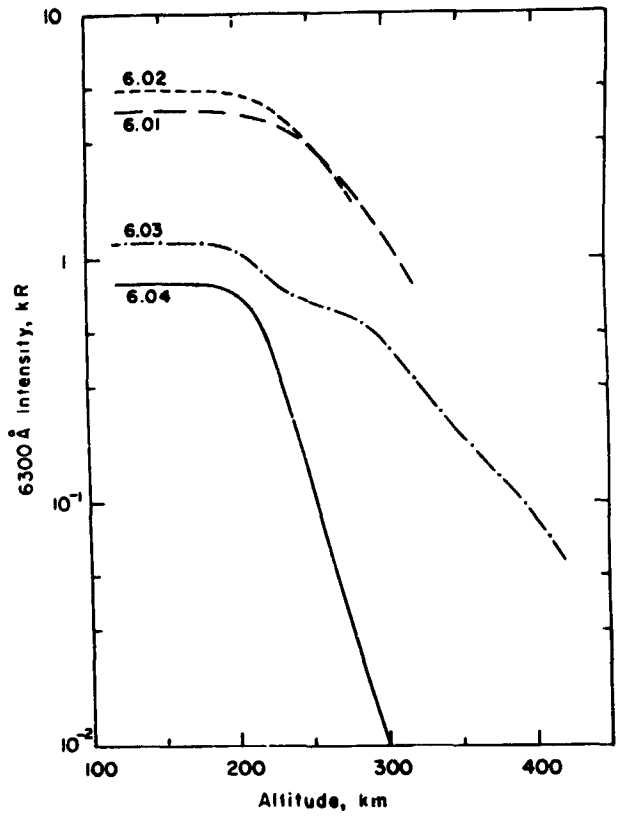


FIGURE 15.—Dependence on altitude of the zenith intensity which results from excitation by thermal electrons for four measured profiles of electron temperature and density (Spencer, Brace, and Carignan, 1962; Brace, Spencer, and Carignan, 1963). The results are for the 750° low O_2 model atmosphere with $\beta = 10^{-10}$ cm^3 sec^{-1}.

about 30 percent for flight 6.02, 20 percent for flight 6.01, 5 percent for flight 6.03, and 0.1 percent for flight 6.04.

Flight 6.01 was fired from Fort Churchill on 16 March 1960, during spread *F* conditions and flight 6.04 was fired from Wallops Island on 3 March 1961, during quiet ionospheric conditions. The red line intensities predicted for thermal excitation are 4.1 kR and 0.79 kR, respectively. The altitude profiles are markedly different, the emission rate for flight 6.04 peaking at 220 km and decreasing rapidly and that for flight 6.01 varying little from 220 km up to at least 320 km.

Flight 6.04 provides an interesting contrast to flight 6.02 which was fired from Fort Churchill on 15 June 1960, also during quiet ionospheric conditions. Flight 6.02 is actually very similar to the spread *F* flight 6.01 with a predicted intensity of 4.9 kR and an emission rate varying little from 225 km up to at least 280 km.

Flight 6.03 was fired from Wallops Island on 3 August 1960, following a period of magnetic disturbance. Below 250 km the results are similar to those for the quiet ionosphere Wallops Island flight 6.04. However, the 6.03 emission rate is enhanced at greater altitudes, with a second maximum at 310 km and a relatively slow decrease above this altitude. The predicted intensity for flight 6.03 is 1.2 kR.

The daytime electron temperature is very sensitive to small disturbances and it appears that the intensity and the altitude dependence of the red line may be still more sensitive. Thus an increase of 10 percent in the electron temperatures measured on flight 6.04 would raise the predicted red line intensity from 0.79 kR to 2.1 kR. The combined profile due to such arbitrary thermal excitation, to Schumann-Runge photodissociation with $T_s = 4610$K and to ionic recombination with $f = 2$

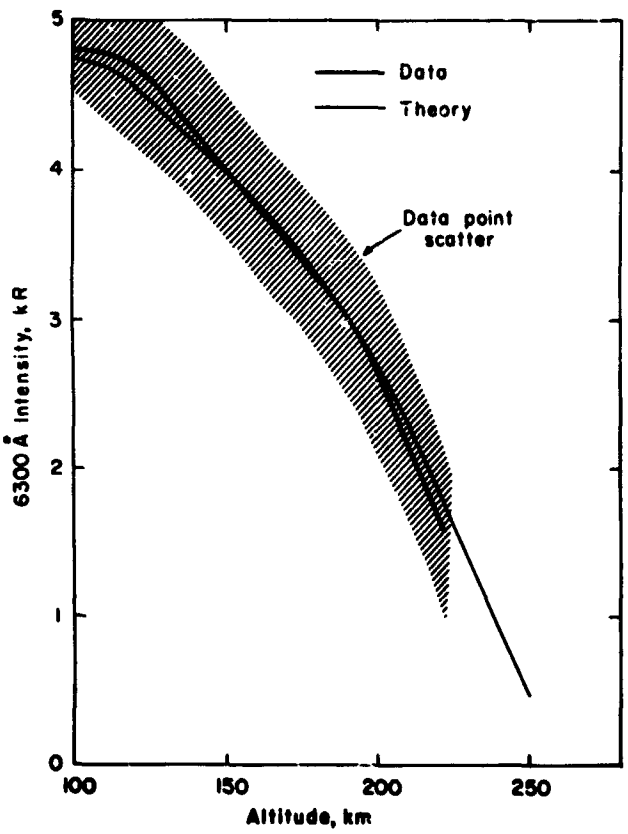


FIGURE 16.—Comparison of the data of Zipf and Fastie (1963) with the theoretical intensity which results from the combination of photodissociation with $T_s = 4610$K, ionic recombination with $f = 2$, and electron excitation given by the profile calculated for flight 6.04 enhanced by a factor of 2.7. The theoretical results are for the 750° low O_2 model atmopshere with $\beta = 10^{-10}$ cm^3 sec^{-1}.

is in harmony with that observed by Zipf and Fastie (1963), as Fig. 16 demonstrates. This model reconciles the observation of Zipf and Fastie with Noxon's (1963) upper limit on the intensity at Fort Churchill if we assume that electron impact excitation was negligible at the time of Noxon's measurement.

The very high intensities of the order of 40 kR observed by Noxon and Goody (1962) and by Jarrett and Hoey (1963) can be ascribed to thermal excitation by electrons with a temperature in the region of 4000K. Such high temperatures have been detected, but only occasionally, on Explorer XVII (Brace and Spencer, 1964) and it is significant that very intense red line emission is also a rare phenomenon (Noxon, 1964).

The Explorer XVII data show an early morning maximum in the electron temperature (Brace and Spencer, 1964), related presumably to the dawn effect (Dalgarno and McElroy, 1964b), and an early morning enhancement of the red line may be a regular feature of its diurnal variation.

Satellite observations (Willmore, Henderson, Boyd, and Bowen, 1964; Brace and Spencer, 1964) show that the electron temperature remains higher than the neutral particle temperature during the night. The temperatures are usually too low to give rise to significant thermal excitation and the intensity associated with the nocturnal rocket flight 6.05 of Brace, Spencer and Carignan (1963) was a fraction of a rayleigh. However, the energy source associated with the heating of the electrons may contribute directly to excitation of the red line and it is interesting to note that Barbier (1964) has suggested that about 20R of red line emission is not directly correlated with the normal ionospheric parameters.

A similar effect may occur during the day.

7.2 Non-Thermal Excitation

A major source of heat during daylight exists in the fast photoelectrons produced by photoionization by solar ultraviolet radiation. The photoelectrons lose energy through a complicated sequence of collision processes, one of which is electron impact excitation of atomic oxygen to the $O(^1D)$ state (Hanson and Johnson, 1961; Hanson, 1963; Dalgarno, McElroy and Moffett, 1963; Dalgarno, 1964b). The resulting altitude profile will be similar to that derived for thermal excitation and the intensity may be of the order of a kilorayleigh in which case the importance of thermal excitation in the interpretation (Fig. 16) of the observed profile (Zipf and Fastie, 1963) must be correspondingly diminished.

8. CONCLUSIONS

Noxon's observation that the intensity of the red line at a zenith angle of 80° is less than 2 kR places severe restrictions upon the contributions from photodissociation of molecular oxygen and from recombination of molecular ions. It appears necessary that the deactivation coefficient for collisions of $O(^1D)$ atoms with O_2 be as large as 10^{-10} cm^3 sec^{-1}, and that the solar flux in the region of 1450 Å be less than the reported values. Then the contributions of photodissociation and of ionic recombination vary from at most 3.4 kR for an overhead sun to at most 1.3 kR for a zenith angle of 90°.

It is difficult to reconcile the limit placed by Noxon with the altitude profile measured by Zipf and Fastie without postulating an additional excitation mechanism effective at high altitude. The additional mechanism may be impact excitation by photoelectrons or it may be thermal excitation by hot electrons. The latter source can tentatively be identified as that responsible for the observed variability of the dayglow intensities. The dayglow red line may therefore provide a means of monitoring the electron temperature and thereby of gaining information about temporal variations in upper atmosphere energy sources. A search for correlations between simultaneous measurements of the red line and the electron temperature would be instructive.

ACKNOWLEDGMENTS

Our study of the emission of the red line in the dayglow was stimulated by Dr. J. F. Noxon's observations and we have enjoyed many discussions with him on the subject. Mrs. O. P. Ko provided valuable assistance in programming and running the computer.

A part of this research was performed while Walker was at Columbia University and was supported by the National Aeronautics and Space Administration under Grant Number NsG–145.

REFERENCES

EARBIER, D., 1957: La lumière du ciel nocturne en été a Tamanrasset. *Compt. Rend.*, **245**, 1559–1561.

———, 1964: Nouvelles observations de la raie rouge du ciel nocturne en Afrique. *Ann. Geophys.*, **20**, 22–33.

———, and J. GLAUME, 1962: La couche ionosphérique nocturne *F* dans la zone intertropicale et ses relations avec l'emission de la raie 6300 Å du ciel nocturne. *Planetary Space Sci.*, **9**, 133–148.

BARBIER, D., F. E. ROACH and W. R. STEIGER, 1962: The summer intensity variation of [01] 6300 Å in the tropics. *J. Res. NBS*, **66D** (1), 145–152.

BATES, D. R., 1948: Theoretical considerations regarding the night sky emission. *Emission Spectra of Night Sky and Aurorae*, London, Physical Society, 21–33.

———, 1955: Charge transfer and ion-atom interchange collisions. *Proc. Phys. Soc. London*, **A68**, 344–345.

———, 1959: Some problems concerning the terrestrial atmosphere above about the 100 km level. *Proc. Roy. Soc. London*, **A253**, 451–462.

———, 1960: The airglow. *Physics of the Upper Atmosphere*, New York, Academic Press, 219–267.

——— , and A. DALGARNO, 1953: The altitudes of the luminous layers in the earth's atmosphere. *J. Atmos. Terr. Phys.*, **4**, 112–123.

——— and ———, 1954: Theoretical considerations regarding the dayglow. *J. Atmos. Terr. Phys.*, **5**, 329–344.

BIONDI, M. A., 1964: Electron-ion and ion-ion recombination. *Ann. Geophys.*, **20**, 34–46.

BRACE, L. H., and N. W. SPENCER, 1964: First electrostatic probe from Explorer XVII. *J. Geophys. Res.* **69**, in press.

———, ———, and G. R. CARIGNAN, 1963: Ionosphere electron temperature measurements and their implications. *J. Geophys. Res.*, **68**, 5397–5412.

BRANDT, J. C., 1958: Oxygen red lines in the airglow. III. The dayglow. *Astrophys. J.*, **128**, 718–723.

BRYANT, R., 1964: Densities obtained from drag on the Explorer 17 satellite. *J. Geophys. Res.*, **69**, 1423–1425.

CADLE, R., 1964: Daytime atmospheric O(^{1}D). *Discussions Faraday Soc.*, **37**, in press.

CARMAN, E. H., and B. P. KILFOYLE, 1963: Relationship between [01] 6300 A zenith airglow and ionospheric parameters f_0F_2 and $h'F$ at Townsville. *J. Geophys. Res.*, **68**, 5605–5607.

CHAMBERLAIN, J. W., 1958: Oxygen red lines in the airglow. I. Twilight and night excitation processes. *Astrophys. J.*, **127**, 54–66.

———, 1961a: The energies in the spectra of the airglow and aurora. *Ann. Geophys.*, **17**, 90–99.

———, 1961b: *Physics of the Aurora and Airglow.* New York, Academic Press, 704, pp.

DALGARNO, A., 1964a: Thermal reactions involving charged particles. *Discussions Faraday Soc.*, in press.

———, 1964b: Corpuscular radiation in the upper atmosphere. *Ann. Geophys.*, **20**, 65–74.

———, and M. B. MCELROY, 1964a: Electron and ion production rates in the upper atmosphere. To be published.

——— and ———, 1964b: Electron temperatures at dawn. To be published.

———, ———, and R. J. MOFFETT, 1963: Electron temperatures in the ionosphere. *Planetary Space Sci.*, **11**, 463–484.

DEMORE, W., and O. F. RAPER, 1964: Deactivation of O(^{1}D) in the atmosphere. *Astrophys. J.*, **139**, 1381–1383.

DETWILER, C. R., D. L. GARRETT, J. D. PURCELL and R. TOUSEY, 1961: The intensity distribution in the ultraviolet solar spectrum. *Ann. Geophys.*, **17**, 9–18.

DONAHUE, T. M., and W. G. FASTIE, 1964: Observation and interpretation of resonance scattering of Lyman α and OI (1300) in the upper atmosphere. *Space Research*, **4**, Amsterdam, North-Holland Publishing Company, 304–324.

FASTIE, W. G., H. M. CROSSWHITE and D. F. HEATH, 1964: Rocket spectrophotometer airglow measurements in the far ultraviolet. *J. Geophys. Res.*, **69**, in press.

FITE, W. L., J. A. RUTHERFORD, W. R. SNOW and V. A. J. VAN LINT, 1962: Ion-neutral collisions in afterglow. *Discussions Faraday Soc.*, **33**, 264–272.

GARSTANG, R. H., 1961: Mutual magnetic interactions and oscillator strengths in the first spectrum of oxygen. *Proc. Cambridge Phil. Soc.*, **57**, 115–120.

HALL, L. A., W. SCHWEIZER and H. E. HINTEREGGER, 1963: Diurnal variation of the atmosphere around 190 kilometers derived from solar extreme ultraviolet absorption measurements. *J. Geophys. Res.*, **68**, 6413–6417.

HANSON, W. B., 1963: Electron temperatures in the upper atmosphere. *Space Research*, **3**, Amsterdam, North-Holland Publishing Company, 282–302.

———, and F. S. JOHNSON, 1961: Electron temperatures in the ionosphere. *Mèmoires Soc. R. Sci. Liège, Ser. 5*, **4**, 390–424.

HARRIS, I., and W. PRIESTER, 1962: Time-dependent structure of the upper atmosphere. *J. Atmos. Sci.*, **19**, 286–301.

HINTEREGGER, H. E., 1961: Preliminary data on solar extreme ultraviolet radiation in the upper atmosphere. *J. Geophys. Res.*, **66**, 2367–2380.

———, and K. WATANABE, 1962: Photoionization rates in the *E* and *F* regions, 2. *J. Geophys. Res.*, **67**, 3373–3392.

JACCHIA, L. G., and J. SLOWEY, 1964: Atmospheric heating in the auroral zones: A preliminary analysis of the atmospheric drag of the Injun 3 satellite. *J. Geophys. Res.*, **69**, 905–910.

JARRETT, A. H., and M. J. HOEY, 1963: A ground-level photographic observation of the day airglow emission of atomic oxygen at 6300 Å. *Planetary Space Sci.*, **11**, 1251–1252.

JASTROW, R., and L. KYLE, 1961: The earth atmosphere. *Handbook of Astronautical Engineering*, New York, McGraw-Hill, Chap. 2, 2–13.

JURSA, A. S., M. NAKAMURA and Y. TANAKA, 1963: Molecular oxygen distribution in the upper atmosphere. *J. Geophys. Res.*, **68**, 6145–6155.

KING-HELE, D. G., 1963: Decrease in upper atmosphere density since the sunspot maximum of 1957–58. *Nature*, **198**, 832–834.

LINDSAY, J. C., 1963: Scientific results of the first orbiting solar observatory. *Trans. Amer. Geophys. Union*, **44**, 722–725.

METZGER, P. H., and G. R. COOK, 1964: A reinvestigation of the absorption cross sections of molecular oxygen in the 1050–1800 A region. *J. Quantitative Spectroscopy Radiative Transfer*, **4**, 107–116.

NAGY, A. F., L. H. BRACE, G. R. CARIGNAN and M. KANAL, 1963: Direct measurements bearing on the extent of thermal nonequilibrium in the ionosphere. *J. Geophys. Res.*, **68**, 6401–6412.

NICOLET, M., 1961: Density of the heterosphere related to temperature. *Res. in Space Sci.*, Smithsonian Inst. Astrophys Observ., Spec. Rpt. 75, 30 pp.

———, and W. SWIDER, 1963: Ionospheric conditions. *Planetary Space Sci.*, **11**, 1459–1482.

NIER, A. O., J. H. HOFFMAN, C. Y. JOHNSON and J. C. HOLMES, 1964: Neutral composition of the atmosphere in the 100- to 200-kilometer range. *J. Geophys. Res.*, **69**, 979–989.

NORTON, R. B., T. E. VAN ZANDT and J. S. DENISON, 1963: A model of the atmosphere and ionosphere in the *E* and *F1* regions. *Proc. Int. Conf. on the Ionosphere*, London, Institute of Physics and Physical Society, 26–34.

NOXON, J. F., 1963: Observation of daytime aurora. *J. Atmos. Terr. Phys.*, **25**, 637–645.

———, 1964: A study of the 6300 Å oxygen line in the day airglow. *J. Geophys. Res.*, **69**, 3245–3255.

———, and R. M. GOODY, 1962: Observation of day airglow emission. *J. Atmos. Sci.*, **19**, 342–343.

PACKER, D. M., 1961: Altitudes of the night airglow radiations. *Ann. Geophys.*, **17**, 67–74.

POKHUNKOV, A. A., 1963: Gravitational separation, composition and structural parameters of the night atmosphere at altitudes between 100 and 210 km. *Planetary Space Sci.*, **11**, 441–449.

PURCELL, J. D., D. M. PACKER and R. TOUSEY, 1960: The ultraviolet spectrum of the sun. *Space Research*, **1**, Amsterdam, North-Holland Publishing Company, 581–589.

SAYERS, J., and D. SMITH, 1964: Ion and charge exchange reactions involving atmospheric gases. *Discussions Faraday Soc.*, **37**, in press.

SCHAEFER, E. J., 1963: The dissociation of oxygen measured by a rocket-borne mass spectrometer. *J. Geophys. Res.*, **68**, 1175–1176.

SEATON, M. J., 1954: Excitation processes in the aurora and airglow, 2. Excitation of forbidden atomic lines in high altitude aurorae. *J. Atmos. Terr. Phys.*, **4**, 295–313.

———, 1956: The calculation of cross-sections for excitation of forbidden atomic lines by electron impact. *The Airglow and the Aurorae*, London, Pergamon Press, 289–301.

———, 1958: Oxygen red lines in the airglow. II. Collisonal deactivation effects. *Astrophys. J.*, **127**, 67–74.

SPENCER, N. W. L. H. BRACE and G. R. CARIGNAN, 1962: Electron temperature evidence for nonthermal equilibrium in the ionosphere. *J. Geophys. Res.*, **67**, 157–175.

TARASOVA, T. M., 1963: Night-sky emission-line intensity distribution with respect to height. *Space Research*, **3**, Amsterdam, North-Holland Publishing Company, 162–172.

TOUSEY, R., 1963: The extreme ultraviolet spectrum of the sun. *Space Science Rev.*, **2**, 3–69.

WALLACE, L., and J. W. CHAMBERLAIN, 1959: Excitation of O_2 atmospheric bands in the aurora. *Planetary Space Sci.*, **2**, 60–70.

WALLACE, L., and R. A. NIDEY, 1964: Measurement of the daytime airglow in the visual region. *J. Geophys. Res.*, **69**, 471–479.

WHITTEN, R. C., and I. G. POPPOFF, 1964: Ion kinetics in the lower ionosphere. *J. Atmos. Sci.*, **21**, 117–133.

WILLMORE, A. P., C. L. HENDERSON, R. L. F. BOYD and P. J. BOWEN, 1964: Electron temperature in the upper *F*-region. *Proc. Roy. Soc. London, A*, in press.

ZIPF, E. C., and W. G. FASTIE, 1963: An observation of day airglow emission at 6300 Å. *J. Geophys. Res.*, **68**, 6208–6209.

N65 - 32163

A LATITUDE SURVEY OF THE NIGHT AIRGLOW*

T. N. DAVIS† AND L. L. SMITH‡

A four-color turret photometer has been used for shipboard observations of the night airglow and the aurora. Zenith observations have been obtained during 1962 close to the 70°W meridian from the northern auroral zone to the Antarctic pack ice. Most of the observations were taken during the Southern Winter.

Maximums in the [OI] 5577A intensity were found near 30–40°N and 40°S, with the Southern Hemisphere level being generally lower than that in the Northern Hemisphere. The [OI] 6300A latitudinal variation was similar to that of [OI] 5577A except that maximums in the latitudinal distribution were found near 15° geomagnetic north and south latitudes. In contrast with the [OI] emissions, the sodium group (NaD doublet and OH bands near 5893A) generally increased to the south with maximums near 20–30°N and 40–50°S geographic latitude. The total 5340A background increased from north to south. When the integrated starlight and zodiacal light components were subtracted from the total background, the resultant terrestrial component showed a pronounced minimum at low latitudes and apparent symmetry about the geomagnetic equator.

I. INTRODUCTION

Photometric observations of the aurora and the night airglow made aboard the U.S.N.S. *Eltanin* during the period March/November, 1962. The *Eltanin* departed from New York in late March and moved northward into the Laborador Sea where photometric observations of aurora were made on four nights After returning to New York, the *Eltanin* departed on May 24 on a trip southward and arrived in Valparaiso, Chile on June 27. Clear skies existed throughout most of this cruise and useful airglow observations were obtained on 17 nights. In the period July/ November 1962, the *Eltanin* made several cruises southward from Valparaiso and useful observations (during clear, moonless dark hours) were obtained on 36 nights. See Figure 1 for the positions of the *Eltanin* when the observations were made.

Observations were made in four narrow wavelength bands centered near 5577, 6300, 5893 and 5340A with a turret photometer built by the Fritz Peak Observatory following a design by Purdy, Megill and Roach (1961). This instrument has a 5° diameter circular field of view and was mounted on a free swinging gimbal to assure observation of the zenith. The gimbal mount was effective in reducing the effect of the ship's motion to the extent that the photometer axis pointed to within 5° of the zenith throughout most of the observations. Observations at a given wavelength were obtained at 5 minute intervals.

The two-color method (Roach and Meinel, 1955) was applied to the 5577, 6300 and 5893A observations in an attempt to remove the contributions due to airglow continuum, zodiacal light, and integrated starlight; the method involves subtraction of these contributions as determined by observation through the 5340A filter. Details of the data reduction and calibration procedures together with tabulations of the results are given in a limited circulation report (Davis and Smith, 1964).

A complication to the determination of the intensities of the [OI] 5577, [OI] 6300 and NaD 5893A emissions arises due to contamination from nearby OH band emissions. Figur 2 shows the filter transmission characteristics and wavelengths

*Published in the *Journal of Geophysical Research*, 70(5):1127–1138, March 1, 1965.

†Goddard Space Flight Center and University of Alaska, College, Alaska

‡Fritz Peak Observatory, National Bureau of Standards, Boulder, Colorado.

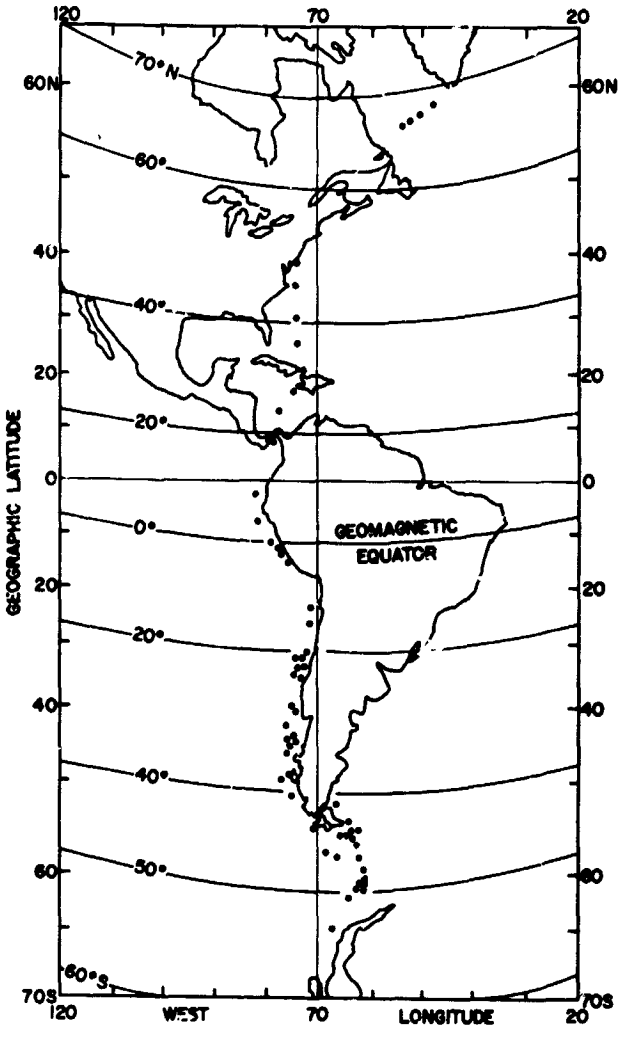


FIGURE 1.—Location of the *Eltanin* on nights when useful nighttime observations were made.

of the [OI] and NaD (doublet) emissions. A correction to the [OI] 5577A intensity is required due to the presence of the OH [7/1] band emissions with the band head at 5562A. Using estimates of the mean zenith intensities of the OH 5562A and [OI] 5577A line emissions given by Roach (1956), 15R and 250R, respectively, or measurements by Krassovsky, Shefov, and Yarin (1962) together with the known transmission characteristics of the 5577A filter, it is estimated that on the average approximately 90 percent of the intensity remaining after correction for continuum sources is due to the [OI] 5577A emission. Although the 6300A filter is quite narrow, contamination from the OH (9/3) band with band head near 6256A is significant. Based on data given by Chamberlain (1961, pp. 555/557) the OH contribution through the 6300A filter is estimated to be in the neighborhood of 10 to 20 Rayleighs. At low levels of [OI] 6300A emission encountered during some of the measurements described below, the OH contamination is especially severe. Serious contamination also occurs through the 5893A filter as a result of the OH (8/2) band near 5386A. The NaD and OH emissions can covary (see Chamberlain, 1961 p. 517) and it is estimated that approximately 70 percent of the line and band emission measured through the 5893A filter is due to the NaD doublet at 5890/5896A. Since the observations reported here are obtained over a wide range of latitude, it must be recognized that there is much uncertainty about the OH contributions. Hence, no attempt has been made to correct the results presented below for OH contamination.

II. RESULTS

Several types of variation occur in the night airglow; these types include latitudinal variation, seasonal variation, diurnal variation, nightly variation and possibly also effects produced by high-altitude explosions. Considering the nature of the data taken so far aboard the *Eltanin*, it is impossible to completely separate the different types of variation. An additional complication arises from inhomogenieties of the data, since on some nights complete observations are obtained, while on others, as little as only an hour of observation may be available.

II. A. Variation of [OI] 5577A

A. 1 *Night-to-night and Diurnal Variations*

The observations show two categories of diurnal variation in 5577A. One type, illustrated in Figure 3A, is one in which little variation in the 5577A intensity occurs during the middle of the night. In the second category are those nights in which an enhancement occurs during some part of the night; see Figures 3B and 3C. Evidently the time at which the enhancement occurs is random, since averaging over a number of nights (Figure 4A) yields a nearly constant level of 5577A emission throughout the night. The data do not indicate any latitude dependence of the character of the diurnal variation, although the quantity of data is not sufficient to rule out such a dependence.

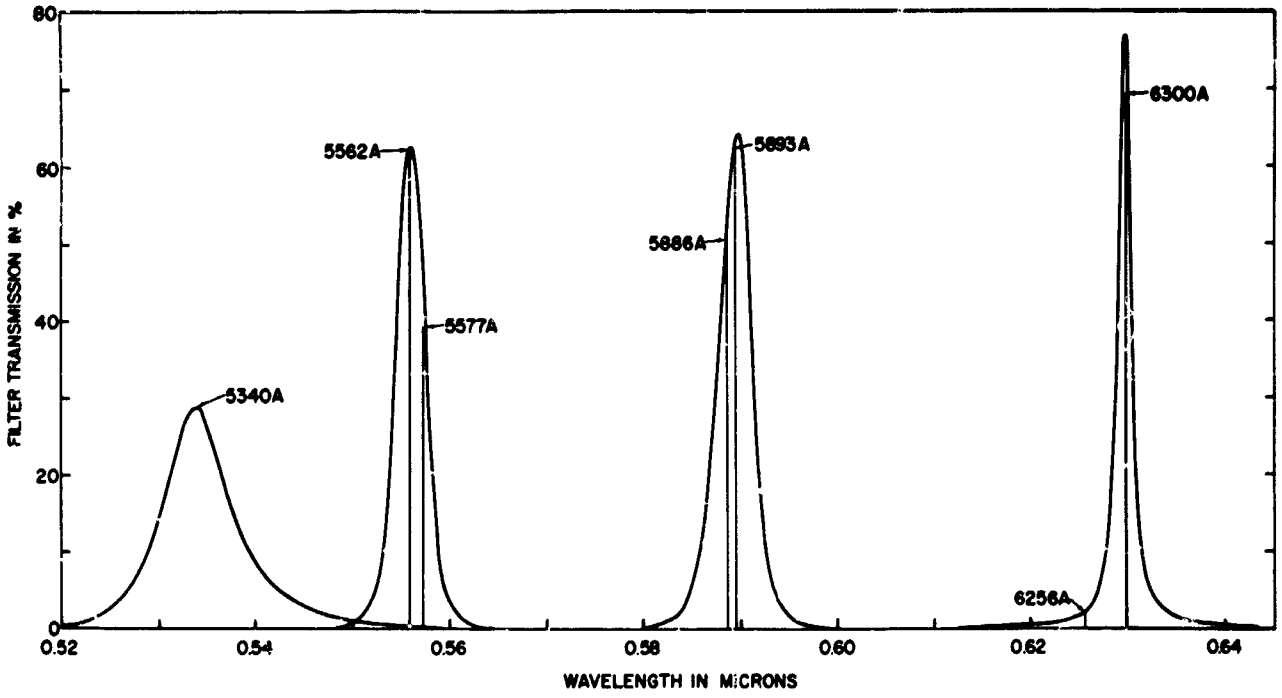


FIGURE 2.—Transmission characteristics of the interference filters and locations of line emissions near the filter peaks.

A. 2 *Latitude and Seasonal Variations*

Figure 5A shows a plot of the nightly average and nightly range of hourly averages of 5577A intensity versus date. The variations here are of several types: latitude, nightly, and seasonal. The same data are plotted in Figure 5B against latitude and those taken during a limited period, May 24/September 1, 1962, are plotted similarly in Figure 5C. An attempt to eliminate part of the nightly variation is shown in Figure 5D, which is compiled by taking the averages of all data in 10 latitude-degree groups. Figures 5C and 5D provide the best indication of both the seasonal and latitude variation in 5577A intensity. The northern hemisphere data are obtained over a short period of time (May 24/June 2) and therefore include only diurnal, nightly and latitudinal variations. These types of variation cannot be separated completely, however Figures 5C and 5D indicate the possibility of a latitudinal maximum near 30/40° geographical latitude and decreasing intensity as the eqator is approached. Moving southward from the equator, there is some increase in the average intensities (Figures 5C and 5D) with a weak maximum occurring near geographic latitude 40°. There appears to be a definite decrease in the 5577A airglow level to the south as far as 65° geographic latitude where the observations end. The data are consistent with a seasonal variation in which there is a maximum summer and minimum in winter.

II. B. Variation of [OI] 6300A

B. 1 *Night-to-night and Diurnal Variations*

Unlike the [OI] 5577A, the [OI] 6300A emission shows a definite variation through the night. The well-known post-twilight decay and the lesser predawn increase are evident in the data taken at all latitudes; see Figure 4B. A nightly variation occurs in [OI] 6300A; for the most part it appears as an over-all change in the level of the emission through the night.

The only exception occurred when the *Ellanin* was near the equator. On the two nights shown in Figure 6, and possibly on a third, an enhancement in [OI] 6300A developed in the hours near midnight. These monochromatic enhancements may be related to the intertropical red arcs described by Barbier, Weill, and Glaume (1961); however, the enhancement is relatively small and apparently restricted in time to the hours near

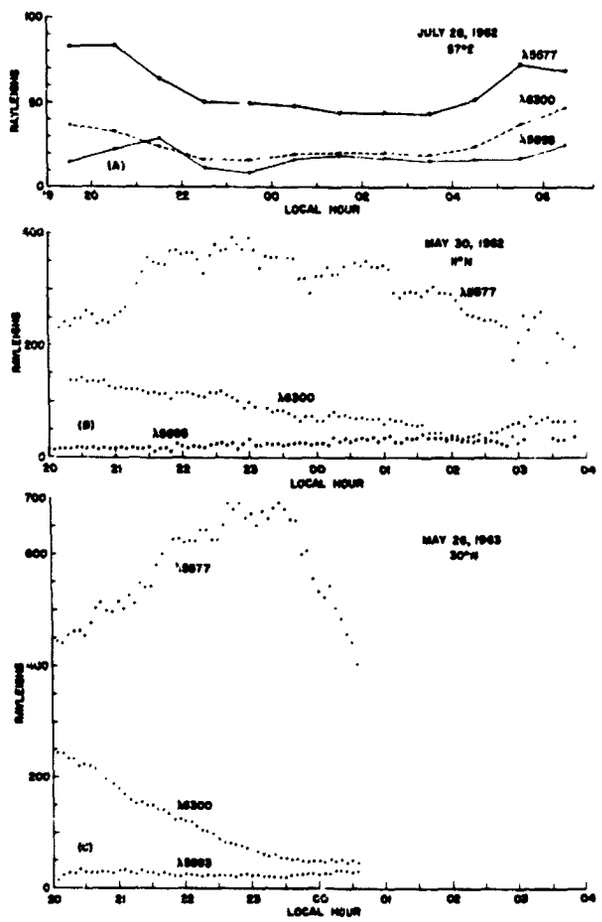


FIGURE 3.—A) Example of a night of relative quiet in the airglow variations (plot of hourly averages); B) and C) are examples of active variations in 5577A.

local midnight. Although the midnight enhancements observed on June 2 and 6, 1962 may not be related to the intertropical red arcs, the zones where the intertropical red arcs appear coincide with the positions of the two maxima on the curve of Figure 7C. There is much similarity between the observed latitudinal distribution of 6300A in the region within 30°N and 30°S magnetic latitude and the F region electron content per vertical column by Duncan (1960). In this calculation, he assumed an initially uniform electron distribution to be modified by electrodynamic lift and subsequent settling down the geomagnetic field lines. Since it has been found that the intensity of the 6300A emission is approximately proportional to the maximum electron density in the F2 region (Huruhata et al., 1959), it appears possible that the processes described by Duncan may contribute to the observed 6300A distribution in the tropic regions.

B. 2 *Seasonal and Latitudinal Variation*

The diagrams showing all data versus latitude (Figure 7A), Southern Hemisphere Winter versus latitude (Figure 7B) and 10-latitude-degree averages (Figure 7C) show generally higher intensities in the Northern Hemisphere and thus may indicate a seasonal variation of [OI] 6300A similar to that of [OI] 5577A; namely, a maximum in summer and minimum in winter. Maximums in the latitudinal distribution of intensity appear near the equator at approximately 15°N and 15°S geomagnetic latitude.

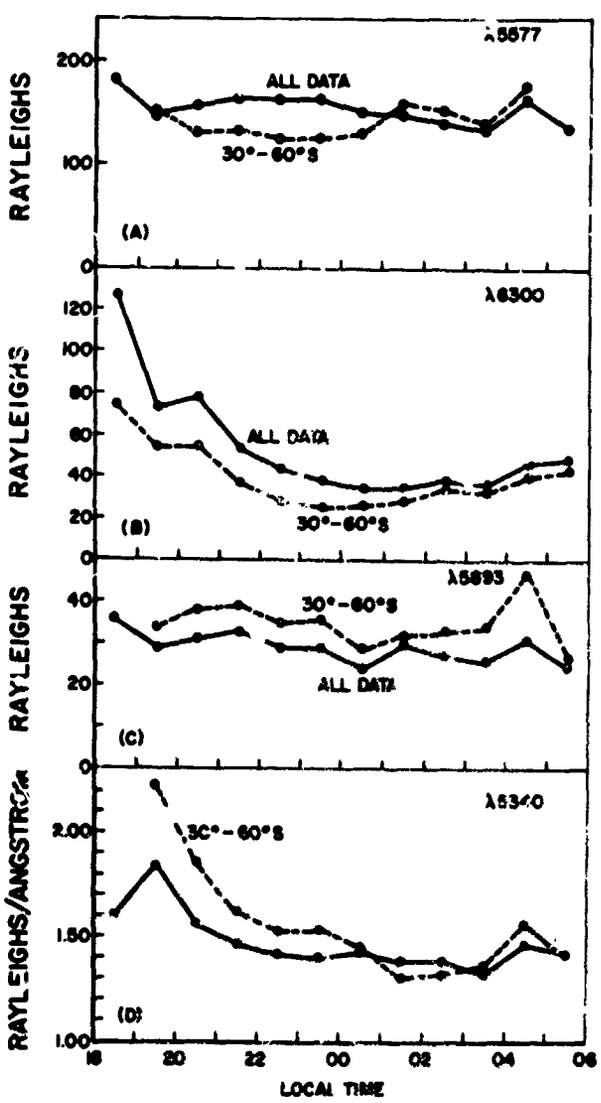


FIGURE 4.—Average diurnal variation of the airglow emissions.

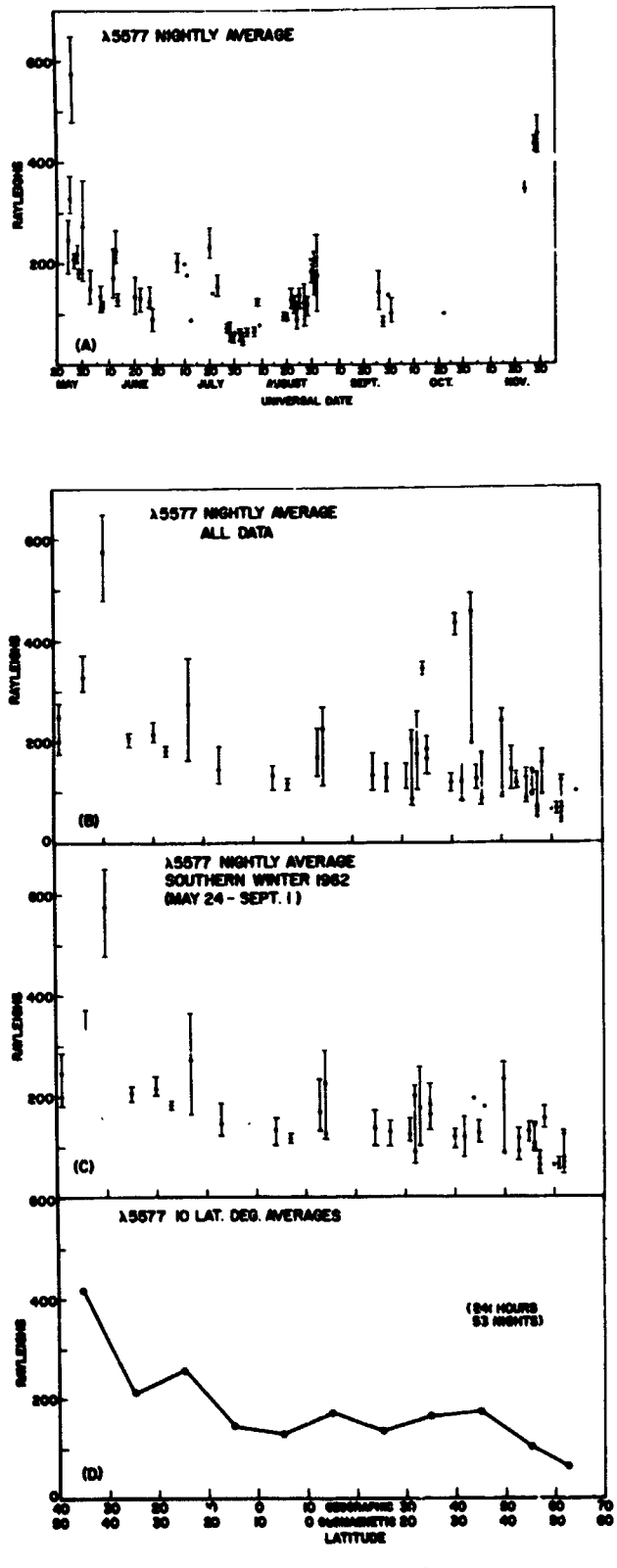


FIGURE 5.—Nightly average and range of hourly average of 5577A: A) versus date, B) all data versus latitude, and C) for data taken May 24 to September 1, 1962. D) ten-degree latitude averages versus latitude, all data.

Another representation of the [OI] 6300A data is shown in Figure 8, where the observations are plotted against latitude and local time and then used to compile isophotes of zenith intensity. This diagram is suggested by a similar one by Barbier (1961) and should be compared with it.

The intensity ratio 6300A/5577A obtained from individual airglow observations is highly variable; ratios averaged over a night's observations range between 0.09 and 0.67. The average ratio of the observations in the region 30/60°S is 0.25.

II. C. Variation of the NaD group

As mentioned previously, the 5893A filter passes both the NaD doublet centered near 5893A and OH emissions including the OH (8/2) emission at 5886A. These emissions will be referred to here as the sodium group. The sodium group intensity remains relatively constant on a given night, and the average diurnal variation curves, Figure 4C, indicate little or no regular variation through the night.

Plots of the sodium group variations versus latitude (Figure 9), show a trend toward higher intensities going from north to south. Superimposed on this trend are irregularities. A maximum appears near 20/30°N, and there is a minimum centered near 0/10°N. Another maximum may exist near 10/20°S, and relatively high intensities are observed over the latitude range 30/[illegible]°S. The intensities observed between 58 and 65°S are relatively low. Overall, the variation with latitude of the sodium group intensity is somewhat less than that of the [OI] 5577A and 6300A emissions. The observed sodium group intensities are consistent with a seasonal variation in which there is a maximum in winter and minimum in summer.

II. D. Variation of the Background near 5340A

Except for some evening twilight contamination or post-twilight enhancement, the background near 5340A averaged over all data remained relatively constant through the night; see Figure 4D. The total background near 5340A, shows a general increase from north to south. This total background includes components from integrated starlight, zodiacal light and airglow continuum. Using integrated starlight intensities given by

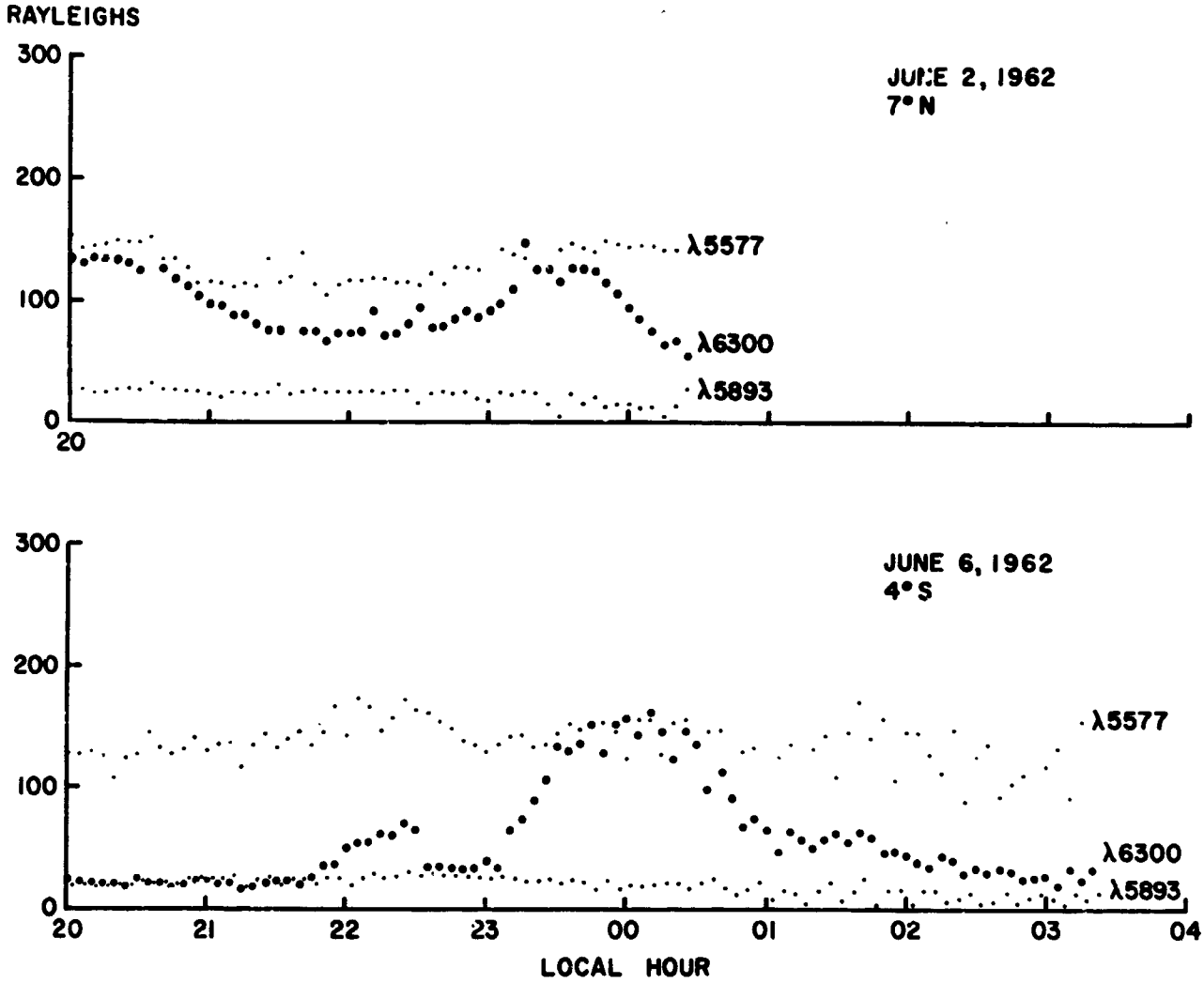


FIGURE 6.—Nights on which enhancements in 6300A occurred.

Roach and Megill (1961) and zodiacal light intensities from Roach and Smith (manuscript in preparation) the contribution of these components to the total background is determined and displayed in Figure 10. The background component due to airglow continuum is obtained by subtraction and is also shown in Figure 10. The airglow continuum determined by this method shows a pronounced minimum at low latitudes and appears to be fairly symmetrical about the goemagnetic equator. For comparison, one-year average values of the airglow continuum obtained by Roach and Smith for two airglow observatories, Fritz Peak, Colorado and Haleakala, Hawaii, are indicated on Figure 10 by the numbers 1 and 2, respectively.

II. E. Auroral Zone Observations

Limited observations of auroras were made during March and April 1962 at the northern auroral zone, however high seas and partial cloudiness during the period of these observations detract from their value. Bright auroral forms were observed in the zenith during part of the observations; 5577A intensities ranged up to 92kR, those of 6300A ranged up to 1.8kR. Although the results may not be meaningful, intensities of the sodium group were not much greater than that observed in the airglow (see Figure 7).

One interesting aspect of the auroral observations is the behavior of the average [OI] 6300/5577 ratio. Individual values of the ratio varied be-

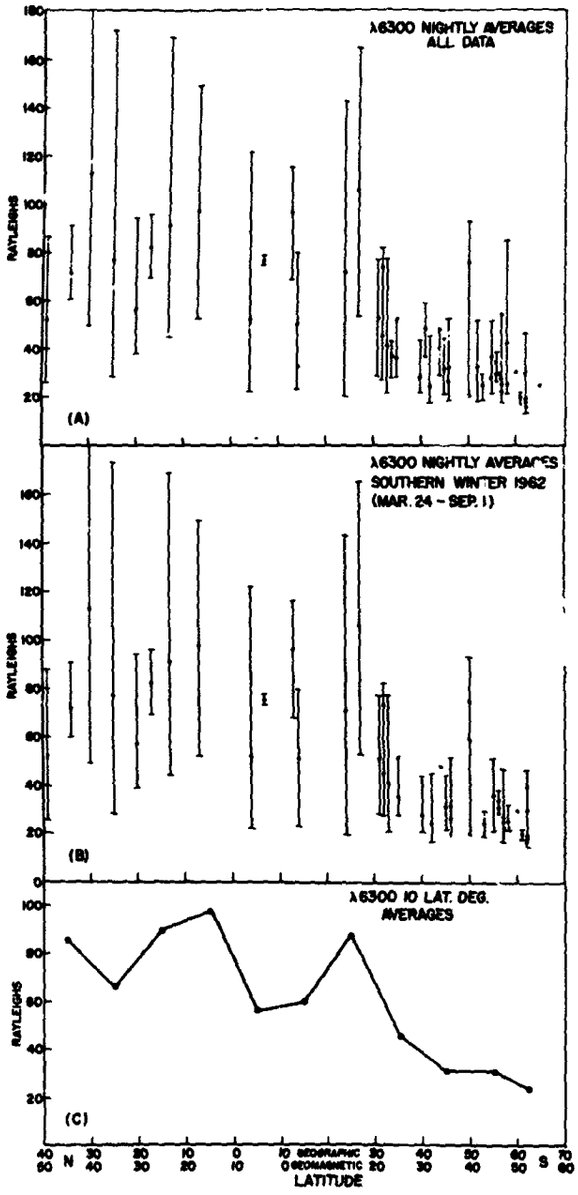


FIGURE 7.—Nightly average and range of hourly averages of 6300A: A) for all data and B) for data taken May 24 to September 1, 1962. C) ten-degree latitude averages versus latitude for all data.

tween 0.007 and 0.5, with the ratio decreasing with increasing 5577A and 6300A intensity: see Figure 11. The higher values of the ratio tend to occur in the hours before midnight, and the lower values after midnight. The tendency for the 6300A/5577A ratio to decrease with increasing 5577A intensity may be due to electron deactivation of the oxygen ^1D level in the brighter

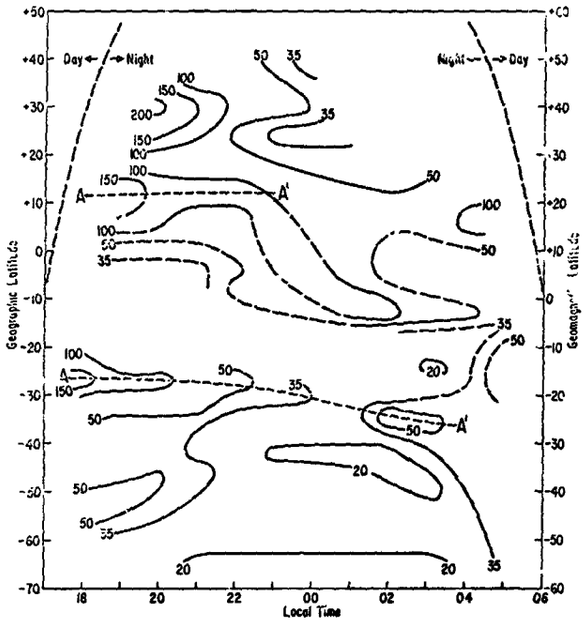


FIGURE 8.—Isophote map giving zenith intensity of 6300A in rayleighs as derived from the *Eltanin* observations during Southern Winter 1962. The dashed lines A–A' indicate possible arc-like zones of enhanced radiation. (Compare with Barbier, 1961)

auroras (Chamberlain 1961, p. 315). Also, molecular deactivation of the ^1D level by collisions with O_2 may be responsible for the results shown in Figure 12 if the brightening of 5577A is associated with increasing auroral exitation at low altitude in the post-midnight hours.

III SUMMARY AND DISCUSSION OF RESULTS

While it is not possible to separate completely the types of variations which influence the results obtained here, these results are compatible with the diurnal, seasonal, and latitudinal variations shown in Table I in summary form. As indicated in the right-hand column of Table I, the observations may be interpreted as suggesting either a rather general seasonal variation or a pronounced hemispherical effect. One aspect of the observations which is evident in the various diagrams but which is not indicated in Table I is the sizable nightly variation observed in each of the airglow components. A comparison of these nightly variations with the global magnetic activity index Kp did not reveal any connection. However, it is interesting that the results of this study do suggest

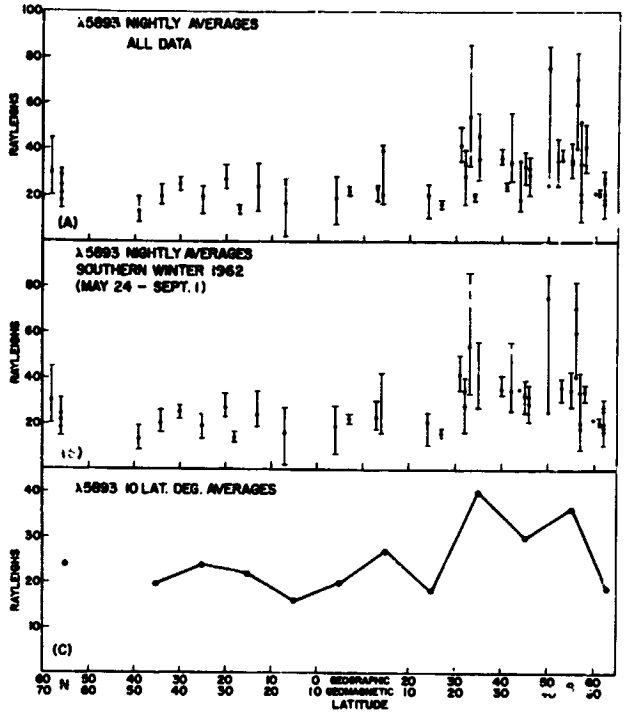


FIGURE 9.—Nightly average and range of hourly averages of the sodium group (NaD 5893A and OH): A) for all data and B) for observations taken May 24 to September 1, 1962. C) ten-degree latitude averages versus latitude, all data.

that the latitudinal distribution of intensity of some airglow emissions may be influenced by the configuration of the geomagnetic field.

The diurnal and nightly variation of 5577A observed on the *Eltanin* in 1962 is somewhat less that that observed in previous years on the *Soya* (Nakamura, 1957 and 1959). Assuming that the calibrations are correct for each set of data, the *Eltanin* observations show lower absolute 5577A intensities as well. Two differences exist in the conditions under which these sets of data were obtained: the observation aboard the *Soya* were made during southern summer and during a period of high solar activity whereas the *Eltanin* observations were made near a minimum in the solar cycle and during southern winter.

Several factors affect the accuracy of the results reported here. One is the uncertainty in the absolute calibration of the comparison source. Another is the problem of OH contamination, which is discussed above. Since the observed 5577A intensities in the southern hemisphere are low, somewhat more than 10 percent of the 5577A intensities reported here may be due to OH. If it is assumed that the NaD and OH emissions

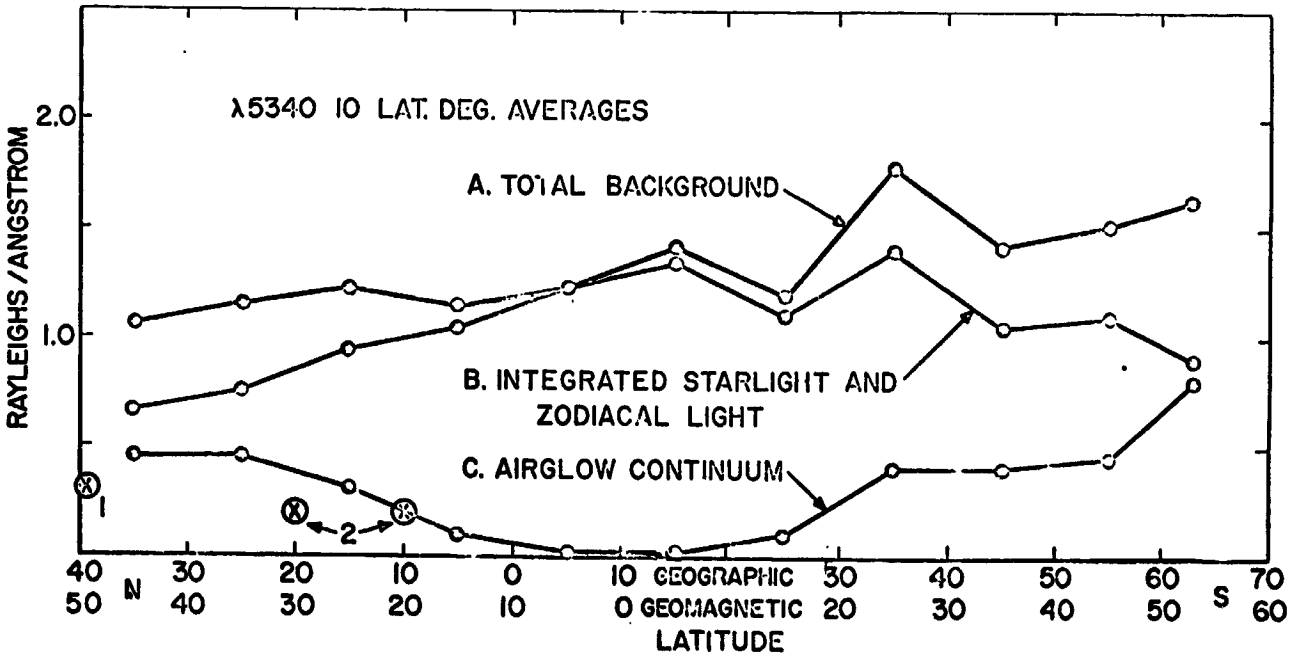


FIGURE 10.—Ten-degree latitude averages of the 5340A background: A) total background, B) the component due to integrated starlight and zodiacal light as computed by Roach and Smith (private communication), and C) the airglow continuum component determined by subtracting curve B from curve A. The point labeled (a) represents the one-year average airglow continuum obtained by Roach and Smith for the Fritz Peak Observation (geographic latitude 40°, geomagnetic latitude 50°) and the points labeled (2) are the same for the Haleakala Observatory (geographic latitude 20°, geomagnetic latitude 10°).

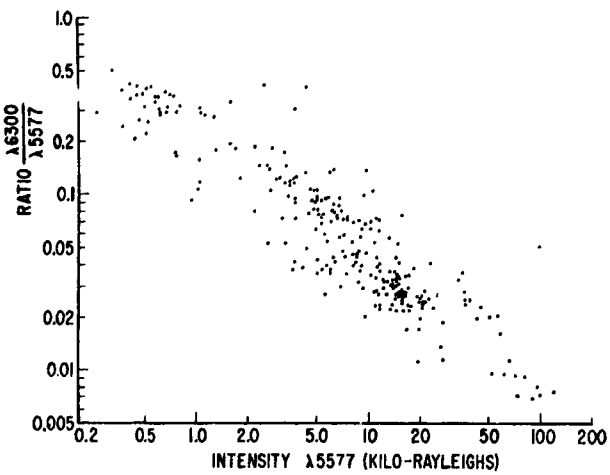


FIGURE 11.—The zenith intensity ratio 6300A/5577A plotted against the zenith intensity of 5577A in the auroral zone aurora.

co-vary, then, at all times, approximately 70 percent of the sodium group intensity observed through the 5893A filter is due to NaD 5893A. It has been found during these observations that these observations that the 2-color method is not a completely satisfactory means of eliminating the astronomical and airglow continuum components from the observations. In using that method it has been assumed that a single proportionality constant relating the contaminants to the 5340A intensity is applicable at all observed light intensities. Observations made during periods of relative stability in the level of the various airglow emissions and in which the astronomical component varied (for example, passage of the Milky Way over the photometer) indicate that the proportionality constant for each wavelength was too low to correct for the astronomical component. On the other hand, if one adopts proportionality constants of sufficient size to give a complete correction when the astronomical components are large, an obvious over-correction occurs at other

TABLE I.—*Summary of* Eltanin *Airglow Observations*

Emission	Diurnal Variation	Latitudinal Variation	Seasonal Variation (or Hemispherical Difference)
[OI] λ5577	No regular diurnal variation irregular activity (sometimes monochromatic) on some nights.	Apparent maximum near 30–40°N, decreasing to minimum near equator, then gradual increase to maximum near 40°S. Decreasing intensity southward as far as 65°S.	Compatible with maximum in summer and minimum in winter. (Or if no such general seasonal variation, the intensities are greater in the northern hemisphere than in the southern hemisphere).
[OI] λ6300	Common post-twilight and predawn variations persist at all latitudes. Monochromatic enhancement observed at local midnight near equator.	Apparent maxima near 15° geomagnetic north and south.	
NaD+ OH near λ5893	No diurnal variation evident.	Apparent maximum near 20–30°N. Shallow minimum near 0–10°N. Irregular increase southward to maximum near 40–50°S and decreasing from 55°S to 65° south.	Compatible with minimum in summer and maximum in winter. (or, as above, higher relative intensity in the southern hemisphere).
Background near λ5340	Possible post-twilight enhancement, otherwise no regular diurnal variation.	Relatively steady increase of total background from 40°N to 65°S. Astronomical components maximize in equatorial region; airglow continuum lowest in equatorial region with apparent symmetry about geomagnetic equator.	

times (the line emission intensities become negative). Consequently, the 2-color method, as used here, must be considered as a very crude means of correction for the atmospheric continuum and astronomical backgrounds. Despite the various uncertainties which influence the absolute values of the emissions reported here, the trends summarized in Table I appear to be real in that they appear in the data prior to and after the application of any corrections.

ACKNOWLEDGMENTS

We are grateful to Dr. Franklin E. Roach for his aid and advice throughout this study. Dr. Manfred H. Rees also has contributed many helpful discussions and several specific suggestions which we have followed. Most of the data presented here were collected through the conscientious shipboard efforts of Mr. F. T. Berkey. Part of the work was accomplished while one of us (T.N.D.) was a Guest Worker at the Fritz Peak Observatory of the National Bureau of Standards. This work was supported by the Office of Antarctic Programs of the National Science Foundation grant number 18704.

REFERENCES

Barbier, D., Les variations d'intensité de la raie 6300A de la luminescence nocturne, *Ann. Geophys.*, 17, 3–15, 1961.

Barbier, D., G. Weill, and J. Glaume, L'emission de la raie rouge du ciel nocturne en Afrique, *Ann Geophys.*, 17, 305–318, 1961.

Chamberlain, J. W., *Physics of the Aurora and Airglow*, Academic Press, New York and London, pp. 315, 517, 555–557, 1961.

Davis, T. N., and L. L. Smith, Latitudinal and seasonal variations in the night airglow, Report No. UAG-R156, Geophysical Institute, College, Alaska, p. 19, 1964.

Duncan, R. A., The equatorial F-region of the ionospher, *J. Atmos, Terr. Phys., 18*, 89–100, 1960.

Huruhata, M. T. Nakamura, H. Tanabe, and T. Tohmatsu, Oxygen red line in the night airglow and ionospheric F2 region, *Rept. Ionos. and Space Res. in Japan*, 13, 283–289, 1959.

Krassovksy, V. I., N. N. Shefov, and V. I. Yarin, Atlas of the airglow spectrum 3000–12400Å, *Planet. Space Sci., 9*, 883–915, 1962.

Nakamura, Jungi, Latitude effect of night airglow, *J. Geophys. Res., 62*, 487–488, 1957.

Nakamura, Junji, Latitude effect of night airglow, *Rept. Ionosphere Res. in Japan, 12*, 419–427, 1959.

Purdy, C. M., L. R. Megill, and F. E. Roach, A new airglow photometer, *J. of Research of the National Bureau of Standards, 650*, 213–216, 1961.

Roach, F. E., Manual for photometric observations of the airglow during the International Geophysical Year, NBS Rept. 5006, National Bureau of Standards, Boulder, Colorado, 1956.

Roach, F. E., and L. R. Megill, Integrated starlight over the sky, *Astrophys. J., 133*, 228, 1961.

Roach, F. E., and A. B. Meinel, The height of the nightglow by the Van Rhijn method, *Astrophysical Journal*, 122, 530–553, 1955.

A65-13158

THE UPPER ATMOSPHERE OF JUPITER*

S. H. GROSS† AND S. I. RASOOL‡

The properties of the upper atmosphere of Jupiter have been theoretically investigated. The vertical temperature profile in the atmosphere above the clouds and the structure of the ionosphere have been computed for two different model atmospheres: Model I corresponds to a H/He mixing ratio of 20 : 1 by volume (Urey, 1959), and Model II has a H/He ratio as low as 0.03 : 1 (Öpik, 1962). Assuming that the atmosphere of Jupiter above the clouds is in radiative equilibrium, the temperature distribution in the upper atmosphere of Jupiter has been computed. With a cloudtop temperature of 153° K and the total optical thickness of the atmosphere in the far infrared of 0.66, the radiation emitted by the planet is 3.1×10^4 ergs cm^{-3} sec^{-1}. This value of radiation flux is about four times higher than the solar flux received by Jupiter. The equilibrium distribution of electron densities in the ionosphere of Jupiter has also been calculated. The maximum electron densities of $\sim 10^6$ electrons/cc are found to occur at the altitudes of 220 km and 110 km above the cloudtop for the two model atmospheres.

INTRODUCTION

Revived interest in planetary astronomy has provided us with a considerable amount of new data on the temperature, pressure, and composition of the upper atmosphere of Jupiter. These data may now be used to derive coherent and satisfactory models of the structure of the Jovian atmosphere.

In this paper we first derive the two extreme model atmospheres for Jupiter which are consistent with the available data on the relative abundances of gases on the planet. We then derive the thermal structure of the upper atmosphere of Jupiter for these two extreme cases, and study the formation of an ionosphere due to the photoionization effect of the solar ultraviolet radiation.

MODELS OF THE JOVIAN ATMOSPHERE

Based on the recent observational evidence, the following assumptions have been made regarding the composition, temperature, and pressure in the Jovian atmosphere.

Composition and Pressure

The atmosphere of Jupiter, above the clouds, contains 150 m-atm of CH_4 and 7 m-atm of NH_3 (Kuiper, 1952). Recently Spinrad and Trafton (1963) estimated the pressure at the cloudtops to be $\sim$3 atm. This pressure value indicates that the amounts of CH_4 and NH_3 quoted above would account for less than 1 percent of the total atmosphere of Jupiter, indicating that more than 99 percent of the atmosphere may be composed of hydrogen and helium with, perhaps, some traces of neon (Gallet, 1962).

The exact abundances of hydrogen and helium in the atmosphere of Jupiter are, at the present time, a subject of considerable controversy (Kuiper, 1952; Urey, 1959; Öpik, 1962; Spinrad and Trafton, 1963; Foltz and Rank, 1963). The controversy lies in the quantitative interpretation of the quadrupole lines of molecular hydrogen observed in the high dispersion spectra of Jupiter (Kiess, Corliss, and Kiess, 1960; and Spinrad and Trafton, 1963). The estimates of the amount of H_2 present in the atmosphere of Jupiter above the clouds vary from 5.5 km atm (Zabriskie, 1960) to $\sim$200 km atm (Foltz and Rank, 1963).

For the calculations of the structure of the upper atmosphere of Jupiter, therefore, two extreme models for the composition have been considered. In the first, the hydrogen/helium ratio by volume is 20:1 (Urey, 1959), while in the second, it is as low as 0.03:1 (Öpik, 1962). The mean molecular weight of the atmosphere of Jupiter above the clouds in these two models

*Published in *Icarus*, 3(4):311–322, November 1964.
†Airborne Instruments Laboratory, Long Island, New York.
‡Goddard Space Flight Center, and New York University, New York.

would be 2.2 and 3.95. respectively. The pressure at the cloudtop, for both cases, has been assumed to be 3 atm.

Temperature

Murray and Wildey (1963) have recently reported a temperature of 128° K measured in the 8–12 μ atmospheric "window," confirming the older measurements (Menzel *et al.*, 1926) which gave a value of 130° K.

Estimates of the temperature at the Jovian cloudtop have also been made by measuring the radiation emitted by the planet in the microwave regions. Thornton and Welch (1963), from their measurements made at $\lambda = 8.35$ mm, estimate a cloudtop temperature of $144 \pm 23°$ K. Radio-brightness temperatures at cm wavelength regions are contaminated with nonthermal radiation from the planet and therefore cannot be used for a reliable estimate of the cloudtop temperatures.

Because of the presence of 700 cm-atm of NH_3 in the atmosphere of Jupiter (Kuiper, 1952), the levels in the atmosphere to which the infrared and the microwave measurements refer are not reliably known. NH_3 has a strong absorption band in the 8–12 μ interval and therefore the temperatures observed in the infrared may correspond to a level high up in the atmosphere. At $\lambda = 8.35$ mm the absorption due to NH_3 is much less, and therefore the level to which the temperature measured at this wavelength would refer may be deep in the atmosphere and near the cloud level.

Several authors (Öpik, 1962; Kuiper, 1952; Gallet, private communication) have given estimates of the cloudtop temperature on the assumption that the clouds are composed of NH_3 crystals and therefore they are in phase equilibrium with the amount of NH_3 vapor above. On this basis, Öpik has suggested a temperature of 156° K at the cloudtop.

VERTICAL DISTRIBUTION OF TEMPERATURE

In order to calculate the vertical distribution of temperature above the clouds, it is assumed that the atmosphere is in radiative equilibrium. With the Eddington's approximation and assuming a frequency-independent absorption (gray atmosphere), the equation of radiative transfer gives the following expression for the distribution of temperature with altitude:

$$T_z^4 = T_e^4\left(\tfrac{1}{2} + \tfrac{3}{4}\tau_z\right). \qquad (1)$$

Here T_z is the temperature at any level z, T_e is the effective blackbody temperature of the planet, and τ_z is the average infrared optical thickness of the atmosphere above the level z.

In the case of Jupiter, it is assumed that $z = 0$ at the cloudtop where $\tau_z = \tau_0 =$ the total optical thickness of the atmosphere above the clouds.

The absorption in the infrared is due to NH_3 and CH_4. In Fig. 1 is shown the absorption spectrum of NH_3 for the wavelengths between 7 and 60 μ, based on the band intensities for the vibration–rotation bands of ν_2 and ν_4 and the line intensities of the rotation lines recently tabulated by Gille (1964). The absorption has been computed for a mean pressure of 1.5 atm and a mean temperature of 150° K for the total amount of NH_3 of 7 m-atm (Kuiper, 1952). The only absorption band of CH_4 which is significant in this spectral interval is the 7.7-μ vibration–rotation band, and for the measured amount of 150 m-atm, the atmosphere of Jupiter is completely opaque in this spectral interval.

The absorption for the total amounts of NH_3 and CH_4 in the atmosphere has been folded into the 150° K blackbody energy curve and the net energy emitted is shown as the dashed line curve in Fig. 1. The mean transmittance of the Jovian atmosphere over the entire spectral region is 52 percent, giving the total optical thickness of the atmosphere $\tau_0 = 0.66$.

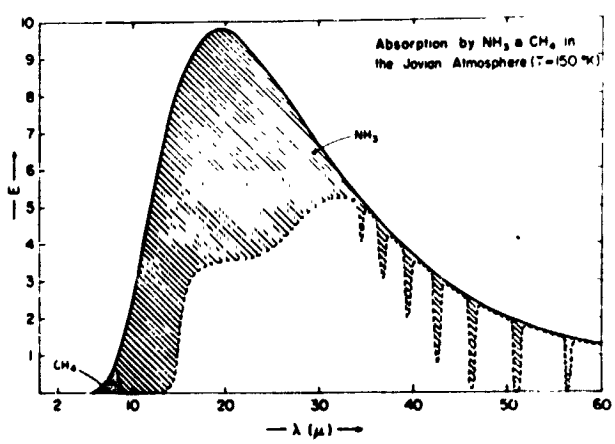


FIGURE 1.—Transmission of infrared radiation through the Jovian atmosphere, composed of 150 m-atm of CH_4 and 7m-atm of NH_3 at 150°K.

Above the cloud level, the opacity (τ_z) will, at any level z, be given by

$$\tau_z = \tau_0 e^{-z/H}, \quad (2)$$

where H is the atmospheric scale height in km, and z is the height in km above the clouds.

We now assume that the 128° K temperature measured by Murray and Wildey (1963) refers to a level high up in the atmosphere where the optical thickness of NH_3 in the 8–12 μ spectral interval has decreased to a value of unity. At this level the total optical thickness of the atmosphere above will be <0.1. We therefore solve Eq. (1) for T_e putting $T_z = 128°$ K and $\tau_z = 0$. The effective temperature of the planet thus calculated is 153° K. Equations (1) and (2) are now combined to obtain the temperature profile of the atmosphere above the clouds.

For the total optical thickness of 0.66, the cloudtop temperature is also found to be 153° K. For the vertical distribution of the temperature, calculations have been made for the two model atmospheres and the results are shown on the left-hand side of Fig. 2.

From this figure it may be noted that the 128° K temperature measured by Murray and Wildey corresponds to the level of 35 or 70 km, depending on which model is being considered. This is consistent with the fact that Murray and Wildey observed a very slight limb darkening, which denotes the existence of a very small temperature gradient at these levels.

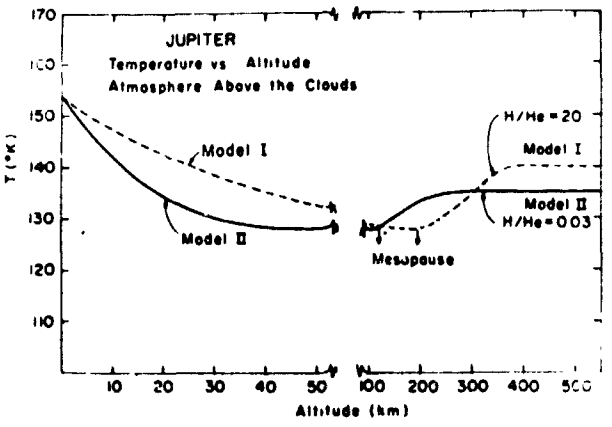


FIGURE 2.—The vertical distribution of temperature in the atmosphere of Jupiter above the clouds for the two model atmospheres.

Also to be noted is the probability that the effective temperature of the planet Jupiter may be about 153° K and not 102° K. The latter should be the effective temperature if the planet were in radiative equilibrium with the incoming solar radiation. The total flux radiated from Jupiter, for an effective temperature of 153° K, is 3.1×10^4 ergs cm^{-2} sec^{-1}. The incoming radiation flux from the Sun for a Jovian albedo of 0.47 is, however, only 0.8×10^4 ergs cm^{-2} sec^{-1}. The difference of 2.3×10^4 ergs cm^{-2} sec^{-1} is probably the radiation being emitted from the interior of the planet.

A rough estimate of the gravitational potential energy released by Jupiter ($3M^2G/5R$, where M and R are the mass and radius of the planet) indicates that if this energy were uniformly distributed over a period of 10^{10} years, the present flux from the interior of the planet could be as high as 3.2×10^5 ergs cm^{-2} sec^{-1}. The rate of the release of energy due to the gravitational contraction of the planet has certainly not remained uniform over this length of time, and the luminosity of the planet is probably much less now than in the earlier period of the planetary history. Our calculations indicate that if the present rate of radiation from the interior of the planet were only 1/12 of its average value, the relatively higher temperatures observed in the atmosphere of Jupiter could be explained by the release of energy due to the gravitational contraction.

In the higher atmosphere of Jupiter, the temperature profile is modified by the deposition of the solar untraviolet energy. We shall therefore assume that the isothermal region extends only up to a level defined as the "mesopause." Above this level the absorption of solar ultraviolet radiation by hydrogen and helium is very significant and part of the energy deposited at that level will heat the atmosphere above the isothermal value of 128° K. This increase in temperature in the "exosphere" gives rise to a positive temperature gradient and permits the energy deposited in the upper atmosphere to be conducted downwards. The "mesopause" is defined as the altitude at which all the energy conducted downwards from the higher atmosphere is radiated away. In the case of Jupiter CH_4 and NH_3 are the radiating molecules.

Following the method described by Bates (1951) and Chamberlain (1962), we find that the mesopause is at a level where the total number density is $\sim 3\times10^{14}$ cm^{-3}. These calculations are made on the assumption that "vibrational relaxation" exists at this altitude, i.e., collisions are rare during the radiative lifetime of the molecules.

The levels at which the number density equals 3×10^{11} cm^{-3} correspond to 104 and 187 km above the clouds for the two model atmospheres as shown in Fig. 2.

Below the level of the mesopause, the atmosphere is thoroughly mixed and the mean molecular weight does not vary with the altitude. Above the level of the mesopause, the atmosphere is assumed to be in diffusive separation and the vertical distribution of each constituent is calculated for its own molecular weight. This assumption is based on the analogy with the Earth, and is supported by the suggestion of Spinrad and Trafton (1963) that strong circulation currents leading to thorough mixing may exist in the atmosphere above the clouds.

The rise in temperature in the atmosphere above the mesopause level is calculated by asuming that 50 percent of the ultraviolet radiation absorbed in the upper atmosphere ($\lambda<912$Å) goes into heating. The method used for these calculations is that described by Nicolet (1959), and for the two model atmospheres, exospheric temperatures of 140° and 135° K, respectively, are obtained (Fig. 2).

This considerably smaller increase in temperature on Jupiter, as compared with the exospheric temperature of the Earth, which is of the order of 1500° K, results because (1) the intensity of solar radiation at Jupiter is 27 times less than at the distance of the Earth; (2) the conductivity for the gaseous mixture on Jupiter is higher than for the O_2–N_2 mixture for the Earth; and (3) the density scale heights in the upper atmosphere are $\sim$5 times less than in the upper atmosphere of the Earth.

IONOSPHERE

Using the temperature distribution in the upper atmosphere of Jupiter shown in Fig. 2, we can now proceed to investigate the properties of an ionospheric layer which may be formed by the photoionization of the atmospheric constituents by the solar ultraviolet radiation.

In an atmosphere mainly composed of hydrogen and helium, the equilibrium electron density in the ionosphere will be determined by a number of ionizing, recombination, and charge exchange processes. The most important of these reactions are shown in Table 1 and have been considered in the present study.

From Table 1 it is readily seen that though only the ions of He and H_2 will be formed by direction photoionization of these gases, H_3^+, H^+, and HeH^+ can also be expected to be produced by charge transfer mechanisms. The reaction (6) to form H_3^+ is very rapid, so that, in equilibrium, little H_2^+ can be expected. HeH^+ is also formed fairly rapidly [reactions (5) and (8)], further decreasing the concentration of H_2^+ and also of He^+ ions. Hydrogen atoms are generated by these reactions at a fairly fast rate and H^+ ions can be produced by direct photoionization. Because of the slow recombination rate of H^+ [reaction (13)], it becomes an important constituent of the ionosphere. We therefore first calculate the distribution of H atoms in the upper atmosphere of Jupiter and then proceed to the computation of equilibrium electron densities in the ionosphere of Jupiter.

Distribution of Hydrogen

Hydrogen atoms generated at higher altitudes are assumed to diffuse downward and recombine by three-body recombination processes [reaction (14)] at lower altitudes, the third body being either H_2 or He depending on the atmospheric model being considered. Since the recombination time is long compared with the time to diffuse through the atmosphere, H atoms may be considered in diffusive equilibrium, distributed exponentially at high altitudes but with the density decreasing rapidly at lower levels where recombination is significant. In equilibrium, the total hydrogen atoms generated each second must equal the total lost due to recombination.

From reactions (5), (6), (8), 10, and (12) of Table 1, it is seen that every ionized H_2 molecule will form two H atoms, and also every ionized He atom will result in two H atoms. The total rate of formation of atomic hydrogen, J(H), can be obtained by integrating the ionization rates of H_2 and He over a vertical column.

TABLE 1.—*Hydrogen and Helium Reactions*

Process	Reaction	Cross section or rate symbol and value	Reference	Remarks
Photoionization	(1) $H_2+h\nu\rightarrow H_2^++e$	$\sigma(H_2)$	*a,b*	$\lambda\leq 804$ Å
	(2) $H+h\nu\rightarrow H^++e$	$\sigma(H)$	*c*	$\lambda\leq 912$ Å
	(3) $He+h\nu\rightarrow He^++e$	$\sigma(He)$	*d*	$\lambda\leq 504$ Å
Charge transfer and ion–molecule reactions	(4) $He^++H_2\rightarrow He+H_2^+$	$\alpha_4=10^{-13}$	*f*	Estimated
	(5) $He^++H_2\rightarrow HeH^++H$	$\alpha_5=10^{-12}$	*e*	Upper limit
	(6) $H_2^++H_2\rightarrow H_3^++H$	$\alpha_6=5\times10^{-10}$	*e*	
	(7) $H_2^++H\rightarrow H^++H_2$	$\alpha_7=10^{-10}$	*f,o*	Not measured, unimportant relative to (6)
	(8) $H_2^++He\rightarrow HeH^++H$	$\alpha_8=10^{-10}$	*e*	
Recombination	(9) $H_2^++e\rightarrow H'+H''$	$\alpha_9=10^{-6}$	*c,f–n*	Upper limit
	(10) $HeH^++e\rightarrow He+H$	$\alpha_{10}=10^{-6}$	*f*	Estimated
	(11) $He^++e\rightarrow He$	$\alpha_{11}=10^{-12}$	*c*	Theoretical, He_2^+ negligible at low pressures
	(12) $H_3^++e\rightarrow H_2+H$	$\alpha_{12}=10^{-8}$	*f*	Estimated
	(13) $H^++e\rightarrow H$	$\alpha_{13}=10^{-12}$	*c*	Theoretical
	(14) $H+H+X\rightarrow H_2+X$	$\alpha_{14}=10^{-32}$	*f,h,p*	X may be H_2 or He
	(15) $H+e\rightarrow H^-+h_\nu$	$\alpha_{15}=10^{-16}$	*q*	Rate to small too be significant

[a] Po Lee and G. L. Weissler, *Astrophys. J.* **115**, 570 (1952).
[b] N. Wainfan, W. C. Walker, and G. L. Weissler, *Phys. Rev.* **99**, 542 (1955).
[c] D. R. Bates, "Atomic and Molecular Processes." Academic Press, New York, 1962.
[d] Po Lee and G. L. Weissler, *Phys. Rev.* **99**, 540 (1955).
[e] C. F. Giese and W. B. Maier II, *J. Chem. Phys.* **39**, 739 (1963).
[f] A. Dalgarno, private communication, 1963.
[g] E. Bauer and T. Wu, *Can. J. Phys.* **34**, 1436 (1956).
[h] F. R. Zabriskie, Dissertation, Princeton University, Princeton, New Jersey, 1960.
[i] M. A. Biondi and S. C. Brown, *Phys. Rev.* **76**, 1697 (1949).
[j] J. M. Richardson and R. B. Nolt, *Phys. Rev.* **81**, 1531 (1951).
[k] L. J. Varnerin, *Phys. Rev.* **84**, 563 (1951).
[l] K. B. Perrson and S. C. Brown, *Phys. Rev.* **100**, 729 (1955).
[m] N. A. Popov and E. A. Afaneseva, *Soviet Phys.–Tech. Phys.* (*English Transl.*) **4**, 764 (1960).
[n] L. B. Loeb, "Handbuch der Physik," Vol. 21, p. 471. Springer-Verlag, Berlin, 1955–60.
[o] C. F. Giese, private communication.
[p] E. Rabinowitch, *Trans. Faraday Soc.* **33**, 283 (1937).
[q] A. Dalgarno and A. E. Kingston, *The Observatory* **83**, 39 (1963).

Assuming that the solar flux of wavelengths <804Å is completely absorbed above the mesopause in the ionization of H_2 and He, then $J(H)=2\times$ the total number of photons incident on a square centimeter column. At the distance of Jupiter, the solar flux of $\lambda<804$Å can be obtained by reducing by a factor of 27 the observed values of the solar flux at the distance of the Earth recently published by Hinterreger (1961). This gives $J(H)=2\times10^9$ atoms cm^{-2} sec^{-1}.

Having obtained the total rate of production of atomic hydrogen, we can now determine the vertical distribution of this gas following the method described by Zabriskie (1960). This method consists of solving the diffusion equation given by Chapman and Cowling (1939) with the following boundary conditions: (1) At higher altitudes the diffusion current equals the rate of formation of H. This is a reasonable assumption since recombination will occur mainly at lower altitudes. (2) At some lower altitude both the diffusion current and the density of atomic hydrogen must go to zero. This is probably also reasonable since recombination will simultaneously reduce both of these quantities.

Using these boundary conditions, Zabriskie derives a simple expression to compute the vertical distribution of atomic hydrogen. His

formulation has been applied to the two model atmospheres that have been studied, and the resulting distributions are shown in Fig. 3.

It is noted in Fig. 3 that the maximum density of atomic hydrogen is of the order of 5×10^9 cm^{-3} and occurs at a level 10–20 km above the mesopause. It is also noted that the density of atomic hydrogen increases rapidly above the mesopause and then decreases exponentially with a scale height of approximately 44 km for the two cases.

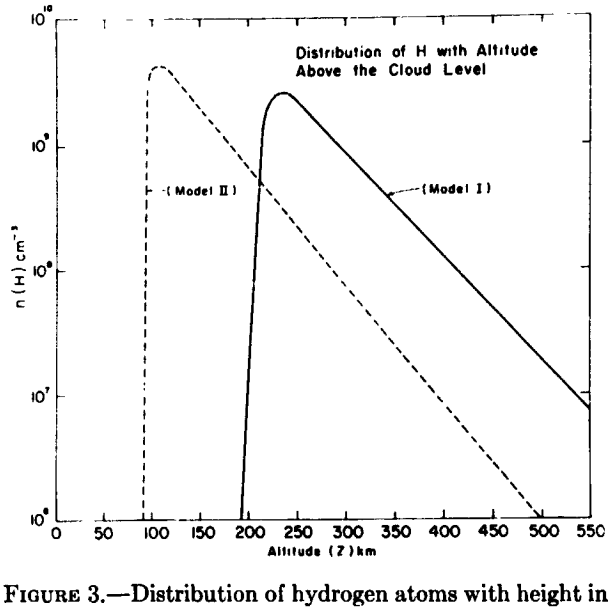


FIGURE 3.—Distribution of hydrogen atoms with height in the upper atmosphere of Jupiter.

Although the two model atmospheres are two extreme cases in which the H/He ratio varies by a factor of 600, it is interesting to note that the equilibrium number density of atomic hydrogen for the two cases is approximately the same. The amount of atomic hydrogen however, even at the level of its peak density, is some 10^4 times less than the total number of particles.

Since it has been assumed that diffusive separation applies above the mesopause, the densities of molecular hydrogen and of helium at these altitudes are readily found from the temperature distribution and they will decrease more rapidly than the density of atomic hydrogen. Consequently, at an altitude of roughly 400 km above the mesopause for Model I, and at 240 km above the mesopause for Model II, atomic hydrogen is the dominant constituent.

Ionization

Having obtained the distributions of He, H_2, and H with altitude, it is now possible to compute the equilibrium ion density distribution as a function of height.

The ionization equations for the reactions listed in Table 1 are as follows:

$$\dot{n}_e = Q_e(H_2)+Q_e(He)+Q_e(H)-\alpha_9 n\ i(H_2^+) - \alpha_{10}n_e n(HeH^+)-\alpha_{11}n_e n(He^+) - \alpha_{12}n_e n(H_3^+)-\alpha_{13}n_e n(H^+) \quad (3)$$

$$\dot{n}(He^+) = Q_e(He)-\alpha_{11}n_e n(He^+) - \alpha_4 n(He^+)n(H_2) - \alpha_5 n(He^+)n(H_2) \quad (4)$$

$$\dot{n}(H_2^+) = Q_e(H_2)-\alpha_9 n_e n(H_2^+)-\alpha_8 n(He)n(H_2^+) - \alpha_6 n(H_2)n(H_2^+)-\alpha_7 n(H)n(H_2^+) + \alpha_4 n(H_2)n(He^+) \quad (5)$$

$$\dot{n}(HeH^+) = \alpha_5 n(H_2)n(He^+)+\alpha_8 n(He)n(H_2^+) - \alpha_{10}n_e n(HeH^+) \quad (6)$$

$$\dot{n}(H_3^+) = \alpha_6 n(H_2)n(H_2^+)-\alpha_{12}n_e n(H_3^+) \quad (7)$$

$$\dot{n}(H^+) = Q_e(H)+\alpha_7 n(H_2^+)n(H)-\alpha_{13}n_e n(H^+) \quad (8)$$

and the total ion density

$$n_e = n(He^+)+n(H_2^+)+n(HeH^+) + n(H_3^+)+n(H^+). \quad (9)$$

In these equations $n(X)$ is the density of constituent X at any altitude z, α_n's are the corresponding values of reaction rates given in Table 1, $Q_e(X)$ is the photoionization rate, and n_e is the electron density in cm^{-3}.

The rate of formation of H^- from reaction (15) is so slow that it has been neglected in the above equations. He_2^+ is also neglected since it is too unstable to contribute significantly. In addition, it is found that there is little He^+ to produce He_2^+ by the three-body reaction.

Since $n(He^+)$ and $n(H_2^+)$ will be small in comparison with $n(H_3^+)$ or $n(HeH^+)$ and n_e with respect to $n(H_2)$ or $n(He)$, we can neglect certain terms in these equations. In Eq. (4) one can neglect the term involving α_{11} and in Eq. (5), the term with α_9. Also reaction (4) is unimportant compared to reaction (6). This is because $n(He^+)$ will not be sufficiently larger than $n(H_2^+)$ to make up for the difference in rate coefficients.

From these equations we find that the contribution of $n(He^+)$ and $n(H_2^+)$ to the total electron

density is relatively small and therefore we can safely neglect these terms in Eq. (9). The total equilibrium electron density is therefore

$$n_e = n(H^+) + n(H_3^+) + n(HeH^+). \quad (10)$$

On solving the resulting set of simultaneous equations and inserting the values of the reaction rates from Table 1, we find

$$n_e = \left\{ 10_8 \left[Q_e(He) + \frac{\{5n(H_2)+n(He)\}Q_e(H_2)}{5n(H_2)+n(H)+n(He)} \right] + 10^{12} \left[Q_e(H) + \frac{n(H)Q_e(H_2)}{5n(H_2)+n(H)} \right] \right\}^{1/2}. \quad (11)$$

In order to evaluate the concentration of electron density as a function of altitude, we first determine the rates of production of H^+, H_2^+, and He^+.

The rate of production of ions is

$$Q_e(X) = \tfrac{1}{4} \int_{\lambda 1}^{\lambda 2} n_z(X)\sigma_\lambda(X)\mathfrak{F}_z(\lambda)d\lambda, \quad (12)$$

where $n_z(X)$ is the number density of the ionizing constituent X at the level z, $\sigma_\lambda(X)$ is the ionization cross section for the constituent at the wavelength λ, and $\mathfrak{F}_z(\lambda)$ is the solar radiation flux at wavelength λ reaching the level z, after having been attenuated by all the ionizing constituents above. The factor 1/4 is due to the fast rotation of Jupiter, so the evaluation of Q_e gives a mean value of the ion production rate over the whole globe. Equation (12) is evaluated by approximating the ionization cross section over the spectral interval and also by taking an average value of the solar radiation $\mathfrak{F}$ over the relevant spectral regions. The values used are shown in the Table 2.

The ionization sections and the corresponding wavelengths for He, H_2, and H are given in Table 2. Although there is a significant overlapping in the relevant spectral regions for the ionization of the individual gases, it may, however, be noted that H_2 is the principal absorber of the far ultraviolet radiation, even when He is the principal constituent, due to the following reasons: (a) The density scale height is large compared to He by a factor of 2, and (b) at altitudes of ~180 km and above, H_2 is the principal constituent. (At much higher altitudes where H is the major constituent, it is optically too thin to be the principal absorber.)

TABLE 2.—*Mean Cross Sections and Fluxes*

$\bar{\sigma}(H_2)$	($\lambda < 504$)	1.7×10^{-18} cm²
$\bar{\sigma}(H_2)$	($804 > \lambda > 504$)	6.1×10^{-18} cm²
$\bar{\sigma}$(He)	($\lambda < 504$)	3.3×10^{-18} cm²
$\bar{\sigma}$(H)	($912 > \lambda > 804$)	5.0×10^{-18} cm²
$\bar{\sigma}$(H)	($804 > \lambda > 504$)	3.0×10^{-18} cm²
$\bar{\sigma}$(H)	($\lambda < 504$)	6.0×10^{-19} cm²
$\bar{\mathfrak{F}}\infty$	($\lambda < 504$)	6.7×10^{8} ph/cm²/sec[a]
$\bar{\mathfrak{F}}\infty$	($804 > \lambda > 504$)	4.0×10^{9} ph/cm²/sec[a]

[a] Hinterreger, 1961.

In Table 2 the values of $\bar{\sigma}$ have been calculated for the relevant spectral interval for the various overlapping bands. $\bar{\mathfrak{F}}\infty$, the total incident solar flux in a wavelength band, is calculated for the distance of Jupiter from the values obtained recently by Hinterreger (1961) at the distance of the Earth.

The ionization cross sections of atomic hydrogen are theoretical and have been obtained from Bates (1962). References for $\sigma(H_2)$ and σ(He) are given in Table 1. From the numerical values listed in Table 2, we calculate the production rates of the H, He, and H_2 ions from Eq. (12), using a mean approximation over the band, and the results are shown in Figs. 4 and 5.

It may be noted that the rate of production of H^+ is 10 times less than the rate of production of H_2^+ in both our model atmospheres. As the recombination rate of H^+ is much slower than that

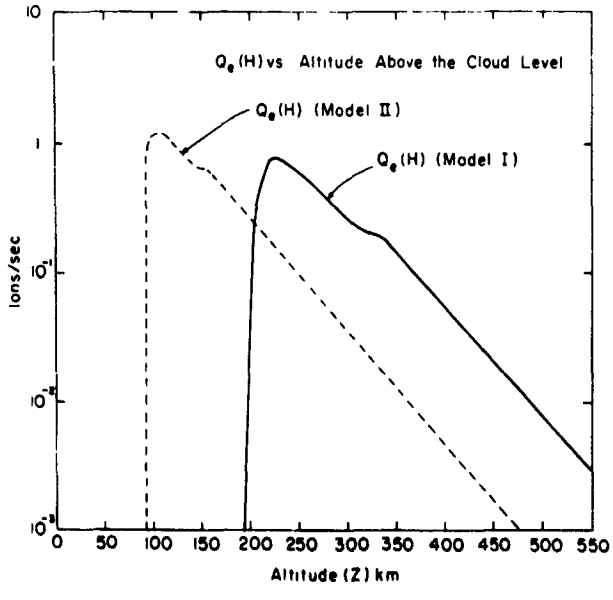


FIGURE 4.—Rate of production of H^+ in the upper atmosphere of Jupiter

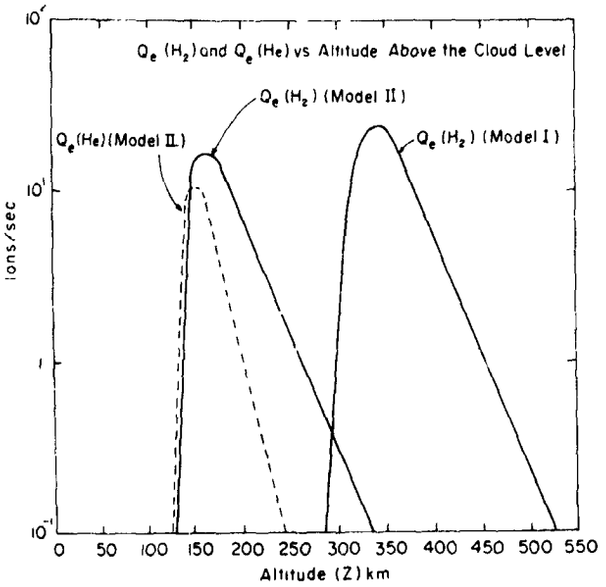


FIGURE 5.—Rate of production of H_2^+ and He^+ for Model I and Model II atmospheres.

of H_2^+ [which is via H_3^+ or HeH^+, or dissociatively via reaction (9), Table 1], it is expected that the atomic hydrogen will be a significant contributor to the total electron density. In the case where He makes up 97 percent of the total atmosphere, Fig. 4 shows that even then the rate of production of He ions is less than H_2^+, though the amount of hydrogen in the atmosphere has been assumed to be only 3 percent. This circumstance is due to the screening of the radiation less than 504 Å by the H_2 above.

Using these computed values of $Q_e(H_2)$, $Q_e(He)$, and $Q_e(H)$, we can now calculate the total electron densities as a function of altitude from Eq. (11). The results are shown in Fig. 6, which is a plot of n_e vs altitude for both models.

When H/He=20, the peak electron density is 7×10^5 cm^{-3} and occurs at ~200 km above the cloud level, with a secondary maximum of about 3×10^5 at ~350 km altitude. In the Model II atmosphere, the maximum electron density occurs at 110 km with a peak value of about 10^6 cm^{-3}. The extent of the ionosphere in the two cases is of the order of ~ 500 km.

The major constituent of the ionosphere for both models is H^+, contributing more than 99 percent of the ions. The maximum concentration of H_3^+ and HeH^+, for the two cases, is ~10^3 to 10^4 cm^{-3}. H_2^+ and He^+ ions have a peak concentration even less, ~10^2 cm^{-3}.

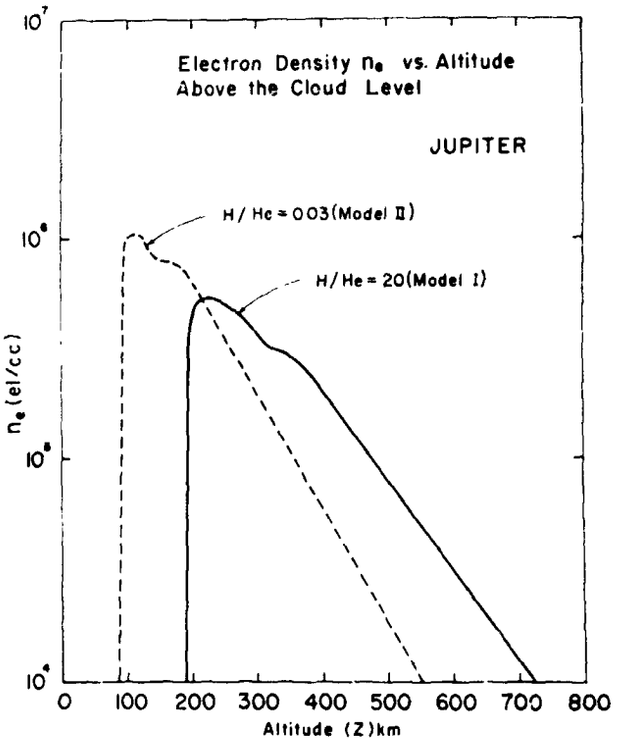


FIGURE 6.—Distribution of total equilibrium electron density as a function of altitude in the upper atmosphere of Jupiter.

RESULTS

The following are some of the important results of the preceding calculations regarding the atmosphere of Jupiter:

(1) If the cloudtop temperature of Jupiter is 153° K and the amounts of NH_3 and CH_4 above the clouds are 700 and 15 000 cm-atm, then the total average optical thickness of the atmosphere in the infrared is 0.66.

(2) Assuming the atmosphere to be in radiative equilibrium, the approximate temperature profile calculated by a simplified version of the equation of radiative transfer for a gray atmosphere indicates that the temperature near the cloudtop decreases with altitude slowly to reach an asymptotic value of 128° K at approximately 35 km for Model II and at approximately 70 km for Model I (Fig. 2). These temperature profiles are consistent with the recently measured temperatures of the planets in the far infrared and in the mm region of the planetary spectrum.

(3) The temperature of the upper atmosphere of Jupiter has been calculated, assuming that 50

percent of the ultraviolet energy deposited in the thermosphere is converted to heat and the energy is conducted downwards to the mesopause to be radiated away by CH_4 and NH_3. The mesopause levels for Model I and II are at 187 km and 106 km, respectively, and the corresponding exospheric temperatures are found to be 140° and 135° K. The results are summarized in Fig. 2.

(4) Calculations of the equilibrium electron density for Model I (H/He=20) result in an ionospheric layer with a maximum density of $\sim 7\times 10^5$ electrons cm^{-3} at an altitude of 200 km above the cloud level. A slightly higher electron density (1.0×10^6 cm^{-3}) is achieved in the second case when H/He=0.03, the peak being at 120 km above the clouds.

(5) The major constituent of the ionosphere in both cases in H^+.

DISCUSSION

Results on the electron density distribution in the upper atmosphere of Jupiter may be compared with the earlier studies of the Jovian ionosphere.

The most comprehensive study to date of the possible structure of the ionosphere of Jupiter has been made by Zabriskie (1960). His main conclusion is that a Jovian ionosphere of peak electron density greater than 10^5 cm^{-3} is highly unlikely. The disagreement of our results with those of Zabriskie is probably due to the following differences in our basic assumptions and approach to the problem:

(1) The exospheric temperatures assumed arbitrarily by Zabriskie were, for his two cases, 600° K and an isothermal atmosphere at 100° K.

(2) Only hydrogen was assumed to be the ionizable constituent of the Jovian atmosphere.

(3) Zabriskie underestimates the ionization of atomic hydrogen by neglecting the ionizing radiation band of 804–912 Å, for which he assumes an ionization cross section of 10^{-20} cm^2. According to Bates (1962), however, it may be $\sim 5.0\times 10^{-18}$ cm^2. This enhances the contribution of H^+ to the total electron density.

(4) The solar flux values for the far ultra-violet adopted by Zabriskie are a factor of 2 less than those recently given by Hinterreger (1961) and used in this study.

(5) Zabriskie neglected the reaction $H_3^+ + H_2 \rightarrow H_3^+ + H$ which contributes significantly in the formation of H.

Rishbeth (1959) has made a very rough study of the ionosphere of Jupiter with the main result that an ionosphere mainly composed of H^+ is formed with a maximum density of 10^6 electrons cm^{-3}. This result is very similar to ours, but it is based on the erroneous assumption that the major fraction of H_2 is directly photodissociated into atomic hydrogen which then ionizes to H^+. From the results of Rishbeth, it is not clear at precisely what altitude this maximum occurs.

The photodissociation of H_2 is a very inefficient process (Bates and Nicolet, 1950) and although we are in agreement with Rishbeth for the maximum electron density, our process of formation of atomic hydrogen is through the dissociative recombination of H_3^+ and H_2^+.

Zhelezniakov (1958) has also briefly discussed the formation of a Jovian ionosphere and quotes a maximum electron density value of 10^7 cm^{-3}. This estimate is based on the assumption that the "effective recombination coefficient in the hydrogen layer" is 4×10^{-14} cm^{-3} sec^{-1}, which is incompatible with the values given in Table 1.

It may be mentioned that in our calculations we have neglected the possible dissociation and ionization of methane and ammonia, which may be quite important at the level of the mesopause and may give rise to a secondary ionization peak at that altitude.

The results of this study may have an important bearing on the validity of those hypotheses which invoke the ionosphere of Jupiter as a direct participant in the generation of the decameter radiation. For example, a mechanism has recently been proposed which required electron densities of the order of 10^8 cm^{-3} in the Jovian ionosphere (Chang, 1963).

ACKNOWLEDGMENTS

It is our pleasure to thank Dr. Alex Dalgarno of Queens College, Belfast, Ireland, and Dr. A. G. W. Cameron of the Goddard Institute for Space Studies for numerous valuable discussions and suggestions during this work. We would also like to thank Professor Edwin Salpeter of Cornell University and Dr. Siegfried Bauer of the Goddard Space Flight Center for critically reading the

manuscript. We are grateful to Allan Liebman who helped us with the computations.

REFERENCES

BATES, D. R. (1951). The temperature of the upper atmosphere. *Proc. Phys. Soc. (London)* **B64**, 805–821.

BATES, D. R. (1962). "Atomic and Molecular Processes." Academic Press, New York.

BATES, D. R., and NICOLET, M. (1950). The photochemistry of atmospheric water vapor. *J. Geophys. Res.* **55**, 301–327.

BAUM, W. A., and CODE, A. D. (1953). A photometric observation of the occultation of σ Arietis by Jupiter. *Astron. J.* **58**, 108.

CHAMBERLAIN, J. W. (1962). The upper atmosphere of the planets. *Astrophys. J.* **136**, 582.

CHANG, D. B. (1963). Amplified whistlers as the source of Jupiter's sporadic decameter radiation. *Astrophys. J.* **138**, 1231.

CHAPMAN, S., and COWLING, T. G. (1939). "The Mathemathical Theory of Non-Uniform Gases." Cambridge Univ. Press, London.

FOLTZ, J. V., and RANK, D. H. (1963). Intensity of the quadrupole band of molecular hydrogen. *Astrophys. J.* **138**, 1319.

GALLET, R. (1962). Private communication.

GILLE, J. (1964). Dissertation, Harvard University, Cambridge, Massachusetts.

HINTERREGER, H. E. (1961). Preliminary data on solar extreme ultraviolet radiation in the upper atmosphere. *J. Geophys. Res.* **66**, 2367.

KIESS, C. C., CORLISS, C. H., and KIESS, H. K. (1960). High dispersion spectra of Jupiter. *Astrophys. J.* **132**, 221.

KUIPER, G. P. (1952). Planetary atmospheres and their origin. *In* "Atmospheres of the Earth and Planets." Univ. of Chicago Press, Chicago, Illinois.

LEE, Po, and WEISSLER, G. L. (1952). Absolute absorption of the H_2 continuum. *Astrophys. J.* **115**, 570.

MENZEL, D. H., COBLENTZ, W. W., and LAMPLAND, C. O. (1926). Planetary temperature derived from water-cell transmissions. *Astrophys. J.* **63**, 177.

MURRAY, B. C., and WILDEY, R. L. (1963). Stellar and planetary observations at 10 microns. *Astrophys. J.* **137**, 692.

NICOLET, M. (1959). La Thermosphère. *Ann. Géophys.* **15**, 1.

ÖPIK, E. J. (1962). Jupiter: chemical composition, structure and origin of a giant planet. *Icarus* **1**, 200.

RISHBETH, H. (1959). The ionosphere of Jupiter. *Australian J. Phys.* **12**, 466.

SPINRAD, H., and TRAFTON, L. M. (1963). High dispersion spectra of the outer planets. I. Jupiter in the visual and red. *Icarus* **2**, 19.

THORNTON, D. D., and WELCH, W. J. (1963). 8.35 mm radio emission from Jupiter. *Icarus* **2**, 228.

UREY, H. C. (1959). The atmospheres of the planets. *In* "Handbuch der Physik," Vol. 52, p. 363. Springer-Verlag, Berlin.

ZABRISKIE, F. R. (1960). Dissertation, Princeton University, Princeton, New Jersey.

ZHELEZNIAKOV, V. V. (1958). On the theory of sporadic radio emission from Jupiter. *Soviet Astron.–A.J. (Engl. transl.)* **2**, 2, 230.

N65-32114

THE INFRARED INTERFEROMETER SPECTROMETER EXPERIMENT (IRIS)*

R. A. HANEL AND L. CHANEY†

ACKNOWLEDGMENTS

In addition to the experimenters, the following people made substantial contributions to the effort presented here:

Frank Bartko, Goddard Space Flight Center

Fred Bartman, University of Michigan

Lawrence Byrne, Goddard Space Flight Center

Grady Nichols, Goddard Space Flight Center

Dr. David Wark, United States Weather Bureau

1. SCOPE

The purpose of this report is

to show how significant information on the atmosphere and surface of Mars can be extracted from the thermal-emission spectrum of the planet

to describe the design of an infrared interferometer capable of measuring the thermal-emission spectrum

to document requirements and constraints of the infrared interferometer experiment on a planetary fly-by spacecraft

2. SCIENTIFIC OBJECTIVES

The objective of the infrared interferometer spectrometer (IRIS)‡ experiment is to obtain information on the composition and structure of the atmosphere and surface of the planet Mars. This information will be extracted from the thermal-emission spectrum within the interval of 2000 cm^{-1} (5 microns) to 500 cm^{-1} (20 microns) and possibly to 300 cm^{-1} (33 microns) with a spectral resolution of 5 cm^{-1}. The IRIS experiment is deliberately aimed at a wide spectral range to permit clear identification of many predictable spectral features and, even more important, to allow a search for unexpected phenomena. For a nominal fly-by distance of 5000 km, the resolved area will be a circle of 250-km diameter, approximately equivalent to the resolution now obtained by photographic techniques with ground-based telescopes.

The wide spectral range of the instrument, combined with the relatively high spectral resolution, makes a variety of scientific studies possible; these can be grouped into investigations concerning the atmosphere, and those concerning the planetary surface. In addition, many other inferences can be made from the results of the measurements. The following discussions indicate the wide scope of scientific investigations made possible by this experiment.

2.1 The Thermal-Emission Spectrum

The thermal-emission spectrum of a planetary atmosphere depends on many parameters, of which the most important are the temperature profile, the surface temperature and emissivity, the concentration of atmospheric constituents, and the surface pressure. Under the condition of radiative equilibrium, the temperature profile is completely defined by the parameters mentioned above as well as the solar zenith angle. In theory, if a greater number of measurements of the specific intensity are obtained than the unknown parameters which exist, then all parameters influencing the thermal-emission spectrum can be determined; however, the measurements must be made in independent spectral regions or at different zenith angles. In practice, the extraction of individual parameters is more difficult, and within the limited accuracy of the instruments, different

*Published as *Goddard Space Flight Center Document X-650-64-204*, July 1964.

†University of Michigan.

‡IRIS is the Greek word for rainbow.

parameters can be determined only within certain limits.

To study the influences of the various parameters on the thermal-emission spectrum and to see how these parameters can be determined from the measurements, some preliminary calculations have been made (Figure 1, Table 1).

The radiative equilibrium temperature distribution for a simulated Martian atmosphere has been calculated, using technique described in Reference 1. Some cases are shown in which the temperature distributions require modification due to convective energy transport. Five individual cases were studied:

Case 1 corresponds to the estimated maximum conditions which could possibly occur near the subsolar point. The surface temperature was assumed to be 320°K, slightly above the maximum measured from the earth in the atmospheric window. The calculated temperature profile is shown in Figure 1.

Case 2 corresponds to the expected nominal condition in the late afternoon local Martian time and $\pm 45°$ latitude.

Cases 3 and 4 have been calculated for higher surface pressures.

Case 5 is representative of the coldest conditions expected during the Martian night.

All calculations were carried out for an N_2–CO_2 atmosphere with trace amounts of H_2O. Other gaseous constituents that may be detected (by techniques described later) may influence the equilibrium temperatures, and will have to be considered in addition to the CO_2 and the small amount of water vapor. Furthermore, these calculations assume only pressure broadening of the spectral lines. In the low-pressure conditions to be expected on Mars, Doppler broadening will be an important factor and must be considered in a more refined treatment.

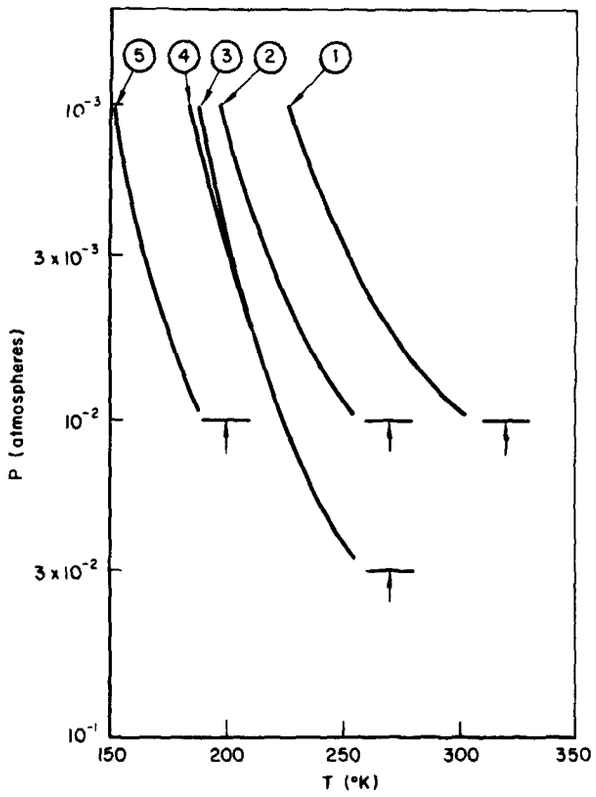


FIGURE 1.—Radiative equilibrium temperatures in a simulated N_2–CO_2 atmosphere with trace amounts of water.

TABLE 1.—*Summary of Assumptions Made in the Calculations of Temperatures for the Five Cases Shown in Figure 1.*

Case	1	2	3	4	5
Solar zenith angle	0°	60°	60°	60°	night
Surface temperature (degrees Kelvin)	320°	270°	270°	270°	200°
Surface pressure (atmospheres)	.01	.01	.03	.03	.01
CO_2 concentration	.25	.25	.10	.25	.25
H_2O concentration	$.4 \cdot 10^{-4}$	$.4 \cdot 10^{-4}$	$.4 \cdot 10^{-4}$	$.4 \cdot 10^{-4}$	$.4 \cdot 10^{-4}$
Surface emissivity	1	1	1	1	1
Surface reflection for near-infrared	.25	.25	.25	.25	.25

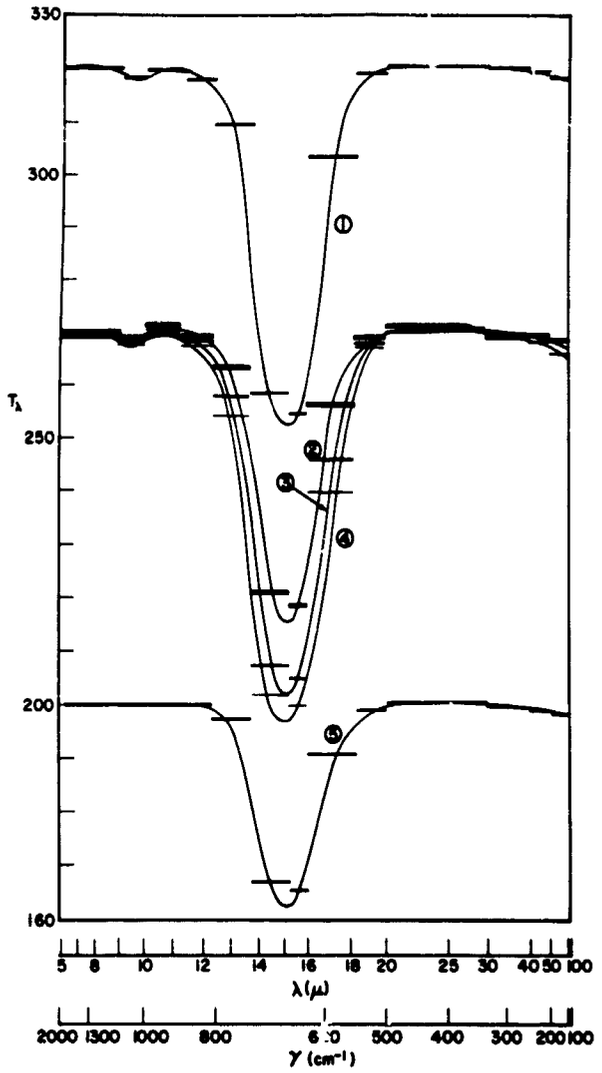


FIGURE 2.—Low-resolution thermal spectra calculated for temperature profiles shown in Figure 1.

Figure 2 shows the corresponding thermal spectra. The spectral resolution of the diagram is very low, varying between approximately 50 and 100 wavenumbers in the most interesting portions. The abscissa is proportional to the fraction of the energy available from a blackbody at 270°K. The wavenumber and the wavelength scales are nonlinear. The strong 15-micron CO_2 band dominates the central part in all spectra. Although the temperature profiles of Cases 3 and 4 are almost the same, the higher abundance of CO_2 in Case 4 causes lower temperatures in the 15-micron band. A comparision between Cases 2 and 4 clearly show the strong influence of the surface pressure, evidence that the surface pressure may be extracted from the emission spectrum.

2.2 Atmospheric Composition

Identification of gases known to be present in the Martian atmosphere (CO_2, H_2O) and estimates of their abundances are among the objectives of the IRIS experiment. But an even more important objective is the possible detection; as well as the establishment of better upper limits, of yet unidentified but suspected constituents (N_2O, NO, NO_2, CH_4, C_2H_4, C_2H_6, NH_3, O_3).

The identification of gases is based on the characteristic absorption spectrum of each gas. Gases in the atmosphere absorb thermal radiation from the surface and, according to Kirchhoff's law, re-emit blackbody radiation as determined by their own temperature. As most atmospheric temperatures will be somewhat lower than surface temperatures, the atmospheric spectra will appear as absorption spectra. However, this is not necessarily true under all conditions; for example, under the condition of a strong temperature inversion in the atmosphere, or in spectral regions where the surface emissivity is low, a reversal may take place and the spectrum of a gas may appear in emission. These complications do not exist in the interpretation of spectra in the visible or near-infrared (say below 2 microns) where thermal emission is negligible compared to reflected solar radiation.

The long-wave infrared spectra may be more difficult to interpret, and the theory of radiative transfer must be used extensively in a proper treatment of the long-wave spectra; but the thermal-emission spectra yield additional information, such as the temperature profile, which will be discussed later. Furthermore, many atmospheric constituents which might be of biological interest, such as CH_4, are relatively complex molecules which show strong vibrational and rotational bands in the middle and far infrared but very little detectable absorption phenomena in the visible part of the spectrum. For these reasons, the analysis of atmospheric contituents in the far infrared is extremely interesting in spite of the difficulties associated with their interpretation.

Figure 3 shows the spectra of constituents which may be expected in the atmosphere of Mars. The spectra are drawn schematically only, without the many details which can be recognized with a resolution of 5 wavenumbers. The abundance of the gases was taken according to the

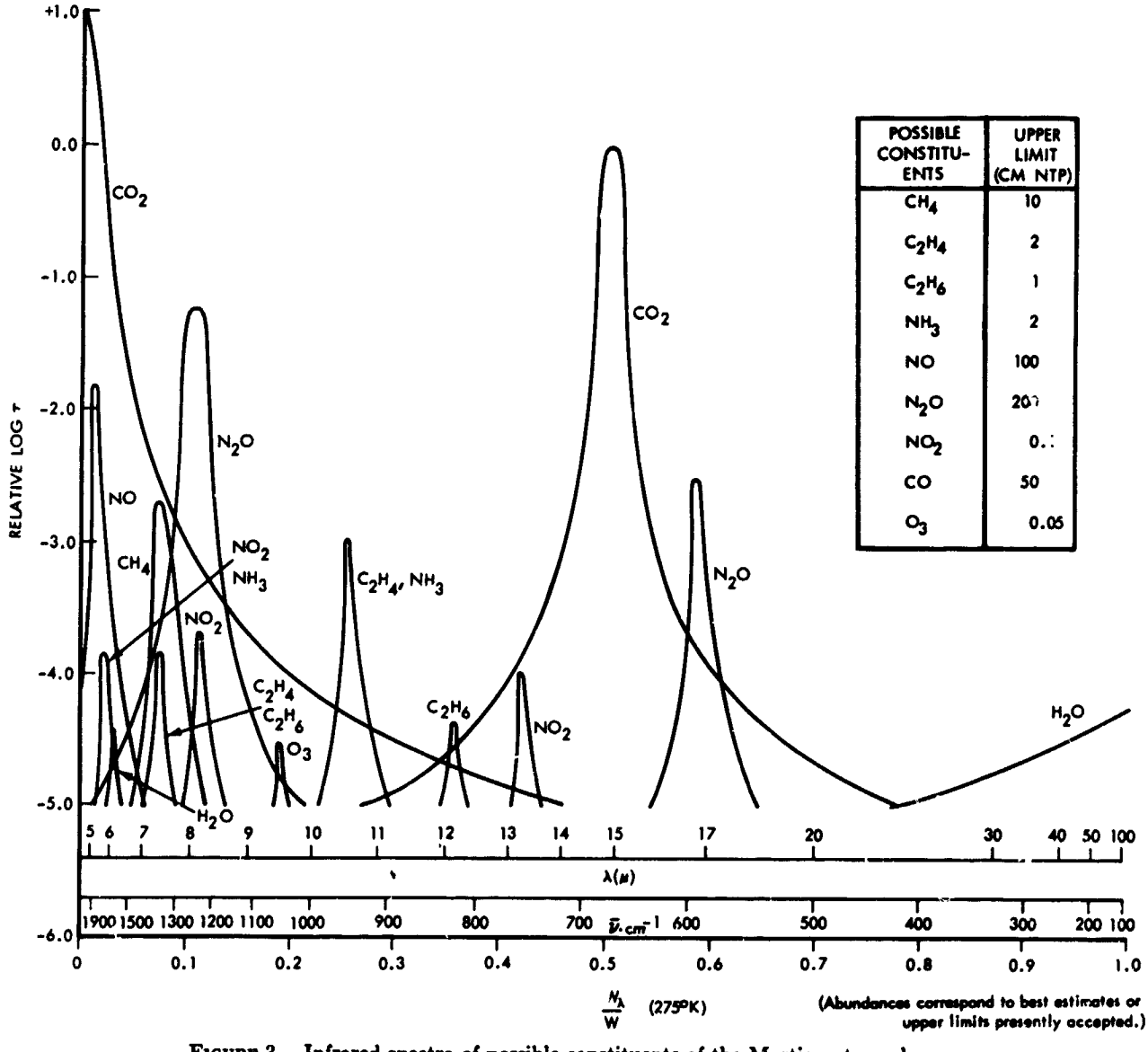


FIGURE 3.—Infrared spectra of possible constituents of the Martian atmosphere.

upper limits presently accepted (Kuiper, 1952; Sinton, 1961)*. The optical depth is shown on a relative scale.

The broad wings of the bands shown in Figure 3 correspond to the absorption continuum arising from the sum of the wings of all individual lines in the band. The situation for the Lorentz line shape is shown. This must be considered an upper limit for a number of reasons: First, as Benedict has shown, the Lorentz shape does not strictly apply far away from the line centers; second, and more important, Doppler broadening will be an important factor which causes a much faster decay of the line shape away from the line center and consequently gives rise to a much weaker continuum than that indicated. The 15-micron CO_2 band and the N_2O band near 8 microns, using the above upper limits and measured band intensities, are strong enough to be totally opaque. Methane (CH_4) and other hydrocarbons (C_2H_4) may also be detected. Water vapor, ozone, and others are more difficult to extract from the spectra But, because water vapor is extremely important both from an atmospheric-physics and from a biological viewpoint, a technique for the possible identification of this as well as other less abundant constituents will be described here.

*References 2 and 3.

Water vapor has strong vibration-rotation bands near 6.3 microns and the rotational band beyond 20 microns. If the spectral response of the instrument is restricted to 20 microns, then only the first bands are available for analysis. Figure 4 shows the transmission of a synthetic atmosphere consisting of the expected amount of CO_2 and water vapor, in the full detail to which the spectra can be generated by the interferometer. In addition, the figure shows spectral transmission of CH_4 and N_2O, both present in a quantity corresponding ot the presently accepted upper limit, and under the expected average atmospheric conditions. As can be seen, the atmosphere is nearly opaque between 600 and 750 wavenumbers (15-micron CO_2 band) and near 1300 (N_2O). Water vapor shows absorption phenomena between about 1300 and 2000 wavenumbers. Even in the strongest parts, the maximum absorption does not exceed a few percent. Consequently, it will be difficult to recognize the presence of water vapor in this band, and even more difficult to determine the abundance quantitatively. Fortunately, however, the water-vapor band spreads over almost 700 wavenumbers and shows considerable structure. This feature allows the use of statistical methods to obtain information about water vapor from almost 140 individual spectral data points derived from a single interferogram. The cross-correlation function of the received spectrum and the known spectrum of water vapor is obtained by

$$C_{(x)} = \int I_\nu W_{(\nu+x)} d\nu.$$

Even under conditions of severe noise, the cross-correlation function shows a maximum for $x = 0$. The correlation technique is a powerful tool for recognizing and estimating the amount of particular constituents in a complicated and possibly noisy spectrum. Figure 5 shows a sample of a cross-correlation function of the water-vapor spectrum with a combination of random noise and water vapor (signal-to-noise ratio equal to one). The same technique may also be employed to assess the amounts or upper limits for other atmospheric constituents.

2.3 Atmospheric Structure

Use of satellite-measured radiation values to determine the atmospheric temperature distribution of the earth's atmosphere has been discussed by Kaplan (1960), Wark (1961), Yamamoto (1961), King (1963), and Fleming, et al. (1964).*

Temperature distribution in the atmosphere is related to the absorption characteristics and the upward radiance through the radiative transfer equation expressed, for example, in terms of the pressure. In the spectral region of strongest

*References 4, 5, 6, 7, and 8.

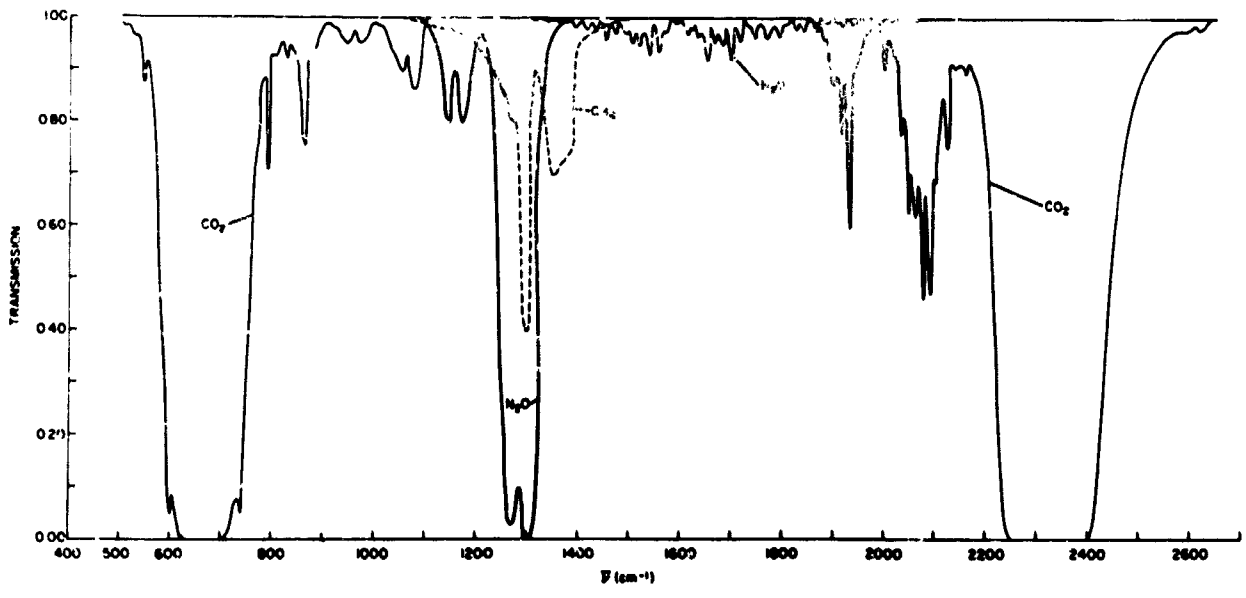


FIGURE 4.—Transmission of a synthetic atmosphere of the expected amount of CO_2 (55m atm) and water vapor (14 precipitable microns) and the upper limits of N_2O and CH_4.

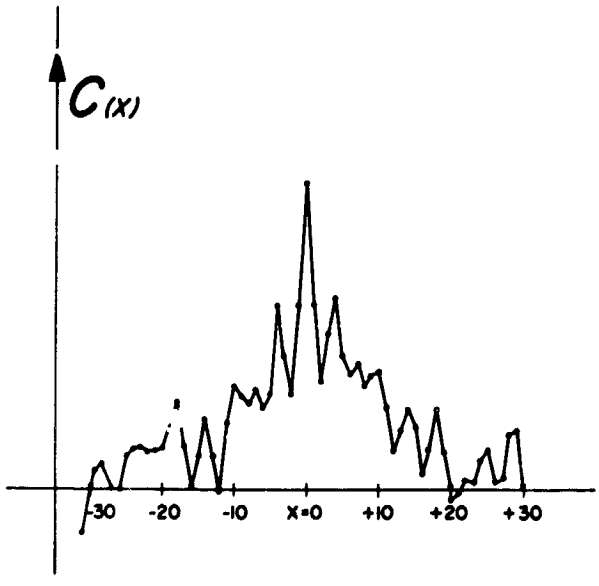


FIGURE 5.—Cross-correlation function of the spectrum of water vapor and a combination of water vapor plus random noise with a signal-to-noise ratio of unity.

absorption, the atmospheric emission arises from the upper portion of the atmosphere. Radiation in spectral regions of weaker absorption originates from lower layers in the atmosphere. Similar considerations can be applied to the extraction of the temperature distribution of the Martian atmosphere.

To illustrate, Figure 6 shows the transmission in selected spectral intervals within the 15μ carbon dioxide band. The transmission data were provided by Wark and Yamamoto. The transmission curves are based upon unpublished laboratory measurements (Yamamoto, Sasamori, Wark and Moore abstract in Publication IAMAP No. 13 on the XIIIth General Assembly of the IUGG, 1963) which assumed that the total amount of CO_2 was 5500 cm atmospheres, and the surface pressure 10 mb (Kuiper, 1964)* which corresponds approximately to the expected conditions on Mars.

Temperature corrections were made in the same manner as in the Meteorological Satellite Laboratory (USWB) reports #10 and #21, using a Mars model atmosphere to calculate effective temperatures T_e (surface temperature 270°K, adiabatic lapse rate to 6.5 mb, radiative equilibrium to 1 mb, extrapolation above.)

Figure 7 shows the atmospheric zones contributing to the outgoing radiance which can be

*Reference 9.

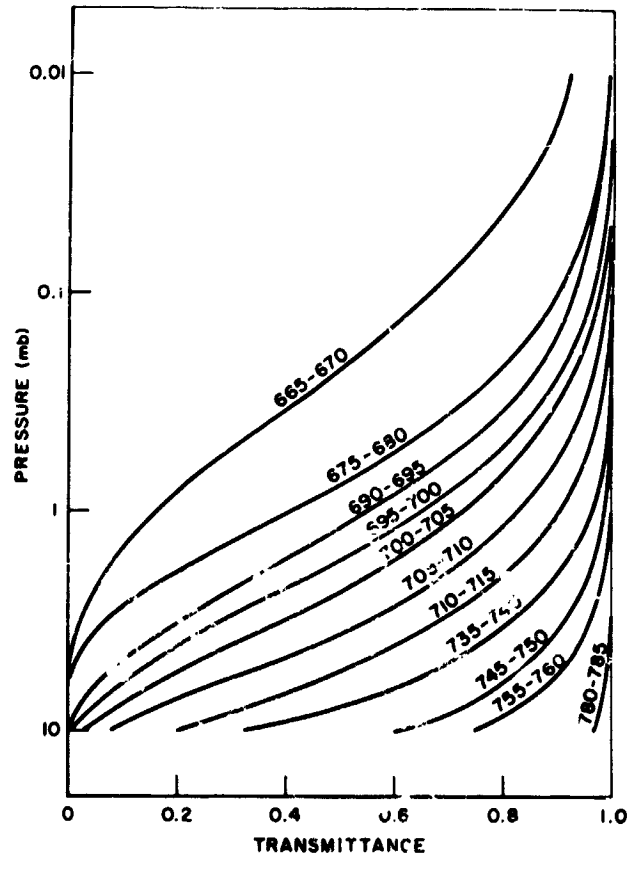


FIGURE 6.—Transmission function of CO_2 in selected spectral intervals in the 15μ band.

detected by the probe. Lorentz broadening only was taken into account; this approach is inaccurate, especially at the lower atmospheric pressures. In general, a somewhat greater absorption will be found than that shown, but the differences will be very small in all but the intervals of greatest absorption. Figure 7 demonstrates that the radiance values recovered by the interferometer allow a determination of the temperature profile. As the interferometer resolution is 5 cm^{-1} over the whole band, redundant measurements are available which increase the accuracy of the temperature determination.

2.4 Surface Studies

After the atmospheric constituents are identified, the effect of atmospheric absorption can be removed from the spectrum in those spectral ranges where the atmosphere is not opaque. The specific intensity emitted by the surface in the non-opaque regions is then obtained,

$$I_\nu(\text{surface}) = \epsilon_\nu B_\nu(T_{surface}).$$

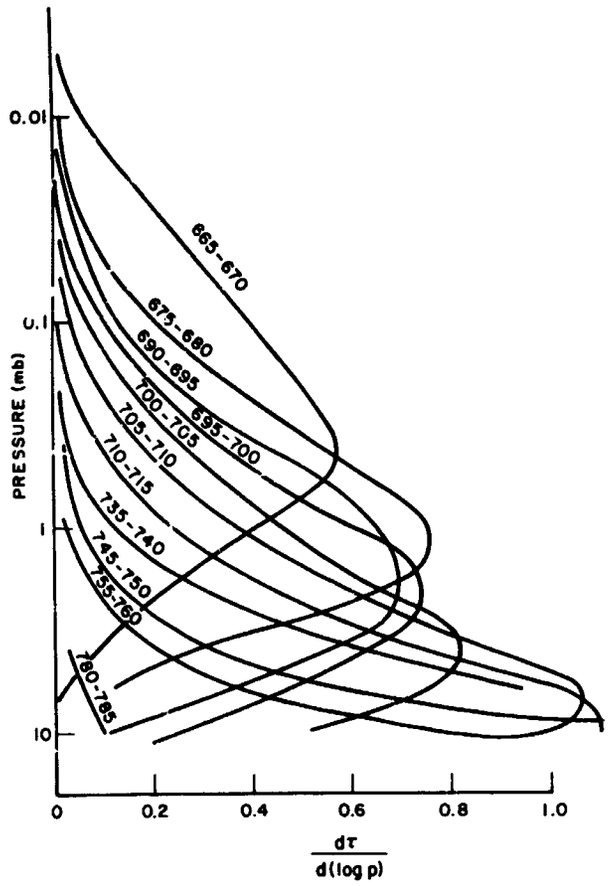


FIGURE 7.—Weighting functions indicating the pressure range from which thermal emission originates.

In the far infrared, most natural surfaces known on earth (rocks, gravel, plants, snow, etc.) have a high emissivity over large spectral ranges. However, the crystalline structure of minerals and the structure of complex molecules create stretching and bending resonances. For example, quartz exhibits strong stretching bands between 800 and 1100 cm^{-1} caused by the Si-O bond; other stretching groups (Si-Si) appear between 600 and 800 cm^{-1}. The same mineral has a bending mode (Si-O-Si) at 430 to 460 cm^{-1}. These resonances give a pronounced rise in the infrared reflectivity which can best be observed on polished samples. The increase in reflectivity is necessarily associated with a reduction in emissivity over the same spectral range.

The sharp rise in reflectivity has been well known since the early days of infrared studies. Rubens used the selective reflectivity of quartz plates to isolate narrow spectral bands in the far infrared. The reststrahlen phenomenon depends somewhat on the temperature of the sample and to a much higher degree on the surface roughness. Lyon (1963)* has studied the emission and reflection spectra of many minerals in great detail. The emissivity of several rocks in solid but unpolished form can be judged from Figure 8.

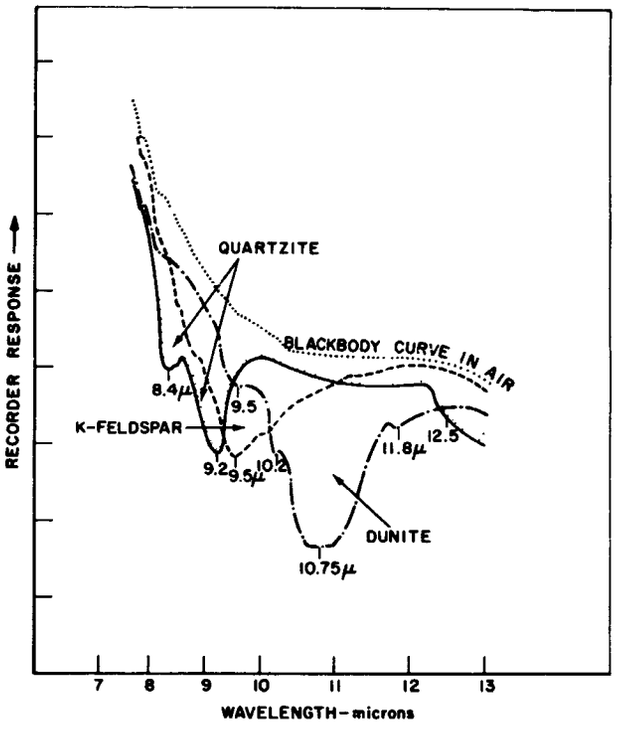


FIGURE 8.—Emissivity in the infrared of rock samples.

Polished samples show an even stronger reststrahlen effect. Grooves on a rough surface act as small cavities and increase the emissivity considerably. Similar phenomena can be recognized in the sandy or powdered form, as shown in Figure 9. Fine dust of a particle size below 1 micron loses the reststrahlen effect almost completely, and the surface emits almost with the characteristics of a graybody or possibly even a blackbody. Scattering effects also become important for these particle sizes. For a grain the size of fine sand, approximately 25 to 50 microns, the reststrahlen phenomenon is weak, but the identification of the material may still be possible, as has been demonstrated also by Bell and Eisner (1956).† Identification of the mineralogical com-

*Reference 10.
†Reference 11.

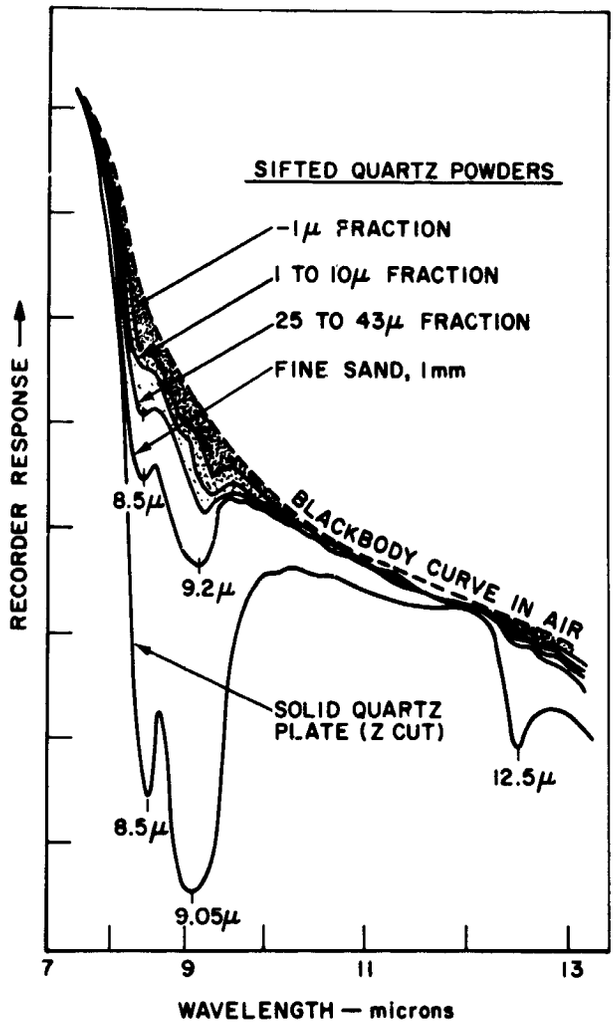


FIGURE 9.—Emissivity in the infrared of quartz and quartz sand of different grain size.

position of the Martian surface by the thermal-emission spectrum is, therefore, very promising, except for areas where fine dust may cover a large portion of the surface. Such areas may exist on Mars, but other areas may expose more solid rocks or gravel or coarse sand. Winds may sweep highlands clear of the finer particles and dust may accumulate in the valleys. The appearance of a pronounced reststrahlen effect would not only allow identification of a particular mineral, but would also be indicative of a coarse surface structure.

The surface emission spectrum may also yield important information on the existence of other chemical substances of inorganic, or possibly even organic, nature if they occur in a sufficiently high concentration. The Sinton bands near 3.5μ, although outside the spectral range of this experiment, are a typical example of unexpected (and so far unexplained) spectral features. The search for reflection or absorption phenomena is especially interesting if the spatial resolution allows measurements in predominantly bright and predominantly dark areas.

A broad program of laboratory and field measurements on earth is an indispensable requirement for proper interpretation of planetary emission spectra. The engineering model of the interferometer under construction will be used for experimental studies in this area.

2.5 Additional Investigations

In addition to the generally well defined tasks already mentioned, a number of inferences may be derived directly or indirectly from the spectra:

Chemistry of the Atmosphere and Surface

Chemical equilibrium between identified gases and others can be used to estimate the abundance of unidentified but suspected constituents; the photochemical equilibrium between CO_2, CO, O_2, and O_3 is one example. Furthermore, the atmospheric gases must be in chemical equilibrium with the surface materials.

Ultraviolet Radiation

Knowledge of the composition of the atmosphere, especially the CO_2 content and upper limits for O_3, allows calculations of the upper limits of the solar ultraviolet spectrum at the surface of the planet. The ultraviolet radiation at the surface is of biological significance.

Relative Humidity in the Atmosphere

The relative humidity near the surface is again of great biological interest. Low temperatures near the morning terminator may allow water vapor to condense in the form of dew or frost; condensation of this nature may be the only source of water available to vegetation if it is present. High temperatures near the subsolar point would cause the water to boil there. At 10 mb, water boils at +7°C. The relative humidity in the atmosphere also governs the condensation of clouds. Atmospheric regions where water or

ice clouds are possible, or likely, can therefore be determined. All these considerations will be greatly aided by theoretical studies of radiative and convective equilibrium conditions in the Martian atmosphere.

Meteorology

The temperature distribution in the atmosphere and the mode of energy transport are intimately connected. The measured temperature profile will give an indication of atmospheric stability. Lapse rates close to the adiabatic one may indicate a convective layer in the atmosphere (troposphere). Theories on the general circulation, or more concisely the Martian meteorology, are intimately related to the temperature profile.

The temperature-versus-pressure data will also be important in investigating the existence and occasional sudden disappearance of the blue haze. If this phenomenon is associated with condensing or sublimating material, the phase diagram of suggested materials must be consistent with the temperature structure derived by the experiment.

The planning of missions designed to land probes on the surface of Mars will benefit also from a better knowledge of the temperature profile, the atmospheric composition, and the surface pressure.

Spectral Emissivity and Surface Temperature

Determination of the spectral emissivity allows a more accurate determination of the true surface temperature. Most materials which show residual ray phenomena in certain parts of the spectrum radiate as nearly blackbodies elsewhere. Therefore, the maximum equivalent blackbody temperature determined over a wide spectral range will be closer to the true temperature of the surface than measurements obtained by radiometers covering a wide or narrow spectral interval. A better interpretation will be possible of radiometric data obtained in the atmospheric window from the earth, or of radiometric measurements performed from fly-by spacecraft. However, special care must be taken in interpreting surface temperatures if the surface is covered with very fine powdery material; this situation will require application of the scattering theory of small particles densely packed, which is quite complicated and not well understood today.

Thermal Inertia of Surface Material

The surface temperature as a function of solar zenith angle, especially near the morning terminator, provides data about the thermal inertia of the surface material. Calculations similar to those performed for the lunar surface can be applied. The great difference in thermal inertias of loose sand and solid rock provides a tool for distinguishing among various surface conditions.

3. THE INTERFEROMETER SPECTROMETER

The instrument, basically a Twyman-Green modification of the Michelson interferometer (Figure 10), departs from the classical laboratory designs in several areas:

The drive mirror is suspended by a parallel leaf-spring system.

The mirror assembly is driven by a moving coil in a strong magnetic field; nearly critical damping is provided by electronic means.

The mirror motion is monitored by an auxiliary interferometer to assure proper sampling times and uniformity in the mirror speed by closed-loop feedback control.

The compensating plate is moved to a position adjacent to the beam-splitter.

The particular configuration of the interferometer (Figure 11) was chosen because of its inherent compactness, large light-gathering power, and design simplicity.

Radiation arriving within the field of view is divided by the beam-splitter into two components which recombine and interfere after reflections on the fixed and moving mirrors respectively. The recombined beam, focused on the detector, is modulated by the motion of the scan mirror proportional to the speed of the mirror and the wavenumber of the incident radiation. The output signal from the detector, called the interferogram, is the Fourier transformation of the spectrum of the incident radiation. The signal is amplified, digitized, and transmitted over the spacecraft telemetering system. On the ground, the spectrum is reconstructed by applying the inverse transformation. The original spectrum can be precisely reconstructed only within certain limits of accuracy of spectral range and of spectral resolution.

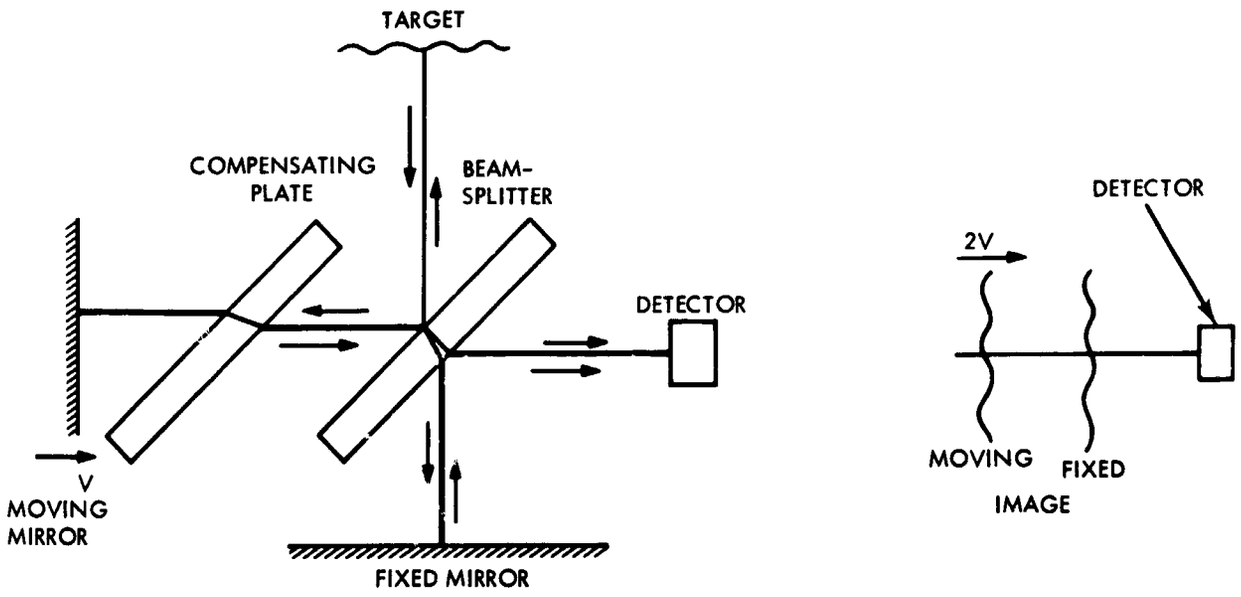


FIGURE 10.—Schematic operation of Michelson interferometer.

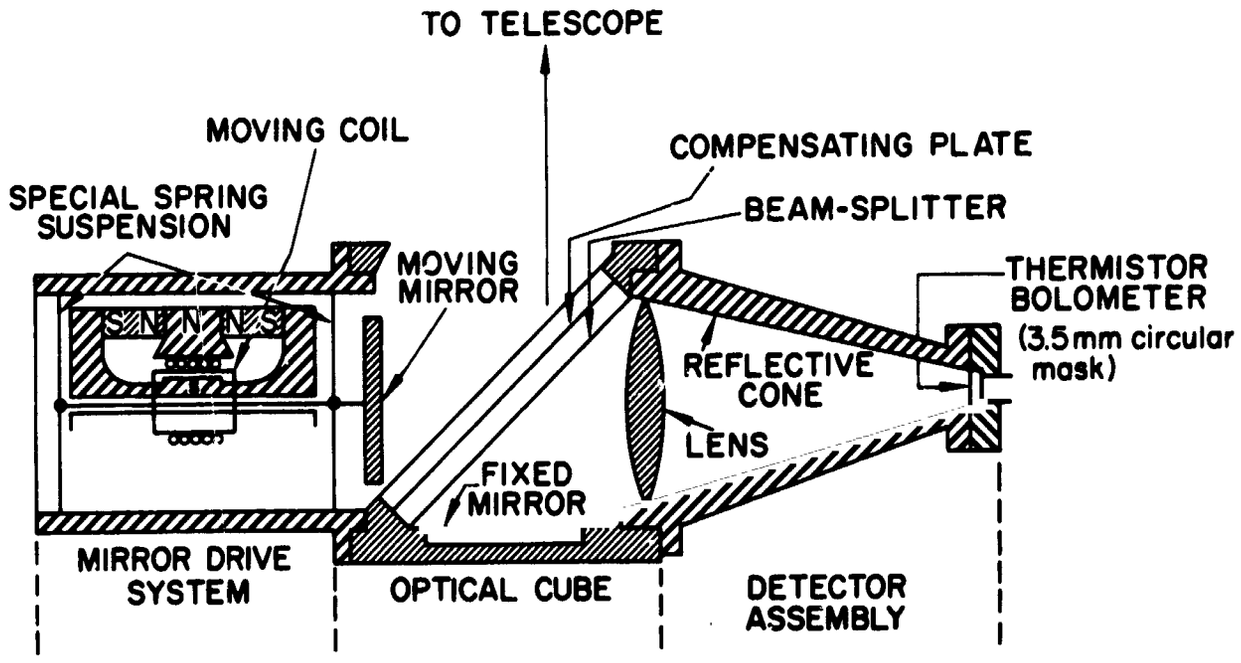


FIGURE 11.—Schematic diagram of interferometer cube, mirror drive system, and detector assembly.

3.1 Theory of Fournier Spectroscopy

Interference spectroscopy, first suggested by Michelson after the development of his original interferometer, was not pursued because of extreme computational difficulties, and further development was forced to wait upon the advent of electronics and high-speed digital computers. In the last decade, interference spectroscopy has made great advances; work has been reported by Fellgett (1958), Gebbie (1959), Mertz 1956), Connes (1961), Strong and Vanesse (1958–59), Loewenstein (1963),* and many others. Excellent summaries may be found in articles by Jacquinot (1960) and by Connes (1961).†

*References 12, 13, 14, 15, 16, 17, and 18.
†References 19 and 20.

The interferometer is shown schematically in Figure 10; the source of radiation will be an extended source which completely fills the field-of-view with radiation passing through the instrument without collimation. The output fringe pattern of the interferometer is focused on the detector, which is masked to eliminate all but the central fringe. The output of the detector is amplified and recorded as a function of path difference, d, between the fixed and movable mirror.

The relationship between the effective path difference, the order of interference, and the radiation wavelength is

$$m\lambda = 2d \cos \theta$$

where

d is the path difference between the two interferometer paths (cm)

m is the order of interference

λ is the wavelength (cm)

θ is the angle of an oblique ray through the interferometer.

This expression shows that, for a given monochromatic signal of wavelength λ, the circular fringe corresponding to a given order m expands away from the center of the pattern as the separation of the mirrors is increased. The detector is mounted behind the mask so that only the central fringe is detected; each spectral element is thereby modulated sinusoidally at a rate inversely proportional to its wavelength. As the source has an appreciable range of spectral elements, the signal detected will be the superposition of many modulated intensities multiplied by a parameter which is the "responsivity" of the system; i.e.,

$$\phi_x = \frac{1}{2}\int_0^\infty K_\nu N_\nu d\nu + \frac{1}{2}\int_0^\infty K_\nu N_\nu \cos 2\pi\nu x \, d\nu$$

where N_ν is the spectral radiance of the source (assumed to be the same at all parts of the field-of-view)

ν is the wavenumber, $\text{cm}^{-1}\left(\nu = \frac{1}{\lambda}\right)$

x is the instantaneous value of the effective path difference of the two beams of the interferometer (cm)

K_ν is a parameter which expresses the effective spectral responsivity of the interferometer, including the characteristics of the optics, the detector, and the electronic amplifiers

The interferogram is then taken to be

$$I_x = 2\left[\Phi_x - \frac{1}{2}\Phi(0)\right]$$

$$= \int_0^\infty K_\nu N_\nu \cos 2\pi\nu x \, d\nu,$$

which is the Fourier cosine transform of $K_\nu \cdot N_\nu$. The product $K_\nu \cdot N_\nu$ can be obtained uniquely from the measured interferogram, taking the inverse transform by digital computation. The relationship between N_ν and the product $K_\nu \cdot N_\nu$ will be established by laboratory calibrations; the validity of the laboratory calibrations is established by in-flight calibrations.

The true spectral distribution $K_\nu \cdot N_\nu$ is found only when the interferogram is recorded for an infinite range of x. For a finite range of x, each spectral component is spread out into a shape known as the instrumental profile. This profile also depends on the obliquity of rays through the interferometer, and thus on the solid angle of the field-of-view. The instrument profile may be modified as desired during computation of the transform; this process, called apodization, usually is done to remove secondary maxima at the expense of resolution. Including all these effects, the expression for the interferogram becomes

$$I_x = \int_0^\infty N\nu \cdot s(\nu, x) \cdot \cos\left[2\pi\nu x + \phi(\nu, x)\right] d\nu,$$

where

$$S(\nu, x) = s(\nu, x) \exp\left[i\phi(\nu, x)\right]$$

is the Fourier transform of the instrument profile (including the parameter K_ν mentioned previously).

The spectral distribution is again obtained by computing the Fourier transform of $I_{(x)}$. Reconstruction of the spectral distribution of the source

N_ν is obviously an approximation of the actual spectral distribution. This approximate solution (i.e., the transform of the interferogram) is the convolution of the actual spectral distribution with the instrument profile; therefore, it is mandatory to determine the form of the instrument profile. This can be done experimentally, because the instrument profile is the apparent spectral distribution given by a monochromatic source. This determination of the instrument profile is an important part of the laboratory calibrations.

Two factors which influence the instrument profile and reduce the quality of the data are the limitation arising from the finite path of the movable mirror, and the effect of oblique rays through the interferometer. The resolution limit imposed by the finite-path limitation is

$$R=\nu L \text{ or } \Delta\nu=\frac{1}{L},$$

(assuming linear apodization to remove secondary maxima in the instrument function). The resolution limit imposed by oblique-ray effects is

$$R\Omega=2\pi, \text{ or } \Delta\nu=\frac{\nu\Omega}{2\pi}.$$

In the above, R is the resolution $\nu/\Delta\nu$, L is the maximum path difference on either side of zero path difference, and Ω is the solid-angle field-of-view of the interferometer.

Advantages of the Interferometer

An analysis by Jacquinot indicates that, where a source whose spectrum is defined by a radiance N_ν, having M spectral elements of interest, is to be analyzed with a resolution R in a time T with an optical system which is detector-noise-limited (as presently in the infrared), then the signal-to-noise ratio is increased by maximizing the time t of observation of each spectral element, the admissible solid angle Ω, and the transmission factor ϵ_ν (the optical efficiency).

An interferometer fulfills these requirements to a high degree. The time of observation is $t=T(m/M)$, where m is the number of spectral elements studied simultaneously. For an interferometer, m equals M, and m/M has the maximum value 1. The luminous energy received by the detector is $N_\nu \cdot A \cdot \Omega$. The quantities A and Ω can both be made large for an interferometer, as long as the resolution limitation $R\Omega=2\pi$ is not exceeded (for comparison, the same resolution limit for a grating spectrometer is $R\Omega<0.1$). The transmission factor ϵ is the only area where the superiority of the interferometer, may be questioned for a given application. The reflectivity and transmissivity of the beam-splitter is of primary importance and it may be difficult to obtain high efficiency from mechanically suitable materials in certain portions of the infrared spectrum.

3.2 Mechanical and Electromechanical Design

The instrument for the IRIS experiment consists of:

- interferometer (cube, detector, mirror drive system, preamplifiers)
- two auxiliary detectors, one for reflected solar radiation, the other for total thermal radiation
- telescope with calibration target (blackbody)
- electronics compartment, containing all electronic circuits except for the preamplifiers in the cube
- mechanical structure to receive the interferometer, the auxiliary detectors, the telescope, and the calibration target, and to provide mechanical strength and thermal protection

The electronics compartment, located separately, is connected to the interferometer and to the auxiliary detectors by a cable.

The mechanical design of the interferometer is modular. The drive unit, the fixed mirror assembly, the detector assembly, and the monochromatic reference source-detector combination are integral components attached to the optical cube, which houses the beam-splitter and provides mechanical strength and thermal conduction to all components.

The scan mirror is one of the most critical items in the design because it must be moved through a distance of 2 mm with a total angular rotation of less than 2 seconds of arc.

The use of parallel leaf springs to support the mirror was chosen in preference to slides on ways, or pivot-point mounts, since it avoids

frictional surfaces undesirable in space applications. The driving force is generated by an electrical current in a moving coil in a strong magnetic field. As Figure 11 shows, the magnetic gap is inside the dual-magnet design, thus minimizing residual magnetic fields on the outside.

Lack of uniformity in drive-mirror speed normally produces catastrophic effects on the wavenumber calibration of interferometers. A resolving power of 400 would require a drive-speed constancy of better than 1 part in 400. Two methods used to eliminate virtually all errors and uncertainties in the mirror motion, and in the subsequent wavelength calibration, are both based on the use of an auxiliary interferometer:

- A monochromatic source (0.5852 micron) feeds the interferometer and generates a sine wave at an auxiliary detector. The frequency of the sine wave is directly proportional to the speed of the mirror. An error signal, derived by a digital comparison of this frequency to a reference frequency generated by a clock, provides fine adjustment of the mirror-drive system. The speed of the drive mirror is therefore feedback-controlled.
- Sampling times of the interferogram are also derived from the reference frequency.

Actually, either of the aforementioned methods would suffice.

3.3 Optical Design

The optical design of the interferometer is conventional. The flat mirrors and the detector determine the limiting apertures. The condition of optimum solid angle $R\Omega = 2\pi$ is fulfilled at the highest wavenumber to be analyzed (2000 cm^{-1}, 5μ). The optimum angle is then 8 degrees (± 4 degrees) for a resolving power of 400 (5 cm^{-1} at 2000 cm^{-1}).

The fixed mirror and the scan mirror are made of gold-coated quartz polished to a flatness better than 0.1μ. Gold was selected because of its stability and its high reflectivity in the infrared. The two-axis mirror mount is machined from a single piece of metal. A differential screw which produces the fine adjustment required also anchors each adjustment plane at three points.

The condensing optic consists of a KRS–5 lens, combined with a gold-coated reflective cone which increases the effective detector area. Dimensions are chosen to provide the 8-degree optimum acceptable angle. This particular optic has sensitivity to almost 40μ, which will be important for the future development of the instrument in extending the long wavelength limit. The f-number of this system will approach the ideal limit of 0.5.

The bolometer in this design is essentially an integral part of the condensing optic. The detector is a 3.5-mm square masked to a circle of 3.5-mm diameter, having a time constant of 1 millisecond.

The beam-splitter is the most critical item in the optical train. All the rays make two passes through the beam-splitter; hence, all the losses arising in this unit are doubled. Although metallic beam-splitters are preferable from the standpoint of uniformity with wavelength, they have the serious disadvantage that an absorption loss from 25 to 35 percent occurs, depending on the metal used. Dielectric beam-splitters cannot be made uniform with wavelength, but over the wavelength span of interest the efficiency will be at least equal to that of a metallic splitter. The dielectric material selected is germanium, which has the additional advantage of being highly reflective for wavelengths shorter than 2μ so that most of the reflected visible radiation will not be collected.

The substrate material selected for the beam-splitter and the compensating plate is IRTRAN–4, basically a glass, which can be optically worked in the same fashion as any optical glass. This is a great advantage, because the beam-splitter must be flat to 1/10 wavelength of the shortest wavelength of interest. The principal disadvantage of IRTRAN–4 is its long wavelength cutoff at 20μ.

Other materials investigated for their suitability as beam-splitters (including KBr, KRS–5, CsI, and plastic pellicles) have higher transmission to long wavelengths, but are less suitable in other respects. Further investigations are in process.

The radiation source for the monochromatic reference signal will be the 0.5852μ neon line, which is quite strong; in conjunction with a suitable interference filter, the signal is suitably monochromatic. Neon was selected because lamps are available in small sizes and very little power is required. The focusing and condensing

optics are all glass of conventional and straightforward design. A silicon photovoltaic cell serves as a detector.

The telescope is designed to accept a beam of parallel radiation with an angular aperture of 2.8 degrees and to fill completely the 8-degree field-of-view of the interferometer. The primary and secondary mirror are electroformed for lightweight construction and then gold-coated.

3.4 Electronic Design

The electronic circuitry for the interferometer package is designed to operate as independently of the spacecraft as possible. The spacecraft supplies the power and command signals necessary to operate the experiment; the electronics subsystem of the experiment processes the interferometer data and presents them to the spacecraft in a digital format. The spacecraft supplies a single command at the beginning of each interferometer frame, but the necessary sequence of events to complete the data frame is generated by means of the interferometer system.

Functions of the interferometer electronics (Figure 12) are:

- To convert the 2400-cps 100v *p-p* spacecraft power into the voltages required to operate the interferometer instrument
- To provide the necessary power and control to drive the mirror
- To amplify the thermistor bolometer output signal
- To sample the bolometer signal, and to present it in digital form to the spacecraft

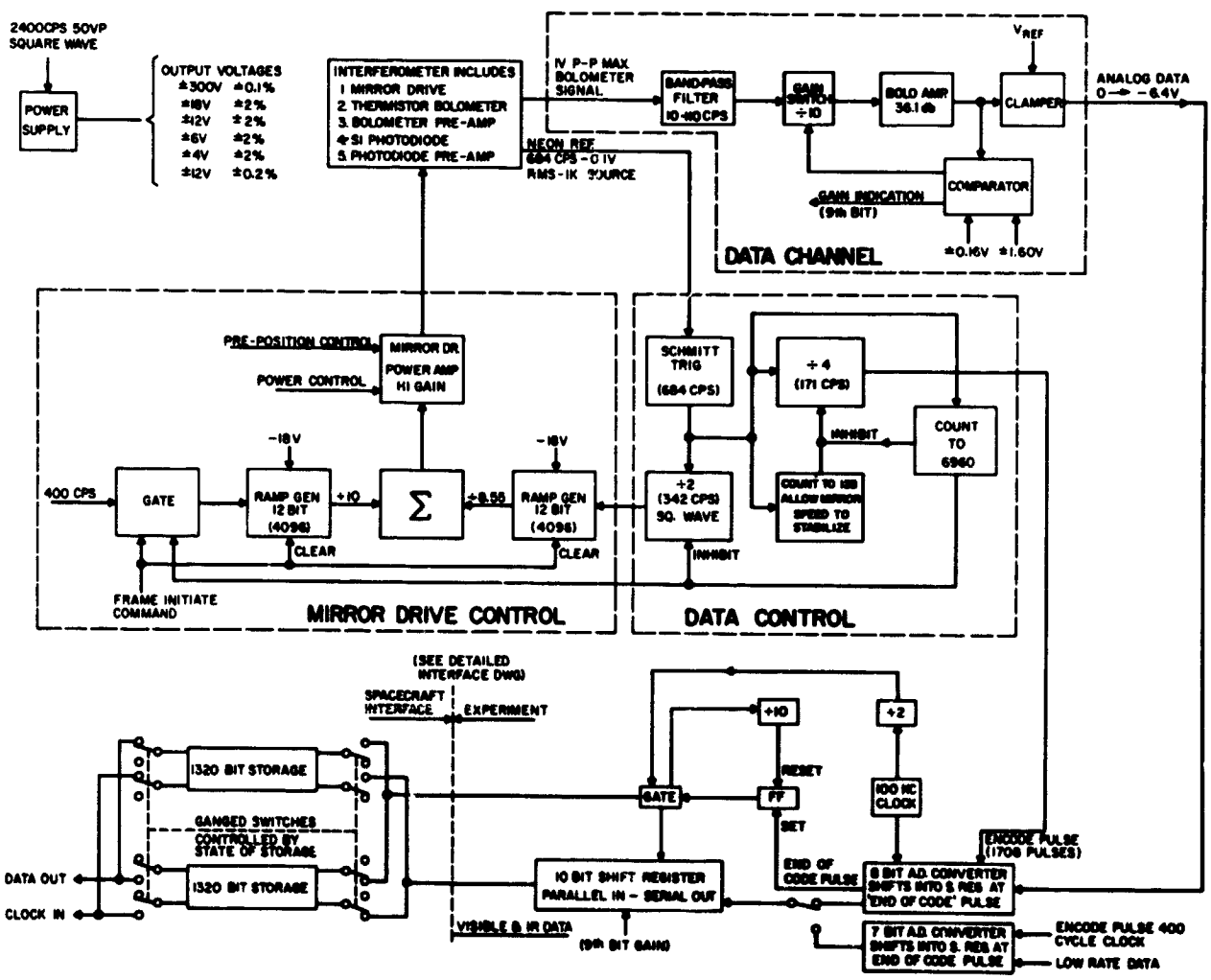


FIGURE 12.—Interferometer electronics, block diagram.

The power supply converts the 2400-cps 100v *p-p* spacecraft power to the voltage levels necessary for operating the interferometer. The power supply includes the necessary regulation (±5 percent) to maintain the voltage to the required accuracy. Three telemetry points are included to monitor:

- Spacecraft input voltage
- 300v ±0.1 percent bolometer bias supply
- 12v ±0.2 percent reference for mirror drive

The power supply is regulated to 1 percent by feedback to the primary of the power transformer; the 300-volt supply has an additional control that maintains 0.1 percent regulation.

The data channel receives the signal from the thermistor bolometer preamplifier. The signal is band-limited between 10 cps and 110 cps in an electrical filter network. This band-limited signal is then coupled to a switched attenuator and amplified.

The attenuator has two positions, controlled by the level of the preamplifier output voltage: 0–db attenuation and 20–db attenuation. The preamplifier output fed into a comparator generates an output to switch in the 20–db attenuator when the preamplifier output exceeds ±3.2 volts, and to switch out the attenuator when the signal falls below ±0.32 volts. The gain change is necessary to accommodate the extremely large dynamic range (2000:1) of the input signal from the bolometer preamplifier. The output of the bolometer preamplifier is maintained between 0 and 6.4 volts, a value which is compatible with the analog-to-digital converter.

The function of the data control channel is to:

- Provide the necessary encode pulses to the analog-to-digital converter at proper mirror positions
- Ensure that data points are not transmitted until the mirror velocity has stabilized
- Allow precisely 1708 data points for each data frame

The count-to-128 inhibits the encode pulse for a sufficient time to allow the mirror velocity to stabilize at 0.2 mm/sec. After 128 counts, the 171-cps squarewave is coupled to the analog-to-digital converter as encode pulses. The count-to-1708 circuit counts the number of encode pulses that have been applied to the analog-to-digital converter. After 1708 samples, the mirror drive circuit is de-energized and reset to zero, preparing the experiment for the next frame-initiate command from the spacecraft.

The analog-to-digital converter generates an 8-bit binary word from the bolometer amplifier output. The encode pulse from the data-control channel initiates the digital-to-analog voltage ramp that is compared with the analog voltage from the data channel. When the reference analog voltage matches the data voltage, the digital-to-analog circuit is stopped and the digital word transferred in parallel into the 10-bit shift register as the 8 least significant bits. The ninth bit is generated by comparator in the data channel designating the amplifier gain; the tenth bit is word synchronization.

At the end-of-code pulse, a gate is opened that allows the 50-kc clock to shift the 10-bit word serially into the spacecraft buffer storage. After 10 counts, the gate between the 50-kc clock and the 10-bit shift register is closed. The tenth bit in the shift register is used for synchronizing the data words in the ground stations.

An additional 7-bit analog-to-digital converter and a set of analog gates are included to sample the low data-rate channels. The analog-to-digital converters are sequentially switched into the spacecraft buffer storage system. The frame-initiate command switches the high data-rate channel into the spacecraft storage, and the end-of-frame pulse switches in the low data-rate channel.

The 400-cps spacecraft clock is coupled to the low data-rate commutator and the 7-bit analog-to-digital converter to serve as switching pulses and encode pulses.

The low data-rate channel contains the housekeeping telemetry and auxiliary measurements of the integrated visible and thermal radiation from the planet; these measurements are transmitted in real time as well as in the low data rate.

The function of the mirror-drive servo-loop is to maintain the mirror velocity constant at 0.2 mm/sec for the 10-second duration of the interferometer data frame. The control system consists of two digital-to-analog voltage ramp generators fed into a summing circuit that drives the mirror drive-power amplifier. The first ramp is a positive-going voltage generated from the 400-cps

spacecraft clock; the second ramp is a negative-going voltage generated from the 342-cps neon signal from the interferometer. The output voltages from the two digital-to-analog circuits are weighted so that no signal is applied to the mirror power amplifier when the mirror velocity is constant at 0.2 mm/sec. The loop characteristics of the servo stabilize the mirror velocity at 0.2 mm/sec ±1 percent within a fraction of a second after the frame-initiate pulse is received from the spacecraft. After the 1708 data points have been sampled, a pulse from the data-control circuit closes the gate between the 400-cps clock and the digital-to-analog circuits, and resets the counters in the digital-to-analog circuits to zero.

The electronics for the IRIS instrument are designed to perform the functions necessary to control the interferometer and process the data presentation to the spacecraft. The design consists primarily of digital electronics, because of the inherent reliability of digital systems. The digital circuits are constructed of Texas Instrument modules. This subsystem is a self-contained instrument requiring a minimum of external control for the operation of the experiment. A reliability survey will be conducted as one of the system evaluation parameters.

3.5 Signal-to-Noise Ratio

The peak signal falling on the detector (for zero path difference) is

$$P=\frac{1}{2}N\cdot A\cdot\Omega\cdot\epsilon,$$

where

P = radiant power (watts)

N = target radiance (watt cm^{-2} ster^{-1})

A = effective aperture (cm^2)

Ω = solid angle (ster)

ϵ = average optical efficiency

The factor ½ accounts for the 50-percent reflection losses inherent in this type of interferometer. The noise at the detector is given by the noise equivalent power (NEP) of a thermistor bolometer, which is approximately

$$NEP=1.6\cdot10^{-10}\sqrt{a}\sqrt{\frac{\Delta f}{\tau}},$$

where

a = area of detector (cm^2)

Δf = noise bandwidth (cps)

τ = detector time constant (sec)

Additional noise will be contributed by the compensating thermistor ($\sqrt{2}$), noise in the bias supply ($\sqrt{2}$), and preamplifier noise ($\sqrt{2}$). For operation at 260 or 270°K, the NEP is lower; however, this improvement will not be considered here.

The peak signal-to-noise ratio in the interferogram, for zero path displacement, is then

$$\frac{S}{N}=\frac{P}{NEP}\approx10^9NA\,\Omega\epsilon\sqrt{\frac{\tau}{2a\Delta f}}.$$

The signal power in the spectral interval $\Delta\nu$ will be approximately equal to the peak power P times the ratio $N_\nu\Delta\nu\epsilon_\nu/N\epsilon$. In addition, a signal-to-noise improvement factor proportional to $\sqrt{n/2}$ (n = number of samples per interferogram) is obtained; therefore,

$$\left(\frac{S}{N}\right)_{\Delta\nu}\approx10^9N_\nu\Delta\nu\epsilon_\nu A\cdot\Omega\sqrt{\frac{\tau n}{2a\Delta f}}.$$

Assuming reasonable numerical values for some parameters,

effective aperture $A=10$ cm^2

solid angle $\Omega=1.57\cdot10^{-2}(8°)$

optical efficiency $\epsilon\approx\epsilon_\nu=0.2$

detector time const. $\tau=1$ ms

detector area $a=.1225$ cm^2

number of samples $n=1708$

spectral element $\Delta\nu=5$ cm^{-1}

noise equivalent bandwidth $\Delta f\approx136$ cps

the peak signal-to-noise ratio becomes

$$\frac{S}{N}=2.5\cdot10^5N_\nu,$$

and for a spectral interval

$$\left(\frac{S}{N}\right)_{\Delta\nu}=3.6\cdot10^7N_\nu.$$

The instrument temperature (270°K) was chosen to give approximately equal signals for the maximum blackbody temperature to be measured (320°K), and for outer space (0°K), which is used as a check of calibration. The maximum radiance between 500 and 200 cm^{-1} is about $8 \cdot 10^{-3}$ watt cm^{-2} $ster^{-1}$ (320°K); therefore, the maximum signal-to-noise ratio is about 2000.

Within the spectral intervals, the signal-to-noise ratios depend largely on spectral radiance. A plot of the N_ν for various blackbody temperatures is shown in Figures 13 and 14; in both cases, the instrument is at 270°K. However, before the final signal-to-noise ratio can be discussed realistically, errors in the coding process must be taken into account.

The most efficient way to encode is to set one coding step equal to the rms value of the noise in the interferogram. Higher coding accuracy yields

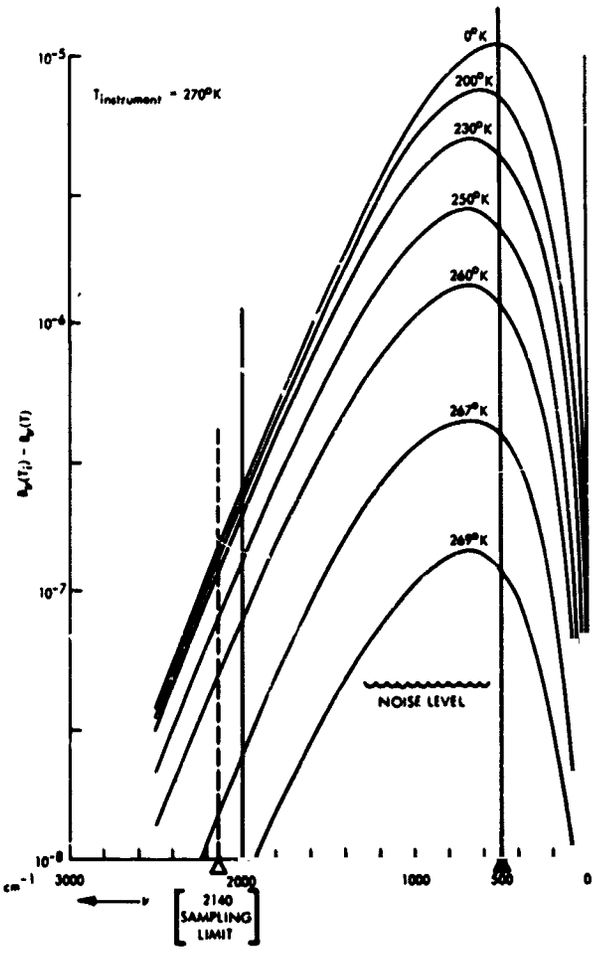


FIGURE 13. — Difference in spectral radiance of a target colder than the interferometer.

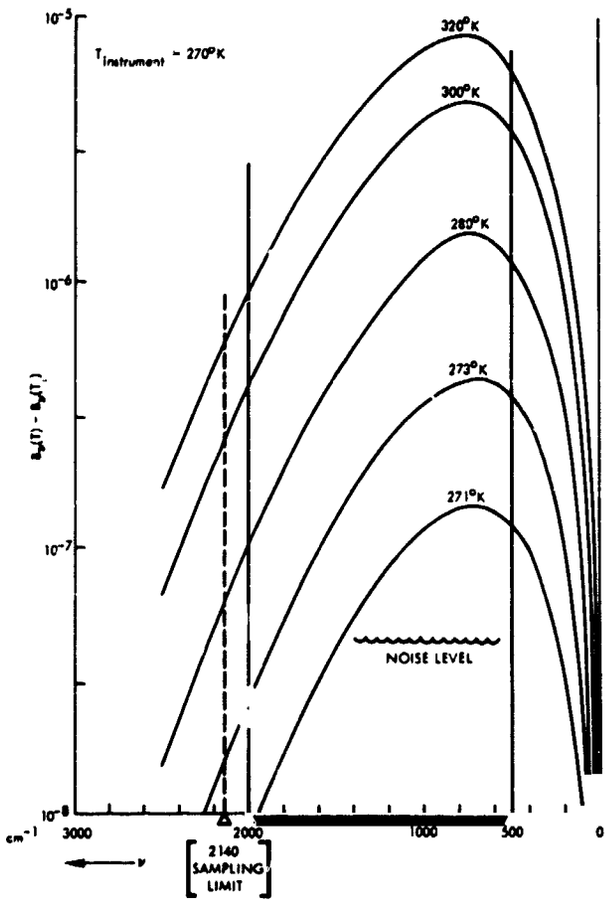


FIGURE 14. — Difference in spectral radiance of a target warmer than the interferometer.

insignificant improvement; lower coding accuracy is undesirable. If encoding is performed to 1 part in 2000 of the maximum possible amplitude in the interferogram, the uncertainties introduced by the coding process will equal errors caused by detector noise. The final signal-to-noise value, including the coding error, becomes

$$\frac{S}{N} = 1.7 \cdot 10^{3} N \text{ and } \left(\frac{S}{N}\right)_{\Delta\nu} = 2.5 \cdot 10^{7} N_\nu.$$

A $(S/N)_{\Delta\nu}$ equal to unity is obtained for $N_\nu = 4 \cdot 10^{-8}$. This value is indicated in Figures 13 and 14. A temperature difference of 1°K should therefore be detectable below 1400 cm^{-1} (above 7 microns); at 1800 cm^{-1} (5.5 microns), 3°K corresponds to the noise, and at 2000 (5 microns) only about 6°K.

These numbers were derived for 5 cm^{-1} resolution. Clearly, numerical smoothing techniques

can be applied and resolving power can be traded against accuracy. This operation is done on the ground after the interferogram has been received, so that considerations of this nature do not influence the design of the instrument.

It remains to be shown how a signal-to-noise ratio of 2000 can be obtained in the coding process. The large central peak in the middle of most interferograms is very pronounced; in most cases, the amplitudes in the rest of the interferogram are considerably smaller. An 8-bit word is used to encode on a scale of plus and minus 128 levels, or a total of 256. One level is equal to the rms value of the noise. Larger amplitudes are divided by 10 electronically and then encoded. The position of the divide-by-10 circuit is indicated by a ninth bit. This effective dynamic compression maintains full accuracy for most points in the interferogram, but only one-tenth of the full accuracy for all points exceeding a tenth of the maximum possible amplitude. However, the one-tenth value in amplitude is exceeded for only a very small number of data points, compared to the total of 1708 points. The slightly reduced accuracy of a few data points can be shown to have negligible influence on the reproducibility of the spectra.

3.6 Auxiliary Measurements

A number of auxiliary measurements will be performed to facilitate better interpretation of the interferograms and to validate the operation of the instrument.

Reflected Solar Radiation

Reflected solar radiation between 0.5 micron and about 1 micron will be measured by a small radiometer, bore-sighted with the optical axis of the interferometer and having the same field-of-view as the interferometer. The purpose of this measurement is to allow correlation of dark and bright areas on the planet with the interferograms. This instrument will use a silicon cell as a detector, with a time constant of about 10 seconds. A 2-percent relative accuracy and a 10-percent absolute accuracy are sufficient for this measurement.

Thermal Radiation

Also bore-sighted with the interferometer will be a thermopile having the same spectral response and the same field of view as the interferometer. The purpose of this instrument is to furnish additional calibration of the interferometer during planetary encounter. The time constant of this instrument will also be 10 seconds. A 2-percent relative accuracy are likewise sufficient for this measurement.

Temperature Sensors

Various temperatures will be measured in the optical cube (the temperature of the bolometer mounting, the drive mirror, the beam-splitter mounting, and a point near the entrance aperture) with an accuracy of 0.5°C. These temperature measurements will aid in the proper interpretation of the interferograms. In addition, the temperature of the primary and secondary mirror will be monitored (±5.0°C). The blackbody used for calibration purposes before and after planetary encounter will be monitored by two lead thermistors to provide redundancy in this measurement. Twelve temperature sensors are provided, including the two temperature sensors in the auxiliary radiometers and the two sensors located in the electronics compartment.

Voltage Measurements

A number of important voltages and execution commands will be monitored to certify proper operation of the instrument. The measured voltages include the bolometer bias voltage, the main power in the optical cube (−12 and −18 volts), and other strategically important voltages in the electronics compartment. A total of 25 auxiliary measurements will be performed.

4. DATA REDUCTION

Computational efforts concerned with the IRIS experiment fall into two major categories: computer simulation of the interferogram, and subsequent analysis to retrieve the optimum amount of information from the received data. The artificially synthesized interferograms will serve as aids for signal-to-noise requirements and in verifying the encoding and digitizing techniques to be applied, as well as in furnishing the basis for the early development of the data analysis techniques to be used. The synthesis of the interferograms has been divided into two phases and the analysis of the data into three phases.

4.1 Synthesis

The purpose of this portion of the study is to simulate the interferogram in order to study the effects of various levels of detector noise, to analyze the possible methods of digital encoding, and to determine the total effect on the ability to recover the underlying frequencies of interest.

The objective of the first phase of the study is a computer program which generates (for specified wavenumbers, instrument and target temperatures, time duration, and sampling rates) the amplitudes of a Fourier series (interferogram) according to the basic instrument characteristics. All computer outputs are formatted, if desired, for plotting offline on an x-y plotter. This part of the program has been completed, and artificial interferograms which were constructed have played an important role in the initial design of the digitization scheme for the instrument.

The first step of the second phase of the interferogram synthesis program is to include the instrument function. The second step is to generate and apply any desired level of detector noise to the signal. The third step is to apply several scaling techniques to find the best one, which will then be used for the actual instrument. Finally, the digitized version of the interferogram will be formatted to become identical to the output of the physical instrument. The result—the final, simulated interferogram—is again formatted for offline plotting.

4.2 Analysis

The analysis program is being developed concurrently with the synthesis program, because it is essential that the actual interferogram provide the best possible analysis of the data, and of the information extracted. In order to accomplish this, the synthesis and analysis process must be a "closed-loop" learning and improvement system. The resulting analysis program should, therefore, be one which is best fitted to analyze the data from the actual instrument.

The purpose of the first phase of the analysis program is to apply the inverse of the "second-order" effects of the synthesis program; that is, given an interferogram in digital form, the inverse operations of scaling digitizing must be applied, and an apodization function applied, before the frequency analysis is begun.

The second phase of the analysis consists of performing a general Fourier analysis to determine the individual frequency components contributing to the overall composite signal. The Fourier analysis is carried out using both cosine and sine terms.

As a third phase, refinements in the analysis are comtemplated. A spectral analysis typically consists of resolving ambiguities and questions that usually cannot be answered by a single analytical technique. The third phase is devoted to the study and development of techniques to be used instead of (or in addition to) the classical, generalized Fourier analysis mentioned above. Techniques to be investigated include auto-regressive analysis—to be used in conjunction with stepwise Fourier analysis—and digital bandpass filters, based on the Laplace transform.

5. INSTRUMENT CALIBRATION

To verify the proper operation of the instrument and to assure accuracy in the received data, an in-flight calibration technique will be used. Following is a summary of the final calibrations to be performed on earth, prelaunch check-of-calibration procedures, and in-flight calibration requirements.

5.1 Primary Calibration

The primary calibration and check of performance of the instrument will be carried out in a vacuum chamber with cryogenic walls at liquid N_2 temperature. The thermostat of the instrument will maintain the proper operating temperature of about 270°K in the optical cube. A blackbody will be placed in the field-of-view whose temperature can be adjusted between +80°K and +320°K. Interferograms will be taken at 10°K increments. Adjustments of gains, thresholds, etc., will be set at this time. Calibration with the instrument's own reference blackbody will also be checked at this time.

Spectral calibration and resolution will be checked by two methods: First, a polyethylene sheet will be inserted in front of the blackbody in vacuum. Polyethylene has numerous absorption bands of precisely known wavelengths. Second, while in air (or in a chamber with walls and instrument at room temperature), a blackbody at a distance of about 10m (1 m in the CO_2 chamber) will be kept 10 to 20 degrees above the instrument

temperature. The strong Q branch of the 15μ CO_2 band, which should be clearly observable in absorption, serves as a wavelength standard. As the mirror speed is derived from the 0.5852-micron neon reference line, this test will verify only proper scan speed. Adjustments will not be necessary.

The field-of-view of the instrument with respect to the mounting interface will be measured on an optical bench. The interferometer will not be operated in the conventional sense for this measurement; the output of the thermistor bolometer will be monitored while a chopped, high-intensity source is moved at a fixed distance. The response will be mapped as a function of the angular position of the source.

5.2 Calibration Checks During Environmental Tests

Before and after vibration, during vacuum thermal tests, etc., check of calibration are performed by exposure to blackbodies. In air, blackbodies above room temperature are used; in vacuum tests, blackbodies above and below operating temperatures (+80°K to 320°K). Also, exposure to the electrically-heated reference blackbody which is part of the instrument is performed regularly.

The spectral response is checked by placement of filters (polyethylene or mylar) in front of a blackbody.

The field-of-view is checked in air only, by exposure to a small hot source placed in front of a room temperature background. Readout takes place in the normal operation mode of the interferometer.

5.3 Check of Calibration in the Launch Area and on Launch Pad

Exposure to the electrically heated, thermostatically controlled reference blackbody will serve as the check of calibration.

5.4 Check of Calibration During Interplanetary Flight

Two complete interferometer frames will be required for each check of calibration. During the first frame, the reference blackbody will be placed within the field-of-view; this is the normal position of the reference blackbody during flight. The second calibration frame will be taken without the blackbody, with the interferometer viewing outer space. After this, the blackbody will be returned into its normal stowed position. In-flight calibration is required at regular intervals of 15 or 30 days, except during the first week, when one calibration every other day is required.

5.5 Calibration at Planetary Encounter

Shortly before and shortly after planetary encounter, a sequence of two calibration frames is required. The pre-encounter calibration should be as close as possible before encounter but definitely not earlier than 1 hour before. Both pre-encounter calibration frames and post-encounter frames should be stored and transmitted together with the data frames. After encounter, the sequence of the calibration frames will be reversed: first, the sky background will be measured as the planet moves out of the field-of-view of the instrument; then the reference blackbody will be brought in, and the final calibration frame will be taken. Thereafter no further data will be required from the interferometer.

6. MISSION PROFILE

6.1 Planetary Transfer

If possible, the instrument will be launched and stowed in the "power on" position. All circuits will be energized and the scan mirror placed in the midrange neutral position. The calibration target (blackbody) will be in front of the telescope so that it seals the instrument thermally; in this position, the blackbody will protect the optical elements during launch and will minimize the heater power necessary to maintain the proper operating temperatures of the instrument (270°K) and the blackbody (285°K). Total electrical power required during this time will not exceed the power required during the data-gathering period (approximately 7 watts). Check of calibration is performed at regular intervals as discussed in 5.4.

The following command and readout sequence constitutes the in-flight calibration:

- Frame-initiate command to obtain one interferogram of the reference target
- Command to move the blackbody out of the field-of-view of the telescope

- Frame command to obtain an interferogram of outer space
- Command to return the blackbody to the normal position in front of the telescope

6.2 Planetary Encounter Mode

The planetary encounter mode consists of calibration before and after encounter, as discussed in 5.5, with an intervening data-gathering period. The following sequence of commands and events constitutes the encounter mode:

- Frame-initiate command to obtain one interferogram of the reference target
- Command to move the blackbody to a position which will clear the field-of-view of the telescope
- Frame command to obtain an interferogram of outer space
- Frame command at least every 30 seconds (or, if possible, every 15 seconds) during planetary encounter
- Frame command after the planet has left the field-of-view
- Command to return the blackbody to the normal position in front of the telescope
- Frame-initiate command to obtain interferogram of reference target

After the encounter mode, the data-gathering part of the experiment will be completed and electrical power may be removed from the instrument.

6.3 Other Experiment Considerations

- Trajectory—No special requirements
- Fly-by distances—A fly-by distance of 2000 km (closest approach to surface) would be most desirable; at this distance, the interferometer can be used without a telescope. Presently the telescope is designed to cover the range between 5000 and 10,000 km. An approach distance between 10,000 and 20,000 km would also be acceptable, but would require a new (larger) design of the telescope.
- Fly-by velocity—Low fly-by relative velocity is desirable primarily to minimize the spatial smear during the 10-second time required for each interferogram.
- Aiming point—The aiming point should be chosen to fulfill the following conditions, listed in sequence of importance:
  (1) Equal chance to observe "dark" and "bright" areas of Mars
  (2) View of the subsolar point and equatorial regions
  (3) Scan across the terminator (morning terminator preferred)
  (4) Observe primarily the illuminated portion of Mars, but to have a chance to see part of the dark side also
- Magnetic cleanliness—No special requirements
- Electrical noise—Only the output circuit of the bolometer and the input to the preamplifier would be sensitive to stray electrical fields. Care will be taken to have proper shielding. The rest of the circuit, which is digital, should be insensitive to stray electrical fields. The noise level which can be tolerated at the electrical interface will be small, but numerical values cannot be specified at this time.
- Acoustical and structural noise—During the 10-second periods of data collection or calibration, the platform should not be exercised. The normal motion of adjustment (hunting) of the spacecraft's control system should not provide problems. The effect of structural noise is not fully known at this time, and specifications of the vibrational environment to which the instrument will be exposed on the spacecraft are urgently needed.
- Interaction and coordination with other experiments—It is strongly suggested that all radiometric experiments be bore-sighted to allow cross-correlation among the data obtained. Simultaneous ground-based observations should be encouraged.
- Spacecraft engineering measurements—No engineering measurements are required from the spacecraft. All auxiliary measurements are part of the interferometer experiment.

7. EXPERIMENT-SPACECRAFT INTERFACE

The interface between the experiment and the spacecraft (Figure 15) consists of power input, commands, clock pulses and data signals.

The frame-initiate commands supplied by the spacecraft consist of pulses occurring every 30 or 15 seconds after planet acquisition. The pulse should be about 2 milliseconds in duration and +2 volts in amplitude, with a risetime of about 10 microseconds from a source impedance of about 100 ohms.

The end-of-frame command supplied by the experiment and coupled into the spacecraft is a pulse with an amplitude of 2.5 volts, and a risetime of less than 5 microseconds from a source impedance of about 5K ohms.

The 400-cps clock consists of pulses of a 4-volt amplitude with a 0.3-microsecond risetime and a source impedance of about 100 ohms.

The 50-kc clock coupled from the experiment to the spacecraft will have a source impedance of about 100 ohms and an amplitude of 2.5 volts. The risetime will be on the order of 1 microsecond.

Data output from the experiment will consist of a series of 10-bit digital words, a "one" corresponding to +2.5 volts and a "zero" corresponding to 0 volts. The source impedance in the data channel is 5K ohms.

The real-time data from the thermal and visible channel will consist of · pulse-width about 10 microseconds in duration, with an amplitude of +2.5 volts and a 1-ms risetime.

Design Parameters for Infrared Interferometer

Spectral range	2000–500 cm^{-1} (5–20μ) (possible extension to 300 cm^{-1} (33μ)
Spectral resolution	5 cm^{-1} over total range
Travel of drive mirror	0.2 cm
Diameter of effective aperture	3.6 cm
Dectector	Thermistor bolometer in conical light pipe
Detector time constant	1–2 ms
Duration of interferogram	10 sec
Electrical frequencies at detector	20–80 cps
Sampling	Every fourth fringe of monochromatic source 5852.5 Å
Sampling rate	171 samples sec^{-1}
Samples per interferogram	1708

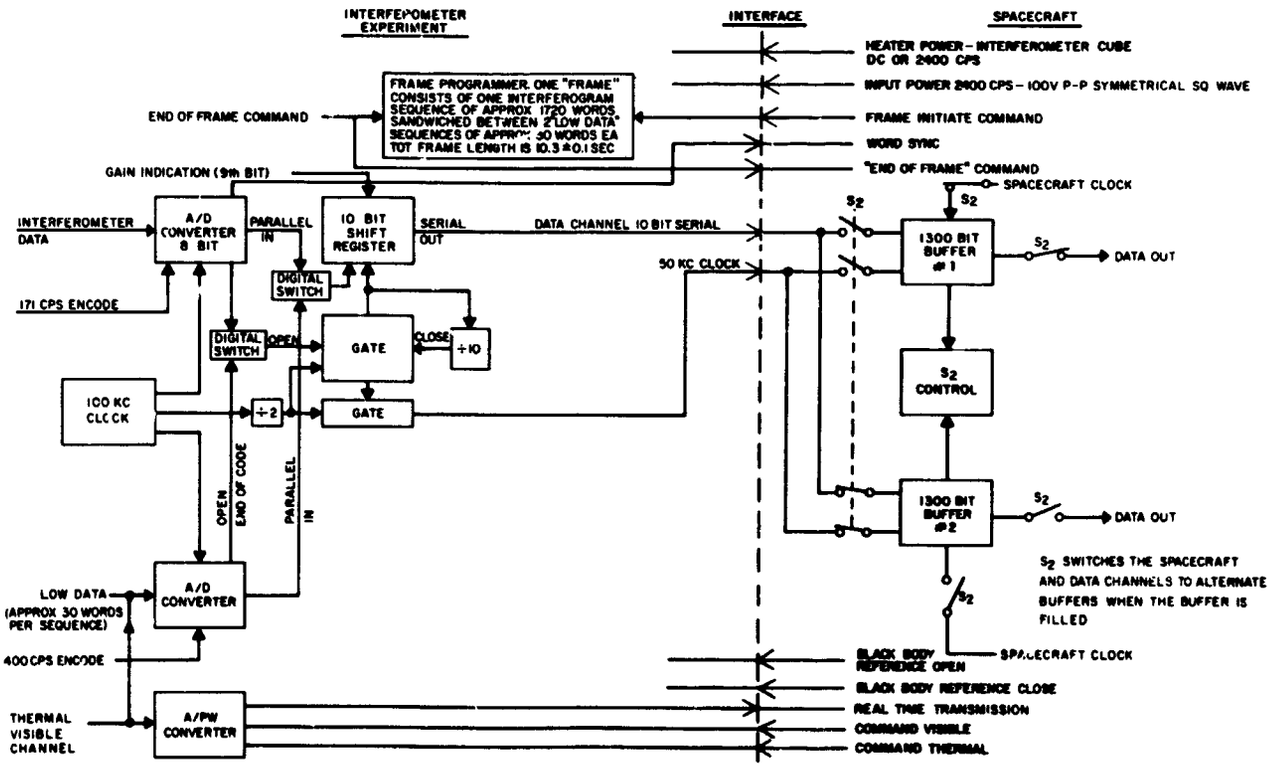


FIGURE 15.—Spacecraft-experiment interface.

Housekeeping information	25 samples per interferogram
Samples per frame	1733 samples
Frame synchronization	3 words
Words per frame	1736
Bits per word	10
Bits per frame	17360
Calibration frames in vincinity of planet	4
Minimum number of data frames for minimum fly-by speed	40 (every 30 sec) 80 (every 15 sec)
Total number of frames	44 (for 30-sec interval) 84 (for 15-sec interval)
Total number of bits to be stored	$.765\times10^6$ (for 30-sec interval) 1.46×10^6 (for 15-sec interval)
Field-of-view of interferometer	1.57×10^{-2} ster ($\sim8°$)
Field-of-view with telescope	2×10^{-3} ($\sim2.8°$) (approx. circle of 250 km diameter at 5000 km)
Calibration during interplanetary flight before and after planetary encounter	
a. space	
n. blackbody	
Monochromatic source	Neon lamp plus interference filter
Monochromatic wavelength	5852.5 Å
Detector for monochromatic light	Silicon junction
Frequency of reference signal	684 cps

REFERENCES

1. Hanel, R. A., and Bartko, F., Radiative Equilibrium in Planetary Atmospheres, NASA TN-G569 (1964).
2. Kuiper, G. P., ed., "The Atmospheres of the Earth and Planets," University of Chicago Press (1952).
3. Sinton, W. M., An Upper Limit to the Concentration of NO_2 and N_2O_4 in the Martian Atmosphere. *Publ. Astron. Soc. Pacific.*, Vol. 73, p. 125 (1961).
4. Kaplan, L. D., "The Spectroscope as a Tool for Atmospheric Sounding by Satellites," *Inst. Soc. of Amer. Conf.* (Sept. 1960).
5. Wark, D. Q., "On Indirect Soundings of the Stratosphere from Satellites," *J. Geophys. Res.*, Vol. 66, p. 77 (1961).
6. Yamamoto, G., "Numerical Methods for Estimating the Stratospheric Temperature Distribution from Satellite Measurements in the CO_2 Band," *J. of Meterol.*, Vol. 18, p. 581 (1961).
7. King, J. I. F., "Meterological Inferences from Satellite Radiometry, I," *J. Atmos. Sci.*, Vol. 20, p. 245 (1963).
8. Fleming, H., Twomey, S., and Wark, D. Q., "The Inference of the Vertical Distribution of Atmospheric Parameters from Spectral Measurements," paper presented at the 25th national meeting of the Amer. Meteorol. Soc., Los Angeles (Jan. 1964).
9. Kuiper, G. P., paper presented at 45h American Geophysical Union meeting, Washington, D. C. (April 1964).
10. Lyon, R. J. P., Evaluation of Infrared Spectrophotometry for Compositional Analysis of Lunar and Planetary Soils. NASA Contract (NASA-49/04) Quarterly Status Report No. 4, Stanford Research Institute (June 1963).
11. Bell, E. E., and Eisner, I. L., Infrared Radiation from the White Sands at White Sands National Monument, New Mexico. *Jour. Opt. Soc. of Amer.*, Vol. 46, p. 303 (1956).
12. Fellgett, P., Spectrometric Interferential Multiples Power Measures Infra-Rouges sur les Etoiles. *J. d'Physique et la Rad.*, Vol. 19, p. 237 (1958).
13. Gebbie, H. A., *J. Phys. Radium*, Vol. 19, p. 230 (1959).
14. Mertz, L., *J. Opt. Soc. Amer.*, Vol. 46, p. 548 (1956).
15. Connes, J., *Rev. d'Opt.*, Vol. 40, p. 45, 101, 151, 213 (1961).
16. Strong, J., *Concepts of Classical Optics*, W. H. Freeman and Co., San Francisco (1958).
17. Strong, J., and Vanesse, G. A., Interferometric Spectroscopy in the Far Infrared. *J. Opt. Soc. Amer.* Vol. 49, p. 844 (1959).
18. Loewenstein, E. V., On the Correction of Phase Errors in Interferograms. *App. Optics*, Vol. 2, p. 491 (1963).
19. Jacquinot, P., *Reports on Progress in Physics*, Vol. 23, p. 263 (1960).
20. Connes, J., Recherches sur la Spectroscopie par Transformation de Fourier. *Rev. d'Opt.*, Vol. 40, pp. 45, 101, 157, 213 (1961).

See also:

Öptik, E. J., The Atmosphere and Haze of Mars, *J. Geophys. Res.*, Vol. 65, p. 3057 (1960).

Rea, D. G., and Welch, W. J., The Reflection and Emission of Electromagnetic Radiation by Planetary Surfaces, *Space Science Review*, Vol. 2, p. 558 (1963).

Steel, W. H., Interferometers Without Collimation for Fourier Spectroscopy. *J. Opt. Soc. Amer.* Vol. 54, p. 151 (1964).

De Vaucouleurs, G., Physics of the Planet Mars, Faber and Faber, London (1954).

Kiselev, B. A. and Parshin, P. F., Some Distortion in Fourier Spectroscopy. *Opt. i. Spektr.*, Vol. 12, p. 311 (1961).

Meissner, K. W., Interference Spectroscopy. *J. Opt. Soc. Amer.*, Vol. 31, p. 405 (1941).

Strong, J., Interferometry for the Far Infrared. *J. Opt. Soc. Amer.*, Vol. 47, p. 354 (1957).

Kuiper, G. P., "Planets and Satellites," Vol. 3 of The Solar System (Kuiper, G. P. and Middlehurst, B. M., ed.), Chicago: Univ. of Chicago Press (1961).

Urey, H. C. "The Atmosphere of the Planets" in Handbuch der Physik (S. Fluegge, ed.), V. 52:363–418, Berlin: Springer-Verlag (1959).

Sagan, C., and Kellogg, W. W., "The Terrestrial Planets," *Ann. Rev., Astron. Astrophys.*, 1:235–266 (1963).

N64-27737

RADIATIVE EQUILIBRIUM IN PLANETARY ATMOSPHERES. I. APPLICATION OF THE STRONG LINE ABSORPTION LAW TO THE ATMOSPHERE OF VENUS*

RUDOLF A. HANEL AND FRANK BARTKO

The interpretation of the data obtained for Venus in the $8-12\mu$ atmospheric window clearly requires some knowledge of the temperature distribution in the planet's upper atmosphere. For this purpose temperatures, net fluxes, and specific intensities were calculated for a CO_2-N_2 atmosphere in radiative equilibrium. The wavelength, temperature, and pressure dependence of the CO_2 absorption, including the excited bands, was considered. Solar radiation absorbed in the clear atmosphere influences the temperature profile distinctly. The effects of various CO_2 concentrations, cloud top pressure levels, and cloud reflectivities on the equilibrium temperatures are also discussed.

LIST OF SYMBOLS

$\bar{A}_i$	Fractional absorptance in the wavelength interval i.
$B_\nu(T)$	Planck function at temperature T (watts cm^{-2} $ster^{-1}$ wave $number^{-1}$).
c	Subscript denoting cloud surface.
$E_n(\tau)$	Exponential integral of order n.
$\pi F_\nu(\tau_\nu)$	Spectral total flux at τ_ν (watts cm^{-2} wave $number^{-1}$).
H	Pressure scale height (H_0=value at 273°K).
i	Index denoting wave number interval.
$I_\nu(\mu, \tau_\nu)$	Specific intensity (watts cm^{-2} $ster^{-1}$ wave $number^{-1}$).
j	Index denoting vertical coordinate.
$J_\nu(\tau_\nu)$	Spectral mean intensity at optical depth τ_ν (watts cm^{-2} wave $number^{-1}$ $ster^{-1}$).
l_ν	Generalized absorption coefficient, $l=m^n$ (cm^{-1} NTP).
m	Parameter defined in Equation 19.
n	Exponent defined in Equations 2 and 19.
P	Pressure (P_0=1 atm).
q	CO_2 volume concentration.
r_{ic}	Cloud reflectivity in wave number interval i.
s	Geometric path length (cm).
S_ν	Solar constant in wave number interval 18 (watts cm^{-2}).
t	Integration variable.
T	Atmospheric temperature (T_0=273°K).
u	Path length of active gas at NTP (cm).
u*	Modified pressure- and temperature-reduced path length (cm). See Equation 16.
X	Parameter proportional to the product of the spectral line intensity and the ratio of path length to line half width.
α	Flux convergence criterion (Equation 40).
β	Parameter proportional to the ratio of the spectral line half-width to the average line spacing.

*Published as *NASA Technical Note* D-2397, August 1964; also presented at the Conference on Atmospheric Radiation, Eppley Laboratories, Newport, R. I.

γ Exponent describing the temperature dependence of excited bands (Equation 13).

$\epsilon_{\nu 0}$ Spectral emissivity of cloud surface.

ζ Cosine of the solar zenith angle.

μ Direction cosine ($\mu = \cos\theta$).

ν Wave number (cm^{-1}).

ρ Atmospheric density.

$\tau_{\mu\nu}$ Optical thickness in direction μ at wave number ν.

ω Solid angle (ster).

INTRODUCTION

The structure of a planet's atmosphere is governed primarily by convective and radiative energy transport. On the earth, convective processes preponderate in the troposhpere and radiative processes in the stratosphere. Similarly, the atmospheres of Mars and Venus should exhibit zones of convective and radiative transfer. Convection probably characterizes the Venus atmosphere below the cloud level, especially if the clouds consist of dust particles, whereas the region above the cloud layer is probably dominated by radiative transfer. The interpretation of infrared measurements of the atmospheres of the terrestrial planets must, therefore, be based on the theory of radiative transfer and appropriately modified for convection where necessary. The objectives of the program discussed in this paper were to establish an analytical tool for the study of planetary atmospheres based on the radiative transfer theory. An initial version of this program has been applied to the atmosphere of Venus. This application must be used with relatively inaccurate estimates of physical parameters which influence the solution of the transfer equation. However, the program permits a parametric study where the number of possible solutions is ultimately restricted by observational constraints.

The equilibrium temperature distribution of an N_2–CO_2 atmosphere was calculated for several volume concentrations of CO_2. The wavelength, temperature, and pressure dependence of the absorption coefficients was taken into account. The effect of direct and diffusely reflected solar radiation absorbed by the atmosphere was also included. An effective surface emissivity was introduced as an additional parameter and permitted a deviation from the usual assumption that the surface radiates like a blackbody in the infrared. In addition to the radiative equilibrium temperatures, the specific intensities were calculated. Net fluxes were obtained by angular integration, taking into account fully the dependence of the intensity on direction. The simplifying procedure of converting parallel beam radiation to an equivalent diffuse radiation field was avoided (References 1–3).

The net flux of a planet can be determined independently from the heat budget by considering the observed albedo and the appropriate value of the solar constant. This restricts the number of theoretically possible solutions, since only those which yield total flux values consistent with the albedo can be admitted. The number of solutions can be restricted further by a comparison of calculated and measured values of the specific intensity at several wavelength intervals and zenith angles.

In the discussion below, the assumptions and limitations of the calculations are stated and the mathematical formulation is presented. An illustration of the effects of the various parameters on the temperature profiles follows, but further conclusions will be discussed elsewhere.

SOLUTION OF THE RADIATIVE TRANSFER EQUATION

In order to calculate the equilibrium temperatures in a finite, nongrey, optically thick atmosphere, the equation of radiative transfer must be solved, subject to appropriate boundary conditions (Reference 4, for example). The assumption of radiative equilibrium holds as long as the atmosphere is stable against convection (Reference 5, for example). Stability, in turn, is maintained as long as the temperature gradient is everywhere less than the adiabatic gradiet.

For the particular application considered here, the surface is taken to be the cloud tops and is assumed to be the demarcation between regions of energy transport dominated by radiation and convection. It is realized that the assumption of a solid surface for the clouds is a severe simplification. A more rigorous treatment would apply

the theory of radiative transfer for a scattering medium (Reference 6, for example).

For an atmosphere in local thermodynamic equilibrium the equation of transfer is

$$\frac{dI_\nu(\mu, \tau_{\mu\nu})}{d\tau_{\mu\nu}} = I_\nu(\mu, \tau_{\mu\nu}) - B_\nu(T) , \quad (1)$$

where $B_\nu(T)$ is the Planck function and $I_\nu(\mu, \tau_{\mu\nu})$ the specific intensity of the radiation field. For radiative transfer calculations dealing with strong molecular band absorption, the optical thickness $\tau_{\mu\nu}$ can be expressed

$$\tau_{\mu\nu}(l_\nu, u^*, \mu) = l_\nu \left(\frac{u^*}{\mu}\right)^n = \frac{\tau_\nu(l_\nu, u^*)}{\mu^n} , \quad (2)$$

where l_ν is a generalized absorption coefficient and u^*/μ a pressure- and temperature-reduced path length. It is convenient to use the optical path in the vertical direction, τ_ν, and the direction cosine μ, instead of the optical path, $\tau_{\mu\nu}$, in an arbitrary direction. The appearance of μ under the exponent n has a significant influence on the transfer problem. The value of n is unity for weak line or grey absorption and one-half or less for the square root or strong line absorption law (Reference 7). In contrast to weak line or grey absorption, the strong line absorption law reduces the thermal coupling between adjacent atmospheric layers and permits a more effective exchange of radiative energy over greater optical distances.

Very general boundary conditions have been selected. At some optical depth, $\tau_{\nu c}$, the atmosphere is assumed to be bounded by a perfectly diffuse reflector of effective emissivity $\epsilon_{\nu c}$, related to the surface reflectivity by

$$\epsilon_{\nu c} = 1 - r_{\nu c} . \quad (3)$$

For positive values of μ this assumption requires the solution of Equation 1 at the boundary $\tau_{\nu c}$ to be

$$I_\nu^+(\tau_\nu = \tau_{\nu c}) = \epsilon_{\nu c} B_\nu(T_c) + r_{\nu c} F_\nu^-(\tau_{\nu c}) + \frac{S_\nu}{\pi} \zeta r_{\nu c} e^{-\frac{\tau_{\nu c}}{\zeta^n}} . \quad (4)$$

The first term on the right represents the thermal emission by the surface of temperature T_c. The second term arises from the nonblack characteristic of the surface; the downward flux incident on the surface is partially reflected back into the atmosphere. The third term describes the reflected solar component.

At the top of the atmosphere where $\tau_\nu > 0$ only solar radiation is assumed incident. For thermal radiation outer space is considered a perfect sink. Hence the upper boundary condition is

$$I_\nu^-(\zeta, \tau_\nu = 0) = \frac{S_\nu}{\pi} \zeta . \quad (5)$$

The solution of the transfer equation subject to the boundary conditions and the absorption law for τ_ν may be verified to be

$$I_\nu^+(\mu, \tau_\nu) = \epsilon_{\nu c} B_\nu(T_c) e^{-\frac{\tau_{\nu c} - \tau_\nu}{\mu^n}} + \int_{\tau_\nu}^{\tau_{\nu c}} B_\nu(t) e^{-\frac{t - \tau_\nu}{\mu^n}} \frac{dt}{\mu^n}$$

$$+ r_{\nu c} e^{-\frac{\tau_{\nu c} - \tau_\nu}{\mu^n}} \frac{2}{n} \int_0^{\tau_{\nu c}} B_\nu(t) E_{\frac{2}{n}}(\tau_{\nu c} - t) dt + \frac{S_\nu}{\pi} r_{\nu c} \zeta e^{-\frac{\tau_{\nu c}}{\zeta^n}} e^{-\frac{\tau_{\nu c} - \tau_\nu}{\mu^n}} , \quad (6)$$

and

$$I_\nu^-(\mu, \tau_\nu) = \int_0^{\tau_\nu} B_\nu(t) e^{-\frac{\tau_\nu - t}{\mu^n}} \frac{dt}{\mu^n} + \frac{S_\nu}{\pi} \zeta e^{-\frac{\tau_\nu}{\zeta^n}} . \quad (7)$$

In Equation 6 $I_\nu^+(\mu, \tau_\nu)$ is the upward intensity whose respective components arise from the surface emission, the atmospheric emission by all layers at optical depths greater than τ_ν, and the downward fluxes diffusely reflected upwards. $I_\nu^-(\mu, \tau_\nu)$ is the downward intensity consisting of the atmospheric emission arising from all layers with optical depths less than τ_ν and the incident solar radiation.

The temperature distribution is obtained from a solution of the equation of radiative equilibrium which equates the total emitted and absorbed radiation for each volume element. For a stratified atmosphere each layer must satisfy the energy balance condition

$$\int_\nu \int_\omega \int_V l_\nu \, e^{-\frac{\Delta\tau}{\mu^n}} B_\nu \, dV \, d\omega \, d\nu = \int_\nu \int_\omega \int_V l_\nu \, e^{-\frac{\Delta\tau}{\mu^n}} I_\nu \, dV \, d\omega \, d\nu \, . \qquad (8)$$

The exponential term represents the attenuation losses due to self-absorption in the layer of finite optical thickness. Integration over all directions is required since I_ν as well as $e^{-\Delta\tau/\mu^n}$ is a function of μ. Furthermore, since absorption and emission may occur in different parts of the spectrum the equality holds only if the integration is taken over all frequencies. This general form of the condition of radiative equilibrium applies to any volume element as well as to different forms of absorption laws. If $\Delta\tau$ is very small in all directions, the familiar form (References 8 and 9) of the radiative equilibrium equation results:

$$\int_\nu l_\nu \, B_\nu \, d\nu = \int_\nu l_\nu \, J_\nu \, d\nu \, . \qquad (9)$$

Substitution of the intensities obtained (Equations 6 and 7) into the equation of radiative equilibrium leads to a Milne-type integral equation in $B_\nu(\tau)$, whose solution is required in obtaining the temperature distribution. The solution is obtained by iteration beginning from an assumed temperature distribution. The net flux is computed after each iteration. When the distribution of the net flux is constant within specified limits, radiative equilibrium is achieved and the calculation proceeds to the final step, the computation of limb darkening.

The components of the upward and downward intensities, the net flux, and the condition of radiative equilibrium will be examined in more detail in the manner required for computer solution. Hence a brief description of the coordinate system and the calculation of τ_ν follows.

COORDINATE SYSTEM

A plane parallel atmosphere is assumed and the computations are carried out in a 33 layer coordinate system (Figure 1). The interface between

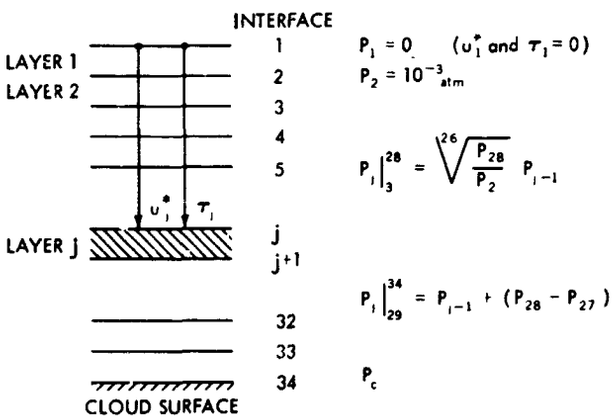


FIGURE 1.—Coordinate system.

layer j–1 and j is called interface j and the quantities u_j^* and τ_j are computed from the top of the atmosphere (interface one, zero pressure) to interface j and the corresponding pressure P_j. In all computations the pressure at interface j=2 was taken equal to 10^{-3} atm. Layer one above this level is included for the sake of completeness. The computed temperature is only an average value and it should not be interpreted that layer one is isothermal. For this region of the atmosphere, local thermodynamic equilibrium may not exist (Reference 10, for eaxmple) and the absorption lay adopted may not strictly apply (Reference 11). In addition, photochemical and other atmospheric processes will have a strong effect on

the temperature in this region. For layers between j = 2 and 28, pressures are calculated according to the recursion formula

$$P_j = \sqrt[26]{\frac{P_{28}}{P_2}}\, P_{j-1} \quad . \tag{10}$$

This mode of pressure slicing yields layers of approximately the same geometrical height but of increasing optical thickness. To avoid a very thick layer adjacent to the cloud boundary, a linear pressure slicing was applied between j = 28 and the surface (j = 34).

ABSORPTION COEFFICIENTS

The calculation of the optical thickness requires a value of the absorption coefficient, which varies with pressure and temperature, and a value for the mass of the absorbing gas. For pressures greater than 10^{-3} atm the individual line shape is determined primarily by collision broadening. The effect of this factor on atmospheric structure has been studied by several authors (Reference 12, for example). For band absorption the concept of optical thickness defined for a single frequency is still very useful, but τ becomes, in general, a complicated function of u*. For all cases studied in this paper, an analytical expression can be derived for this function, based on the so-called strong line absorption law (Reference 7). This approximation applies to CO_2 within the pressure ranges of interest, namely from 10^{-3} to 1 atm for CO_2 concentrations exceeding 0.01 part per volume.

The mass of active gas is usually expressed in terms of a path length. An element of the path length is defined by the product of the geometric path length ds and the normalized density of the active gas

$$du = q\frac{\rho}{\rho_0}\, ds \quad . \tag{11}$$

Since $\rho = MP/RT$,

$$du = q\,\frac{P}{P_0}\,\frac{T_0}{T}\, ds \quad . \tag{12}$$

In the domain of collision broadening, the pressure and temperature dependences of the half-width of absorption lines are usually considered by defining a reduced path length equal to the product of du, the pressure, and the inverse square root of the temperature. The temperature dependence of the excited bands may be included by an exponential term which represents the increase of the population of the lower vibrational energy state with temperature (References 13, for example). An additional temperature dependence of the rotational levels, within each vibrational state, has the effect of altering the band shape. However, this effect is much smaller than that due to changes in the population of the vibrational state. The reduced path length, with these effects, is then

$$du^* = q\left(\frac{P}{P_0}\right)^2 \left(\frac{T_0}{T}\right)^{3/2} e^{\gamma\left(\sqrt{250} - \sqrt{T}\right)}\, ds \quad . \tag{13}$$

The numerical value of γ was computed from Reference 14 for each spectral region, and the adopted values are listed in Table 1. The value

Table 1

Values used in the Computation of CO_2 Absorption for Various Spectral Intervals.

i	ν (cm^{-1})	m	n	γ
1	0-200	0	0.50	0
2	200-300	0	0.50	0
3	300-400	0	0.50	0
4	400-495	0	0.50	0
5	495-550	7.8×10^{-5}	0.60	0
6	550-625	4.6×10^{-3}	0.42	0
7	625-660	9.0×10^{-1}	0.38	0
8	660-720	4.9×10^{-1}	0.42	0
9	720-810	9.5×10^{-4}	0.40	0
10	810-880	6.5×10^{-5}	0.55	2200
11	880-920	8.6×10^{-6}	0.56	2760
12	920-1000	4.9×10^{-6}	0.54	2330
13	1000-1100	4.4×10^{-5}	0.52	2290
14	1100-1400	0	0.50	0
15	1400-2000	0	0.50	0
16	2000-2600	6.6	0.50	0
17	2600-8000	1.1×10^{-3}	0.50	0
18	1700-8000 (solar)	4.9×10^{-7}	0.27	0

μds equals the vertical path element dh. Thus, by using the hydrostatic equation

$$\frac{dh}{H} = -\frac{dP}{P} \quad . \tag{14}$$

the reduced path length in a direction μ can be expressed by

$$du^*(\mu) = -\frac{1}{\mu} qH_0 \left(\frac{T_0}{T}\right)^{1/2} e^{\gamma\left(\frac{1}{250} - \frac{1}{T}\right)} \frac{P\,dP}{P_0^2} \quad (15)$$

For the remainder of this paper, only the reduced path length in the vertical direction, $u^*(\mu=1)$, is used. The reduced path length from a level P to outer space is a function of the wavelength as well as the temperature profile,

$$u_\nu^* = qH_0 \int_0^P \left[\frac{T_0}{T(P)}\right]^{1/2} e^{\gamma_\nu \left(\frac{1}{250} - \frac{1}{T(P)}\right)} \frac{P\,dP}{P_0^2} \quad (16)$$

For the numerical computations Equation 16 was reformulated in terms of the adopted coordinate system,

$$u_{i,j}^* = u_{i,j-1}^* + qH_0 \left(\frac{T_0}{T_{j-1}}\right)^{1/2} e^{\gamma_i \left(\frac{1}{250} - \frac{1}{T_{j-1}}\right)} \frac{P_j^2 - P_{j-1}^2}{2P_0^2} \quad (17)$$

A constant temperature within each layer is assumed. The scale height H_0 for standard temperature is a function of the molecular weight and, therefore, of the fractional concentration of each gas; adopted values are given in Table 2.

Table 2

Pressure Scale Height in the Atmosphere of Venus as a Function of the CO_2 Concentration.

q(CO_2 per volume)	H_0 (cm)	q H_0 (cm)
1.0	5.901×10^5	5.901×10^5
0.75	6.491×10^5	4.865×10^5
0.5	7.212×10^5	3.606×10^5
0.25	8.114×10^5	2.029×10^5
0.1	8.771×10^5	8.771×10^4
0.05	9.015×10^5	4.508×10^4
0.01	9.210×10^5	9.210×10^3

Since CO_2 is the primary absorbing constituent, an accurate representation of its absorption spectrum as a function of pressure, temperature, and path length is required. The absorption spectrum of CO_2 can be separated into three important regions. The near infrared region from 1 to about 6μ is most important in regard to the absorption of solar radiation but less significant in regard to the thermal emission spectrum of the clouds (Reference 15). The 9.4 and 10.4μ excited bands of CO_2 are within the second region considered in this calculation. They lie in a spectral region for which planetary observations exist (References 16–18), although they play only a minor role in the determination of the atmospheric structure. Any realistic interpretation of the data from References 16–18 must necessarily represent this part of the spectrum adequately. The strong bands near 15μ determine the atmospheric structure in the radiative regime; the interval of 12–20μ is the third region considered.

The spectrum should be divided into many intervals for accurate representation, but the number of intervals must be limited to avoid unreasonably long computing times. As a compromise, the 18 intervals listed in Table 1 were selected.

For the pressure and temperature ranges considered, the individual line shapes are controlled primarily by collision broadening. For high concentrations of CO_2 ($q \geq 0.01$) the distribution of the lines and their spacings in the major bands indicates appreciable overlapping. The square root law would overestimate the absorption in such intervals; but the strong line approximation is a generalization of this law which is valid even when overlapping occurs. This approximation is expressed in terms of a single variable u^* and is well suited to extrapolation of the absorption to large path lengths and low pressures (Reference 7).

For each spectral interval, the fractional absorption was computed from the expression

$$\bar{A}_i = 1 - e^{-\tau_i} \quad (18)$$

where

$$\tau_i = (m_i u_i^*)^{n_i} \quad (19)$$

The m_i are similar to the generalized absorption coefficients and the n_i are constants determined by the degree of overlapping of the individual lines and the slope of the curve of growth.

Wherever possible, the spectral intervals were chosen to coincide with the intervals used by Burch et al (Reference 13). The coefficients m_i and n_i were determined graphically by a fit of the strong line curve with applicable data points taken from this source. Where experimental data were not available the calculated tables of CO_2 absorption compiled by Stull et al. (Reference 14) were used. A sample is shown in Figure 2 and the various values of m and n are listed in Table 1.

The application of the strong line approximation is valid whenever the central regions of the various lines become opaque. For each spectral interval a check of the validity of the strong line approximation for the conditions considered was carried out. A relatively weak absorbing region (720–810 cm^{-1}) of the 15μ CO_2 band may serve as an illustration. If the approximation holds for this spectral interval, then it will be valid for stronger absorbing regions. According to Plass (Reference 7) the strong line approximation applies for all values of β to an accuracy of 10 percent, if $X > 1.63$. For this interval

$$\left.\begin{aligned} \beta &= 0.023\ P\ , \\ X &= 0.01\ \frac{u}{P}\ . \end{aligned}\right\} \qquad (20)$$

According to Equations 12 and 14, u/P equals qH and defines the CO_2 concentration (see Table 2). As can be seen, the strong line approximation is valid within an accuracy of 10 percent for CO_2 concentrations in excess of 10^{-3} and more accurate for concentrations in excess of 10^{-2}, which is the range of interest for Venus.

SPECIFIC INTENSITY

The formulas derived in the section concerned with the solution of the radiative transfer equation must be expressed in terms of the adopted coordinate system. The specific intensity is a function of ν, τ, and μ and is conveniently expressed by

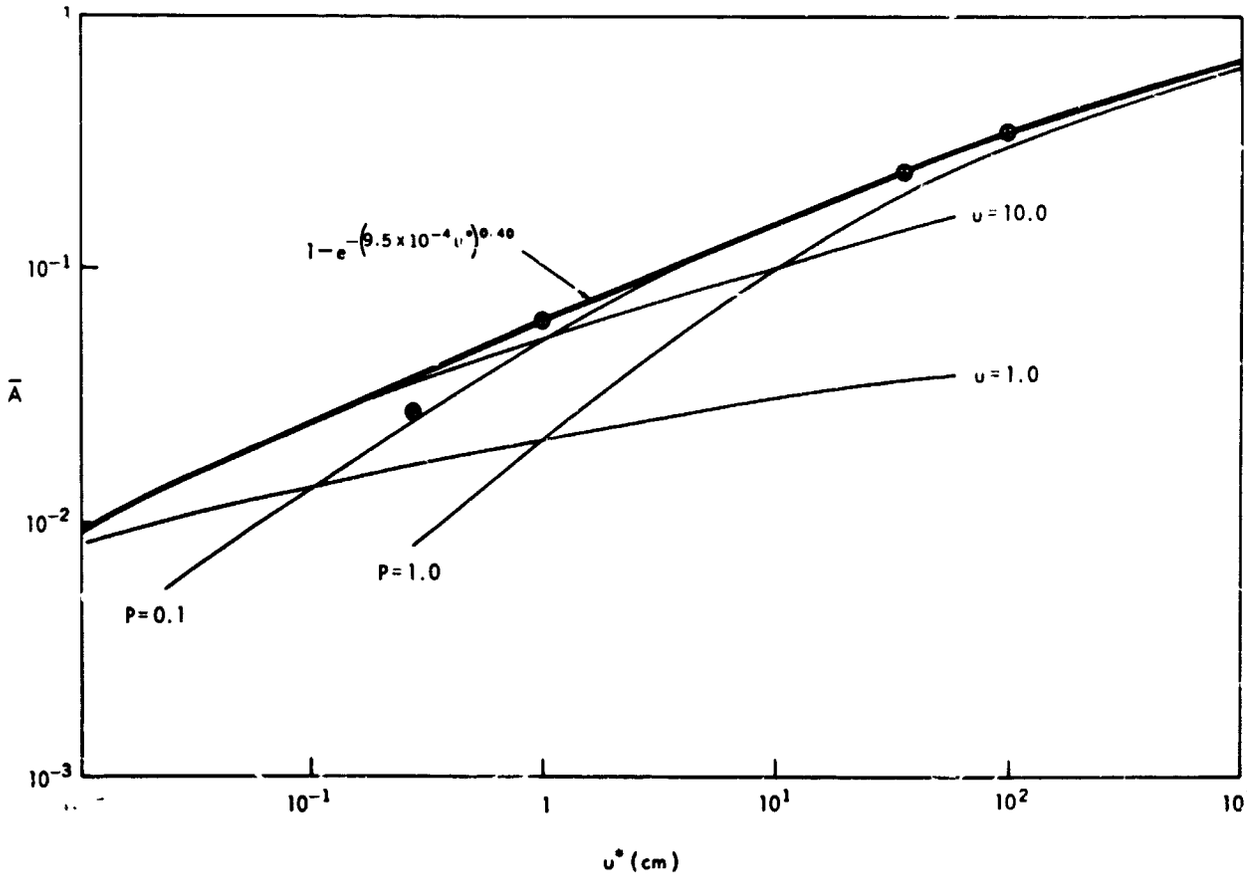


FIGURE 2.—Illustration of strong line approximation fit for the spectral interval 720–810 cm^{-1}, as derived from Stull et al. (Reference 14). Data points taken from Burch et al. (Reference 13) and slightly modified are included for comparison.

upward and downward vectors. As mentioned before, the solution of the transfer equation appropriate here requires four components for the upward intensity:

1. Thermal emission from the cloud surface
2. Direct thermal emission from the atmosphere below interface j
3. Downward thermal emission from the whole atmosphere, reflected upward by the clouds
4. Downward solar radiation reflected upward by the clouds.

The downward intensity includes values for:

5. Direct emission from the atmosphere above interface j
6. Direct solar radiation.

These six components are shown in Figure 3 and discussed below.

Surface Emission

For a frequency interval i and an interface j, the thermal radiation emitted, which is attenuated between the sufrace and the interface j, is given by

$$I_{ij} = \epsilon_{ic} B_{ic} e^{-\frac{\tau_{ic} - \tau_{ij}}{\mu^n}} . \qquad (21)$$

The designation of emissivity applies directly for a solid or liquid surface, but some qualifications regarding the interpretation of ϵ_{ic} are necessary for a scattering medium such as a cloud surface. At this level the gas and cloud particles have a

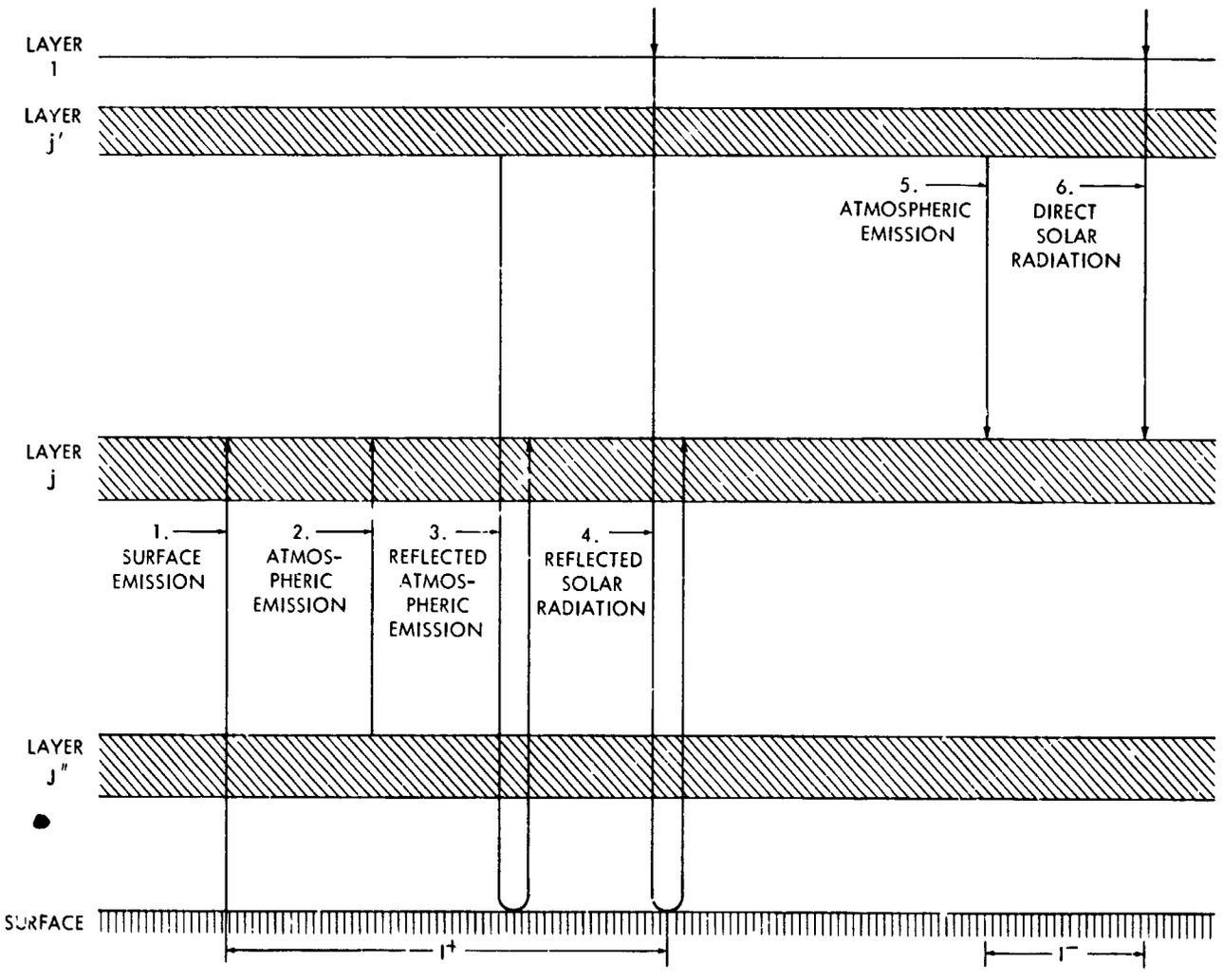


FIGURE 3.—Schematic diagram of the radiation components considered.

temperature T_c. For those spectral intervals where the atmospheric gas is opaque, the optical properties of the cloud particles are not important. In regions of the spectrum where the gas between particles is transparent the optical properties of the particles and their concentration become very important. For these intervals ϵ_c is defined as a boundary value, which indicates how much smaller the emerging radiation is compared with a blackbody at temperature T_c. The cloud emissivity was taken as unity in the strong absorption bands and set equal to a constant value in the transparent regions. In general, ϵ_{ic} may be a complicated function of wavelength and μ but this dependence is ignored.

Emission from the Atmosphere below Interface j

The contribution from each layer is proportional to the Planck function, an attentuation factor, and the optical thickness,

$$I_{ij} = \sum_{j'=j}^{33} B_{ij'} \, e^{-\frac{\tau_{ij'} - \tau_{ij}}{\mu^n}} \, \frac{\Delta\tau_{ij'}}{\mu^n} \quad . \tag{22}$$

Emission from the Atmosphere Reflected by the Clouds

This component is the product of the downward flux at the cloud surface, the reflectivity, and the attenuation factor of the atmosphere between the surface and τ_j. Since constant temperature is assumed within a layer, the flux integral (see below) can be integrated for each layer by using the recursion formula for exponential integrals,

$$E_n(\tau) \, d\tau = dE_{n+1}(\tau) \quad . \tag{23}$$

The reflected intensity becomes

$$I_{ij} = r_{ic} \, e^{-\frac{\tau_{ic} - \tau_{ij}}{\mu^n}} \, \frac{2}{n} \sum_{j'=1}^{33} B_{ij'} \left[E_{\frac{2}{n}+1}\left(\tau_{ic} - \tau_{i,\,j'+1}\right) - E_{\frac{2}{n}+1}\left(\tau_{ic} - \tau_{ij'}\right) \right] \quad . \tag{24}$$

Reflected Solar Radiation

Similarly, the reflected solar intensity is calculated from

$$I_{18,\,j} = \frac{S}{\pi} \, \zeta \, r_{18,\,c} \, e^{-\left(\frac{\tau_c}{\zeta^n} + \frac{\tau_c - \tau_j}{\mu^n}\right)} \quad . \tag{25}$$

Within the spectral range of 1.25 to 5.9μ the reflectivity of the clouds for solar radiation was taken equal to 0.6 which is consistent with recent observations (Reference 19). Again clouds were considered to be perfect diffusers.

Direct Emission from Atmospheric Layers above Interface j

This contribution is analogous to the emission from lower layers:

$$I_{ij} = \sum_{j'=1}^{j-1} B_{ij'} \, e^{-\frac{\tau_{ij} - \tau_{ij'}}{\mu^n}} \, \frac{\Delta\tau_{ij'}}{\mu^n} \quad . \tag{26}$$

Direct Solar Radiation

The direct solar radiation is given by the solar constant for 1.25 to 5.9μ appropriate for the distance of Venus (0.047 watt cm^{-2}):

$$I_{18,\,j} = \frac{S}{\pi} \, \zeta \, e^{-\frac{\tau_j}{\zeta^n}} \quad . \tag{27}$$

FLUX

The flux is derived conventionally by integration of the components of specific intensity over μ (Reference 9, for example):

$$\pi F = \int I \, \mu \, d\omega = 2\pi \int I \, \mu \, d\mu \quad . \tag{28}$$

Integration leads to exponential integrals of the form of

$$\int_0^1 e^{-\frac{\tau}{\mu^n}} \, \mu \, d\mu = \frac{1}{n} \, E_{\frac{2}{n}+1}(\tau) \quad , \tag{29}$$

$$\int_0^1 e^{-\frac{\tau}{\mu^n}} \, \frac{d\mu}{\mu^{n-1}} = \frac{1}{n} \, E_{\frac{2}{n}}(\tau) \quad . \tag{30}$$

The results are expressed by exponential integrals tabulated for integral order numbers (Reference 20). For all wavelength intervals concerned with thermal radiation (i = 1 to 17) n is close to 1/2; for integration, n was taken equal to 1/2, which leads to exponential integrals of the fifth and fourth orders respectively. For the interval concerned with solar radiation (i = 18) the exponent n is 0.27; for integration, n was taken equal to 1/4, which leads in this case to exponential integrals of the ninth and eighth orders respectively.

Analogous to the intensity, the upward flux consists of four components, and the downward flux consists of two.

The individual components of the upward flux F_{ij}^{+} are

1. $$4\epsilon_{ic} B_{ic} E_5(\tau_{ic} - \tau_{ij}) \quad , \tag{31}$$

2. $$4 \sum_{j'=j}^{33} B_{ij'} \left[E_5(\tau_{ij'} - \tau_{ij}) - E_5(\tau_{i,j'+1} - \tau_{ij}) \right] \quad , \tag{32}$$

3. $$16 r_{ic} E_5(\tau_{ic} - \tau_{ij}) \sum_{j'=1}^{33} B_{ij'} \left[E_5(\tau_{ic} - \tau_{i,j'+1}) - E_5(\tau_{ic} - \tau_{ij'}) \right] \quad , \tag{33}$$

4. $$8 \frac{S}{\pi} r_{18,c} \zeta e^{-\frac{\tau_c}{\zeta^n}} E_9(\tau_c - \tau_j) \quad . \tag{34}$$

The components for the downward flux F_{ij}^{-} are

5. $$4 \sum_{j'=1}^{j-1} B_{ij'} \left[E_5(\tau_{ij} - \tau_{i,j'+1}) - E_5(\tau_{ij} - \tau_{ij'}) \right] \quad , \tag{35}$$

6. $$\frac{S}{\pi} \zeta e^{-\frac{\tau_j}{\zeta^n}} \quad . \tag{36}$$

The net flux ($\pi F_\nu = \pi F_\nu^{+} - \pi F_\nu^{-}$) is calculated for each layer j, for each wavelength interval i, and for the sum over all spectral intervals.

CONDITION OF RADIATIVE EQUILIBRIUM

The general condition of radiative equilibrium (Equation 8) will now be applied to an atmosphere consisting of many layers of finite optical thickness. For an element of area dA in a layer of optical thickness $\Delta\tau_\nu$, the element of reduced volume is

$$dV = \mu \, dA \left(\frac{\Delta u^*}{\mu} \right)^n \quad . \tag{37}$$

In this expression μdA represents the projection of the element of area in a direction μ, and $(\Delta u^*/\mu)$ is the effective element of the radiating gas. For convenience and consistency with the equation for τ (Equation 2), l_ν in Equation 8 may be included with Δu^*. Upon inserting Equation 37 into Equation 8, the factor $(m\Delta u^*/\mu)^n e^{-\Delta\tau/\mu^n}$ becomes just the increment of absorptance of the layer. Upon integration from 0 to $\Delta\tau$ the radiative equilibrium condition becomes

$$\int_\nu \int_\mu I_\nu(\mu) \left(1 - e^{-\frac{\Delta\tau}{\mu^n}} \right) \mu \, d\mu \, d\nu = \int_\nu B_\nu \int_\mu \left(1 - e^{-\frac{\Delta\tau}{\mu^n}} \right) \mu \, d\mu \, d\nu \quad . \tag{38}$$

The computed intensities were inserted into Equation 38 and integration over μ was carried out analytically. In terms of the adopted coordinate system, the radiative equilibrium condition is

$$
\begin{aligned}
&4\sum_{i=1}^{17} \epsilon_{ic} B_{ic} \left[E_5(\tau_{ic} - \tau_{i,j+1}) - E_5(\tau_{ic} - \tau_{ij})\right] \\
&+4\sum_{i=1}^{17}\sum_{\substack{j'=1 \\ j'\neq j}}^{33} B_{ij'} \left[E_5\left(|\tau_{ij} - \tau_{i,j'+1}|\right) - E_5\left(|\tau_{i,j+1} - \tau_{i,j'+1}|\right) + E_5\left(|\tau_{i,j+1} - \tau_{ij'}|\right) - E_5\left(|\tau_{ij} - \tau_{ij'}|\right)\right] \\
&+16\sum_{i=1}^{17} r_{ic} \left[E_5(\tau_{ic} - \tau_{i,j+1}) - E_5(\tau_{ic} - \tau_{ij})\right] \sum_{j'=1}^{33} B_{ij'} \left[E_5(\tau_{ic} - \tau_{i,j'+1}) - E_5(\tau_{ic} - \tau_{ij'})\right] \\
&+\frac{S}{\pi}\zeta\left\{e^{-\frac{\tau_{18,j}}{\zeta^{1/4}}}\left(1 - e^{-\frac{\Delta\tau_{18,j}}{\zeta^{1/4}}}\right) + 8e^{-\frac{\tau_{18,c}}{\zeta^{1/4}}} r_{18,c}\left[E_9(\tau_{18,c} - \tau_{18,j+1}) - E_9(\tau_{18,c} - \tau_{18,j})\right]\right\} \\
&= 2\sum_{i=1}^{17} B_{ij}\left[1 - 4E_5(\Delta\tau_{ij})\right] \quad .
\end{aligned}
\tag{39}
$$

The terms on the left side are arranged in the adopted sequence; the contribution from the surface, then the contribution from the atmosphere (direct and reflected), and finally solar radiation, (direct and reflected).

Radiation from the atmospheric layers above and below layer j is combined into one expression by the use of absolute value signs in the arguments of the exponential integrals. The right side represents the emission from layer j and the left side represents radiation absorbed by layer j.

The numerical solution of Equation 39 was carried out on an IBM 7094 computer. For each layer j and temperature T (between 100 and 350°K) the i wave number intervals of the right side of Equation 39 were summed by using the appropriate value of $\Delta\tau_{ij}$. A functional relationship between the total thermal emission E_j from a layer at temperature T_j was tabulated. The thermal emission from a layer depends on the Planck function and the optical thickness; both are functions of the temperature. The optical thickness depends on temperature through u^* (Equation 16). With this set of E vs. T tables for each layer, the calculation of the temperature profile continued as follows.

First the energy absorbed by a layer j (left side of Equation 39) was computed from an estimate of the temperature profile. After the absorbed energy was obtained, the computer searched the previously computed table for a temperature for layer j which balanced Equation 39. In other words, it determined the temperature at which layer j could reradiate the absorbed energy. This was repeated for each layer and the resulting temperature distribution was checked for constancy of the net flux. If the flux criterion was not satisfied, the computed temperature distribution was re-inserted and the iteration procedure continued.

The convergence criterion used in terminating the calculation was

$$
\left|\frac{F_{(\text{layer } 10)} - F_{(\text{layer } j)}}{F_{(\text{layer } 10)}}\right| \leq \alpha \quad . \tag{40}
$$

where α was chosen according to the desired accuracy. In general α was taken equal to 0.005. Obviously, higher accuracy would require more iterations. Once the convergence criterion was satisfied, the limb function was calculated in the final step,

$$\frac{I_i^{\uparrow}(j = 1, \mu)}{I_i^{\uparrow}(j = 1, \mu = 1)} \quad (41)$$

The limb function was computed for each spectral interval i and for a range of specified μ, by using the computed equilibrium temperature distribution. This expression (41) includes all upward components of the specific intensity.

The iteration process converges rapidly for low surface pressures, where, for a praticular layer, radiation from the surface dominates radiation from other atmospheric layers. For optically dense atmospheres the convergence process takes much longer. These effects are demonstrated in Figures 4 and 5 where samples of the iteration process for low and high surface pressures and for low and high initial temperature estimates are shown. More general studies of the convergence of iterative processes are available (Reference 21, for example).

RESULTS

As stated earlier, a portion of the atmosphere of Venus was selected for analysis with the formulated program. Several sets of good observational material presently exist which can serve as a basis for comparison with the computed results.

The atmosphere of Venus above the clouds is more amenable to analysis and consequently is better understood than the region between the surface and the upper cloud layer (References 22–27). However, starting from essentially the same set of observational data, various students of Venus have reached different conclusions. Estimates of the cloud top pressure range from 0.007 to 1 atm (References 23 and 28–30). Large discrepancies exist in estimates of the water vapor content (References 31–33), and the abundance of

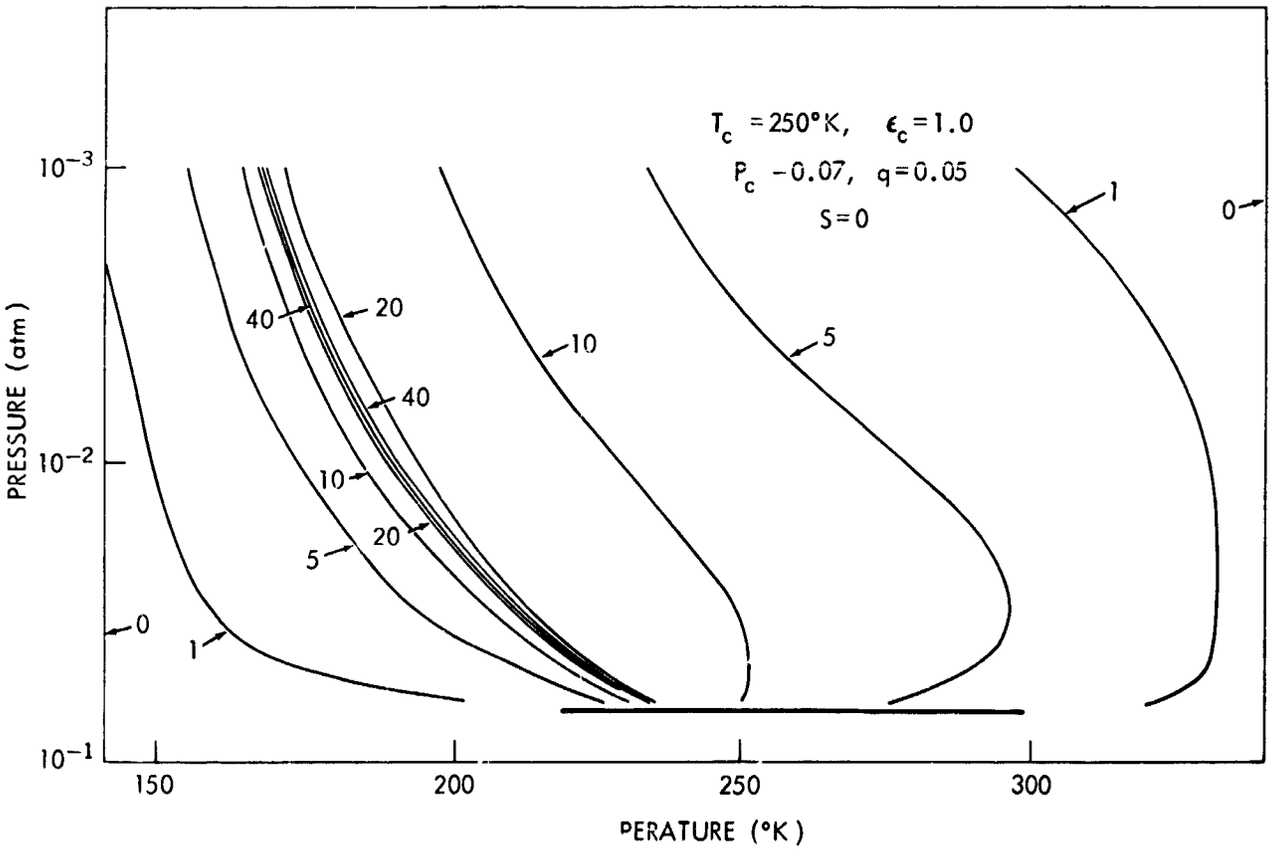


FIGURE 4.—Sample of the convergence of the iteration pr r an optically thin atmosphere. The numerals adjacent to each curve in the figure indicate the iteration numbe corresponding to each temperature distribution sl n.

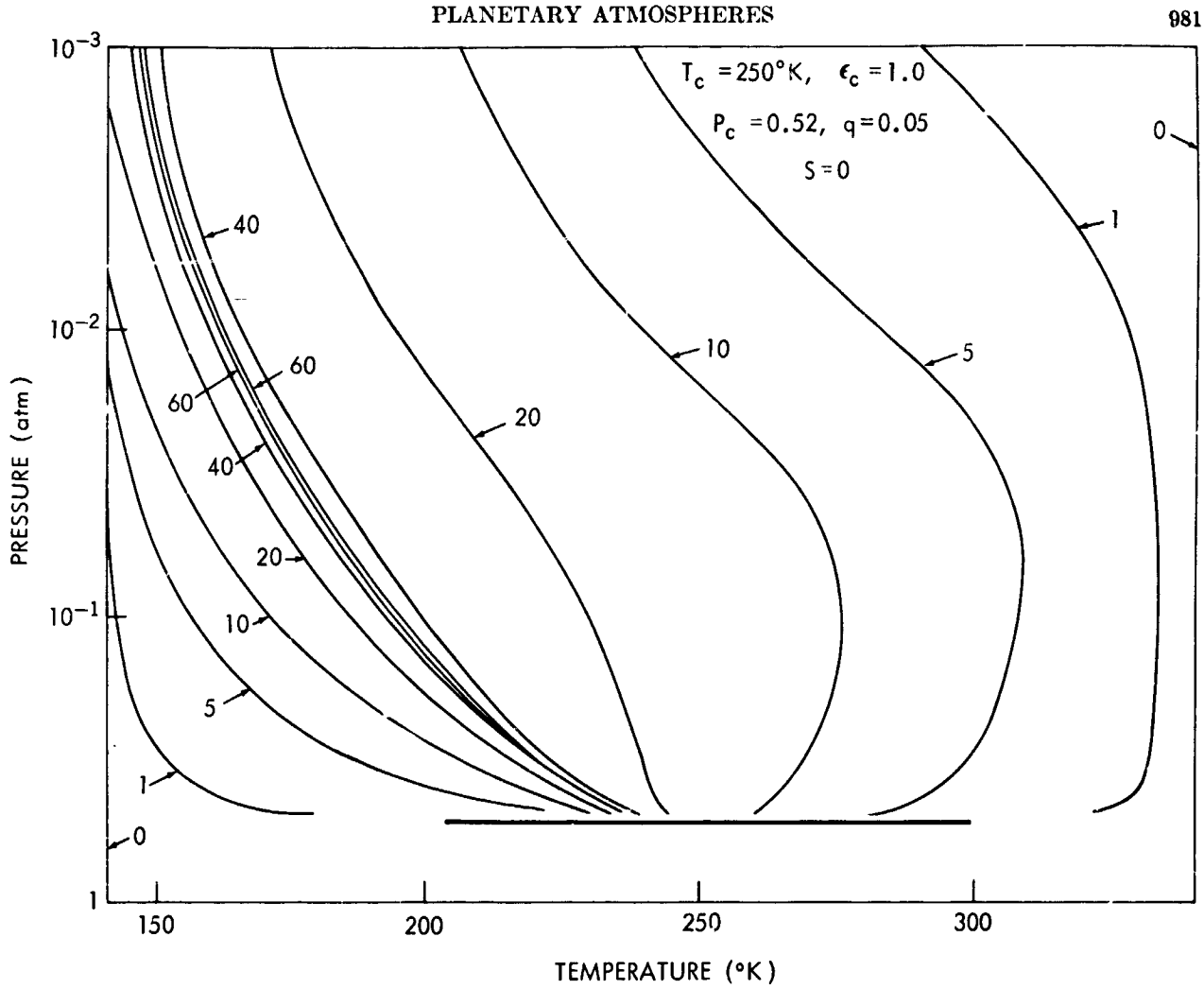


FIGURE 5.—Sample of the convergence of the iteration process for a more optically dense atmosphere than that considered in Figure 4. The numerals adjacent to each curve in the figure indicate the iteration number corresponding to each temperature distribution shown.

CO_2 (References 34–36) is still subject to discussion. The reasons for the discrepancies are manyfold, but the solution to the problem will depend on better observations and more refined interpretation of existing data.

The interpretation of the intensities measured within the atmospheric window by Sinton and Strong (Reference 16), Murray, Wildey and Westphal (Reference 17), and by Chase, Kaplan, and Neugebauer (Reference 18) clearly requires some knowledge of the temperature profile in the upper atmosphere of Venus. So far, only the black-transparent model used by Mintz (Reference 37) and the grey model used by Rasool (Reference 38) and Ostriker (Reference 39) have been applied. None of these calculations consider the strong temperature dependence of the 9.4 and 10.4μ CO_2 excited bands or the solar heating in the clear atmosphere.

It is instructive to note that early application of the grey atmospheric model to the earth's atmosphere by Humphreys (Reference 40), Gold (Reference 41), and Emden (Reference 42) yielded only crude agreement between calculations and reality. Subsequent treatments (References 5 and 43–47, for example) which included the wavelength dependence of the H_2O, CO_2, and O^3 absorption were necessary for satisfactory temperature profiles. Similarly, a more exact and detailed study of the atmosphere of Venus should give better results.

Highly developed nongrey atmospheric models (Reference 48, for example) used to describe the earth's stratosphere exist but have not been

applied to Venus. However, recent work by Arking (Reference 49) and Ohring (Reference 46) with nongrey models have been applied to the atmosphere of Mars. For Venus, the discrepancies in estimates of composition, cloud top pressure, and other parameters necessary for accurate calculations have probably discouraged the application of these models. Furthermore, CO_2 absorption coefficients for relatively large concentrations have become available only recently (References 13, 14, and 50). However, it is felt that a parametric study with a variety of values for cloud top pressure and temperature and for CO_2 concentration should improve our understanding of the physical conditions in the atmosphere of Venus. Results of the calculations can be compared with existing and forthcoming measurements, and the range of values for importar parameters then, hopefully, can be narrowed.

The effect of solar heating, carbon dioxide concentration, and assumed cloud top pressure, temperature, and emissivity, on the atmospheric temperature above the clouds can now be discussed. A number of conclusions can be drawn from this parametric study. However, the attempt to match, in detail, the calculated intensities, fluxes, and limb functions with the observations is still underway.

The pronounced effect of solar heating in the clear atmosphere is apparent in Figure 6. A realistic interpretation of planetary observations of a partially illuminated disk requires further averaging over zenith angles. The temperature maximum disappears for larger zenith angles and the temperature distribution approaches the

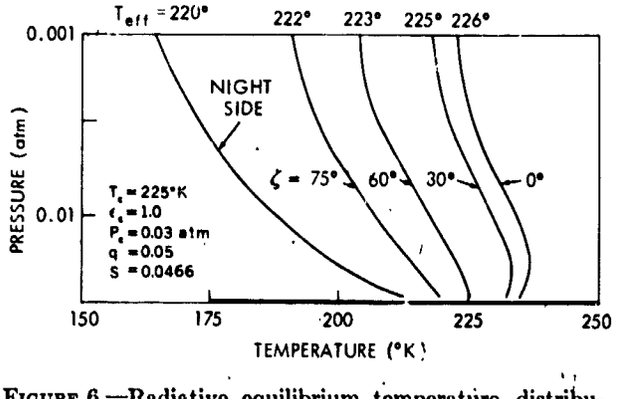


FIGURE 6.—Radiative equilibrium temperature distribution in the Venus atmosphere, illustrating the effects of solar heating for various solar zenith angles.

equilibrium condition on the dark side. Three factors make the large amount of solar heating understandable. Relative to the earth, the solar constant is nearly double; secondly, the high reflectivity of the clouds gives rise to strong reflected radiation and consequently to an additional heat source; and finally, the CO_2 concentration is much higher. The high CO_2 concentration increases the ability of the layers to cool very effectively, but at long wavelengths many spectral regions are already saturated. This atmospheric greenhouse effect is very pronounced. The near infrared absorption region considered in this discussion, spectral interval number 18 (1.25–5.9μ), would saturate at much higher pressures (between 3 and 10 atmospheres). Observations at the 8–12μ window and at 3.75μ (Reference 51) indicate essentially the same brightness temperatures in the sunlit and dark hemispheres. This is not in contradiction to the calculated large temperature differences in the clear atmosphere since the observations exist only for relatively transparent spectral regions. Larger differences are expected in moderately strong absorption regions, but unfortunately data are not available there.

The equilibrium temperatures with and without solar heating for various cloud top pressures are illustrated in Figure 7. For low cloud top pressures and consequently low optical thicknesses, the surface can be observed more readily and hence the effective temperature approaches the cloud top temperature. The discontinuity in the temperature profile near the lower boundary is characteristic of radiative equilibrium considerations in a bounded, finite atmosphere and has been discussed by several authors (References 5 and 52, for example). A more realistic representation of the cloud top which does not assume a "solid surface" will remove the discontinuity and limit the gradient to an adiabatic one. If necessary, allowance for convective as well as radiative transport of energy must be made. It can also be seen from Figure 7 that some of the calculated gradients approach and even exceed the adiabatic one near the cloud top surface. Convection would be induced and an upward transport of cloud particles would occur.

A curve which indicates saturation temperatures for CO_2 is included in Figure 7. For the conditions considered, condensation and possible

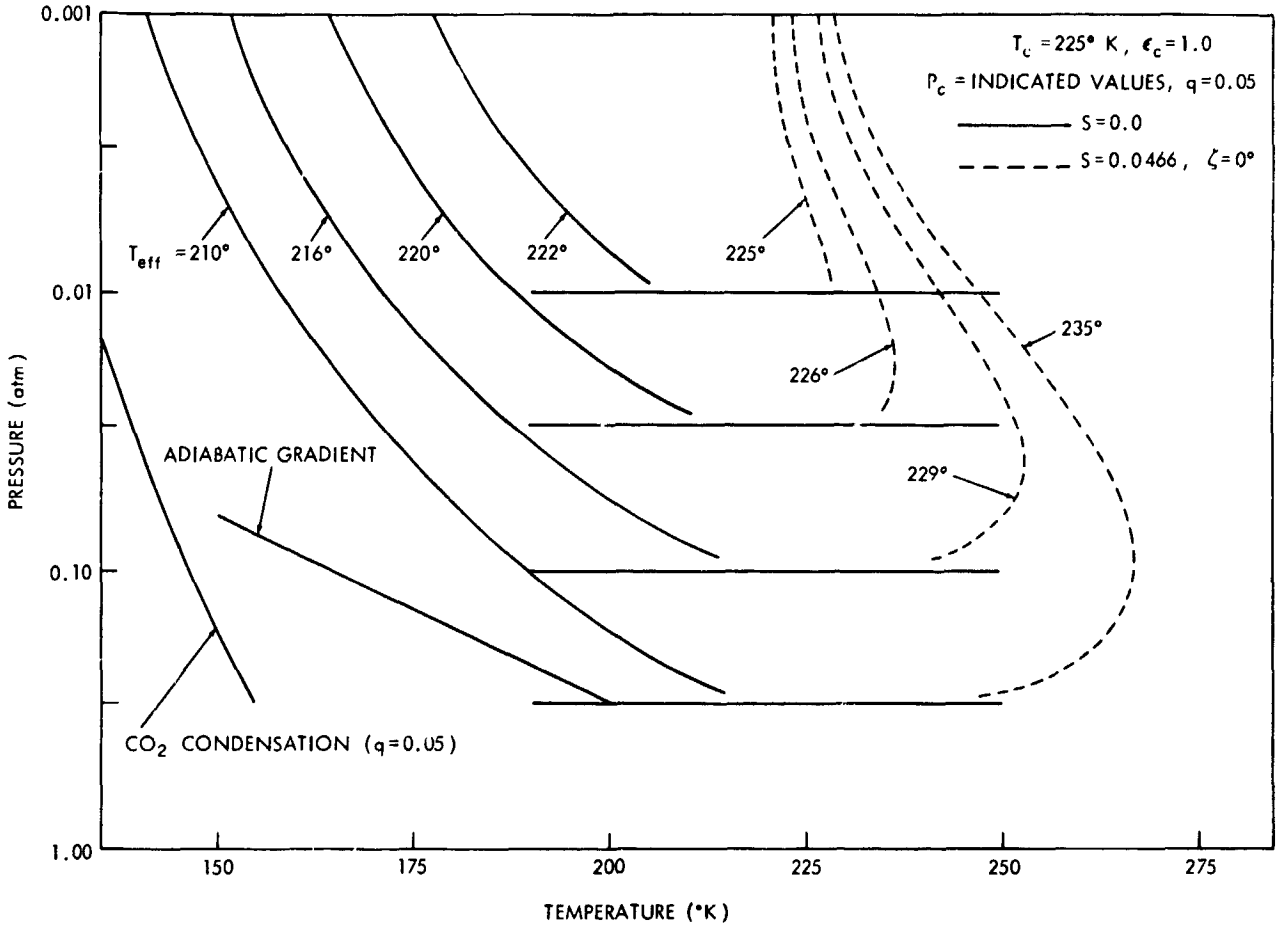


FIGURE 7.—Radiative equilibrium temperature distributions as a function of cloud top pressure level, for the dark and sunlit hemispheres.

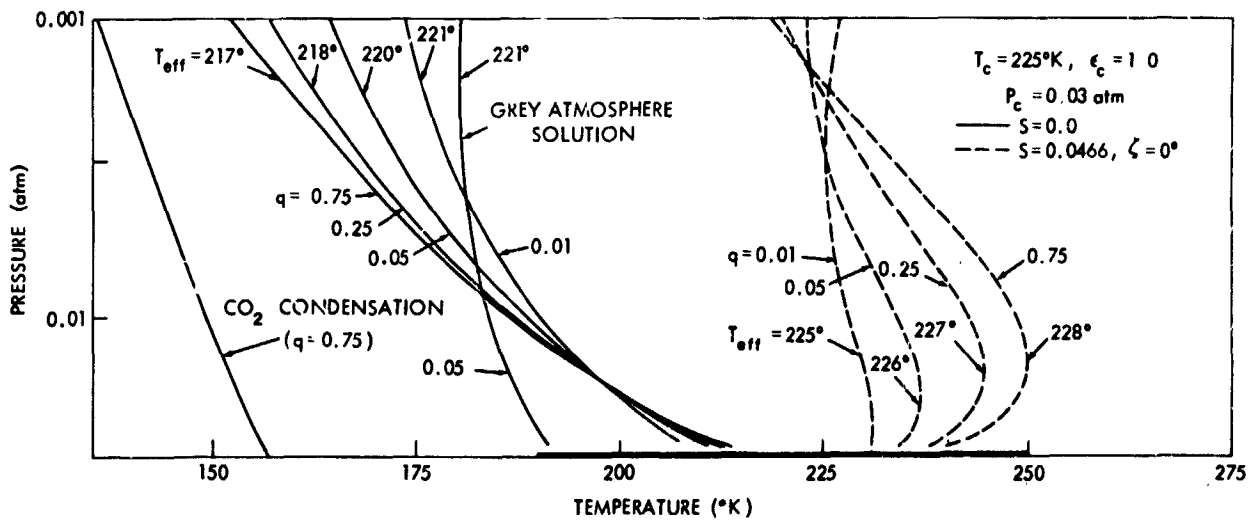


FIGURE 8.—Radiative equilibrium temperature distributions showing the effects of various CO_2 concentrations, for the dark and sunlit hemispheres.

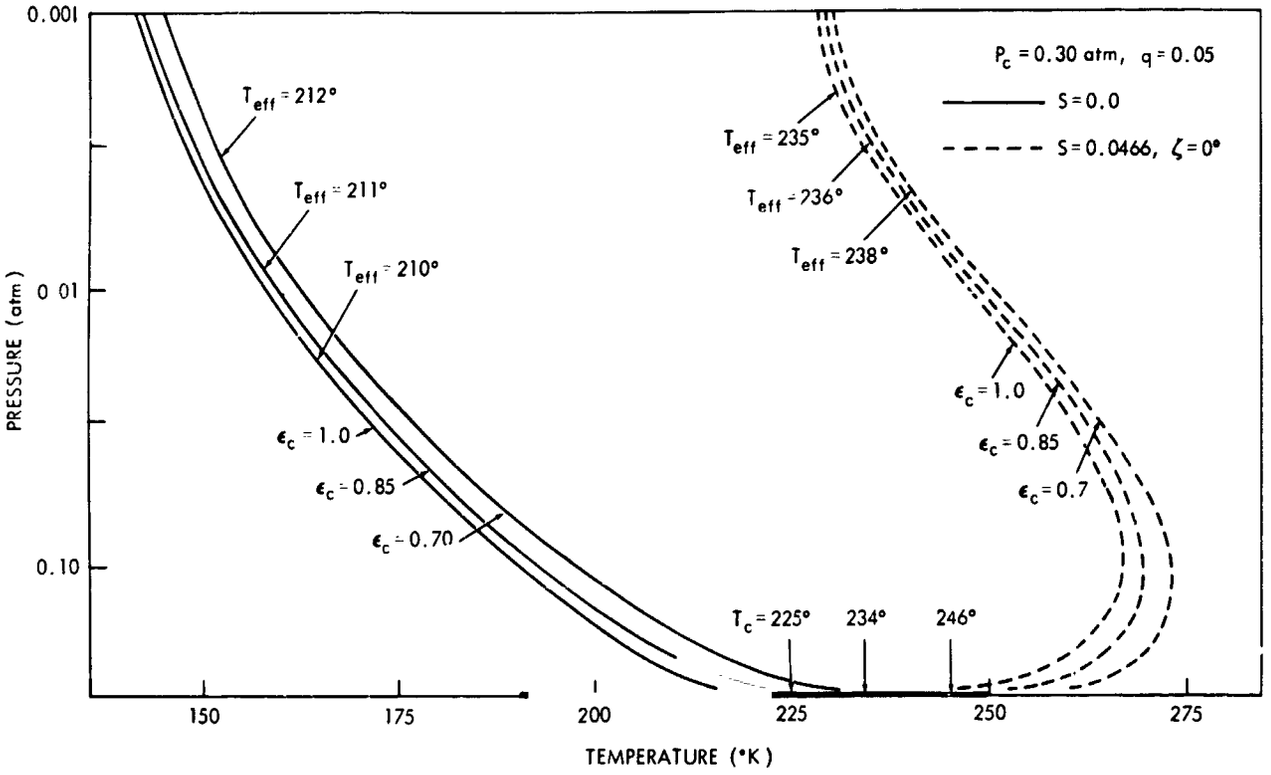


FIGURE 9.—Radiative equilibrium temperature distributions illustrating the effect of cloud top emissivity and temperature, for an almost constant net flux, for the dark and sunlit hemispheres.

formation of CO_2 clouds is unlikely. The wide spread and the slope of temperature profiles shown in Figure 7 make it difficult to extrapolate the temperatures, in detail, to much lower pressures. Caution must be exercised, especially in extrapolations to the low pressures for the occultation measurements $\approx 2.6 \times 10^{-6}$ atm (Reference 53).

The effects of CO_2 concentration on the temperature distribution (Figure 8) are different for the dark and sunlit hemispheres of the planet. On the dark side, the temperatures depend very little on the CO_2 concentration, with the exception of very high layers which can cool more effectively for the higher concentrations. The effects of solar heating have already been discussed.

The temperature distribution of a grey atmospheric model is included for comparison and clearly illustrates the difference between the isothermal character of the grey model and the larger gradients obtained by applying nongrey absorption.

The final illustration (Figure 9) demonstrates the effect of nonblackness for the cloud top surface. The total net flux is kept approximately the same for all cases by adjusting the surface temperature upwards as the emissivity is lowered. As discussed before, the value of ϵ was set equal to unity for the intervals i = 7, 8 and 16, where CO_2 absorption is strongest. In all other intervals ϵ was set equal to the value designated in the figure. The effect on the resulting temperature distribution is small.

In summary then, radiative transfer calculations, which include pressure, temperature, and wavelength dependence of molecular absorption, are a strong analytical tool for the exploration of planetary atmospheres. The full capability of this tool has by far not been exhausted. Further analysis of the spectral and angular dependence of the emitted energy is required for more specific conclusions.

ACKNOWLEDGMENTS

The authors would like to express their gratitude to the following people for their contributions to the effort presented here: Mr. A. F. Simmons, for an expertly written computer program, Mr. T. Mahuron and Miss L. Coury for numerous calculations, and Mr. W. R. Bandeen and Dr. F. Möller for helpful criticism and discussion.

REFERENCES

1. ELSASSER, W. M., and CULBERTSON, M.F., "Atmospheric Radiation Tables," *Meteorol. Monographs* 4(23):1–43, August 1960.
2. KAPLAN, L. D., "On the Calculation of Atmospheric Transmission Functions for the Infrared," *J. Meteorol.* 9(2):139–144, April 1952.
3. PLASS, G. N., "Parallel-Beam and Diffuse Radiation in the Atmosphere," *J Meteorol.* 9(6):429–436, December 1952.
4. KOURGANOFF, V., "Basic Methods in Transfer Problems; Radiative Equilibrium and Neutron Diffusion," Oxford: Clarendon Press, 1952.
5. GOODY, R. M., "The Thermal Equilibrium at the Tropopause and the Temperature of the Lower Stratosphere," *Roy. Soc. Proc.* Ser. A, 197(1051): 487–505, July 7, 1949.
6. CHAMBERLAIN, J. W., and KUIPER, G. P., "Rotational Temperature and Phase Variation of the Carbon Dioxide Bands of Venus," *Astrophys. J.* 124(2):399–450, September 1956.
7. PLASS, G. N., "Useful Representations for Measurements of Spectral Band Absorption," *J. Opt. Soc. Amer.* 50(9):868–875, September 1960.
8. MILNE, E. A., "Thermodynamics of the Stars," *Handbuch der Astrophysik* 3(1):65–255, 1930.
9. CHANDRASEKHAR, S., "Radiative Transfer," Oxford: Clarendon Press, 1950.
10. GOODY, R. M., "Physics of the Stratosphere," Cambridge Univ. Press, 1958.
11. PLASS, G. N., and FIVEL, D. I., "Influence of Doppler Effect and Damping on Line-Absorption Coefficient and Atmospheric Radiation Transfer," *Astrophys. J.* 117(1):225–234, January 1953.
12. STRONG, J., and PLASS, G. N., "The Effect of Pressure Broadening of Spectral Lines on Atmospheric Temperature," *Astrophys. J.* 112:365–379, November 1950.
13. BURCH, D. E., GRYVNAK, D., and WILLIAMS, D., "Infrared Absorption by Carbon Dioxide," Ohio State University Research Foundation, Columbus, report SR-2, January 1961.
14. STULL, V. R., WYATT, P. J., and PLASS, G. N., "The Infrared Absorption of Carbon Dioxide," Air Force Systems Command, Los Angeles, California, report SSD-TDR-62-127, V. III, January 1963.
15. YAMAMOTO, G., "Direct Absorption of Solar Radiation by Atmospheric Water Vapor, Carbon Dioxide and Molecular Oxygen," *J. Atmos. Sci.* 19(2):182–188, March 1962.
16. SINTON, W. M., and STRONG, J., "Radiometric Observations of Venus," *Astrophys. J.* 131(2):470–490, March 1960.
17. MURRAY, B. C., WILDEY, R. L., and WESTPHAL, J. A., "Infrared Photometric Mapping of Venus through the 8–14 Micron Atmosphere Window," *J. Geophys. Res.* 68(16):4813–4818, August 1963.
18. CHASE, S. C., KAPLAN, L. D., and NEUGEBAUER, G., "The Mariner 2 Infrared Radiometer Experiment," *J. Geophys. Res.* 68(22):6157–6170, November 1963.
19. MOROZ, V. I., "The Infrared Spectrum of Venus (1–2.5μ)," *Sov. Astronom. AJ* 7(1):109–115, July–August 1963.
20. PAGUROVA, V. I., "Tables of the Exponential Integral," London: Pergamon Press, 1961.
21. STONE, P. H., "Approximate Integral Equations for the Temperature in Non-Gray Model Atmospheres," *Astrophys. J.* 137(2):628–640, February 1963.
22. UREY, H. C., "The Atmospheres of the Planets," *in: Handbuck der Physik* (S. Elügge, ed.), V. 52:363–418, Berlin: Springer-Verlag, 1959.
23. OPIK, E. J., "The Aelosphere and Atmosphere of Venus," *J. Geophys. Res.* 66(9):2807–2819, September 1961.
24. KUIPER, G. P., "Planets and Satellites, V. 3, The Solar System," (Kuiper, G. P., and Middlehurst, B. M., ed.) Chicago: Univ. of Chicago Press, 1961.
25. REA, D. G., "Molecular Spectroscopy of Planetary Atmospheres," *Space Sci. Rev.* 1(2):159–196, October 1962.
26. SAGAN, C., and KELLOGG, W. W., "The Terrestrial Planets," *Ann. Rev. Astron. Astrophys.* 1:235–266, 1963.
27. ROZENBERG, G. V., "Investigations of the Atmosphere of the Planet Venus by Optical Methods," *Sov. Phys. Doklady*, 8(1):1–2, July 1963.
28. KAPLAN, L. D., "A Preliminary Model of the Venus Atmosphere," Jet Propulsion Laboratory, California Institute of Technology, Pasadena, report JPL TR-32-379, December 12, 1962.
29. SAGAN, C., "The Radiation Balance of Venus," California Institute of Technology, Pasadena, report JPL TR-32-34, September 15, 1960.
30. ANDERSON, C. E., and EVANS, D. C., "The Nature of the Venus Cloud System," Institute of Aerospace Science paper, National Summer Meeting, Los Angeles, California, June 18–22, 1962.
31. STRONG, J., Ross, M. D., and MOORE, C. B., "Some Observations on the Atmosphere of Venus and the Earth during the Strato Lab IV Balloon Flight," *J. Geophys. Res.* 65(8):2526, August 1960.
32. SPINRAD, H., "A Search for Water Vapor and Trace Constituents in the Venus Atmosphere," *Icarus* 1(3):266–270, October 1962.
33. DOLLFUS, A., "Observations of Water Vapor on the Planet Venus," *C. R. Acad. Sci.* 256(15):3250–3253, April 8, 1963 (In French).
34. DUNHAM, T., JR., "Spectroscopic Observations of the Planets at Mount Wilson," in: "The Atmospheres of the Earth and Planets," (G. P. Kuiper, ed.) 288–305, Chicago: Univ. of Chicago Press, 1952, pp 288–305.
35. KAPLAN, L. D., "A New Interpretation of the Structure and CO_2 Content of the Venus Atmosphere," *Planet. Space Sci.* 8(1):23–29, October 1961.
36. SPINRAD, H., "Spectroscopic Temperature and Pressure Measurements in the Venus Atmosphere," *Astron. Soc. Pacific Publ.* 74:187–201, June 1962.

37. MINTZ, Y., "Temperature and Circulation of the Venus Atmosphere," *Planet. Space Sci.* 5(2):141–152, June 1961.

38. RASOOL, S. I., "Structure of Planetary Atmospheres," *AIAA J.* 1(1):6–19, January 1963.

39. OSTRIKER, J. P., "Radiative Transfer in a Finite Atmosphere," *Astrophys. J.* 138(1):281–290, July 1963.

40. HUMPHREYS, W. J., "Vertical Temperature-Gradients of the Atmosphere, Especially in the Region of the Upper Inversion," *Astrophys. J.* 29:14–32, January 1909.

41. GOLD, E., "The Isothermal Layer of the Atmosphere and Atmospheric Radiation," *Roy. Soc. Proc.* Ser. A, 82(551):43–70, February 1909.

42. EMDEN, R., "Über Strahlungsgleichgewicht und Atmosphärische Strahlung. Ein Beitrag zur Theorie der Oberen Inversion," *Sitzungsberichte der mathematisch–physikalischen Klasse*, Akademie der Wissenschaften, Munchen, Part 1: 55–142, January–March 1913.

43. SIMPSON, G. C., "Studies in Terrestrial Radiation," *Roy. Meteorol. Soc. Mem.* 2(16):69–95, March 1928.

44. KING, J. I. F., "Line Absorption and Radiative Equilibrium," *J. Meteorol.* 9(5):311–321, October 1952.

45. YAMAMOTO, G., "Radiative Equilibrium of the Earth's Atmosphere, I. The Grey Case," *Science Reports of the Tohoku University, 5th Series, Geophysics* 5(2):45–57, October 1953.

46. OHRING, G., "A Theoretical Estimate of the Average Vertical Distribution of Temperature in the Martian Atmosphere," *Icarus*, 1(4):328–333, January 1963.

47. KONDRATI'EV, K., "Teplovoi Rezhim Verkhnikh Sloev Atmosfery," ("Thermal Conditions of Upper Strata of the Atmosphere"). Leningrad: Gidrometeoizdat, 1960.

48. MANABE, S., and MÖLLER, F., "On the Radiative Equilibrium and Heat Balance of the Atmosphere," *Monthly Weather Rev.* 89(12):503–532, December 1961.

49. ARKING, A., "Non-Grey Convective Planetary Atmospheres," Eleventh International Astrophysical Symposium, Liege, July 9–12, 1962, 7:180–189, 1963.

50. HOWARD, J. N., BURCH, D. E. and WILLIAMS, D. E., "Near-infrared Transmission Through Synthetic Atmospheres," Ohio State University Research Foundation, Columbus, report SR-1, December 1954.

51. SIMPSON, W. M., *Proc. Eleventh Internat. Astrophys. Sym.*, Liege, July 9–12, 1962.

52. MÖLLER, F., "Über das Strahlungsgleichgewicht in der Nähe der Erdoberfläche," *Meteorol Rundsch.* 13(5): 134–139, September–October 1960.

53. MENZEL, D. H., and DE VAUCOULEURS, G., "Final Report on the Occultation of Regulus by Venus, July 7, 1959," Harvard College Observatory, Cambridge, report SR-1, January 10, 1961.

MIDDLE ULTRAVIOLET DAY RADIANCE OF THE ATMOSPHERE*

J. P. HENNES, W. B. FOWLER AND L. DUNKELMAN

The day radiance of the atmosphere has been measured, photoelectrically, at two middle ultraviolet wavelengths by rocket-borne photometers. Filters and collimators provided an effective field of view of 1.4×10^{-2} steradians, and bandpasses of approximately 100A at 2600A and 230A and 2200A. At a height of 146 km, nadir radiance values of about 0.5 ergs/sec cm² ster 100A were obtained at both wavelengths. There is good agreement between the radiance values measured and those which have been calculated on the basis of single Rayleigh scattering in the presence of ozone. From the radiance data an atmospheric diffuse reflectivity of about 8×10^{-4} at 2600A has been calculated. This may be contrasted with 1957 rocket observations of Mars and Jupiter which yielded albedos at 2700A of 0.94 and 0.26 respectively.

I. INTRODUCTION

If the earth's sunlit side is viewed from space it exhibits a bright appearance in the visible portion of the spectrum. In contrast, the appearance of the sunlit earth in the ultraviolet wavelength region below about 3100A may be expected to be rather dim and relatively uniform. In the visible region the upward flux out of the atmosphere is made up of Rayleigh scattered sunlight, sunlight reflected directly from clouds or from the earth's surface, light scattered from atmospheric dust, and a relatively small amount of resonant and fluorescent scattered sunlight or day airglow. In the middle ultraviolet region from 2100–3000A, however, because of the totally absorbing ozone region located from approximately 10 to 40 km, the day radiance will be produced only by Rayleigh scattering of sunlight in the thin upper atmosphere above the ozone region and by fluorescent scattered sunlight or ultraviolet dayglow. The combination of a thin atmosphere and strong ozone absorption produces a correspondingly low albedo. The absence of contributions from reflections by clouds and surface features and the strong wavelength dependence of Rayleigh scattering will produce a relatively uniform appearance.

The spherical albedo of a planet is the ratio of the total flux emitted in all directions by the planet to the total flux incident on the planet from the sun. Both quantities are being measured over the same wavelength interval. The earth's visual, spherical albedo has been experimentally determined, by studies of earth light on the moon, to range from 0.29–0.56 depending on the season and on which side of the earth is facing the moon. The average value of these visual albedos is given as 0.39–0.40 (Danjon, 1954; Dzhasybekova, Kazachevskii, and Kharitonov, 1960). The major part of the earth's visual albedo comes from contributions by reflection from clouds. If atmospheric scattering is considered separately, assuming no clouds, calculations (Kano, 1958) show the spectral albedos due to Rayleigh scattering to be: infrared region—.0075; visible region—0.096; and near ultraviolet region—0.23. The weighted sum of these various spectral albedos for a Rayleigh atmosphere is given as 0.066 by Kano (1958) for the entire spectral region $\lambda>2900$A, and has been independently calculated by Coulson (1959) to be p. 069. The contribution in these spectral regions from airglow is negligible except over narrow wavelength intervals near the more pronounced dayglow emission lines. The uses and difficulties of dayglow measurements have been summarized recently by Chamberlain (1963).

*Published as *Goddard Space Flight Center Document* X-613-64-70, April 1964; also published in the *Journal of Geophysical Research*, 69(13):2835–2840, July 1, 1964.

The calculations which have been made for scattered solar radiation in the near ultraviolet have included multiple scatterings, ozone distributions, and upward and downward fluxes (Sekera and Dave, 1961a, and Larsen, 1959).

In the middle ultraviolet region calculations have assumed single Rayleigh scattering with ozone absorption (Ban, 1962; Hubbard, 1963; Green, 1964; and Hrasky and McKee, 1964). Multiple scattering or fluorescence has not been included. The middle ultraviolet day radiance must originate in scattering above the ozone region. This insures a relatively small flux since even the high altitude ozone distribution is strongly absorbing in this spectral region. Dalgarno (1962, 1963) and Chamberlain and Sobouti (1962) have derived equations to show the effect of non-Rayleigh resonant and fluorescent scattering near atomic and molecular resonance lines, and point out that this effect could be significant in the ultraviolet where small Rayleigh scattered fluxes are expected.

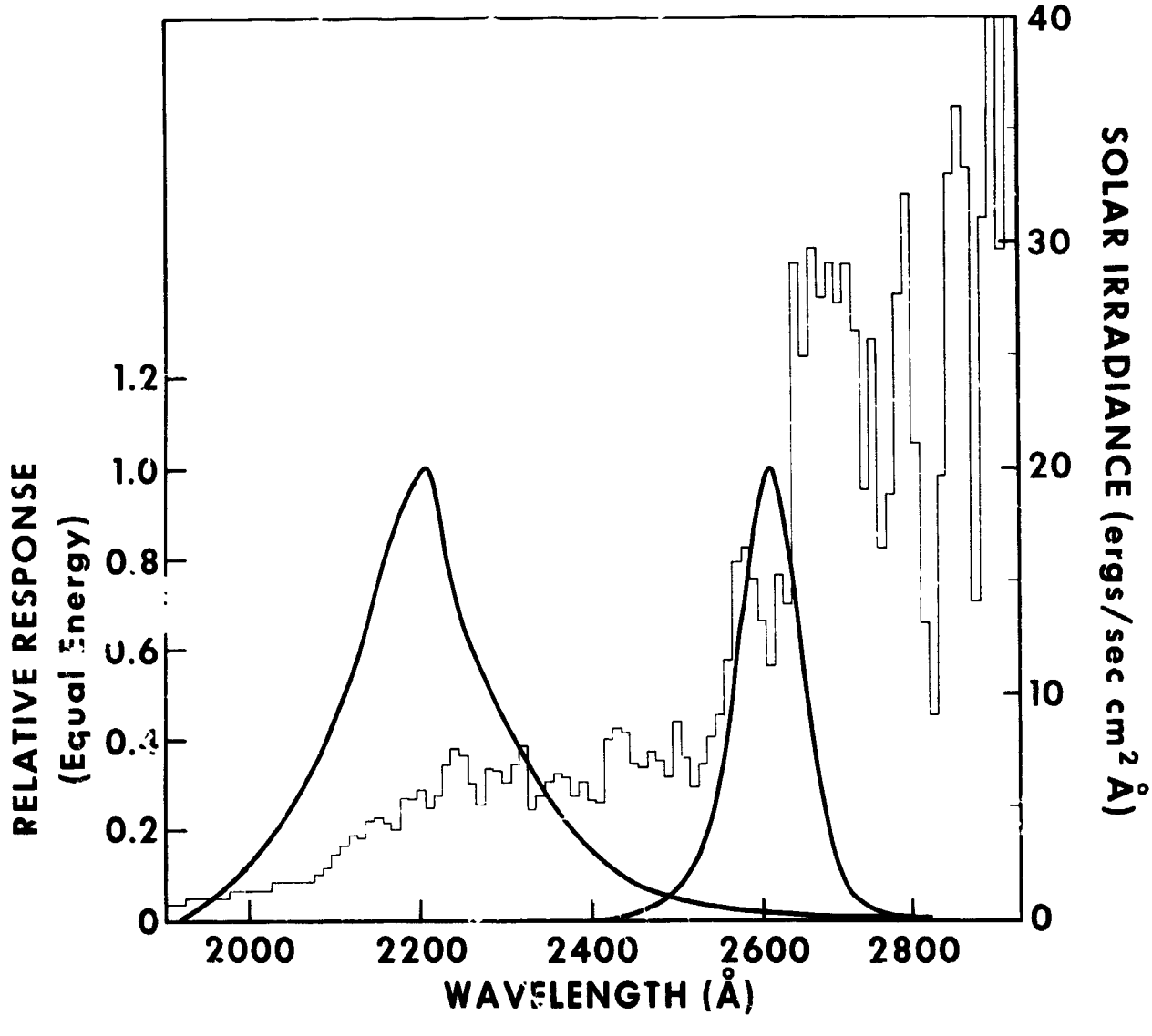


FIGURE 1.—Photometer response curves. Relative response of both photometers is shown on an equal energy basis. The solar spectrum, averaged over 10A intervals, is also shown. (Wilson et al., 1954; Malitson et al., 1960). The 2200A photometer response is provided by a combination of an interference filter and 4 mm $NiSO_4(H_2O)_6$, crystal. The 2600A photometer combines an interference filter, Corning 7-54, 4 mm $NiSO_4(H_2)_6$, Cation-X, and 2 mm Pb doped KCl:KBr crystal (Childs. 1961).

II. THE EXPERIMENT

At 1155 EST (1655 UT) on 8 August 1962 two middle ultraviolet photometers were included in an Aerobee rocket (NASA 4.60) launched from Wallops Island, Virginia. Measurements of the ultraviolet day radiance were taken continuously as the rocket climbed to a peak altitude of 150 km. Each photometer consisted of a mechanical collimator, a filter, and a photomultiplier sensitive only to ultraviolet radiation. The collimator produced an effective field of 1.4×10^{-2} ster (about a 7 degree square field). The filters were of the type described by Childs (1961). The phototubes were type EMR 541F–05 photomultipliers with sapphire windows and cesium-tellurium cathodes (Dunkelman, Fowler, and Hennes, 1962).

The relative spectral response of these photometers is given in Figure 1 on an equal energy basis (amps/watt). Also included in Figure 1 is the solar spectrum, averaged over 10A intervals, taken from data reported by Wilson, et al., (1954) and Malitson, et al., (1960). The effective wavelengths of the photometers were at 2217A and 2610A with effective bandwidths of 230A and 100A respectively. The contribution from the long wavelength region beyond 2800A has been examined and found to be much less than the magnitudes of the radiance measurements made.

During the flight the rocket pitched through a wide range of zenith angles. Its motion, however, was confined by an attitude control system to a single plane having an azimuth of 175° passing through the zenith and the sun. The sun was at a zenith angle of 22°. The photometers, which were pointed out the side of the rocket approximately in the pitch plane and at an angle of 122° with the rocket axis, swept from the nadir up to a zenith angle of about 80° in both the north and south sky at various times.

III. RESULTS

The data are shown in Figure 2 which give (curve a) the zenith angle of the photometers as a function of both flight time and altitude. The aspect data was obtained from solar sensing aspect cells, magnetometers, and the attitude control system error signals and pitch-rate signals. During the flight the rocket's attitude control system malfunctioned to the extent that, although controlled in two axis, the vehicle pitched through much larger angles than those for which the aspect instruments had been designed. Pitch rate signals were integrated to find approximate zenith angles outside of the region of aspect sensor operation. These calculations could be checked by fitting with the aspect signals as the rocket periodically moved back throu the sensor region. The photometer data, reduced from the original telemetry record at one second intervals, are shown in curves (b) and (c). Relative intensity is plotted, with the data normalized to unity for nadir (180°) values. The scatter in the relative intensity values is produced by photometer noise introduced because the signal level was near the bottom of a three-decade logarithmic scale. The photometer sensitivity in this exploratory measurement had been set low to avoid the possibility of saturation due to either unexpectedly large reflectivities or unusually intense dayglow or auroral emissions. Lines have been added to clarify the peaks beyond 260 seconds.

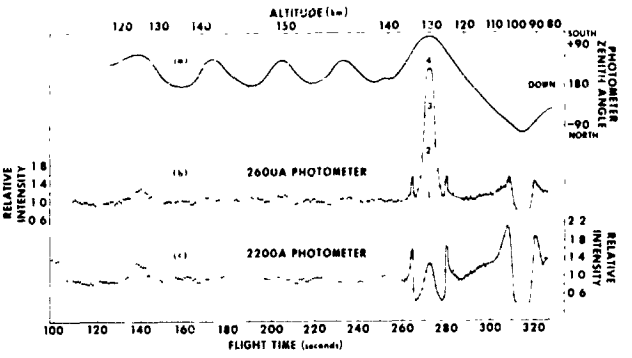


FIGURE 2.—Rocket flight data. Curve (a) is the zenith angle of the photometer direction as a function of time. Curves (b) and (c) are a sampling of the reduced telemetry data plotted in terms of the relative intensity received by the photometer. The points show scatter produced by photometer noise and a small amount of structure as a function of zenith angle. The peaks after 260 seconds correspond to the horizon limb brightening. The large peak at 273 seconds, which is extraneous, is produced by sunlight reflected off the photometer entrance with the photometers pointed above the horizon in the south. Without this sunlight effect the record would appear as it does at 310–320 seconds with the photometers pointed above the horizon in the north.

In Figure 3 the data are replotted as a function of zenith angle. Data are shown only for 10 degree intervals of zenith angle. Part of the spread in the points is due to photometer noise,

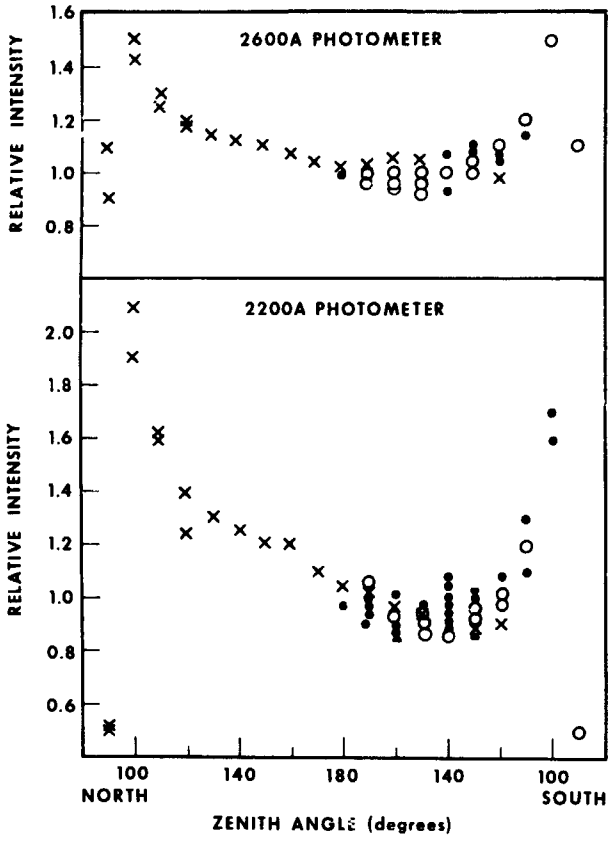


FIGURE 3.—Ultraviolet day radiance as a function of zenith angle. The data in Figure 2 are replotted here as a sampling of telemetry data taken at 10 degree intervals. The crosses indicate data taken at altitudes less than 120 km. Circles indicate data taken at altitudes above 120 km. Solid circles are single points, open circles represent several data points (2–5) occupying the same position. The limb brightening is evident.

part is due to combining data taken at the same zenith angles but from different altitudes.

The values of day radiance measured during this flight are notable for the lack of features seen over a large range of zenith angles. Even the horizon brightening is only a factor of two times the nadir radiance. The photometer field of 1.4 $\times 10^{-2}$ steradians would, of course, diminish the effect of any increased radiance that extended over only a relatively narrow field.

The peaks at about 272 seconds in Figures 2 represent direct sunlight reflected off the photometer entrance. The photometer was pointed up in the southern sky at that point at a zenith angle of about 70 degrees. The magnitude of the reflected solar signal under those conditions was determined by laboratory measurements to be about 10^{-5} that of direct sunlight. Note that this scattered light signal is still larger than the earth radiance signal. The problem of making measurements in the presence of direct sunlight is emphasized by this result.

The magnitude of the radiances detected by the photometers is expressed in Table 1. The nominal effective wavelength and the effective band width are given. The third column gives the earth's radiance, at the nadir, with the sun at 22° zenith angle, assuming a uniform and equal distribution across the wavelength region described by the effective bandwidth. The equivalent photon emission rate in kilo-rayleighs per 100A is also given. The day radiance values determined by these measurements have uncertainties of about 10–15 percent arising from noise in the photometer signal and uncertainties in ultraviolet radiometric standards.

The last column gives the diffuse spectral reflectivity of the earth's sunlit atmosphere averaged over each of the wavelength intervals. This number is arrived at by assuming the scattered sunlight has the same spectral distribution as the incident sunlight throughout each bandwidth (see Figure 1). The values of solar irradiance are taken from the sources cited. The uncertainties in the calculated reflectivities are greater than

TABLE 1.—*Ultraviolet Day Radiance Measurements of the Earth Taken 8 August 1962, at 146 km, with Photometers Pointed at the Nadir; Solar Zenith Angle 22°.*

Nominal Wavelength (A)	Effective Bandwidth (A)	Radiance (ergs/sec cm² ster 100A)	Emission (kilorayleigh/100A) Rate	Diffuse Reflectivity
2600	100	0.5	8×10^{2}	8×10^{-4}
2200	230	0.5	7×10^{2}	3×10^{-3}

those for our observed radiance values by the amount of uncertainty in published absolute spectral solar intensities. If we make the further assumption that the entire earth is a diffuse Lambert reflector, then the diffuse reflectivity of about 8×10^{-4} at 2600A becomes the earth's spherical albedo. This number can be contrasted with the earth's total observable spherical albedo of 0.39–0.40, and with the calculated Rayleigh scattered albedo contribution in the near ultraviolet of 0.23 as given in the introduction.

Of interest also is a comparison of the earth's low middle ultraviolet effective albedo with measurements of the relatively high effective albedos of Mars and Jupiter made in 1957 with a photometer of about 300A bandwidth centered at 2700A (Boggess and Dunkelman, 1959). They obtained values of about 0.24 for Mars and 0.26 for Jupiter.

One other known experimental observation has been made of the middle ultraviolet day radiance. Friedman, Rawcliffe, and Meloy (1963) have reported a satellite borne photometer measurement of the day radiance with a spectral pass band of 140A centered at 2550A. Their result, with the sun at a zenith angle of 49° and the detector pointed at the nadir with a field of 1.2×10^{-4} steradian, showed an atmospheric radiance of 2.0 ± 0.3 ergs/sec cm^2 ster 100A.

Table 2 gives the various measured and calculated middle ultraviolet day radiance values. The calculated values tend to support our results rather well. It should be borne in mind that the data were all taken or calculated using different parameters of altitude, solar zenith angle, ozone distribution, season, etc. so that the results cannot be compared too directly. The relatively large factor of four between the experimental results of Friedman, Rawcliffe and Meloy and ourselves is possibly explained by either a large change in the high altitude ozone distribution between the time of the two measurements or the difficulties in maintaining calibration during the preparations for and launching of a satellite. The similarity of the various calculations and our measurements indicate that the simple approach to ultraviolet atmospheric radiance calculations, involving use of ozone distributions and single scattering with plane atmospheres (Ban, 1962; Hubbard, 1963; and Green, 1964) or spherical atmospheres (Hrasky and McKee, 1964) is adequate for broadband measurements. If higher spectral resolution were to be considered the agreement might be strongly affected by resonant or fluorescent scattering.

The calculated results depend very strongly on ozone distribution, especially in the region around

TABLE 2.—*Comparison of Measured and Calculated Nadir Ultraviolet Day Radiance Values. Radiance in ergs/sec cm² ster 100A.*

Source	2600A Region		2200A Region		Solar Zenith Angle
	Average Radiance Over Interval[a]	Wavelengths (A)	Average Radiance Over Interval[a]	Wavelengths (A)	
This work[b]	0.5	2560–2660	0.5	2100–2330	22°
Friedman, et al. (1963)[b]	2.0	2480–2620			49°
Ban (1962)[c]	0.5	2550–2650			0°
Hubbard (1963)[c]	0.4[d]	2550–2650			0°
Green (1964)[c]	0.5[e]	2550–2650	0.7	2100–2330	0°
Hrasky & McKee[c] (1964)	0.4	2550–2700	0.5	2000–2100	22°

(a) Values estimated or rounded off from data published by the referenced authors.
(b) Measured values.
(c) Calculated values.
(d) A range of 0.05—0.8 is estimated depending on the ozone distribution chosen. The radiance listed is taken from Figure 14 representing a "standard" ozone curve.
(e) A range of 0.15—0.6 is estimated depending on the ozone distribution chosen. The radiance listed is taken from Figure 6 representing an analytical fit to the same "standard" ozone curve as used by Hubbard.

the peak of the ozone absorption curve at 2600A where the very high altitude ozone has a dominant role. The ozone distribution at higher altitudes is relatively unknown, however, so that appreciable uncertainties may be introduced into the calculated values. At the shorter wavelength region around 2200A and at longer wavelengths beyond 2800A the ozone is much less absorbing and incident solar radiation will penetrate more deeply into the atmosphere, thus permitting more scattering and producing higher albedos. By use of satellite-borne photometers with narrow spectral bandpasses, well removed from strong dayglow emission lines, and situated at various middle ultraviolet wavelengths from 2100 and 3000A, a profile of high altitude ozone distribution should be obtainable. Such an idea was proposed in 1957 by Singer and Wentworth (1957), who considered only wavelengths at 2800 and 3000A. Twomey (1961) carried out further calculations in this same wavelength region. The calculations in the near ultraviolet above 3000A have been considered in detail by Sekera and Dave (1961b). The altitudes above 50 km would require, however, use of wavelengths below 2800A where analysis should be much easier. The seasonal and latitudinal variations in upper atmosphere ozone, so important to atmospheric heating processes, would thus be available.

ACKNOWLEDGEMENT

We wish to express our appreciation to William Russell, Jr. of this Center who graciously provided support and the space for this experiment on one of the rockets under his supervision.

REFERENCES

BAN, G. R., "Scattered Radiance in the Middle Ultraviolet Region," in *Symposium on Military Applications of Ultraviolet Radiations*, Report LAS-TR-199-37, University Chicago, Laboratory for Applied Science, 1962.

BOGGESS III, A., and L. DUNKELMAN, "Ultraviolet Reflectivities of Mars and Jupiter," Astrophys. J., **129**, 236, 237, 1959.

CHAMBERLAIN, J. W., "Airglow and the Physics of Upper Atmospheres," Science, **142**, 921–924, 1963.

CHAMBERLAIN, J. W. and Y. SOBOUTI, "Fluorescent Scattering in Planetary Atmospheres. I. Basic Planetary Considerations," Astrophys. J., **135**, 925–937, 1962.

CHILDS, C. B., "Broad Band Ultraviolet Filters," J. Opt. Soc. Amer., **51**, 895–897, 1961.

COULSON, K. L., "Characteristics of the Radiation Emerging from the Top of a Rayleigh Atomsphere. II. Total Upward Flux and Albedo," Plant Space Sci., **1**, 277–284, 1959.

DALGARNO, A., "Spectral Reflectivity of the Earth's Atmosphere, III: The Scattering of Light by Atomic Systems," Geophysics Corp. of America Report 62–28–A, 1962.

DALGARNO, A., "Rayleigh Scattering Near an Absorption Line," J. Opt. Soc. Amer., **53**, 1222, 1963.

DANJON, A., "Albedo, Color, and Polarization of the Earth," in *The Earth as a Planet*, edited by G. P. Kuiper, pp. 726–738, Univ. of Chicago Press, Chicago, 1954.

DUNKELMAN, L., W. B. FOWLER, and J. P. HENNES, "Spectrally Selective Photodetectors for the Middle and Vacuum Ultraviolet," Appl. Optics, **1**, 695–700, 1962.

DZHASYBEKOVA, E. K., V. M. KAZACHEVSKII, and A. V. KHARITONOV, "Determination of the Earth's Albedo," Astron. Zh., **37**, 131–134, 1960; translated in Soviet Astronomy–AJ, **4**, 125–128, 1960.

FRIEDMAN, R. M., D. RAWCLIFFE, and G. E. MELOY, "Radiance of the Upper Atmosphere in the Middle Ultraviolet," J. Geophys. Res., **68**, 6419–6423, 1963.

GREEN, A. E. S., "Attenuation by Ozone and the Earth's Albedo in the Middle Ultraviolet," Appl. Optics, **3**, 203–208, 1964.

HRASKY, W. C., and T. B. McKEE, Langley Research Center, NASA, Hampton, Virginia, private communication, 1964.

HUBBARD, E. L., "Calculation of the Intensity of Light Scattered in the Atmosphere in the Wavelength Region 2300A to 3100A with Absorption by Oxygen and Ozone Considered," Report Sciences, Sept. 1963.

KANO, M., "The Albedo of the Earth's Atmosphere," Papers in Meteorology and Geophysics, Tokyo, **9** (2), 113–122, 1958.

LARSEN, S. H. H., "On the Scattering of the Ultraviolet Solar Radiation in the Atmosphere with the Ozone Absorption Considered," Geofys. Publ., **21** (4), 1959.

MALITSON, H. H., J. D. PURCELL, R. Tousey, and C. E. MOORE, "The Solar Spectrum from 2635 to 2085A," Astrophys. J., 132, 746–766, 1960.

SEKERA, Z. and J. V. DAVE, "Diffuse Transmission of Solar Ultraviolet Radiation in the Presence of Ozone," Astrophys. J., **133**, 210–227, 1961a.

SEKERA, Z. and J. V. DAVE, "Determination of the Vertical Distribution of Ozone from the Measurement of Diffusely Reflected Ultraviolet Solar Radiation," Planet. Space Sci., **5**, 122–136, 1961b.

SINGER, S. F., and R. C. WENTWORTH, "A Method for the Determination of the Vertical Ozone Distribution from a Satellite," J. Geophys. Res., **62**, 299–308, 1957.

TWOMEY, S., "On the Deduction of the Vertical Distribution of Ozone by Ultraviolet Spectral Measurements from a Satellite," J. Geophys. Res., **66**, 2153–2162, 1961.

WILSON, N. L., R. TOUSEY, J. D. PURCELL, and F. S. Johnson, "A Revised Analysis of the Solar Spectrum from 2990–2635A," Astrophy. J., **119**, 590–612, 1954.

LYMAN-α RESONANT SCATTERING USING RANDOM STOKES VECTORS*

S. O. KASTNER

N64-28742

The case of relatively few scatterings of Lyman-α resonance radiation in a uniform hydrogen gas is treated by using Stokes vectors as random variables in a Monte Carlo type of calculation. Doppler redistribution of frequencies and the doublet nature of the line are taken into account.

INTRODUCTION

In radiative transfer problems in which the number of scatterings is low, the nature of the individual scattering process becomes important, i.e. whether isotropic or anisotropic. Lyman-α resonance scattering is an example in which one line component scatters isotropically, the other anisotropically.

We describe here a use of random Stokes vectors in a Monte Carlo type of calculation, to obtain the fluxes of Lyman-α radiation at various levels in a uniform (density and temperature) atmosphere of hydrogen atoms. The method incorporates Doppler redistribution of frequencies, and takes account accordingly of the fact that a photon may be scattered by either of the components of the Lyman-α doublet in the course of its flight. The polarization at each level could also be obtained from the same calculation, if required, by adding up the Stokes vectors. A non-uniform atmosphere was initially attempted, but pointed up a problem in that the mean free path of a photon becomes anisotropic and the anisotropy varies with level.

The interest in this approach stemmed from the possibility of its application to upper atmosphere night glow and similar phenomena in which the intensities, frequencies and polarization of the radiation may be sampled directly by rocket instruments as a function of altitude.

DESCRIPTION OF GENERAL APPROACH

A convenient way to characterize a beam of photons is by the Stokes vector, whose four components describe its intensity and state of polarization completely. The scattering of a photon may then be described by a scattering matrix operating on its Stokes vector to produce a new Stokes vector. If the photon suffers a subsequent scattering, the process is repeated on the new stokes vector, after first taking account of the fact that the coordinate system has changed. McMaster[1] gives an example of this in double Compton scattering.

The scattering matrix for resonant scattering is given in a general form by Chandrasekhar.[2] This is specialized to the Lyman-α case by putting in the appropriate atomic constants. There are two scattering matrices involved in Lyman-α scattering, because two transitions are involved, one isotropic and the other anisotropic.

Successive scatterings of this kind are set up in the present calculation. An incident photon (specified Stokes vector) is injected into the plane parallel atmosphere of H atoms. Its frequency is randomly picked within an assumed Doppler profile, and its direction of incidence may be either randomly picked from an isotropic distribution or chosen at a fixed angle.

The photon encounters an H atom every mean free path λ, where

$$\lambda = \frac{1}{N\sigma},$$

N being hydrogen number density, σ the cross-section for resonant scattering. However, it is not actually scattered if its frequency relative to the H atom is not the same as either one of the two Lyman-α components. It continues on until it does encounter an H atom such that its relative

*Published as *Goddard Space Flight Center Document X-614-64-165.*

frequency lies within the natural width of either of the two Lyman-α components; its Stokes vector is then multiplied by the corresponding scattering matrix. The direction of the scattered photon is decided on the basis of the intensity distribution given by the first two components of the Stokes vector.

(The atom absorbing a photon at any frequency within the natural width is made to re-emit the photon at only the center frequency. This corresponds to an arbitrary kind of redistribution. Actually, when a given atom is illuminated by a line broader than its natural width, its scattering redistributes the radiation over the natural width Heitler.[3] The natural width is negligible compared to the line width here, however.)

This process is repeated for the scattered photon so that successive scatterings are undergone, and the photon describes a random walk in the medium. The medium is divided into 100 zones, and a count of positive and negative flux is kept for each zone.

When the photon escapes from the prescribed region, either forward or backward, a new photon is injected.

DETAILS OF CALCULATION

[a] Photon Path

The medium is taken to extend from $z=0$ to $z=1$ and is infinite in the x, y directions. Photons are injected at $z=0$. The i-th collision-free path of a given photon has the direction cosines (l_i, m_i, n_i). The first scattering results in a scattering direction (l_2, m_2, n_2) which makes an angle θ with the incident direction (l_1, m_1, n_1) and has azimuthal angle φ (Cashwell and Everitt,[4] Ch. VII): θ and φ are chosen as in section (c) below. In subsequent scatterings the scattering angle θ is obtained by the procedure of section (d).

[b] The Lyman-α Scattering Matrix

The general resonant scattering matrix is given by Chandrasekhar[2] as:

$$T=\begin{vmatrix} (\frac{3}{4}E_1\cos^2\theta+\frac{1}{2}E_2) & \frac{1}{2}E_2 & 0 & 0 \\ \frac{1}{2}E_2 & (\frac{3}{4}E_1+\frac{1}{2}E_2) & 0 & 0 \\ 0 & 0 & \frac{3}{2}E_1\cos\theta & 0 \\ 0 & 0 & 0 & \frac{3}{2}E_3\cos\theta \end{vmatrix}$$

in which θ is the scattering angle and the constants E_1, E_2, E_3 depend on the particular energy levels involved.

In the case of Lyman-α, the transitions involved in resonant scattering are $(1s^2S\frac{1}{2}-2P^2P\frac{1}{2})$ and $(1s^2S\frac{1}{2}-2p^2P\frac{3}{2})$. For the first, $j=0$, initial $j=\frac{1}{2}$, and for the second $j=+1$, initial $j=\frac{1}{2}$. The constants E_1, E_2, E_3 are therefore respectively (0, 1, $\frac{2}{3}$) and ($\frac{1}{2}$, $\frac{1}{2}$, $\frac{5}{6}$).

[c] The First Scattering

The incident Stokes vector, which is multiplied by the scattering matrix T, is:

$$S_0=\begin{vmatrix} I_l^0 \\ I_r^0 \\ U^0 \\ V^0 \end{vmatrix}$$

(The relation between this set of components and the alternative set used by McMaster is discussed by Van de Hulst.[5] The component V does not mix with the others in the scattering process so that we can work with the first three components only throughout the calculation. It will be seen that U does get mixed with I_l and I_r in successive scatterings, however.

The scattered photon has its Stokes vector given then by

$$S_1=TS_0$$

If S_0 is the Stokes vector for natural (unpolarized) light,

$$S_0=\begin{vmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 0 \end{vmatrix}$$

the first scattering gives a Stokes vector S_1 which is a function only of the scattering angle θ:

$$S_1=\begin{vmatrix} \frac{3}{4}E_1\cos^2\theta+\frac{1}{2}E_2 \\ \frac{3}{4}E_1+\frac{1}{2}E_2 \\ 0 \end{vmatrix} \equiv \begin{vmatrix} I_l' \\ I_r' \\ U' \end{vmatrix}$$

The probability of the photon actually being scattered at an angle θ_1 is taken to be proportional to the intensity sum $I_l'+I_r'$; θ_i is chosen randomly within this probability distribution by the Von Neumann device described by Cashwell and

Everitt.[4] The azimuth angle φ of the scattered photon is chosen at random within the interval $(0-2\pi)$.

[d] Subsequent Scatterings

Because the plane of the second scattering is rotated in general by an angle χ about the direction of incident propagation with respect to the plane of the first scattering, the Stokes vector S_1 is transformed by a matrix $M(\chi)$, where

$$M(\chi)=\begin{vmatrix} \cos^2\chi & \sin^2\chi & \frac{1}{2}\sin 2\chi \\ \sin^2\chi & \cos^2\chi & -\frac{1}{2}\sin 2\chi \\ -\sin 2\chi & \sin 2\chi & \cos 2\chi \end{vmatrix}$$

into a vector:

$$S_1'=\begin{vmatrix} \cos^2\chi(\frac{3}{4}E_1\cos^2\theta+\frac{1}{2}E_2)+\sin^2\chi(\frac{3}{4}E_1+\frac{1}{2}E_2) \\ \sin^2\chi(\frac{3}{4}E_1\cos^2\theta+\frac{1}{2}E_2)+\cos^2\chi(\frac{3}{4}E_1+\frac{1}{2}E_2) \\ -\sin 2\chi(\frac{3}{4}E_1\cos^2\theta+\frac{1}{2}E_2)+\sin 2\chi(\frac{3}{4}E_1+\frac{1}{2}E_2) \end{vmatrix} \equiv \begin{vmatrix} S_1'(1) \\ S_1'(2) \\ S_1'(3) \end{vmatrix}$$

multiplying this by T gives the resultant scattered Stokes vector S_2:

$$S_2=\begin{vmatrix} (\frac{3}{2}E_1\cos^2\theta+\frac{1}{2}E_2)S_1'(1)+(\frac{1}{2}E_2)S_1'(2) \\ \frac{1}{2}E_2\,S_1'(1)+(\frac{3}{2}E_1+\frac{1}{2}E_2)S_1'(2) \\ (\frac{3}{2}E_1\cos\theta)S_1'(3) \end{vmatrix} \equiv \begin{vmatrix} S_2(1) \\ S_2(2) \\ S_2(3) \end{vmatrix}$$

photons. Fig. 1 is a simplified flow chart of the program.

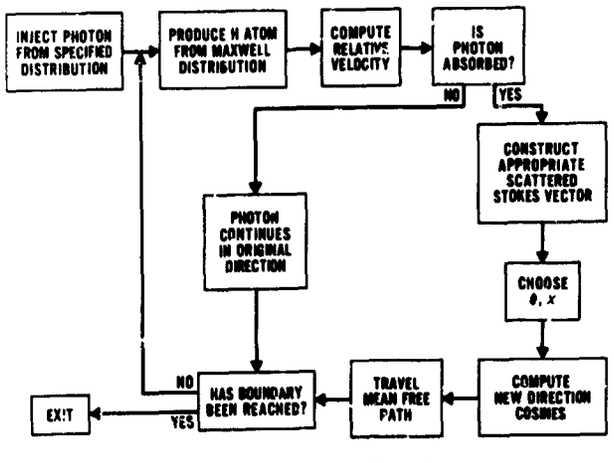


FIGURE 1.—Simplified flow chart.

To find an actual angle θ_2 resulting from the second scattering, the Von Neumann device is again used on the probability distribution

$$p^\theta=S_2(1)+S_2(2)$$

after first randomly choosing χ in the interval $(0-2\pi)$. (The assumption is thereby made that the plane of the second scattering is randomly related to that of the first).

[e] Computer Program

A computer program was written by C. Wade to carry out the calculation. It includes provision for injecting either a unidirectional beam at any specified angle, or isotropically incident

RESULTS AND DISCUSSION

A uniform plane-parallel atmosphere of hydrogen atoms at temperature 500°K, and of 500 km thickness, was taken as the medium on which to run the calculation. Two densities were successively assumed, 1.0×10^5/cc and 2.0×10^5/cc, and with each density two different isotropically incident radiation profiles were used, corresponding to Doppler temperatures T_D of 500°K and 5000°K. The curves of Fig. 2 were obtained for the four cases, each representing results for about 5000 photons. (A computer run for 5000 photons was about one half hour long). The ordinate represents the relative number of photons passing forward through a given plane in the atmosphere specified by the abscissa. Since all photons crossing a given plane are counted, the number corre-

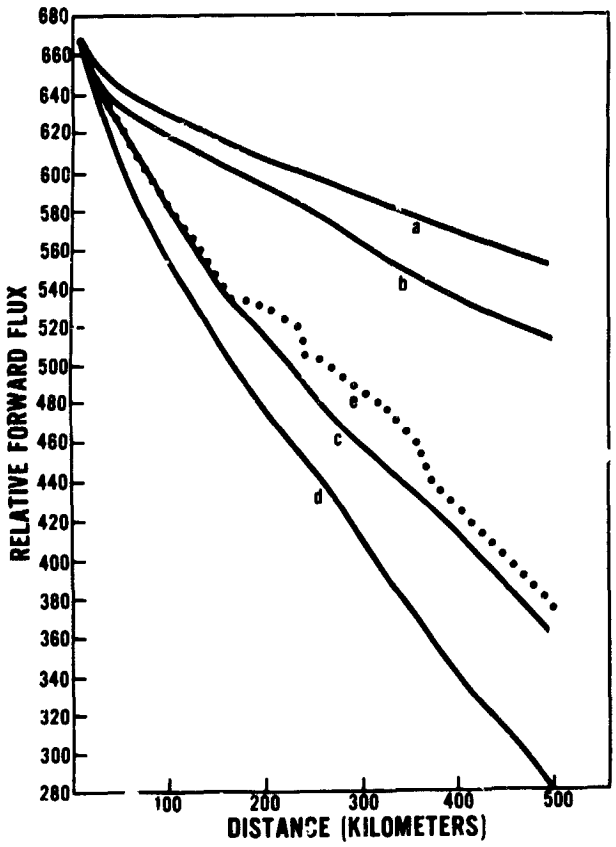


FIGURE 2.—Relative flux distributions.

sponds to the hemispheric flux πF of transfer theory, or to photometric radiancy.

The dotted curve represents the results of case (*c*) for a smaller sample of 1000 photons. The curve (*c*) for 5000 photons departs somewhat from the dotted curve, showing that "statistical equilibrium" has not yet been completely reached. Further runs were not made, however, because the models considered here are not of interest in themselves.

The average number of scatterings per photon in each case was: (*a*) 0.67, (*b*) 1.28, (*c*) 1.96, (*d*) 3.43. The reflection coefficients or albedos were: (*a*) 0.18, (*b*) 0.23, (*c*) 0.46, (*d*) 0.58. Broadening of the incident frequency profile gives greater transparency of the atmosphere to the radiation, as expected, while the two-fold increase in atmospheric density gives greater opacity.

An analytic treatment of the problem would involve the solution of a system of transport equations,[6] with a kernel which incorporates both the resonant scattering matrix and the Doppler redistribution in frequency.[7] The difficulty of this analysis precludes a complete check of the results, but at the same time is an argument for the convenience of the present method.

It may be that the use of a simpler scattering process, i.e. scattering by a single $1s-2p$ transition, would give essentially the same results.

SUMMARY

The use of Stokes vectors as random variables appears to be an effective way of handling some radiative transfer problems. We have applied the method here to Lyman-α scattering in a uniform gas of hydrogen atoms, of density low enough that few scatterings occur. It may be possible, with further work, to extend it to the case of a non-uniform atmosphere.

The same calculation can also, if required, give the frequency, profile and the polarization of the radiation at any level if the frequencies and Stokes vectors are recorded.

REFERENCES

1. W. H. MCMASTER, Am. Jour. of Phys. *22*, 351 (1954).
2. S. CHANDRASEKHAR, Radiative Transfer, p. 51. Dover Publications: New York (1960).
3. W. HEITLER, The Quantum Theory of Radiation, 3rd ed., p. 201. Oxford Univ. Press, Oxford (1954).
4. CASHWELL and EVERETT, The Monte Carlo Method, Pergamon Press (1959).
5. H. C. VAN DE HULST, Light Scattering by Small Particles, p. 41. John Wiley and Sons, Inc., New York (1962).
6. U. FANO, L. V. SPENCER and M. J. BERGER, in Handbuch der Physik *38/2*, p. 684. Springer-Verlag, Berlin (1959).
7. L. G. HENYEY, Proc. Nat. Acad. Sci. *26*, 50 (1940).

NG4-23372

INFRASONIC WAVES FROM THE AURORAL ZONE*

KAICHI MAEDA AND TOMIYA WATANABE†

Pulsating aurorae are proposed as a source of the infrasonic waves associated with geomagnetic activity reported by Chrzanowski et al. One of the most plausible generation mechanisms of these long period pressure waves is the periodic heating of the air around 100 km, corresponding to the auroral coruscation reported by Campbell and Rees. In order to show the energetic relationship between source input and pressure change at sea level, some theoretical calculations are performed by using a simple model of auroral distribution in the isothermal atmosphere.

INTRODUCTION

The purpose of this paper is to interpret the origin of the strange traveling atmospheric waves observed at the ground during intervals of high geomagnetic activity (Reference 1). In particular, discussion centers on possible modes of atmospheric oscillations caused by the periodic bombardment of auroral particles in the polar mesosphere.

Trains of the waves are detected by a system of four microphones placed on each corner of a quadrant roughly 8 km square, located north of Washington, D. C. The presence of a traveling wave is established when the same wave forms can be found on all four records with certain time shifts between them. These time displacements are used to determine the direction of wave propagation and the horizontal phase velocity of the waves. The periods of these infrasonic waves are usually 20 to 80 sec, but occasionally 100 to 300 sec waves are recorded. The pressure amplitude ranges from about 1 to 10 dynes/cm^2.

One of the peculiar features of these waves is the change of arrival direction with time of day. The general trend is from the northeast in the evening, from the north about midnight, and from the northwest in the morning. The shift back to northeast is somewhat discontinuous. These waves occur less frequently in the daytime.

Since auroral activity predominates at about midnight local time, the time dependence of the appearance of infrasonic waves during intervals of high geomagnetic activity can be explained by assuming that the source of this kind of wave is located somewhere in the auroral region, as can be seen from Figure 1. The left side of this figure shows the diurnal variation of the arrival of sound waves during magnetic storms, reported by Chrzanowski et al. (Reference 1). This assumption is consistent with the finding that the enhancement of the wave intensity at Washington, D. C. is generally delayed several hours after the increase of geomagnetic activity, except in the case of very severe magnetic storms. During these storms the delay is less than 30 minutes, indicating southward spread of the source of these waves near Washington, D. C.

In attempting to understand these peculiar pressure waves, it is proposed that they originate from a certain domain of the ionospheric region in the auroral zone which is heated periodically by a severe bombardment of auroral particles. It is thought that this periodic precipitation of auroral particles occurs simultaneously with geomagnetic pulsations.

INTERRELATIONS BETWEEN GEOMAGNETIC FLUCTUATIONS, PULSATING AURORAE, AND INFRASONIC WAVES

The auroral luminosity often fluctuates with incoming auroral particles. The fluctuations of these particles, mostly electrons, can be explained by either a periodic change in the acceleration mechanism of incident particles or by the change of mirror heights of trapped particles in the earth's atmosphere, following the variations of the field intensity of the earth's magnetosphere. The variations of geomagnetic field intensity and of auroral brightness are therefore closely related. A clear example of the correspondence between

*Published as *NASA Technical Note* D-2138, June 1964.
†University of British Columbia.

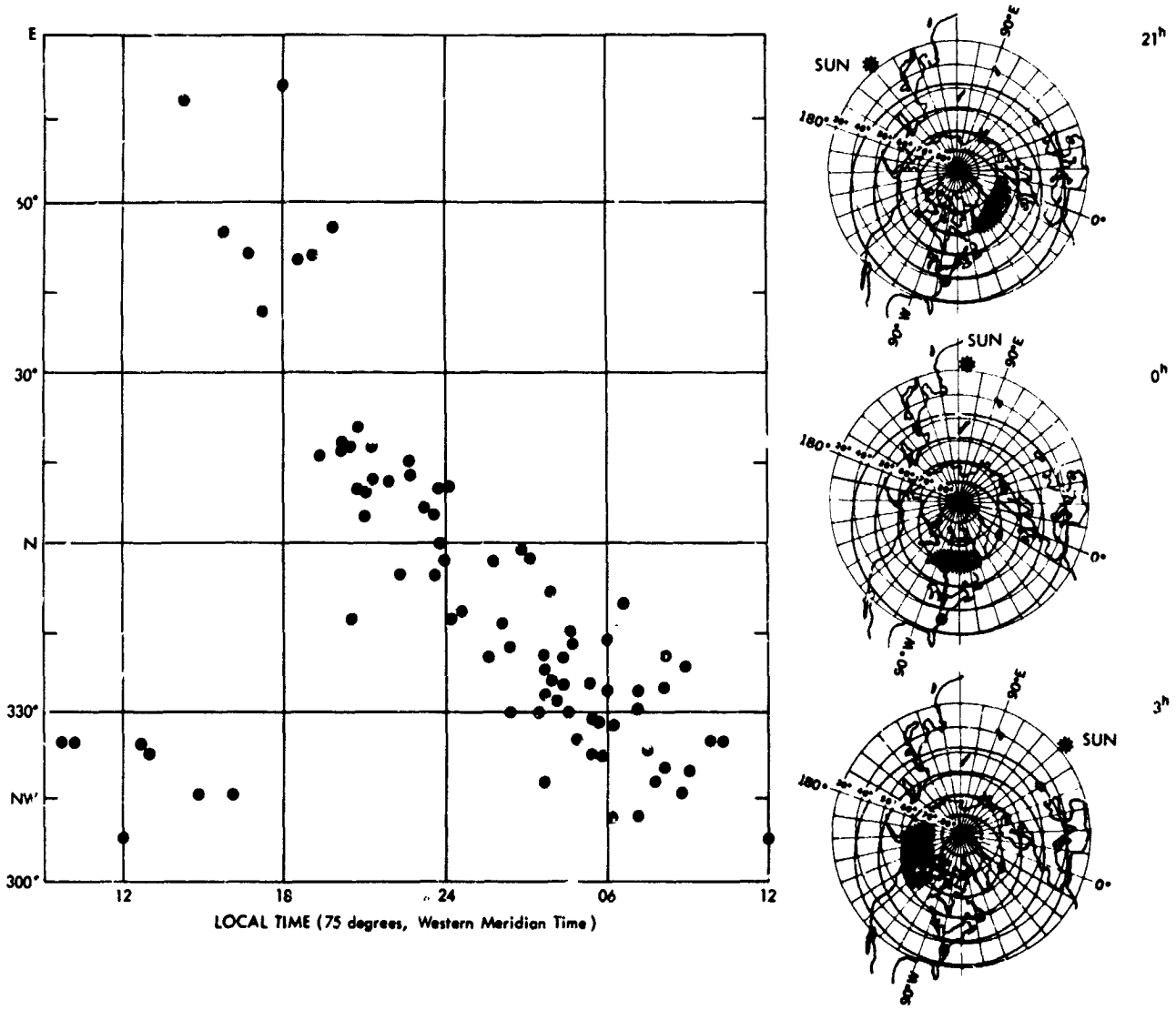


FIGURE 1.—Diurnal variation of the arrival direction of infrasonic waves during magnetic storms observed at the National Bureau of Standards, Washington, D.C. The three figures on the right indicate the shifts of the source of the pressure waves, corresponding to the movement of auroral activity.

the pulsating aurorae and the rapid variation of geomagnetic intensity observed at the ground has been given by Campbell (Reference 2).

According to the analysis of the space probe data given by Sonnet et al. (Reference 3), and that of Explorer X (1961 κ1) given by Heppner et al. (References 4 and 5), the region near the geomagnetic equator beyond about 10 earth radii is occasionally greatly disturbed. Such disturbances may be propagated along the magnetic lines of force as hydromagnetic waves, and are transformed into electromagnetic waves when they reach the conducting ionosphere (References 6–8). After penetrating the ionospheric region as electromagnetic waves, they are almost perfectly reflected at the earth's surface (Reference 9). This produces standing hydromagnetic waves along magnetic lines of force. The simplest mode of such standing waves is the fundamental mode, whose unique node is on the geomagnetic equatorial plane, with two loops of oscillation on the ends of the line of force on the earth, one in each hemisphere. This mode of oscillation has been investigated theoretically by Dungey (Reference 10), who called it the normal mode of torsional oscillations of the magnetic field in the earth's cavity. The eigen period of these oscillations increases rapidly with the latitude of the magnetic

line of force intersecting the earth's surface, varying from several tens of seconds in the subauroral region to about 10 minutes in the polar region (References 11–13).

In addition to the above mode, there is another mode of oscillation responsible for geomagnetic pulsations of shorter periods. As noted by Dessler (Reference 11), the velocity of Alfvèn waves decreases very rapidly with decreasing height in the exospheric region below around 2,000 km. Therefore, because of the continuity of energy flow Alfvèn waves of certain periods coming from the outer exosphere can be expected to be intensified. In this mode of oscillation, the layer of maximum Alfvèn velocity, approximately between 1500 and 3000 km, becomes the node of oscillation, and the loop of oscillation is near the earth's surface. This mode of hydromagnetic oscillation corresponds to the geomagnetic pulsations with periods from about 1 to several seconds, which appear more frequently in nighttime than in daytime (Reference 15).

In considering the above facts, it might be speculated that a possible mechanism for the production of infrasonic waves during auroral activity would be the penetration of Alfvèn waves, including modified Alfvèn waves and retarded sound waves (Reference 13), through the ionosphere. However, as will be shown later, these contributions are very small compared with the pressure disturbances produced by the periodic heating of the lower ionosphere caused by auroral bombardments.

As shown by Heppner (Reference 16) auroral activity predominates around midnight local time and the active region extends towards lower latitudes with increasing activity; pulsating aurorae appear at this phase of auroral activity. In other words, among several types of auroral displays pulsating aurorae appear with the largest disturbance in energy and they occur in fairly low latitudes.

According to Campbell and Ress (Reference 17) the peak of pulsating aurorae is around 100 km, the bottom is at 90 km, the effective thickness is of the order of 20 km, and the most frequent period is from 6 to 10 sec. From the direct measurement of auroral particles by means of rocket borne detectors, the energy flux of auroral particles, mostly electrons, is of the order of several tens of ergs/cm²–sec at weak aurora and increases by more than a factor of 50 at bright aurora (Reference 18). Thus, the energy flux in a strong pulsating aurora can be estimated to be of the order of 10^3 ergs/cm²–sec or more. This figure is consistent with the estimate given by Chamberlain (Reference 19), based on measurements of auroral luminosity.

On the other hand, the energy flux of hydromagnetic waves deduced from the magnetic pulsation data is less than 10 ergs/cm²–sec below 200 km. The energy flux of hydromagnetic waves increases with height. However, the contribution to pressure waves in the lower atmosphere decreases with the increasing height of the source, as will be shown later.

Another evidence for the present idea is the very good correspondence between the appearance of pulsating aurorae and that of infrasonic waves (Figure 2). The occurrence of pulsating aurorae shown on the right of Figure 2 is taken from the visoplot of auroral activity, reported by IGY

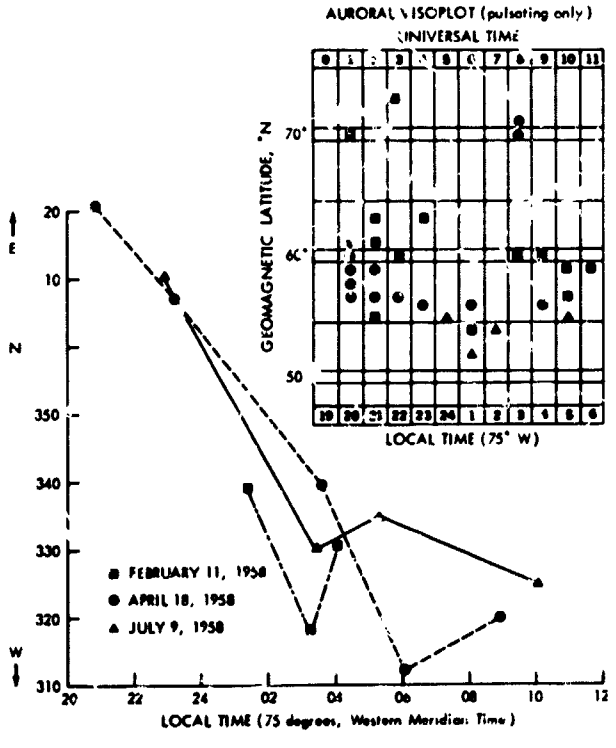


FIGURE 2.—Diurnal variation of the arrival direction of infrasonic waves during specific magnetic storms detected at the National Bureau of Standards, Washington, D. C. (Chrzanowski et al., 1960), and corresponding auroral activities observed in the northern hemisphere.

World Data Center A, on days which correspond to the events, reported by Chrzanowski et al. (Reference 1), shown on the left of this figure. Two occurrences of infrasonic waves in 1957 were reported, but since no visoplots of auroral activity are available before the IGY, they are omitted from this figure. The arrival direction of these waves might deviate from their true direction because of strong wind systems in high altitudes. Therefore, the appearances of pressure waves do not necessarily match the pulsating auroral events in detail. However, it is quite clear that there is a close correspondence between infrasonic waves and pulsating aurorae.

MATHEMATICAL TREATMENT

It is obvious that ray theory is not applicable in this problem, because the wavelength is of the order of an atmospheric depth or more for the infrasonic waves observed.

The excitation and propagation of long period pressure waves in the atmosphere have been investigated by several workers, mainly in two fields of geophysics, i.e., meteorology and ionospheric physics. The former consists of the study of atmospheric oscillations (References 20–24), and the study of micro-barometric disturbances (References 25–36). The latter is mostly concerned with the investigation of ionospheric disturbances and has been discussed by Martyn (Reference 37), Sen and White (Reference 38), White (References 39 and 40) and Hines (References 41 and 42).

Although the subjects considered in the noted references are quite different, the mathematical treatments are essentially the same. They are based on a differential equation of velocity divergence, which is derived from three fundamental equations, i.e., the equations of motion, continuity, and of first law of thermodynamics. The theoretical aspects of the problems in geophysics, including oceanography, are reviewed in References 43–46. Mathematical technique has been developed to solve these problems, introducing the so-called field variables instead of simple hydrodynamical variables.

In the present calculations, however, the classical method based on the equations of velocity divergence will be used, because of the convenience in comparing the results with observed data.

Notations and Fundamental Equations

Preliminary notation will now be given:

- c velocity of sound in the atmosphere in cm/sec, $c^2=\gamma gH$ where $H=RT/g$ is the scale height of the isothermal atmosphere,
- D/Dt Eulerian derivative, $\partial/\partial t+U\nabla$,
- $\mathbf{f}$ resultant of external forces except gravity, dynes/gm,
- $\mathbf{g}(0, -g)$ acceleration of gravity, $g=980$ cm/sec^2,
- k horizontal wave number corresponding to λ, in cm^{-1}, i.e., $2\pi/\lambda$,
- ℓ vertical wavelength of a pressure wave in cm,
- p, ρ, T small departures from static values of pressure, density, and temperature, functions of x, z, and t, in dynes/cm^2, gm/cm^3, and °K,
- p_0, ρ_0, T_0 static values of pressure, density, and temperature, functions of z only,
- p_s, ρ_s, T_s static values of pressure, density, and temperature of air at seal level.
- p_T, ρ_T, T_T total pressure, density, and temperature, i.e., p_0+p, $\rho_0+\rho$, and T_0+T,
- $q'(x, z)$ rate of net accession of heat, erg/gm–sec; $s(x, z)=(\gamma-1)q$ where $q=\rho_T q'$ in ergs/cm^3–sec,
- R gas constant of air, $B/M=2.87\times10^6$ ergs/gm–°C, where B is the universal gas constant, 8.314×10^7 ergs/mol–°K, and M is the molecular weight of air, approximately 28.97,
- $\mathbf{U}(u, w)$ velocity vector, where u is the horizontal (southward) and w is the upward component of air flow in cm/sec,
- x, z horizontal (southward) and vertical (upward) coordinates,
- γ ratio of the specific heats of air, $C_p/C_v\approx1.4$, where C_p and C_v are the specific heat of air at constant pressure and constant volume, respectively,
- η entropy of air in ergs/gm–°K,
- λ horizontal wavelength of a pressure wave in cm,
- σ frequency of the pressure wave corresponding to τ, i.e., $2\pi/\tau$,

τ period of pressure wave in sec,
$\chi(x, z)$ the divergence of velocity in sec^{-1}, i.e., $\chi = \partial u/\partial x + \partial w/\partial z$,
Ω Coriolis vector in rad/sec.

In Eulerian notation, the equation of motion is

$$\frac{D\mathbf{U}}{Dt} + \frac{1}{\rho}\nabla p + \mathbf{g} + \Omega \times \mathbf{U} = \mathbf{f} \quad . \tag{1}$$

and the equation of continuity is

$$\frac{D}{Dt}\left(\frac{1}{\rho}\right) - \frac{1}{\rho}\nabla \mathbf{U} = 0 \tag{2}$$

The variations of pressure due to thermal excitations can be derived from the first law of thermodynamics, which gives the change of entropy (References 43–45),

$$\frac{D\eta}{Dt} = \frac{q'}{T} \quad . \tag{3}$$

By making use of the perfect gas function for entropy (Reference 46) it can be shown that Equation 3 is equivalent to

$$\frac{Dp}{Dt} - c^2 \frac{D\rho}{Dt} = s(x, z, t) \quad . \tag{4}$$

where

$$s(x, z, t) = (\gamma - 1)\rho q'(x, z, t) = (\gamma - 1) q(x, z, t) \quad .$$

Since, in the present problem, the period of oscillation is less than a few minutes, the Coriolis force due to the earth's rotation is negligible, and all other external forces can be assumed to be zero. The equation of motion for the present problem is then simply

$$\frac{D\mathbf{U}}{Dt} + \frac{1}{\rho_0}\nabla p = \mathbf{g} \quad . \tag{6}$$

One-Dimensional Model

For the one-dimensional case atmospheric motion has only the z component of velocity, w; therefore, in its linear approximation form, the equation of motion (Equation 6) is

$$\rho_0 \frac{\partial w}{\partial z} = \frac{-\partial p}{\partial z} - \rho g \quad . \tag{7}$$

The equation of continuity in its linear approximation form is

$$\frac{\partial \rho}{\partial t} + \rho_0 \frac{\partial w}{\partial z} - \frac{\rho_0}{H} w = 0 \quad . \tag{8}$$

And the linear approximation form of the entropy equation (Equation 4) is:

$$\frac{\partial p}{\partial t} = -\rho_0 c^2 \frac{\partial w}{\partial z} + \rho_0 g w + (\gamma - 1) q \quad , \tag{9}$$

where $q = q(z, t)$ is the periodically changing heat source, which can be assumed to be

$$q(z, t) = \begin{cases} q(z)\, e^{i\sigma t} & \text{for } z \geq 0 \quad , \\ 0 & \text{for } -z_0 \leq z < 0 \quad , \end{cases} \tag{10}$$

where

$$q(z) = \frac{q_0 e^{-z/h}}{h} \quad ; \tag{11}$$

h is a constant and q_0 is the maximum rate of heat generation in an atmospheric column with unit cross section, in ergs/cm²-sec; $z = 0$ is taken as the height of the base of heating; and the value at the earth's surface is given by $z = -z_0$.

By eliminating p and ρ from the above equations, the following differential equation is obtained:

$$\frac{\partial^2 w}{\partial z^2} - \frac{1}{H}\frac{\partial w}{\partial z} - \frac{1}{c^2}\frac{\partial^2 w}{\partial t^2} = \frac{\gamma - 1}{\rho_0 c^2}\frac{\partial q}{\partial z} \quad . \tag{12}$$

In the region $z \geq 0$, where the heat source exists,

$$w = \frac{i}{k_2 - k_1} e^{ik_1 z}\left[\int_0^z e^{-ik_1 \xi} Q(\xi)\, d\xi + C\right] + \frac{i}{k_1 - k_2} e^{ik_2 z}\left[\int_0^z e^{-ik_2 \xi} Q(\xi)\, d\xi + D\right] \quad , \tag{13}$$

where k_1 and k_2 are given by

$$\left.\begin{aligned} k_1 &= \sqrt{\frac{1}{l^2} - \frac{1}{4H^2}} - i\,\frac{1}{2H} \\ k_2 &= -\sqrt{\frac{1}{l^2} - \frac{1}{4H^2}} - i\,\frac{1}{2H} \end{aligned}\right\} \quad (14)$$

and

$$\frac{1}{l} = \frac{\sigma}{c} ,$$

$$Q(\xi) = \frac{\gamma - 1}{\rho_0 c^2}\frac{dq(\xi)}{d\xi} . \quad (15)$$

C and D are integration constants, to be determined by the boundary conditions. The t[illegible] dependent factor, $e^{i\sigma t}$, is dropped in Equation 13 and will be always dropped hereafter.

For the region $-z_0 \leq z < 0$, where no heat source exists,

$$w = \frac{i}{k_2 - k_1} e^{ik_1 z} A + \frac{i}{k_1 - k_2} e^{ik_2 z} B ; \quad (16)$$

where A and B are constants. The condition that $w = 0$ at the earth's surface gives

$$B = e^{i(k_2 - k_1)z_0} A . \quad (17)$$

The vertical velocity w should be continuous at $z = 0$, and therefore

$$A - B = C - D . \quad (18)$$

The pressure p also should be continuous at $z = 0$, so that the following condition is to be satisfied:

$$\rho_c c^2 \left(\frac{dw}{dz}\right)_{z=0+\epsilon} - (\gamma - 1)\, q(0) = \rho_c c^2 \left(\frac{dw}{dz}\right)_{z=0-\epsilon} , \quad (19)$$

where $\epsilon \to 0$ and

$$\rho_c = (\rho_0)_{z=0} , \quad (20)$$

the atmospheric density at $z = 0$.

The above equations give

$$k_1 C - k_2 D - k_1 A + k_2 B = \frac{k_1 - k_2}{\rho_c c^2} (\gamma - 1)\, q(0) . \quad (21)$$

For $z \to \infty$, there should be no wave downward, thus from Equation 13

$$C = -\int_0^\infty e^{-ik_1 \xi} Q(\xi)\, d\xi . \quad (22)$$

From Equations 18 and 19,

$$A = C - \frac{\gamma - 1}{\rho_c c^2} q(0) . \quad (23)$$

To find the pressure variation at the earth's surface, we must obtain $(dw/dz)_{z=-z_0}$, which is given by using Equations 16 and 17;

$$\left(\frac{dw}{dz}\right)_{z=-z_0} = e^{-ik_1 z_0} A . \quad (24)$$

Then Equation 9 gives the pressure variation at the earth's surface,

$$p_{z=-z_0} = \frac{i}{\omega} \rho_s c^2 e^{-ik_1 z_0} A . \quad (25)$$

where ρ_s is the atmospheric density at the earth's surface. By substituting Equation 22 into Equation 23,

$$|A| = \frac{(\gamma - 1)\, q_0}{\rho_c\, c^2} \frac{1}{l} \left(1 - \frac{h}{H} + \frac{h^2}{l^2}\right)^{-1/2} , \quad (26)$$

provided that the angular wave frequency σ is larger than a critical frequency σ_A,

$$\sigma_A = \frac{\gamma g}{2c} . \quad (27)$$

By using Equations 14

$$\left|e^{-ik_1 z_0}\right| = e^{-z_0/2H} . \quad (28)$$

Thus the final expression of the amplitude of pressure variations at the ground, caused by the periodic disturbance in the upper atmosphere q(z, t), is

$$p_s = \frac{\gamma - 1}{c} q_0 \left(\frac{\rho_s}{\rho_c}\right)^{1/2} \left(1 - \frac{h}{H} + \frac{h^2}{l^2}\right)^{1/2} . \quad (29)$$

The values for p_s/q_0 which may be derived from this equation are plotted vs. h in Figure 3 for several values of angular frequency, σ. The full lines and dashed lines stand for the scale heights of 8 km and 6.8 km, respectively, for an isothermal atmosphere. The latter gives the proper ratio of $(\rho_c/\rho_s)^{1/2}$ as compared with the observed atmospheric densities. From this figure, it can be seen that periodic heating corresponding to a flux of the order of 100 ergs/cm²–sec produces pressure waves with amplitudes of the order of 1 dyne/cm² at the ground.

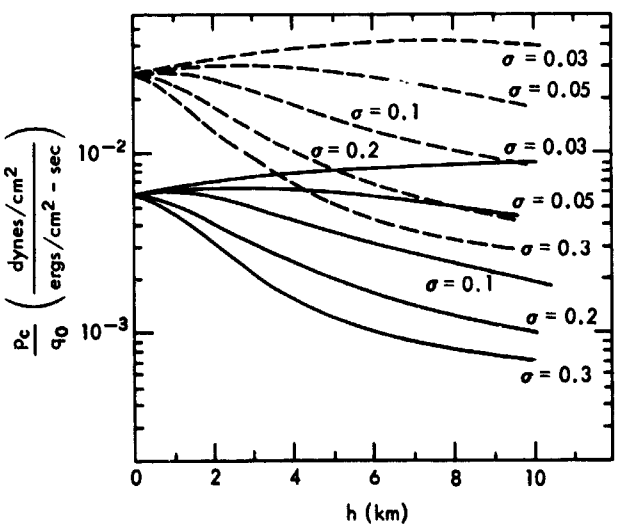


FIGURE 3.—Relative intensity of pressure waves on the ground, produced by periodic auroral heating, vs. the scale height of the heat source, for the one-dimensional model with an isothermal atmosphere.

Two-Dimensional Model

Equation of Velocity Divergence

From Equation 6 the equation of motion for the two-dimensional case is

$$\rho_0 \frac{\partial u}{\partial t} = -\frac{\partial p}{\partial x} , \quad (30)$$

$$\rho_0 \frac{\partial w}{\partial t} = -\frac{\partial p}{\partial z} - g\rho . \quad (31)$$

The equation of continuity (Equation 2) can be written in a first order approximation as

$$\frac{\partial p}{\partial t} + w \frac{\partial \rho_0}{\partial z} = -\rho_0 \chi , \quad (32)$$

where $\chi = \chi(x, z, t)$ is the velocity divergence. By using Equation 32, Equation 4 can be written as

$$\frac{\partial p}{\partial t} = \rho_0 gw - \rho_0 c^2 \chi + s . \quad (33)$$

Under the assumption that the time variations of u, w, p, ρ, and s are proportional to a factor $e^{i\sigma t}$, the following relations between u, w, and p are obtained from Equations 30–33:

$$-\sigma^2 u = \frac{\partial}{\partial x}\left(c^2 \chi - gw\right) - \frac{1}{\rho_0}\frac{\partial s}{\partial x} , \quad (34)$$

$$-\sigma^2 w = c^2 \frac{\partial \chi}{\partial z} - \gamma g \chi + g \frac{\partial u}{\partial x} - \frac{1}{\rho_0}\frac{\partial s}{\partial x} , \quad (35)$$

$$i\sigma p = \rho_0 gw - c^2 \rho_0 \chi + s . \quad (36)$$

By eliminating u, w, and p from these equations, the following differential equation for the velocity

divergence $\chi(x, z)$ is obtained:

$$\nabla^2 \chi + \frac{1}{c^2}\left(\frac{dc^2}{dz} - \gamma g\right)\frac{\partial \chi}{\partial z} - \frac{g}{c^2\sigma^2}\left[\frac{dc^2}{dz} + (\gamma - 1)g\right]\frac{\partial^2 \chi}{\partial x^2} + \frac{\sigma^2}{c^2}\chi$$

$$= -\frac{1}{c^2\sigma^2}\left[\frac{g}{\rho_0(z)c^2}\left(\frac{dc^2}{dz} + \gamma g\right)\frac{\partial^2 s}{\partial x^2} - \frac{\sigma^2}{\rho_0 c^2}\left(\frac{dc^2}{dz} + \gamma g\right)\frac{\partial s}{\partial z} - \frac{\sigma^2}{\rho_0}\nabla^2 s\right] \quad . \quad (37)$$

The derivation of this equation is given in Appendix A.

The Diagnostic Diagram

If there is no thermal excitation, the right side of Equation 37 is zero, and the solution of this homogenous differential equation corresponds to the free oscillation of the atmosphere on the nonrotating earth.

We will now consider pressure waves traveling horizontally in this flat atmosphere. If u, w, p, and ρ are proportional to a factor $e^{i(\sigma t - kx)}$, the following is the equation for the vertical change of $\chi(\sigma, z)$:

$$\frac{d^2\chi}{dz^2} + \frac{1}{c^2}\left(\frac{dc^2}{dz} - \gamma g\right)\frac{d\chi}{dz} + \left[\frac{\sigma^2}{c^2} - k^2 + \frac{gk^2}{\sigma^2 c^2}\left(\frac{dc^2}{dz} + \gamma g - g\right)\right]\chi = 0 \quad .$$

Since $c^2 = \gamma g H(z)$, this may be written

$$\frac{d^2\chi}{dz^2} + \frac{1}{H}(H' - 1)\frac{d\chi}{dz} + \left[\frac{\sigma^2}{\gamma g H} - k^2 + \frac{k^2 g}{\sigma^2 H}\left(H' + \frac{\gamma - 1}{\gamma}\right)\right]\chi = 0 \quad ,$$

where $H' = dH/dz$.

For simplicity, consider the case of an isothermal atmosphere, where $H' = 0$. Equation 39 becomes

$$\frac{d^2\chi}{dz^2} - 2N\frac{d\chi}{dz} + M^2\chi = 0 \quad , \quad (40)$$

where the constant $N = 1/2H$ and

$$M^2 = \frac{\sigma^2}{c^2} - k^2 + \frac{k^2 g^2}{c^2\sigma^2}(\gamma - 1) \quad . \quad (41)$$

This differential equation has the solution:

$$\chi(\sigma, z) = e^{Nz}\left(Ae^{-\mu z} + Be^{+\mu z}\right) \quad , \quad (42)$$

where A and B are constants,

$$\mu^2 = N^2 - M^2 \geq 0 \quad , \quad (43)$$

for the noncellular solution, and

$$\mu = i\eta \quad , \qquad \eta^2 = M^2 - N^2 > 0 \quad , \quad (44)$$

for the cellular solution. As was shown by Pekeris (Reference 27) for the noncellular solution in an isothermal atmosphere (Equation 42) the term with $e^{\mu z}$ must vanish, i.e, $B = 0$. Otherwise the kinetic energy of noncellular waves, which is proportional to $\rho_0(z)\chi^2$, would diverge. Furthermore, because of the condition that the vertical component of velocity must vanish at the ground, we get only two possible types of free oscillation, specified by the following characteristic relations:

Type I: $$k = \frac{\sigma}{c} \quad , \qquad V = U = c \quad , \quad \text{and} \quad \mu = \frac{g(2 - \gamma)}{2c^2} \quad , \quad (45)$$

Type II: $$k = \frac{\sigma^2}{g} \quad , \qquad V = \frac{g}{\sigma} \quad , \quad U = \frac{1}{2}V \quad , \quad \text{and} \quad \mu = \frac{\sigma^2}{g} - \frac{\gamma g}{2c^2} \quad . \quad (46)$$

where $V=\sigma/k$ and $U=d\sigma/dk$, the phase velocity and group velocity of noncellular horizontal waves. Type I corresponds to Lamb's wave (Reference 20), which exists at all frequencies; the waves of type II can be propagated only above a critical value σ_c,

$$\sigma_c = \frac{g}{c}\sqrt{\frac{\gamma}{2}} \qquad (47)$$

(the corresponding period is about 250 sec), and the wave is dispersive.

It should be noted that, because of the decrease of atmospheric density, the amplitude of the Lamb wave (pressure variation) decreases with altitude by a factor $\exp(-gz/c^2)$, preventing the propagation of waves in any except horizontal directions (Reference 46). On the other hand, for the cellular solution η stands for the wave number in the vertical direction and Equation 44 is equivalent to

$$\eta^2 c^2 = \frac{(\sigma^2-\sigma_A^2) + k^2c^2(\sigma_B^2-\sigma^2)}{\sigma^2} \qquad (48)$$

where

$$\sigma_A = \frac{g\gamma}{2c}$$

$$\sigma_B = \frac{g(\gamma-1)^{1/2}}{c} \qquad (49)$$

Brunt's frequency, σ_B, is the frequency of the vertical oscillation of a free air parcel in the atmosphere, changing adiabatically. This expression was also derived by Väisälä as a stability parameter of the atmosphere (Reference 46). The frequency σ_A can be called the atmospheric resonance sound frequency (Reference 47).

The curve for $\eta^2=0$ shown in Figure 4 (a plot of σ vs. k) consists of two branches, A and B. The curve A starts from the σ-axis at $\sigma=\sigma_A$ and is asymptotic to the line C, which corresponds to the solution of the noncellular Lamb wave, $\sigma=kc$. The other curve B passes through the origin and is asymptotic to the horizontal line $\sigma=\sigma_B$. Every point in the two regions where

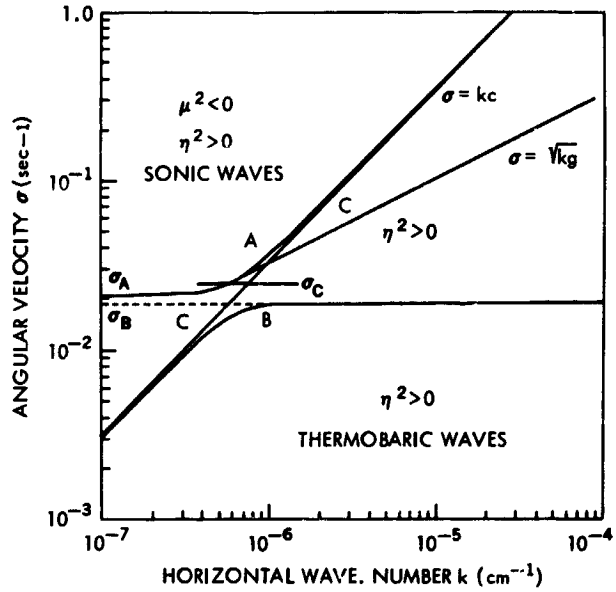


FIGURE 4.—Diagnostic diagram for an isothermal atmosphere with scale height H = 8 Km (T_0 = 273°K).

$\eta^2>0$, one surrounded by curve A and the other surrounded by curve B, leads to an eigen solution of Equation 39. In the third region (between the other two), where $\eta^2<0$, only the points on the line $\sigma=$ kc and $\sigma=\sqrt{kg}(\sigma>\sigma_c)$ lead to eigen solutions. Figure 4 is called the diagnostic diagram of the isothermal atmosphere, and the waves corresponding to the two domains in which $\eta^2>0$ are named Sonic (mode A) and Thermobaric (mode B), respectively (Reference 46). Sonic and thermobaric waves are called acoustic and internal gravity waves, respectively, by Hines (Reference 42).

Intensity of Pressure Waves at the Ground

Since the source of excitation s(x, z) is limited to a certain area inside the auroral zone, the right side of Equation 37 is not uniform with respect to x. Therefore, this differential equation is not in general separable with respect to the variables x and z, and the following assupmtions are made in solving the equation:

1. The atmosphere is isothermal, with scale height H, in km.
2. The distribution of the heat source is uniform along the y direction (this is assumed at the beginning to reduce the problem to two dimensions), but it is limited horizontally in the x direction within $\pm\lambda_0$ (i.e., $-\lambda_0 \le x < \lambda_0$) and is

extended vertically above a certain height z (i.e., $z \geq z_0$).

3. The time variation of the heat source is periodic with an angular frequency σ, and the same phase within the domain indicated in item 2 above.

The first assumption reduces Equation 37 to the following form:

$$\left(1 - \frac{\sigma_B^2}{\sigma^2}\right)\frac{\partial^2 \chi}{\partial x^2} + \frac{\partial^2 \chi}{\partial z^2} - \frac{1}{H}\frac{\partial \chi}{\partial z} + \frac{\sigma^2}{c^2}\chi = \frac{1}{\rho_0(z)c^2}\left[\left(1 - \frac{g}{\sigma^2 H}\right)\frac{\partial^2 s}{\partial x^2} + \frac{\partial^2 s}{\partial z^2} + \frac{1}{H}\frac{\partial s}{\partial z}\right] . \quad (50)$$

Assumptions 2 and 3 can be written as

$$s(x, z, t) = s(z)\, e^{i\sigma t}\, \theta(z - z_0)\left[\theta(x + \lambda_0) - \theta(x - \lambda_0)\right] , \quad (51)$$

where

$$s(z) = \frac{s_0}{h} e^{-(z - z_0)/h} = \frac{(\gamma - 1) q_0}{h} \exp\left[\frac{-(z - z_0)}{h}\right] , \quad (52)$$

and $\theta(\xi)$ is a unit step function of ξ,

$$\theta(\xi) = \begin{cases} 1 & \text{for } \xi \geq 0 , \\ 0 & \text{for } \xi < 0 . \end{cases} \quad (53)$$

It should be noted that q_0 is the maximum rate of heat release in the atmosphere in an air column of unit cross-section (in ergs/cm²-sec).

Since the atmosphere is assumed to be isothermal, the density of air in equilibrium at height z is

$$\rho_0(z) = \rho_s e^{-z/H} , \quad (54)$$

where ρ_s is the atmosphere density at the earth's surface and H is the scale height.

In order to solve Equation 50 under the conditions listed, the following Fourier transforms are applied with respect to x,

$$X(k, z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ikx} \chi(x, z)\, dx , \quad (55)$$

$$S(k, z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ikx} s(x, z)\, dx . \quad (56)$$

Since the heat source s(x, z) vanishes outside of the auroral zone, both s(x, z) and $\chi(x, z)$ must vanish at $x = \pm\infty$. Thus

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\partial^2 \chi}{\partial x^2} e^{ikx}\, dx = -k^2 X(k, z) , \quad (57)$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\partial^2 s}{\partial x^2} e^{ikx}\, dx = -k^2 S(k, z) . \quad (58)$$

With these transforms, Equation 50 is written

$$\frac{d^2 X}{dz^2} - 2N\frac{dX}{dz} + M^2 X = F(z, k) , \quad (59)$$

where

$$F(z, k) = \frac{1}{c^2 \rho_0(z)}\left[\frac{d^2 S}{dz^2} + \frac{1}{H}\frac{dS}{dz} - k^2\left(1 - \frac{g}{\sigma^2 H}\right)S\right] . \quad (60)$$

and N and M^2 are the same as in Equations 40 and 41. The solution of Equation 59 can be written:

$$X(k, z) = e^{n_1 z}\left[C_1 - \frac{1}{2\mu}\int_0^z F(z', k)\, e^{-n_1 z'}\, dz'\right] + e^{n_2 z}\left[C_2 + \frac{1}{2\mu}\int_0^z F(z', k)\, e^{-n_2 z'}\, dz'\right] , \quad (61)$$

where $F(z', k)$ is given by Equation 60 and n_1 and n_2 are the roots of the characteristic equation

$$n^2 - 2Nn + M^2 = 0 . \quad (62)$$

We assume

$$n_1 = N - \mu , \quad (63)$$

$$n_2 = N + \mu , \quad (64)$$

where

$$\left.\begin{aligned} \mu &= \sqrt{N^2 - M^2} \geq 0 \qquad \text{for } N^2 \geq M^2 , \\ \mu &= i\eta, \quad \eta^2 = M^2 - N^2 > 0 \qquad \text{for } N^2 < M^2 \end{aligned}\right\} . \quad (65)$$

The integration constants C_1 and C_2 are determined by the following two boundary conditions:

1. The vertical component of velocity vanishes at the ground,

$$w(x, z = 0) = 0 . \quad (66)$$

2. The kinetic energy of the waves at $z = \infty$ is either zero or remains finite. In the latter case, the vertical component of the disturbance should never be propagated downward at $z = \infty$.

The pressure change at the ground can be obtained from $\chi(x, z=0)$, which is given by the inverse Fourier transform of $X(k, z=0)$, i.e.,

$$\chi(x, 0) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-ikx}\, X(k, 0)\, dk , \quad (67)$$

where $X(k, 0)$ is the solution of Equation 59 at $z=0$ satisfying the above conditions. It is given by Equation B15 in Appendix B.

Since $w(x, 0)=0$, then $s(x, 0)=0$ and Equation 36 reduces to

$$p(x, 0) = -\frac{c^2 \rho_s}{i\sigma}\chi(x, 0) , \quad (68)$$

where $\chi(x, 0)$ is given by the inverse Fourier transform of $X(k, 0)$. The evaluations of $p(x, 0)$ are given in Appendix C and the results are shown in Figures 5–10. Various values are given in Tables 1 and 2. The symbols p_r and p_i signify the real and imaginary parts of the cellular mode, respectively.

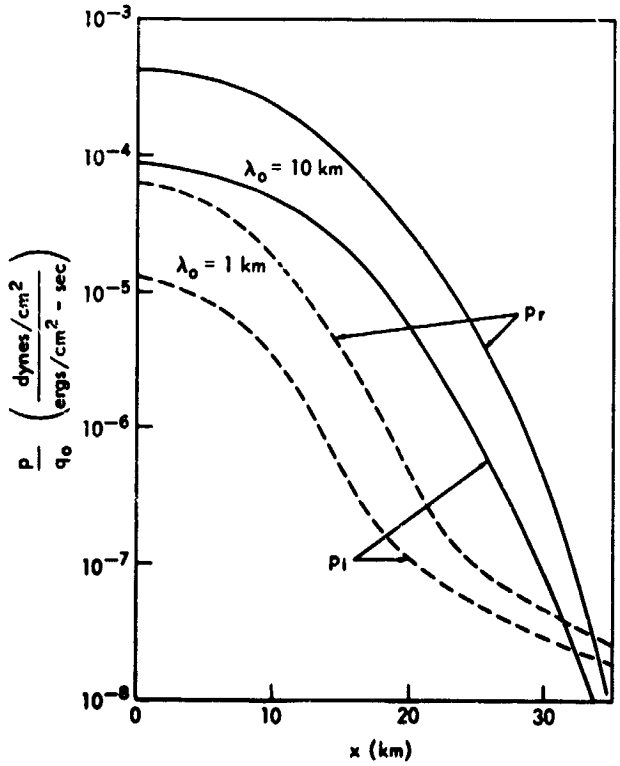


FIGURE 5.—A plot of the intensity of the cellular waves p_r and p_i on the ground, as a function of the input energy flux q_0 for $\tau = 30$ sec and $h = H$, vs. x, the horizontal distance from the center of the source. Full lines and dashed lines correspond to $\lambda_0 = 10$ km and $\lambda_0 = 1$ km, respectively.

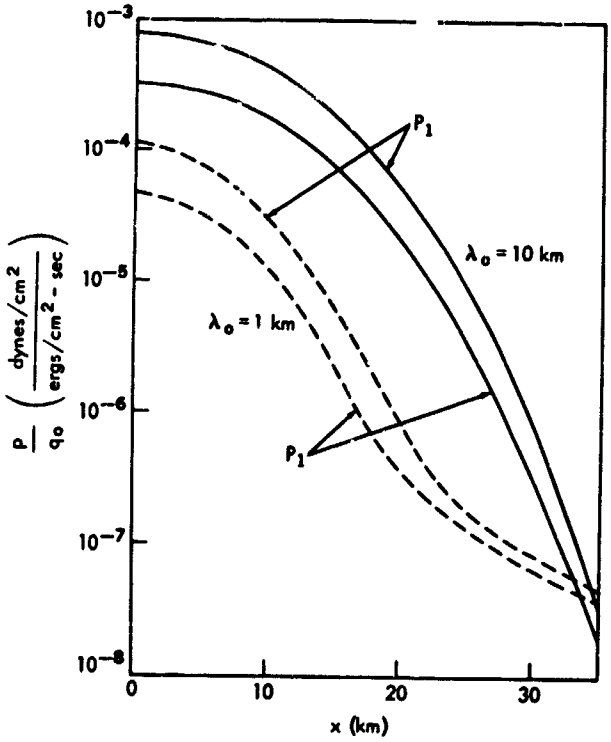


FIGURE 6.—The same plot as Figure 5 except that $h = \frac{1}{2}H$ instead of $h = H$.

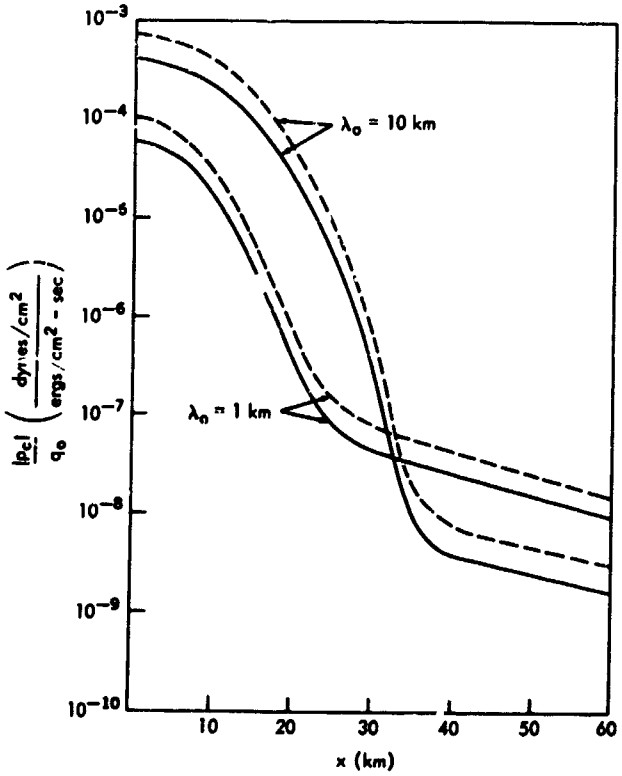


FIGURE 8.—The same plot as Figure 7 except that $\tau = 30$ sec.

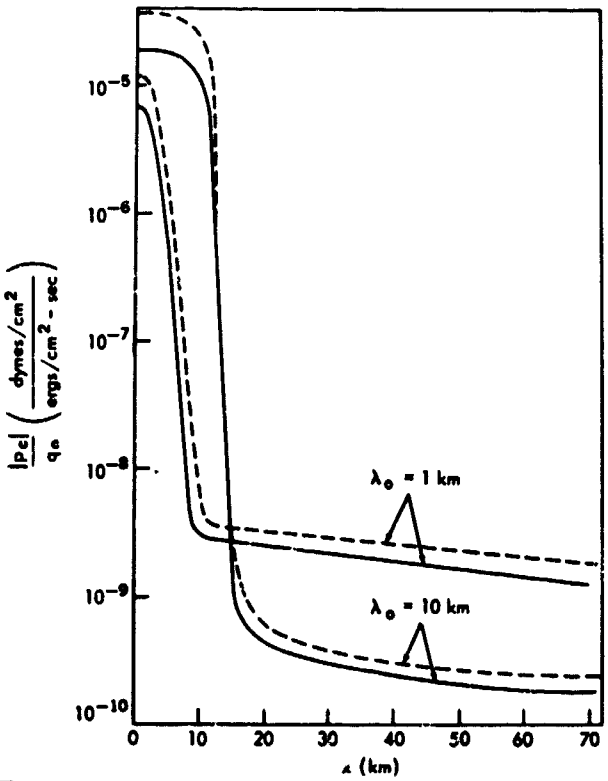


FIGURE 7.—A plot of the intensity of cellular waves produced by periodic auroral heating $|p_e| = \sqrt{p_r^2 + p_t^2}$, vs the horizontal distance from the center of the source for $\tau = 10$ sec. Full lines and dashed lines stand for $h = H$ and $h = \frac{1}{2}H$, respectively.

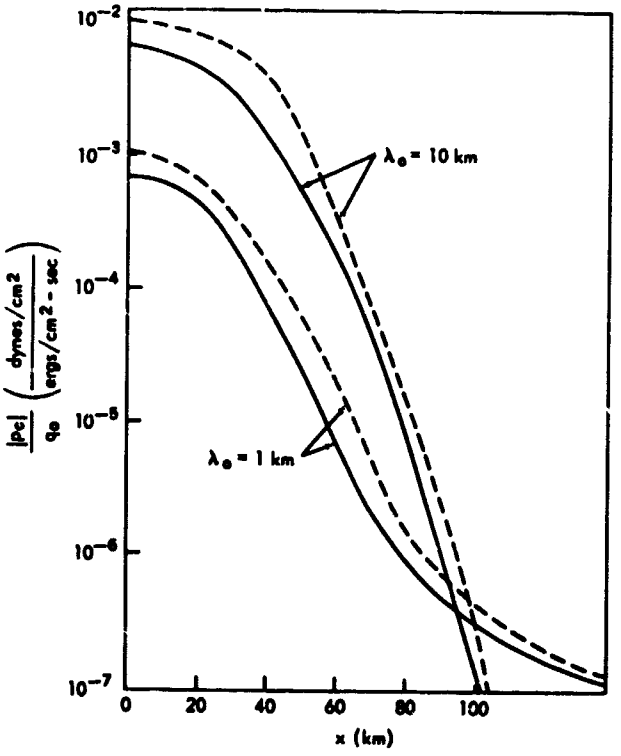


FIGURE 9.—The same plot as Figure 7 except that $\tau = 100$ sec.

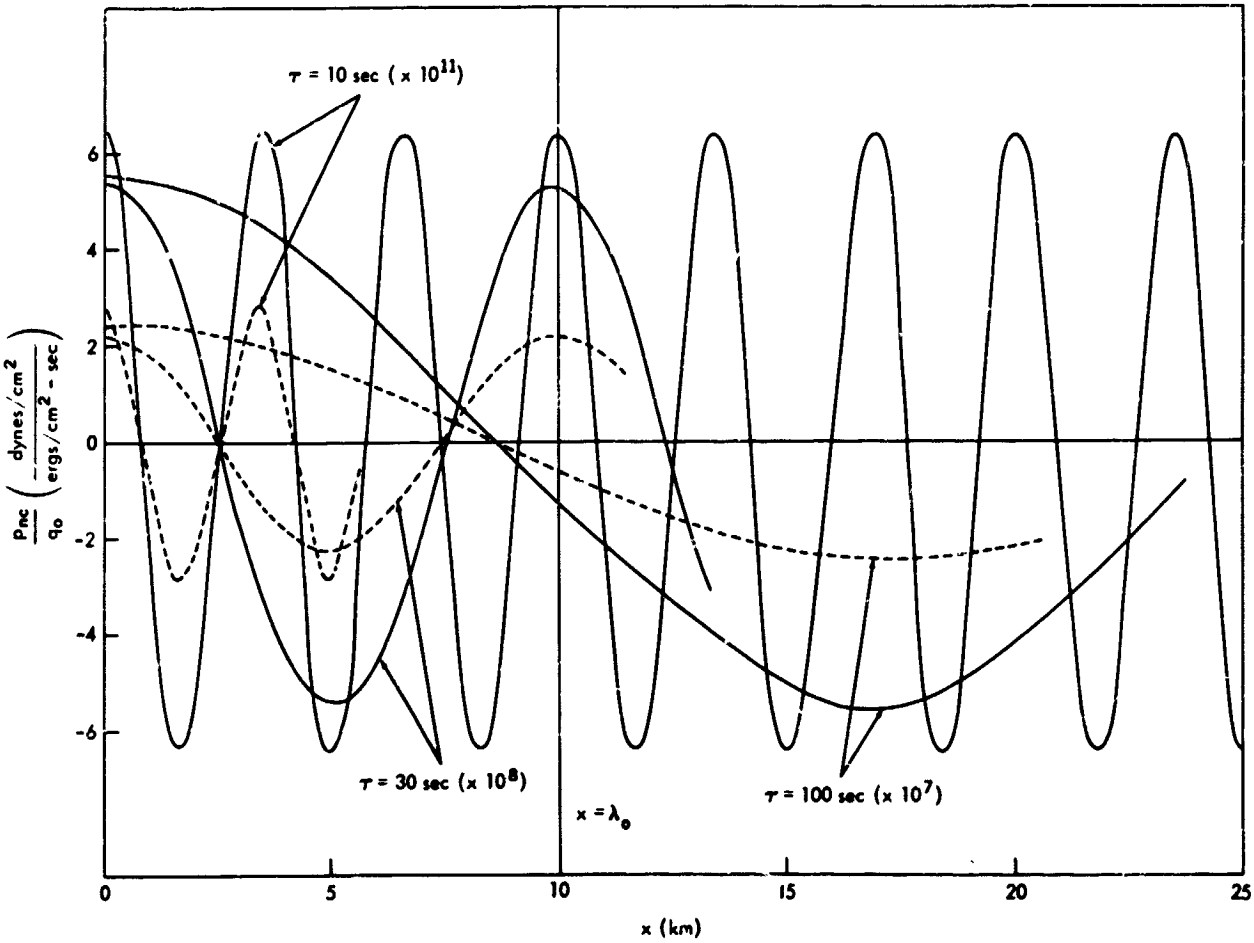


FIGURE 10.—Intensity distribution of noncellular waves on the ground, p_{nc}, in the units of q_0, vs. horizontal distance x from the center of the source of width $\lambda_0 = 10$ km. Notice that the vertical scales are magnified by a factor attached to each curve. Full lines and dashed lines stand for $h = H$ and $h = \frac{1}{2}H$, respectively.

Table 1

Intensity of the Cellular Mode of Infrasonic Waves on the Ground Below the Center of the Source in Units of q_0.

h = H	p_r	p_i	$\lvert p_c \rvert$	h = 1/2H	p_r	p_i	$\lvert p_c \rvert$
λ_0 = 1 km				λ_0 = 1 km			
τ = 10 sec.	6.74×10^{-6}	4.40×10^{-7}	6.75×10^{-6}	τ = 10 sec.	1.33×10^{-5}	1.79×10^{-6}	1.34×10^{-5}
30 sec.	6.15×10^{-5}	1.21×10^{-5}	6.28×10^{-5}	30 sec.	1.10×10^{-4}	4.79×10^{-5}	1.23×10^{-4}
100 sec.	5.46×10^{-4}	4.26×10^{-4}	6.92×10^{-4}	100 sec.	2.77×10^{-4}	1.02×10^{-3}	1.06×10^{-3}
λ_0 = 10 km				λ_0 = 10 km			
τ = 10 sec.	1.82×10^{-5}	1.20×10^{-6}	1.83×10^{-5}	τ = 10 sec.	3.62×10^{-5}	4.80×10^{-6}	3.65×10^{-5}
30 sec.	4.29×10^{-4}	8.57×10^{-5}	4.38×10^{-4}	30 sec.	7.91×10^{-4}	3.32×10^{-4}	8.58×10^{-4}
100 sec.	5.25×10^{-3}	4.11×10^{-3}	6.67×10^{-3}	100 sec.	2.68×10^{-3}	9.85×10^{-3}	1.02×10^{-2}
λ_0 = 100 km				λ_0 = 100 km			
τ = 10 sec.	1.82×10^{-5}	1.20×10^{-6}	1.83×10^{-5}	τ = 10 sec.	3.62×10^{-5}	4.80×10^{-6}	3.65×10^{-5}
30 sec.	4.84×10^{-4}	1.01×10^{-4}	4.94×10^{-4}	30 sec.	8.92×10^{-4}	3.73×10^{-4}	9.68×10^{-4}
100 sec.	1.43×10^{-2}	1.14×10^{-2}	1.83×10^{-2}	100 sec.	7.62×10^{-3}	2.72×10^{-2}	2.83×10^{-2}

Table 2

Values for p_{nc} · $p_{nc}/|p_c|$, and x_0, the Horizontal Distance of the first Nodal Lines from the Center of the Source (horizontal width assumed to be λ_0 = 10 km).

h = H	p_{nc}	$p_{nc}/\|p_c\|$	x_0 (km)	h = 1/2H	p_{nc}	$p_{nc}/\|p_c\|$	x_0 (km)
τ = 10 sec	1.28×10^{-9}	7.0×10^{-5}	0.85	τ = 10 sec	1.11×10^{-9}	3.03×10^{-5}	0.85
30 sec	3.24×10^{-8}	7.4×10^{-5}	2.56	30 sec	2.73×10^{-8}	3.18×10^{-5}	2.56
100 sec	1.10×10^{-5}	1.66×10^{-3}	8.31	100 sec	1.02×10^{-5}	9.8×10^{-4}	8.30

ATTENUATION

Because of the viscosity and thermal conductivity of air, acoustic waves in the atmosphere attenuate. The attenuation coefficient $\alpha(\tau)$, in cm^{-1}, for a wave of period τ is given approximately by (Reference 48)

$$\alpha(\tau) = \frac{4\pi^2}{\tau^2}\frac{1}{c^3}\left(\frac{4}{3}\nu + \frac{\gamma-1}{\gamma}a^2\right), \quad (69)$$

where the coefficient of kinematic viscosity ν is approximately

$$\nu = \frac{1.7 \times 10^{-4}}{\rho(z)} \quad (70)$$

in cm^2/sec, and the coefficient of thermal conductivity a^2 is

$$a^2 = \frac{2.1 \times 10^{-5}}{\rho(z)}, \quad (71)$$

in cm^2/sec.

Since the air density ρ decreases exponentially with height, a^2 and ν increase exponentially with altitude. The so-called attenuation factor,

$$f_a(\tau, z) = \exp\left[-\int_0^z \alpha(\tau, z')\,dz'\right], \quad (72)$$

is shown in Figure 11 as a function of height, for $\tau = 10$ sec, 30 sec, and 100 sec.

It should be noted that the relative amplitude of the pressure wave grows as it propagates upward by a factor $e^{z/2H}$, where H is the scale height, whereas the absolute amplitude decreases by a factor $e^{-z/2H}$, because of the exponential decrease of air density (Reference 49). Therefore, if the amount of excitation energy is the same, the absolute intensity of the pressure wave at the ground increases with the base of the

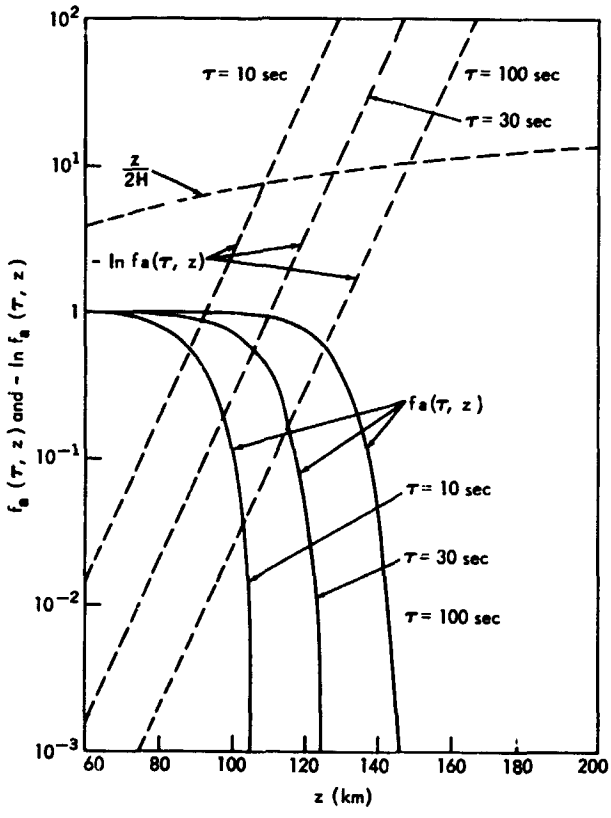


FIGURE 11.—Attenuation factor $f_a(\tau, z)$ vs. altitude of the source z in km and $-\ln f_a(\tau, z)$ vs. z:

$$-\ln f_a(\tau, z) = \int_0^z \alpha(\tau, z')\,dz',$$

a dashed line is also used to indicate the exponential amplification factor $\ln (\rho_z \rho_0)^{\frac{1}{2}} \approx z/2H$, where the scale height H is assumed to be 8 km.

excitation level. This is shown by Equation 29 and Equations C9 and C10 with a factor $e^{z_0/2H}$ for p_s, p_c ($z=0$), and $p_{nc}(z=0)$, and by a curved dash line in Figure 11.

However, because of the steep increase of the kinematic viscosity of air with altitude, the yield of pressure wave excitation drops sharply above a certain altitude for a given period (or frequency)

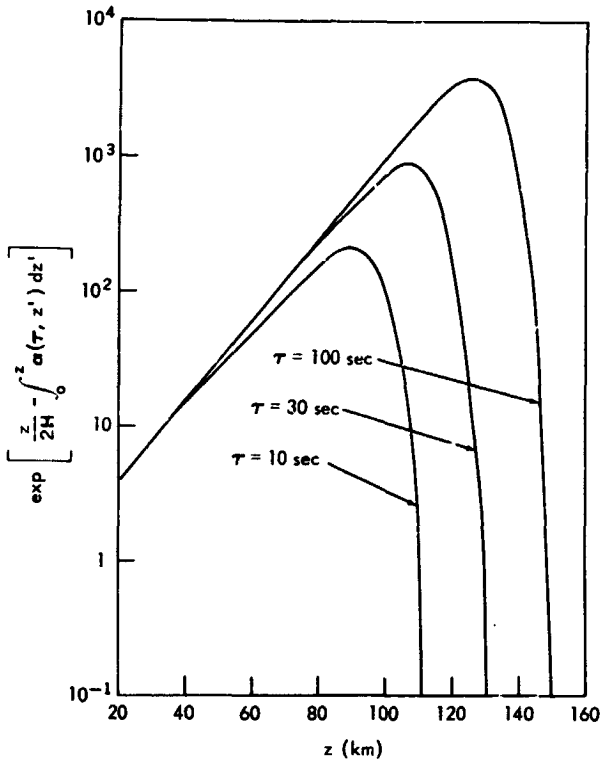


FIGURE 12. —Relative yield of the source to the infrasonic waves in the isothermal atmosphere

$$\sqrt{\frac{\rho(z)}{\rho_0}}\, f_a(\tau, z) \approx \exp\left[\frac{z}{2H} - \int_0^z \alpha(\tau, z')\,dz'\right]$$

vs. the altitude of the source z, where the scale height H is assumed to be 8 km.

of the wave. This is shown in Figure 12 for wave periods of $\tau = 10$, 30, and 100 sec. It can be seen that there is an effective height of excitation of atmospheric acoustic waves for a given period and that this height increases with the period.

As mentioned in the introduction, the excitation of pressure waves by hydromagnetic waves is only effective above the height of auroral activities. Figure 12 shows that waves produced at such a height are greatly attenuated before they reach the ground. Therefore, the contribution of hydromagnetic waves to the excitation of internal atmospheric waves is practically negligible. A more quantitative discussion will be given in the next section.

CONDITIONS FOR WAVE FORMATION

According to previous calculations, a pressure variation at the ground of the order of 1 dyne/cm² can be expected if the maximum rate of heat generation is of the order of 100 ergs/cm²-sec and the layer of periodic heating is around 100 km altitude, with a thickness less than 10 km. Electrons with an energy of the order of 100 kev will lose most of their energy within a layer of the order of 10 km thickness around the height of 100 km (Reference 19, p. 290). According to Chamberlain, the rate of heat generation is the order of 60 ergs/cm²-sec in a bright aurora. Therefore, if the rate of heat generation changes periodically with this order of amplitude, barometric oscillations of the order of 1 dyne/cm² at sea level can be expected from the sources in the upper atmosphere.

However, several other conditions must be satisfied in order that the energy brought into the upper air by auroral electrons can be converted efficiently to pressure waves in the atmosphere. At first, the time τ_1, in which an electron arrives at auroral height from outer space, should be smaller than the period, τ, of the waves concerned. If $\tau_1 \gg \tau$, the phase of the time variation of the source differs from place to place, and the resultant pressure wave originating from those different sources is weakened by superposition.

Since the auroral electrons, whose velocity is of the order of 10^9 cm/sec, lose their energy within a layer of 10 km thickness, $\tau_1 <$ approx. 10^{-4} sec. This is much smaller than the period of the acoustic waves considered here.

As discussed by Hanson and Johnson (Reference 50), electrons impinging into the upper atmosphere lose their energy mostly by inelastic collisions with neutral air particles until ≈ 2 ev, which is the lowest excitation energy of atomic oxygen (^{1}D state). The time τ_2 for ^{1}D excitation collisions with oxygen atoms is of the order of 10^{-3} sec at 100 km and of the order of 1 sec at 400 km.

Below 2 ev, the electrons lose energy in the upper atmosphere mainly by elastic collisions with ambient electrons. The time for those low energy electrons to equilibrate with ambient electrons, τ_3, is obtained from the expression for the rate of energy loss of a fast electron immersed in a thermalized plasma (Reference 50),

$$\tau_3 \approx \frac{5.7 \times 10^3\, E^{3/2}}{N_e}, \qquad (73)$$

where E is the electron energy, i.e., $E \approx 2$ ev, and

N_e is the electron concentration in cm^{-3}. The equilibration time, τ_3, given by this expression is about 10^{-1} sec at 100 km and of the order of 10^{-2} sec above 350 km. Thus $\tau_2+\tau^3$ is much smaller than the periods of the acoustic waves under discussion

It should be noted that elastic collisions with neutral particles dominate over those with ambient electrons below the F2 maximum (around 300 km). The time constant τ_4 for these elastic collisions consists of τ (O) and $\tau(N_2)$, the time constants for the loss of excess electron energy to atomic oxygen and to molecular nitrogen, respectively (Reference 50), and is given by

$$\tau_4 = \frac{2.08 \times 10^{11} E^{-1/2}}{n(O) + 2.36 n(N_2)} , \qquad (74)$$

where n(O) and $n(N_2)$ are the concentration per cm^3 of atomic oxygen and of molecular nitrogen, respectively. For E=2 ev, τ_4 is of the order of 0.1 sec at 100 km and increases with height. It is of the order of 100 sec at 350 km.

As shown in the previous section, the excitation of atmospheric pressure waves above 200 km is not important; the time constant τ_4 does not destroy the condition of wave formation. In other words, below 200 km $\tau_1+\tau_2+\tau_3+\tau_4 <$ approx. τ.

Another condition necessary for wave formation is that the time constant for cooling in a certain domain of auroral activity must be much longer than the period of oscillation. If the initial temperature T_0 is assumed to be horizontally uniform within the domain of the source, $-\lambda_0 \leq x \leq \lambda_0$, then the temperature at the center after t(sec) is

$$T(t) = (T_0 - T_a)\, \Phi\left(\frac{\lambda_0}{2\sqrt{a^2 t}}\right), \qquad (75)$$

where

$$\Phi(y) = \frac{2}{\sqrt{\pi}} \int_0^y e^{-x^2}\, dx , \qquad (76)$$

and a^2 is the thermal diffusivity (coefficient of thermal conduction) of air. T_a is the temperature outside the source.

The time, t_a, necessary to reduce the initial temperature difference between the inside and outside of the source region can be estimated from

$$\left.\begin{aligned} \Phi\left(\frac{\lambda_0}{2\sqrt{a^2 t_a}}\right) &= \frac{1}{2} , \\ t_a &\approx \frac{\lambda_0^2}{a^2} . \end{aligned}\right\} \qquad (77)$$

Since the thermal diffusivity of air a^2 at a height of 100 km is of the order of 10^6 cm^2/sec (see Equation 71), the minimum source width necessary to satisfy the condition for thermal oscillation is of the order of 30m for a period of 10 sec, and of the order of 100m for a period of 100 sec. The source width considered in the present calculation is significantly larger than these widths. In other words, the time constant for source cooling is sufficiently long for pressure wave production.

OTHER POSSIBLE MECHANISMS

Now the possibility of other suitable mechanisms for the generation of microbarometric oscillations during magnetic disturbances will be considered.

Periodic Heating Due to the Absorption of Hydromagnetic Waves

As stated in the introduction, hydromagnetic waves from the magnetosphere lose part of their energy in the lower exosphere and in the ionosphere. The remaining energy, which leaks through the ionosphere, is observed as fluctuations of the geomagnetic field intensity or geomagnetic pulsations.

The rate of energy dissipation of hydromagnetic waves in the ionosphere increases with increasing frequency (References 6, 51, and 52). According to Watanabe, the dissipation of the hydromagnetic wave energy in the ionosphere is negligible for waves with periods longer than 20 sec. In the auroral region, the amplitude of geomagnetic pulsations sometimes exceeds several tens of gammas (i.e., giant pulsations) and the periods are sometimes longer than several tens of seconds. The intensity of the incident wave is usually smaller for higher frequencies (Reference 15). As shown by the power density of the small scale fluctuations of magnetic field intensity observed

in the earth's magnetosphere between 5 and 15 earth radii (Reference 3), the frequency spectrum of hydromagnetic waves, which are regarded as the origins of geomagnetic pulsations observed at the earth's surface, is a decreasing function of frequency above the ionosphere. In any case, the upper limit of the incident wave amplitude may be taken as 100γ. The energy flux associated with these hydromagnetic waves is then of the order of several ergs/cm^2-sec^1, under the assumption that the Alfvèn wave velocity above the ionosphere is of the order of 10^8 cm/sec. The rate of heat generation by absorption of these hydromagnetic waves in the upper ionosphere has been estimated by several authors and their results are shown in Table 3.

The rate of heat generation is generally less than 1 erg/cm^2-sec, except when the period is 1 sec. Therefore the amount of heat generated by hydromagnetic waves penetrating the ionospheer is smaller than that due to the auroral particles. Table 3 also contains estimates of the thickness of the layer of heat generation. In every case the thickness is much larger for hydromagnetic waves than for auroral particles.

As shown in Figure 12, the effective height of pressure wave generation by periodic heating of the upper air is limited, and it is lower for the shorter period. Therefore, generation of acoustic waves by the attenuation of hydromagnetic waves is less efficient than generation by the periodic heating due to auroral particles, even if the energy flux of the incident wave is increased to be the same as the latter.

Pressure Waves Due to the Impacts of Auroral Particles

A particle coming into the atmosphere loses its energy by transferring its downward momentum to the air particles. A pressure wave can be generated if the flux of the particle changes periodically with time.

The upper limit of the pressure intensity, from this process, can be estimated by assuming that all the incoming particles would stop in a very short time within a very thin layer. By assuming the average energy of incident electrons is 6 kev and the maximum flux is of the order of 10^{10}cm^{-2}-sec^{-1}, the maximum pressure exerted upon the thin layer of virtual shock absorber is estimated to be of the order of 4.10^{-8} dynes/cm^2. The corresponding intensity at the earth's surface is of the order of 10^{-5} dyne/cm^2, if all electrons stop at an altitude of 100 km.

Penetration of Hydromagnetic Waves Through the Ionosphere

As mentioned above, most geomagnetic pulsations are due to hydromagnetic oscillations in the exosphere. These oscillations are related to electromagnetic oscillations in the space between the earth's surface and the lower boundary of the ionosphere. Any oscillation mode in which the compression of atmospheric matter is involved gives rise to a variation in the density, and consequently a pressure variation. Therefore, it may be possible that geomagnetic pulsations and microbarometric oscillations have the same origin,

Table 3

Rate of Heat Generation Due to Attenuation of Hydromagnetic Waves with a 100γ Incident Amplitude.

Source	Period of HM waves (sec)	Rate of Heat Generation in an Air Column of unit cross section (erg/cm^2-sec)	Thickness of Heating Layer (km)	Altitude of the Center of the Heat Generating Layer (km)
Dessler (Reference 53)	1	1.3	100	170
Francis and Karplus (Reference 51)	6.3 0.3	0.7 1.6	50 100	125 180
Akasofu (Reference 52)	1 10 100	9.7 0.17 0.028	100 200 300	225 240 250

i.e., hydromagnetic oscillations of the earth's exosphere.

A simple situation can be found at the geomagnetic equator, where geomagnetic pulsations appear sometimes in the N–S component of the geomagnetic field, with amplitudes of the order of several tens of gammas and periods of several minutes (Reference 54). The origins of this kind of pulsation presumably exist in the outer boundary of earth's magnetosphere near the geomagnetic equatorial plane, and it may be propagated as a modified Alfvèn wave. A modified Alfvèn wave is a transverse wave with respect to changes in the electric and magnetic fields. On the other hand, it is a longitudinal wave when viewed as fluid motion (Reference 55). An electromagnetic wave as well as a pressure wave should be in the space between the earth's surface and the lower boundary of the ionosphere. The energy flux associated with an incident modified Alfvèn wave is roughly $(B^2/8\pi)\ V_A$, where V_A is the group velocity of the modified Alfvèn waves and can be taken as the Alfvèn wave velocity at a higher portion of the ionosphere. B is the amplitude of the wave, i.e., the intensity of magnetic fluctuation above the ionosphere.

By assuming conservation of energy flux, the upper limit of the amplitude of the pressure wave at the earth's surface, p_e, can be estimated as follows:

$$\frac{1}{2}\frac{p_e^2}{c_s \rho_s} = \frac{B^2}{8\pi} V_A ,$$

where c_s is the sound wave velocity at the earth's surface, and ρ_s is the air density at the earth's surface. By taking $V_A = 3.10^7$ cm/sec (corresponding to daytime at the sunspot maximum activity) $B = 30\gamma$, $c_s = 3.10^4$ cm/sec, and $\rho_s = 1.25 \times 10^{-3}$ gm/cm^3, we find $p_e \approx 3$ dynes/cm^2. In this estimate the conversion factor between the energy of incident Alfvèn waves and that of secondary pressure waves is assumed to be unity. This factor must actually be very small because of reflections and energy dissipations of incident waves at the upper part of the ionosphere. If it is not small, infrasonic waves could appear in the equatorial region during strong magnetic disturbances. Since the occurrence of aurorae in these regions is negligible, this might provide a method of direct detection of modified Alfvèn waves coming into the earth's atmosphere from the magnetosphere.

CONCLUSIONS

We have shown that one of the most plausible mechanisms for pressure wave generation during geomagnetic disturbances is the periodic heating of the polar ionosphere by auroral particles, observed as pulsating aurorae. As emphasized by Campbell (Reference 56) the main energy source for this type of auroral activity is not necessarily incident auroral particles, but a flow of secondary electrons called the electro-jet. In this respect periodic heating by these intermittent electric currents is essentially the same as the so-called Joule heating discussed by Cole (References 57 and 58).

The following conclusions can be drawn from the present calculation:

1. From Figures 3, 5–9, and 13 it can be seen than an incident energy flux of more than 100 ergs/cm^2-sec will produce acoustic waves observable at the ground, provided the periods are longer than about 10 seconds.
2. The relative intensity of the pressure wave at the ground is higher when the heating is concentrated within a thin layer than when it is distributed over a wide range of altitudes (Figures 3, 5, and 6).
3. The intensity ratio between the inside of the source region and the outside is smaller for long periods than for short periods, as expected.
4. The gradient of intensity around the boundary of the source is steeper when the width of the source is wide.
5. Inside the source the ratio of noncellular to cellular wave intensity is of the order of 10^{-5} for $\tau = 10$ sec and 10^{-3} for $\tau = 100$ sec. Although the noncellular intensity exceeds the cellular at large distances from the source, the contribution of the noncellular wave to the observed intensity is negligible, because both waves attenuate in long distance propagation.
6. According to the present calculations, which are based on an isothermal atmosphere, the intensity of acoustic waves more than several hundred km from the region of

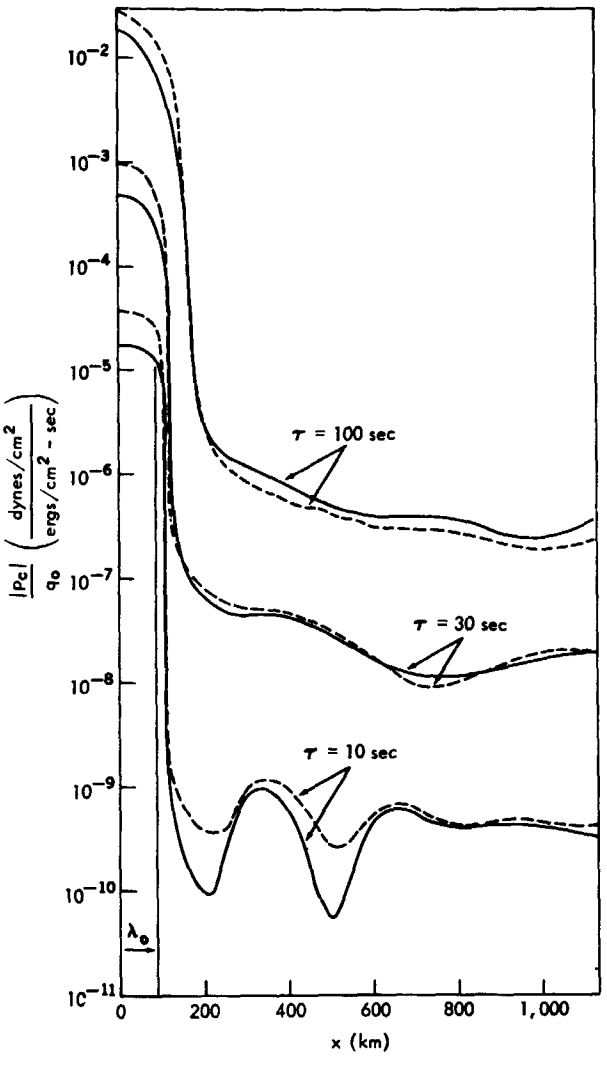


FIGURE 13.—Intensities of cellular waves on the ground in the units of q_0 vs. the horizontal distance x from the center of the source, of which the width λ_0 is 100 km, for τ=10 sec, 30 sec and 100 sec. Full lines and dashed lines stand for h=H and h=½H, respectively.

auroral activity is negligible. For the real atmosphere, therefore, the horizontal propagation of acoustic waves through the ducts around the mesopause and the stratopause is rather important in explaining the diurnal variation of the arrival direction of these waves during the period of high geomagnetic activity, as shown in Figure 1.

7. It should be noted that auroral activities are not necessarily within the auroral zone but, rather, extend toward lower latitudes (nearly to 50°N), when these sonic waves are observed during the periods of magnetic disturbances (as can be seen from the auroral visoplot shown in Figure 2).

To show the position and structure of sonic ducts as well as attenuation of the waves requires the actual atmospheric temperature distributions. This is discussed elsewhere (Reference 59).

Finally, it should be noted that the energy flux of acoustic waves at the ground S in ergs/cm²-sec is

$$S = Ec \quad , \tag{78}$$

where

$$E = \frac{p_s^2}{2\rho_s c^2} \quad ; \tag{79}$$

p_s and ρ_s are the maximum aplitude of pressure change in dynes/cm² and the static density of air atat sea level, and c is the sound velocity in cm/sec.

Since ρ_s is of the order of 1.25×10^{-3} gm/cm³, the energy flux corresponding to $p_s=1$ dyne/cm² is approximately 1.4×10^{-2} erg/cm²-sec. To show the energy relation between the input power and the observed output intensity, as given in Figures 3, 5–10, and 13, the above relation must be used.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to Dr. J. M. Young and his co-workers at the National Bureau of Standards in Washington, D. C., who gave detailed information on their observations of infrasonic waves during magnetic disturbances. Thanks are also due to Dr. G. D. Mead for his comments, and Mr. E. Monasterski for his help on the computer calculations.

REFERENCES

1. CHRZANOWSKI, P., GREEN, G., LEMMON, K. T., and YOUNG, J. M., "Traveling Pressure Waves Associated with Geomagnetic Activity," *J. Geophys. Res.* 66(11):3727–3733, November 1961.
2. CAMPBELL, W. H., "Natural Electromagnetic Energy Below the ELF Range," *Nat. Bur. Stand. J. Res.* 64D(4):409–411, July–August 1960.
3. SONETT, C. P., JUDGE, D. L., SIMS, A. R., and KELSO, J. M., "A Radial Rocket Survey of the Distant Geomagnetic Field," *K. Geophys. Res.* 65(1):55–68, January 1960.
4. HEPPNER, J. P., NESS, N. F., SCEARCE, C. S., and SKILLMAN, T. L., "Explorer 10 Magnetic Field Measurements," *J. Geophys. Res.* 68(1): 1–46, January 1, 1963.

5. HEPPNER, J. P., NESS, N. F., SKILLMAN, T. L., and SCEARCE, C. S., "Magnetic Field Measurements with the Explorer X Satellite," in: *Goddard Space Flight Center Contributions to the 1961 Kyoto Conf. on Cosmic Rays and the Earth Storm*, ed. by F. B. McDonald, NASA Technical Note D-1061, June 1962, pp. 1-20.
6. WATANABE, T., "Electrodynamical Behaviour and Screening Effect of the Ionosphere," *Sci. Rept. Tohoku Univ.* Ser. 5, 9:81-98, 1957.
7. WATANABE, T., "Law of Electric Conduction for Waves in the Ionosphere, *J. Atmos. Terres. Phys.* 24:117-125, February 1962.
8. PIDDINGTON, J. H., "The Transmission of Geomagnetic Disturbances Through the Atmosphere and Interplanetary Space," *Geophys. J.* 2(3):173-189, September 1959.
9. SCHOLTE, J. G., and VELDKAP, J., "Geomagnetic and Geoelectric Variations," *J. Atmos. Terres. Phys.* 6(1):33-45, January 1955.
10. DUNGEY, J. W., "The Propagation of Alfven Waves Through the Ionosphere," Penn. State Univ. Sci. Rept. 57, 1954.
11. KATO, Y., and WATANABE, T., "Further Study on the Cause of Giant Pulsations," *Sci. Rept. Tohoku Univ.* Ser. 5, 8:1-10, 1956.
12. OBAYASHI, T., and JACOBS, J. A., "Geomagnetic Pulsations and the Earth's Outer Atmosphere," *Geophys. J.* 1(1):53-63, March 1958.
13. OBAYASHI, T., "Geomagnetic Storms and the Earth's Outer Atmosphere," *Rept. Ionosphere Res. Japan* 12(3):301-335, September 1958.
14. DESSLER, A. J., "The Propagation Velocity of World-Wide Sudden Commencements of Magnetic Storms," *J. Geophys. Res.* 63(2):405-408, June 1958.
15. JACOBS, J. A., and WATANABE, T., "Propagation of Hydromagnetic Waves in the Lower Exosphere and the Origin of Short Period Geomagnetic Pulsations," *J. Atmos. Terres. Phys.* 24:413-434, June 1962.
16. HEPPNER, J. P., "A Study of Relationships Between the Aurora Borealis and the Geomagnetic Disturbances Caused By Electric Currents in the Ionosphere," Canad. Defense Res. Board Rept. DR-135, 1958.
17. CAMPBELL, W. H., and REES, M. H., "A Study of Auroral Coruscations," *J. Geophys. Res.* 66(1):41-55, January 1961.
18. MCILWAIN, C. E., "Direct Measurement of Particles Producing Visible Auroras," *J. Geophys. Res.* 65(9): 2727-2747, September 1960.
19. CHAMBERLAIN, J. W., "Physics of the Aurora and Airglow," New York: Academic Press, 1961.
20. LAMB, H. "Hydrodynamics," 6th ed., Cambridge, England: The University Press, 1932.
21. BJERKNES, V. F. K., BJERKNES, J., SOLBERG, H., and BERGERON, T., "Physikalische Hydrodynamik," Berlin: J. Springer, 1933.
22. WILKES, M. V., "Oscillations of the Earth's Atmosphere," Cambridge, England: The University Press, 1949.
23. WILKES, M. V., "The Thermal Excitation of Atmospheric Oscillations," *Roy. Soc. Proc.* A207:358-370, July 6, 1951.
24. SAWADA, R., "On the Role of Thermal Excitation in the Atmospheric Tides," *Geophys. Mag.* 26(4):267-281, August 1955.
25. TAYLOR, G. I., "Waves and Tides in the Atmosphere," *Roy. Soc. Proc.* A126:169-183, December 2, 1929.
26. PEKERIS, C. L., "The Propagation of a Pulse in the Atmosphere," *Roy. Soc. Proc.* A171:434-449, July 7, 1939.
27. PEKERIS, C. L., "The Propagation of a Pulse in the Atmosphere. II," *Phys. Rev.* 73:145-154, January 15, 1948.
28. [illegible]ORER, R. S., "The Dispersion of a Pressure Pulse in the Atmosphere," *Roy. Soc. Proc.* A201:175-186, March 22, 1950.
29. GOSSARD, E. E., and MUNK, W., "On Gravity Waves in the Atmosphere," *J. Met.* 11:259-269, 1954.
30. GOSSARD, E. E., "Vertical Flux of Energy into the Lower Ionosphere from Internal Gravity Waves Generated in the Troposphere," *J. Geophys. Res.* 67(2):745-757, February 1962.
31. YAMAMOTO, R., "Study of the Microbarographic Waves (I), Theory of the Microbarographic Waves (I)," *Met. Soc. Japan J.* 34:235-243, December 1956.
32. YAMAMOTO, R., "A Study of Microbarographic Waves (II), Theory of the Microbarographic Waves (II) (continued), *Met. Soc. Japan J.* 35:321-326, 1957.
33. PRESS, F., and HARKRIDER, D., "Propagation of Acoustic-Gravity Waves in the Atmosphere," *J. Geophys. Res.* 67(10):3889-3908, September 1962.
34. WESTON, V. H., "The Pressure Pulse Produced by a Large Explosion in the Atmosphere," *Canad. J. Phys.* 39(7):993-1009, July 1961.
35. WESTON, V. H., "The Pressure Pulse Produced by a Large Explosion in the Atmosphere. II," *Canad. J. Phys.* 40(4):431-445, April 1962.
36. WESTON, V. H., "Gravity and Acoustical Waves," *Canad. J. Phys.* 40(4):446-453, April 1962.
37. MARTYN, D. F., "Cellular Atmospheric Waves in the Ionosphere and Troposphere," *Roy. Soc. Proc.* A201:216-234, March 22, 1950.
38. SEN, H. K., and WHITE, M. L., "Thermal and Gravitational Excitation of Atmospheric Oscillations," *J. Geophys. Res.* 60(4):483-495, December 1955.
39. WHITE, M. L., "Gravitational and Thermal Oscillations in the Earth's Upper Atmosphere," *J. Geophys. Res.* 61(3):489-499, September 1956.
40. WHITE, M. L., "Thermal and Gravitational Atmospheric Oscillations—Ionospheric Dynamo Effects Included," *J. Atmos. Terres. Phys.* 17(3):220-245, February 1960.
41. HINES, C. O., "Hydromagnetic Resonance in Ionospheric Waves," *J. Atmos. Terres. Phys.* 7(1-2): 14-30, August 1955.
42. HINES, C. O., "Internal Atmospheric Gravity Waves at Ionospheric Heights," *Canad. J. Phys.* 38(11): 1441-1481, November 1960.

43. ECKART, C., "The Thermodynamics of Irreversible Processes. I. The Simple Fluid," *Phys. Rev.* 58(3): 267–269, August 1, 1940.
44. ECKART, C., "The Thermodynamics of Irreversible Processes. II. Fluid Mixtures," *Phys. Rev.* 58(3): 269–275, August 1, 1940.
45. ECKART, C., and FERRIS, H. G., "Equations of Motion of the Ocean and Atmosphere," *Rev. Mod. Phys.* 28(1):48–52, January 1956.
46. ECKART, C., "Hydrodynamics of Oceans and Atmospheres," New York: Pergamon Press, 1960.
47. TOLSTOY, I., "The Theory of Waves in Stratified Fluids Including the Effects of Gravity and Rotation," *Rev. Mod. Phys.* 35(1):207–230, January 1963.
48. RAYLEIGH, J. W. S., "The Theory of Sound, Vol. II, Cambridge Trans," 1929, p. 315.
49. SCHRÖDINGER, E., "Zur Akustik der Atmosphäre," *Phys. Zeits.* 18:445–483, October 1, 1917.
50. HANSON, W. B., and JOHNSON, F. S., "Electron Temperature in the Ionosphere," *Mem. Roy. Soc. Sci. Liege* 4:390–423, 1960.
51. FRANCIS, W. E., and KARPLUS, R., "Hydromagnetic Waves in the Ionospher," *J. Geophys. Res.* 65(11): 3593–3600, November 1960.
52. AKASOFU, S., "On the Ionospheric Heating by Hydromagnetic Waves Connected with Geomagnetic Micropulsations," *J. Atmos. Terres. Phys.* 18(2–3): 160–173, June 1960.
53. DESSLER, A. J., "Ionospheric Heating by Hydromagnetic Waves," *J. Geophys, Res.* 64(4):397–401, April 1959.
54. MATSUURA, N., and NAGATA, T., "Turbulence in the Upper Atmosphere," *Rept. Ionosphere Res. Japan,* 12:147–149, 1958.
55. VAN DE HULST, H. C., "Interstellar Polarization and Magneto-Hydrodynamic Waves," in: *Problems of Cosmical Aerodynamics, Proceedings of the Symposium on the Motion of Gaseous Masses of Cosmical Dimensions,* Paris: August 16–18, 1949, Dayton, Ohio: Central Air Documents Office, 1951, pp. 45–56.
56. CAMPBELL, W. H., "Theory of Geomagnetic Micropulsations," AGU annual meeting at Washington, D. C., April 28, 1962.
57. COLE, K. D., "A Source of Energy for the Ionosphere," *Nature* 194:75, April 7, 1962.
58. COLE, K. D., "Atmospheric Blow-Up at the Auroral Zone," *Nature* 194:761, May 26, 1962.
59. MAEDA, K., "Acoustic Heating of the Polar Night Mesosphere," NASA Technical Note D-1912, November 1963.

Appendix A

DERIVATION OF DIFFERENTIAL EQUATION 37

By differentiating Equations 30 and 31 with respect to t and using Equation 33

$$\frac{\partial^2 u}{\partial t^2} = \frac{\partial}{\partial x}\left(c^2 \chi - gw\right) - \frac{1}{\rho_0}\frac{\partial s}{\partial x} \quad . \tag{A1}$$

$$\frac{\partial^2 w}{\partial t^2} = \frac{\partial}{\partial z}\left(c^2 \chi - gw\right) - \left[(\gamma - 1)\, g + \frac{dc^2}{dz}\right]\chi - \frac{1}{\rho_0}\frac{\partial s}{\partial z} \quad . \tag{A2}$$

When the time dependences of u, w, p, ρ, and s are proportional to a common factor $e^{i\delta t}$, a straightforward calculation leads to Equations 34, 35, and 36 from Equations A1, A2, and 33, respectively.

Differentiating Equations A1 and A2 with respect to x and z, respectively, and adding them gives

$$\frac{\partial^2 \chi}{\partial t^2} = c^2 \nabla^2 \chi - g\frac{\partial \zeta}{\partial x} + \left(\frac{dc^2}{dz} - \gamma g\right)\frac{\partial \chi}{\partial z} - \left(\frac{dc^2}{dz} + \gamma g\right)\frac{1}{\rho_0 c^2}\frac{\partial s}{\partial x} - \frac{1}{\rho_0}\nabla^2 s \quad . \tag{A3}$$

where

$$\zeta = \frac{\partial w}{\partial x} - \frac{\partial u}{\partial z} \quad . \tag{A4}$$

Similarly, by differentiating Equations A1 and A2 with respect to z and x, respectively, and subtracting the former from the latter,

$$\frac{\partial^2 \zeta}{\partial t^2} = -\left[\frac{dc^2}{dz} + (\gamma - 1)\, g\right]\frac{\partial \chi}{\partial x} + \frac{1}{\rho_0 c^2}\left(\frac{dc^2}{dz} + \gamma g\right)\frac{\partial s}{\partial x} \quad . \tag{A5}$$

The elimination of ζ from Equations A3 and A5 results in

$$\frac{\partial^4 \chi}{\partial t^4} = c^2 \nabla^2 \frac{\partial^2 \chi}{\partial t^2} + g\left(\frac{dc^2}{dz} + (\gamma - 1)\, g\right)\frac{\partial^2 \chi}{\partial x^2} - g\frac{\partial}{\partial x}\left[\frac{1}{\rho_0 c^2}\left(\frac{dc^2}{dz} + \gamma g\right)\frac{\partial s}{\partial x}\right]$$

$$+ \left(\frac{dc^2}{dz} - \gamma g\right)\frac{\partial}{\partial z}\left(\frac{\partial^2 \chi}{\partial t^2}\right) - \left(\frac{dc^2}{dz} + \gamma g\right)\frac{1}{\rho_0 c^2}\frac{\partial}{\partial z}\left(\frac{\partial^2 s}{\partial t^2}\right) - \frac{1}{\rho_0}\nabla^2\left(\frac{\partial^2 s}{\partial t^2}\right) \quad . \tag{A6}$$

Since it is assumed that u, w, p, ρ, and s are proportional to $e^{i\delta t}$, the elimination of the time derivative from Equation A6 gives Equation 37.

Appendix B

SOLUTION OF EQUATION 59 AT z=0

Applying the Fourier transform to Equations 34 and 35 gives

$$-\sigma^2 U = ikc^2 X - ikgW - \frac{ik}{\rho_0} S \quad , \tag{B1}$$

$$-\sigma^2 W = c^2 \frac{dX}{dz} - \gamma g X + ikgU - \frac{ik}{\rho_0} \frac{dS}{dz} \quad . \tag{B2}$$

where U and W are the Fourier transforms of u and w. Solving these bilateral equations with respect to U and W yields

$$\left(\sigma^4 - k^2 g^2\right) U = -ikgc^2 \frac{dX}{dz} + ik\left(\gamma g^2 - \sigma^2 c^2\right) X + \frac{ik\sigma^2}{\rho_0(z)} S + \frac{kg}{\rho_0(z)} \frac{dS}{dz} \quad , \tag{B3}$$

$$\left(\sigma^4 - k^2 g^2\right) W = -c^2\sigma^2 \frac{dX}{dz} + g\left(\gamma\sigma^2 - k^2 c^2\right) X + \frac{k^2 g}{\rho_0(z)} S + \frac{\sigma^2}{\rho_0(z)} \frac{dS}{dz} \quad . \tag{B4}$$

Since $W(z=0)=0$ and $dS/dz|_{z=0}=S(z=0)=0$, from the above equations

$$\left.\frac{dX}{dz}\right|_{z=0} = m(\sigma) X(0) \quad , \tag{B5}$$

where

$$m(\sigma) = g\left(\frac{\gamma}{c^2} - \frac{k^2}{\sigma^2}\right) \quad . \tag{B6}$$

Similarly, the Fourier transform of Equation 33 is

$$i\sigma P(k, z) = g\rho_0(z) W - c^2 \rho_0(z) X(k, z) + S(k, z) \quad , \tag{B7}$$

where

$$P(k, z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ikx} p(x, z)\, dx \quad .$$

SOLUTION FOR X[z=0], DIRECT CALCULATION

From Equation 61

$$\frac{dX}{dz} = n_1 e^{n_1 z}\left[C_1 - \frac{1}{2\mu} \int_0^z F(z') e^{-n_1 z'} dz'\right] + n_2 e^{n_2 z}\left[C_2 + \frac{1}{2\mu} \int_0^z F(z') e^{-n_2 z'} dz'\right] \quad . \tag{B8}$$

Using Equation B5, from Equation B8 we get

$$C_1 = -\frac{n_2 - m(\sigma)}{n_1 - m(\sigma)} C_2 \quad . \tag{B9}$$

If we write

$$A_1 = C_1 - \frac{1}{2\mu} \int_0^z F(z') e^{-n_1 z'} dz'$$

and

$$A_2 = C_2 + \frac{1}{2\mu} \int_0^z F(z') e^{-n_2 z'} dz'$$

then from Equations B8 and B9, by assuming $h<2H$,

$$A_1 e^{-\mu z} = \begin{cases} C_1 e^{-\mu z} & \text{for } 0 \le z \le z_0 \ , \\ \left[C_1 + Q_0 \frac{\sigma^4 - k^2 g^2}{2\mu c^2 \sigma^2} \frac{1}{n_1 + \beta} C^{-(n_1+\beta)z_0} \right] e^{-\mu z} + Q_0 \frac{\frac{\beta}{h} + \frac{k^2 g}{\sigma^2 H} - k^2}{2\mu\, n_1 + \beta} C^{-\left[(1/h)-(1/2H)\right]z} & \text{for } z_0 \le z < \infty \ , \end{cases} \quad \text{(B10)}$$

and

$$A_2 e^{+\mu z} = \begin{cases} C_2 e^{\mu z} & \text{for } 0 \le z \le z_0 \ , \\ \left[C_2 - Q_0 \frac{\sigma^4 - k^2 g^2}{2\mu c^2 \sigma^2} \frac{1}{n_2 + \beta} e^{-(n_2+\beta)z_0} \right] e^{\mu z} - Q_0 \frac{\frac{\beta}{h} + \frac{k^2 g}{\sigma^2 H} - k^2}{2\mu\, n_2 + \beta} e^{-\left[(1/h)-(1/2H)\right]z} & \text{for } z_0 \le z < \infty \ , \end{cases} \quad \text{(B11)}$$

where

$$Q_0 = \sqrt{\frac{2}{\pi}} \frac{\sin \lambda_0 k}{k} \frac{(\gamma - 1) q_0}{\rho_s h c^2} e^{z_0/h} \ . \quad \text{(B12)}$$

and

$$\beta = \frac{1}{h} - \frac{1}{H} \ . \quad \text{(B13)}$$

The term including $e^{\mu z}$ represents the waves whose amplitudes increase with height. The corresponding kinetic energies diverge to infinity provided that $N^2 > M^2$. On the other hand, if $N^2 < M^2$ this term represents inward-going waves.

Because of the boundary conditions listed in the section "Mathematical Treatment," the terms including $e^{\mu z}$ must vanish. Therefore

$$C_2 = C_1 \frac{\sigma^4 - k^2 g^2}{2\mu c^2 \sigma^2} \frac{1}{n_2 + \beta} e^{-(n_2+\beta)z_0} \ . \quad \text{(B14)}$$

Putting this constant into Equation 61, for $z=0$, gives

$$X(k, 0) = \sqrt{\frac{2}{\pi}} \frac{(\gamma - 1) q_0}{h \sigma^4 c^2 \rho_0} e^{(z_0/2H) - \mu z_0} \frac{\sin \lambda_0 k}{k} \frac{\sigma^4 - k^2 g^2}{(n_1 - m)(n_2 + \beta)} \ . \quad \text{(B15)}$$

The continuity of $W(k, z)$ and $P(k, z)$ at $z = z_0$ is satisfied by the above choice of constants, i.e., it can be shown that

$$\lim_{\epsilon \to 0} W(z_0 - \epsilon) = \lim_{\epsilon' \to 0} W(z_0 + \epsilon') \ .$$

$$\lim_{\epsilon \to 0} P(z_0 - \epsilon) = \lim_{\epsilon' \to 0} P(z_0 + \epsilon') \ .$$

SOLUTION FOR X[k,0] BY MEANS OF THE LAPLACE TRANSFORM

By applying the Laplace transform with respect to z,

$$\tilde{X}(k, s) = \int_0^\infty e^{-sz} X(k, z)\, dz \tag{B16}$$

Equation 59 becomes

$$\left(s^2 - 2Ns + M^2\right) \tilde{X} - (s - 2N) X(0) - \left.\frac{dX}{dz}\right|_{z=0} = Q_0 \frac{e^{-(s+\beta)z_0}}{s+\beta} \left[s^2 - 2Ns - k^2\left(1 - \frac{g}{\sigma^2 H}\right)\right].$$

By making use of the relation shown by Equation B5, this is

$$\tilde{X}(k, s) = \frac{(s - 2N - m)}{(s - n_1)(s - n_2)} X(0) + \frac{Q_0 e^{-(s+\beta)z_0}}{\beta + s} \left[1 - \frac{\sigma^4 - k^2 g^2}{(s - n_1)(s - n_2) c^2 \sigma^2}\right] \tag{B17}$$

where n_1 and n_2 are given [illegible] and 64; $m = m(\sigma)$ and Q_0 are given by Equations B6 and B12.

By making use of the inverse [illegible] of Equation B16 the solution to Equation 59 is

$$X(k, z) = \left(\frac{2N - m - n_1}{2\mu} e^{n_1 z} - \frac{2N - m - n_2}{2\mu} e^{n_2 z}\right) X(0) + Q_0 e^{-\beta z} \theta(z - z_0)$$

$$- Q_0 \frac{\sigma^4 - k^2 g^2}{2\mu c^2 \sigma^2} e^{-\beta z_0} \left[\frac{e^{n_2(z - z_0)}}{\beta + n_2} - \frac{e^{n_1(z - z_0)}}{\beta + n_1} + \frac{e^{-\beta(z - z_0)}}{(\beta + n_1)(\beta + n_2)}\right] \theta(z - z_0) . \tag{B18}$$

Because of the boundary conditions at $z = \infty$, the term proportional to e^{nz} must vanish, i.e., from Equation B18

$$- \frac{2N - m - n_2}{2\mu} X(0) - Q_0 \frac{\sigma^4 - k^2 g^2}{c^2 \sigma^2} \frac{e^{-(\beta + n_2) z_0}}{2\mu(\beta + n_2)} = 0 . \tag{B19}$$

This gives the same solution as Equation B15. It should be noted that the above solution satisfies the conditions of continuity of w(x, z) and p(x, z) at $z = z_0$, although u(x, z) and s(x, z) are not continuous at $z = z_0$.

Appendix C

EVALUATION OF p[x, 0]

As can be seen from Equation B15 X(k, 0) is an even function of k. Therefore, the substitution of Equation B15 into Equation 67 gives

$$\chi(x, 0) = \frac{2}{\pi} \frac{(\gamma - 1) q_0 \sigma^2}{h k_g^2 \rho_s c^4} e^{z_0/2H} \int_0^\infty e^{-\mu z_c} \frac{\cos xk \sin \lambda_0 k}{k} \frac{k_g^2 - k^2}{(n_1 - m)(n_2 + \beta)} dk \qquad (C1)$$

where

$$k_g = \frac{\sigma^2}{g} . \qquad (C2)$$

Since μ is a real function of k for $k \geq k_c$ and complex for $0 \leq k < k_c$, where

$$k_c = \frac{\sigma}{c} \sqrt{\frac{\sigma^2 - \sigma_A^2}{\sigma^2 - \sigma_B^2}} . \qquad (C3)$$

Equation C1 can be written

$$\chi(x, 0) = \chi_c(x, 0) + \chi_{nc}(x, 0) , \qquad (C4)$$

where

$$\chi_c(x, 0) = \frac{2}{\pi} \frac{(\gamma - 1) q_0 \sigma^2}{h k_2^2 \rho_s c^4} e^{z_0/2H} \int_0^{k_c} e^{-i\eta z_0} \frac{\cos xk \sin \lambda_0 k}{k} \frac{k_g^2 - k^2}{(n_1 - m)(n_2 + \beta)} dk , \qquad (C5)$$

and

$$\chi_{nc}(x, 0) = \frac{2}{\pi} \frac{(\gamma - 1) q_0 \sigma^2}{h k_2^2 \rho_s c^4} e^{z_0/2H} \int_{k_c}^\infty e^{-\mu z_0} \frac{\cos xk \sin \lambda_0 k}{k} \frac{k_g^2 - k^2}{(n_1 - m)(n_2 + \beta)} dk . \qquad (C6)$$

The integrand of $\chi_{nc}(\chi, 0)$ has two singular points σ/c and σ^2/g, corresponding to the two types of free oscillations of the noncellular mode which were pointed out by Pekeris* and have been discussed earlier. Because of a steep exponential term $e^{-\mu z_0}$, however, contributions of these singularities to the integral is not as important as the contribution from the narrow band near $k \geq k_c$, where the exponential term is nearly unity.

The integrand of x_c has no singularities but oscillates by the term $e^{-i\eta z_0}$. The main contribution arises from a narrow domain near $k \leq$ approx. k_c. Equation 68 can be written

$$p(x, 0) = p_c(x, 0) + p_{nc}(x, 0) , \qquad (C7)$$

where

$$p_c(x, 0) = p_r + ip_i , \qquad (C8)$$

*Pekeris, C. L., "The Propagation of a Pulse in the Atmosphere. II," *Phys. Rev.* 73:145–154, January 15, 1948.

$$\left|p_c(x,0)\right| = \frac{2}{\pi}\frac{(\gamma-1)q_0\sigma}{hk_g^2c^2}e^{z_0/2H}\left|\int_0^{k_c} e^{-i\eta z_0}\frac{\cos xk\sin\lambda_0 k}{k}\frac{k_g^2-k^2}{(n_1-m)(n_2+\beta)}\,dk\right| = \sqrt{p_r^2+p_i^2}\,; \quad \text{(C9)}$$

and

$$p_{nc}(x,0) \approx \frac{2}{\pi}\frac{(\gamma-1)q_0\sigma}{hk_g^2c^2}e^{z_0/2H}\int_{k_c}^{k_g} e^{-\mu z_0}\frac{\cos xk\sin\lambda_0 k}{k}\frac{k_g^2-k^2}{(n_1-m)(n_2+\beta)}\,dk\,. \quad \text{(C10)}$$

N66 37966

ATMOSPHERIC DENSITIES FROM EXPLORER 17 DENSITY GAUGES AND A COMPARISON WITH SATELLITE DRAG DATA*

GEORGE NEWTON, RICHARD HOROWITZ AND WOLFGANG PRIESTER†

The purpose of this letter is to present the initial atmospheric densities directly measured by the ionization gages on the Explorer 17 satellite. The satellite has been described [*Horowitz*, 1963] and the gauges and their response to the atmosphere have also been reported [*Newton et al.*, 1963]. The data presented in this letter were obtained with two types of ionization vacuum gauges, the hot-filament thermionic type (Bayard-Alpert) and the magnetron type (Redhead). The absolute accuracy of these data is believed to be ±35 percent and the repeatability is ±20 percent. A more complete discussion of the experiment will be included in a paper now in preparation [*Newton et al.*, 1964].

The density data from 47 operations of the satellite in the altitude range 255 to 330 km for local times between 0700 and 2100 hours are plotted as a function of altitude in Figure 1. These data were recorded by the five northern midlatitude minitrack stations:

Blossom Point, Md.
Grand Forks, Minn.
Mojave, Calif.
St. Johns, Newfoundland
Winkfield, England

At the times the measurements were made, the daily planetary geomagnetic index (A_p) was between 0 and 50, and the 10.7-cm solar index ($F_{10.7}$) was between 70 and 100×10^{-22} w/m^2 cps. It is seen from Figure 1 that there is considerable variation in the atmospheric density at a given altitude (that is, greater than a factor of 3 at the altitude of 280 km) which is due in part to local time, geomagnetic activity, and solar activity effects.

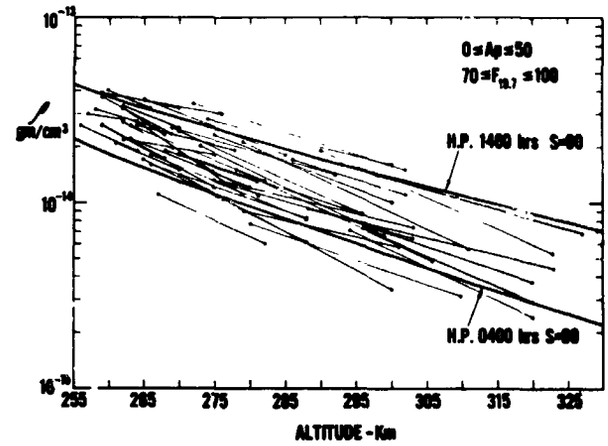


FIGURE 1.—Atmospheric density versus altitude measured by the Explorer 17 density gauges during 47 operations of the satellite. The densities were measured for times between 0700 and 2100 hours, with A_p varying between 0 and 50 and with $F_{10.7}$ varying between 70 and 100×10^{-22} w/m^2 cps.

To compare the directly measured densities with densities inferred from satellite drag, it is desirable to reduce the data to geomagnetically quiet conditions. In this regard, consider Figure 2, which shows $\Delta \log \rho$, the logarithm to the base 10 of the ratio of the measured density to the Harris and Priester model density selected for the appropriate time and altitude, plotted versus the daily geomagnetic index (A_p). These points represent 120 density-altitude profiles for altitudes between 255 and 600 km and most local times; the data were recorded over the same five minitrack stations as were the data in Figure 1. The darkened symbols represent an average of two or more

*Published in the *Journal of Geophysical Research*, 69(21):4690–4692, November 1, 1964.

†Goddard Space Flight Center; on leave from Bonn University.

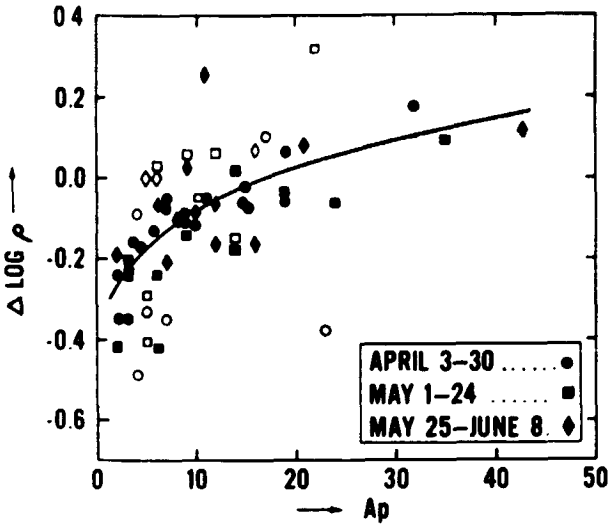


FIGURE 2.—Daily average of the logarithms of the ratios of the gauge measured density to the Harris and Priester model density, plotted against daily geomagnetic A_p index. Solid symbols represent averages of measurements obtained from two or more passes of the satellite over the telemetry stations. Open symbols represent single passes. The data have been adjusted to correspond to $F_{10.7}=83$. The semiannual variation of the atmospheric density has been accounted for in an approximate way by decreasing the model densities by 18% for May and June 1963.

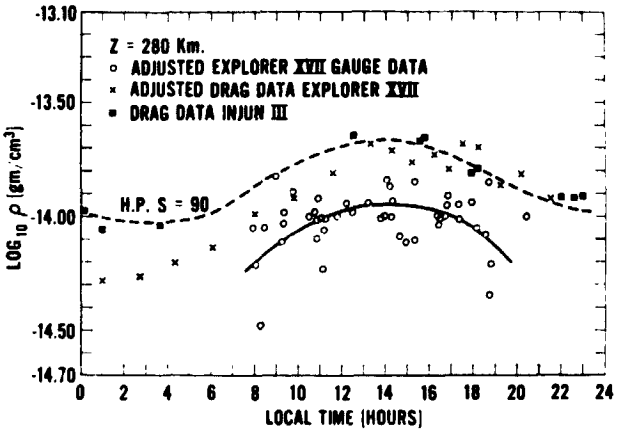


FIGURE 3.—Logarithms of densities normalized to an altitude of 280 km versus local time. Injun 3 data are selected for geomagnetically quiet conditions. Explorer 17 data are adjusted to $A_p=2$ using relation shown in Figure 2 and for an $F_{10.7}$ corresponding to 83.

passes within one day and the light symbols represent one pass. The passes have been individually corrected to an $F_{10.7}$ of 83 using *Roemer's* [1963] formula. The semiannual variation of the atmospheric density has been accounted for in an approximate way by decreasing the model densities by 18 percent for May and June 1963. A visual average of the data shows (*a*) that the relation between Δ log ρ and A_p is nonlinear, the steepest slope being applicable to the smaller A_p values, and (*b*) that the atmospheric density values are more sensitive to geomagnetic disturbances than has previously been reported. Very recently, *Jacchia and Slowey* [1964*a*] found independently that for near-quiet geomagnetic conditions the reaction of the atmosphere to variations in A_p is considerably larger than that expected on the basis of their previous analyses.

Figure 3 shows the logarithm to the base 10 of the gauge densities at the average time and altitude of the pass plotted versus local time and normalized to 280 km using the Harris and Priester model for $S=90$ as a differential altitude transformer. Shown for comparison are the densities obtained from drag observations of Injun 3 [*Jacchia and Slowey*, 1964*b*] and Explorer 17 [*Bryant*, 1964], and the Harris and Priester model with $S=90$. The drag data have also been normalized to the 280-km altitude.

The Injun 3 drag data, indicated by squares, were selected for quiet geomagnetic conditions ($A_p\pm2$) for the time interval of February 18 through June 30, 1963, when the latitude of the satellite varied between $-40°$ and $+60°$. The Explorer 17 drag data, indicated by crosses, correspond to the time interval from April 3 to July 6, 1963, when the satellite perigee was between $+58°$ and $-20°$.

The relation shown in Figure 2 has been used to adjust both the Explorer 17 drag-determined and gauge-measured densities to geomagnetically quiet conditions ($A_p=2$). The Explorer 17 densities also have been normalized to a 10.7-cm solar index of 83, a small adjustment compared to those associated with the A_p. The A_p adjustment resulted in: (*a*) a decrease of approximately a factor of 2 in the apparent scatter between the gauge-measured density values, (*b*) a lowering of the average of the gauge-measured densities by a factor of 0.7, and (*c*) a lowering of Bryant's drag densities by approximately 60 percent.

It is seen in Figure 3 that the atmospheric densities determined by satellite drag techniques appear to be systematically higher by a factor of 2 than the adjusted gauge densities. This difference is greater than the combined uncertainties

assigned to the separate sets of data and would seem to be due to a systematic error in one or both of the measurements. While the difference is not regarded as a serious discrepancy between the two techniques, it is large enough to require a re-examination of the measurements made by each.

A more detailed presentation of the atmospheric density data from these five minitrack stations and their geophysical interpretation is in preparation [*Newton et al.*, 1964]. The presentation will include (1) data of considerably greater altitude and local-time coverage and (2) density scale-height computations. In addition, data from the other eight minitrack (perigee 255 km, initial apogee 920 km, 58° inclination) are currently being analyzed and will be reported.

ACKNOWLEDGMENTS

We wish to thank Mr. Nelson W. Spencer, Goddard Space Flight Center, for helpful discussions in the preparation of this letter.

One of us (W. Priester) is grateful to the National Academy of Sciences—National Research Council for a Senior Research Associateship.

REFERENCES

Bryant, R., Densities obtained from drag on the Explorer 17 satellite, *J. Geophys. Res.*, *69*, 1423–1425, 1964.

Harris, I., and W. Priester, Heating of the upper atmosphere, *Space Res.*, *3*, 53–75, 1963.

Horowitz, R., S-6, an aeronomy satellite, *Advan. Astronautical Sci.*, *12*, Western Periodicals Company, North Hollywood, California, 1963.

Jacchia, L. G., and J. Slowey, Temperature variations in the upper atmosphere during geomagnetically quiet intervals, *Smithsonian Inst. Astrophys. Obs. Rept. 152*, June 1964*a*.

Jacchia, L. G., and J. Slowey, Atmospheric heating of the auroral zones: A preliminary analysis of atmospheric drag of the Injun 3 satellite, *J. Geophys. Res.*, *69*, 905–910, 1964*b* (*SAO Spec. Rept. 136*, September 1963).

Newton, G. P., R. Horowitz, and W. Priester, Explorer 17 pressure gauge experimental data and interpretation, in preparation, 1964.

Newton, G. P., D. T. Pelz, G. E. Miller, and R. Horowitz, Response of modified Redhead magnetron and Bayard-Alpert vacuum gauges aboard Explorer 17, *Transactions of the Tenth National Vacuum Symposium*, edited by George H. Bancroft, pp. 208–212, The Macmillan Company, New York, 1963.

Roemer, M., Die Dichte der Hochatmosphäre und Ihre Variationen während der Phase abklingender Sonnenaktivitat, 1958–1962, *Veroeffentl. Univ. Sternwarte Bonn*, *68*, 1963.

STRATOSPHERIC TEMPERATURE PATTERNS BASED ON RADIOMETRIC MEASUREMENTS FROM THE TIROS VII SATELLITE*

W. NORDBERG, W. R. BANDEEN, G. WARNECKE AND V. KUNDE

The TIROS VII meteorological satellite carried a sensor to map the emitted terrestrial radiation in the 15 micron carbon dioxide band. The spectral response of the sensor ranged from 14.8 to 15.5 microns. The radiation observed by the satellite is a function of the temperatures of the atmosphere varying along the optical path. Atmospheric layers at various altitudes contribute to the observed radiation in varying amounts. More than 96 percent of this contribution stems from altitudes above 10 km and more than 60 percent from a layer ranging in altitude from 18 to 35 km; thus, the radiation measurements were interpreted in terms of average temperatures of a major portion of the stratosphere.

Temperature patterns were analyzed from June 1963 to March 1964. The measurements demonstrate on a global scale, the varying patterns of stratospheric temperatures and circulation. Smooth temperature gradients were found to coincide closely with latitudinal circles on both summer hemispheres. Temperatures increased at high latitudes toward the summer pole (250° K) and rapidly decreased toward the winter pole (200° K).

A fairly uniform region with temperatures generally around 230° K extended between 25° latitude of the summer hemisphere and 40° latitude of the winter hemisphere during solstices. During both equinoxes, the large latitudinal temperature gradients at high latitudes disappeared and differences of about 10–15° K existed with latitudes as well as with longitude. Winter temperature patterns in both hemispheres exhibit strong temperature gradients between the pole and 40°, but a completely uniform pattern never existed in either winter hemisphere. A warm area over the Indian Ocean, though initially small, seemed to play a similar role in the southern winter as the Aleutian warm center during the northern winter. The morphology of a stratospheric warming which occurred over southwestern Asia and penetrated to Europe immediately after the tropical belt had cooled by about 10° K was observed.

I. INTRODUCTION

A radiometer aboard the TIROS VII satellite measured infrared radiation emitted mainly by carbon dioxide in the earth's atmosphere at wavelengths ranging from 14.8 to 15.5 microns. The instrument in this measurement was very similar to the 5 channel radiometers flown on previous TIROS satellites and described by Bandeen, et al. (Reference 1), except that one of the optical channels was modified to have a spectral response as shown in Figure 1a. The initial objective of radiation measurements in this region of the spectrum was to study the characteristics of the earth's horizon as described by Hanel, et al. (Reference 2) and later reported by Bandeen, et al. (Reference 3). Now, with continuous observations over a period of more than six months, covering daily the global zone from 65° N to 65° S, it is possible to use the radiation measurements to describe a number of interesting features of the atmosphere. It will be shown that most of the radiation reaching the satellite in this spectral region is emitted within a defined altitude interval in the stratosphere whose average temperature can be derived from the observed radiation intensities. Quasi-global maps of these average temperatures for different seasons not only reflect the expected variations in the mean stratospheric temperatures, but also reveal the development and behavior of major synoptic systems in that region of the atmosphere. A complete description

*Published as *Goddard Space Flight Center Document X-651-64-115*, May 1964.

†On leave from Institut für Meteorologie und Geophysik der Freien Universität Berlin, Berlin, Germany, as National Academy of Sciences-National Research Council Research Associate with National Aeronautics and Space Administration.

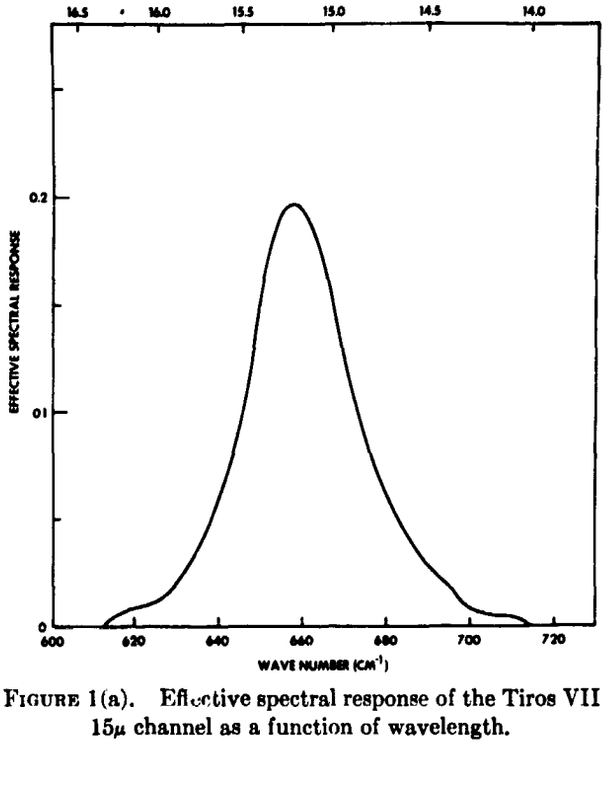


FIGURE 1(a). Effective spectral response of the Tiros VII 15μ channel as a function of wavelength.

of the 5 channel radiation experiment on TIROS VII as well as a critique and summary of all data resulting from this experiment will be contained in the TIROS VII Radiation Data Catalog and Users' Manual (Reference 4).

II. RELATION BETWEEN RADIATION AND STRATOSPHERIC TEMPERATURES

The spectral response of the instrument was chosen to coincide as closely as possible with the 15 micron absorption band of carbon dioxide. The radiance N from the earth and atmosphere measured by the radiometer whose spectral response function $\phi(\lambda)$ is shown in Figure 1a can be expressed as:

$$N=\frac{1}{\pi}\int_0^{\infty}\int_{\lambda=\lambda_1}^{\lambda_2}\phi(\lambda)B(\lambda, T)\frac{\partial\tau}{\partial s}(\rho, p, T, \lambda)d\lambda ds \qquad (1)$$

where

ρ = carbon dioxide density

B = Planck function

s = distance along radiation path

λ_1 and λ_2 = wavelengths between which $\phi\neq 0$

p = atmospheric pressure

T = atmospheric temperature

τ = atmospheric beam transmissivity

It is reasonable to assume that a constant ratio $r=0.000471$ exists between the densities ρ for carbon dioxide and ρ_{air} for air: $\rho=r\,\rho_{air}$. Furthermore, a number of typical distributions for temperature with height h for surface pressures of 1013.25 mb were chosen so that ρ_{air}, p and T could be expressed as functions of h where ρ_{air} and p were computed as a function of h from T by the hydrostatic equation. A relationship between s and h, $s=s(h)$, was established through the choice of several angles at which the satellite at approximately 635 km viewed the earth. The integration of (1) over λ was then performed for seven nadir angles, 0°, 29°, 40°, 47°, 52°, 55°, and 58°. The function $\tau(\lambda, h)$ was determined using, in principle, the method described by Hanel, et al. (Reference 2). That method is based on the general treatment by Elsasser (Reference 5) of the absorption within the 15 micron CO_2 band. Thus,

$$\int_{\lambda_1}^{\lambda_2}\phi(\lambda)B(h, \lambda)\frac{\partial\tau}{\partial s}(h, \lambda)\frac{\partial s}{\partial h}d\lambda=\Psi(h). \qquad (2)$$

Hence,
$$\bar{N}=\frac{1}{\pi}\int_0^{\infty}\Psi(h)dh. \qquad (3)$$

$\bar{N}$ was then computed for each of the assumed temperature profiles and for each nadir angle from (3). The assumed temperature distributions with height and the resulting distribution functions $\Psi(h)$ are shown in Figures 2a through 2c. In accordance with previous practices of reporting TIROS radiation measurements (References 4, 6–8) $\bar{N}$ may be expressed by the temperature T_{BB} of a blackbody whose radiant emittance $\bar{W}$ is given by:

$$\bar{W}=\int_{\lambda_1}^{\lambda_2}\phi(\lambda)B(T_{BB}, \lambda)d\lambda=\pi\bar{N} \qquad (4)$$

Indeed, since the radiometers were calibrated using such blackbodies at various temperatures, the measurement of T_{BB} may be considered as the

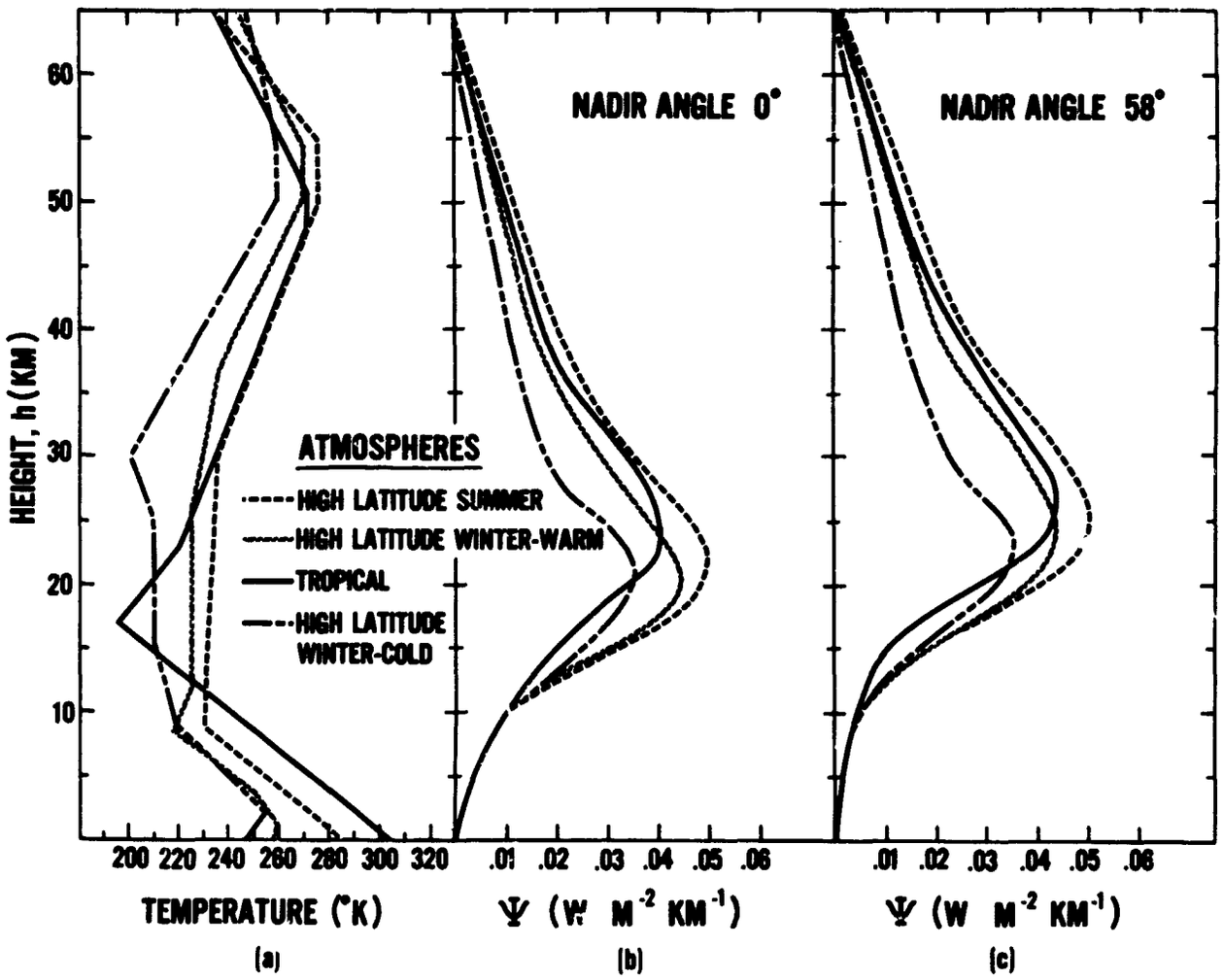


FIGURE 2(a).—Typical temperature profiles based on proposed supplements to the "U. S. Standard Atmosphere 1962" for 60° North summer, 60° North winter (warm and cold) and 15° North. The "warm" and "cold" temperature profiles for 60° N ("high latitude winter") can be considered typical at these latitudes depending on the state of the stratosphere in these regions. The supplements to the U. S. Standard Atmosphere 1962 are in the process of preparation by the U. S. Committee for Extension of the Standard Atmosphere and were summarized by Cole, et al. Reference 16).

FIGURE 2(b).—Weighting functions ψ (h), applying to the measured outgoing radiance $\bar{N}$; nadir angle = 0°

FIGURE 2(c).—Weighting functions ψ (h), applying to the measured outgoing radiance $\bar{N}$; nadir angle = 58°.

primary quantity while N requires a derivation by means of equation (4). The values of T_{BB} corresponding to the temperature profiles of Figure 2a are summarized in Table I.

The radiation measurement expressed by T_{BB} thus represents a mean temperature of the atmosphere in which the vertical temperature distribution is open to choice and where the temperature at each altitude must be given a different weight regarding its contribution to the measured mean. The weighting function is given by Ψ (h), and although Ψ (h) depends to a certain degree on the

TABLE I.—*Computed Mean Temperatures (T_{BB}) Expected to be Observed by 15 Micron Radiometer for Temperature profiles shown in Figures 2a at Nadir Angles of 0° and 58°.*

Nadir Angle	Tropical	High Lat. Summer	High Lat. Winter Cold	High Lat. Winter Warm
0°	227° K	238	213	229
58°	231	240	214	231

assumed vertical temperature profile, it may be seen that for most realistic temperature profiles at a nadir angle $\eta=0$ (Figure 2b), temperatures between 20 and 25 km are weighed most heavily with about 65 percent of the weight being concentrated in the region from 15 to 35 km and more than 96 percent of the weight lying at altitudes above 10 km. It may, also, be seen from Figure 2c that the maximum weight shifts to somewhat higher altitudes for a nadir angle of 58°. But, since this shift amounts to less than 5 km from $\eta=0$ to $\eta=58°$, it is not possible to make a precise quantitative determination of the variation of temperature with altitude from measurements of T_{BB} at various nadir angles. However, it will be shown in the section below that from such measurements one may infer qualitatively whether the temperature is increasing, decreasing, or constant with height over the altitude range where the Ψ functions reach their maxima. Furthermore, with a prudent and realistic choice of assumed vertical temperature profiles, such as shown in Figures 2a, one can easily determine from the T_{BB} measurements which one of these profiles represents the best fit qualitatively to actual conditions at a certain location on the globe and; more important, one may determine from continuous observations how the choices of such "best fits" vary with time. Thus, the maps of T_{BB} measurements plotted for the "quasi-global" (65° N to 65° S) zone for different times of the year, presented in Figures 3–12, enable the synoptic analysis of phenomena related to the temperature structure in the stratosphere between about 15 and 35 km. The occurrence of such events as explosive warmings, the establishment of the Aleutian anticyclone in winter, the onset of the seasonal circulation reversals and similar events may be categorized as such phenomena. Although the TIROS VII radiometer views the entire global zone from 65° N to 65° S during the course of 12 hours, data from only about 60 percent of the orbits during any one

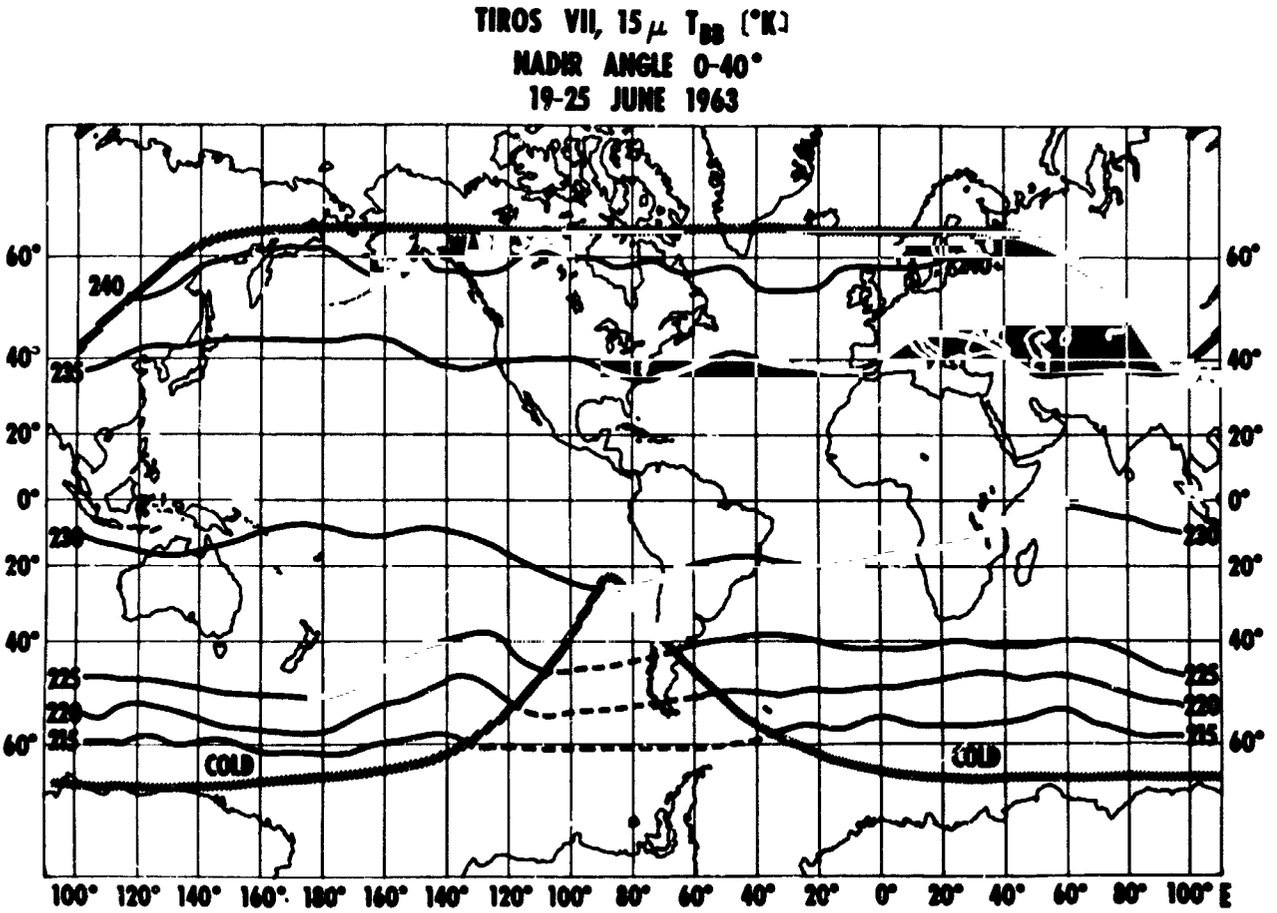


FIGURE 3(a).—Quasi-global map of 15μ equivalent blackbody temperatures averaged over the week 19–25 June 1963; nadir angle 0°–40°. Radiation data cannot be obtained outside the area enclosed by the dashed border.

day can be acquired because only stations within the United States are capable of receiving the data which are stored aboard the satellite for no longer than one orbital period. This causes two significant gaps in the data displayed in the maps, Figures 3–12. One such gap exists at 90° W just off the west coast of South America, the other is at 90° E over the central Soviet Union.

As will be stated in the TIROS VII Radiation Data Catalog and Users' Manual (Reference 4), the precision of each individual measurement is probably not much better than ±3°K. However, one may assume that the physical properties of the stratosphere remain generally constant for several days and are uniform over distances of less than 500 km. Since the field of view of the radiometer is about 5 degrees, covering an area of about 50×50 km on the surface of the earth for small nadir angles, and since in general measurements for the entire quasi-globe are obtained at least once every 24 hours, up to about 1000 measurements may result within one week within an area of about 500×500 km. Therefore, a precision of better than 1°K can be estimated for the temperature values shown in the maps, Figures 3–9 and 11–12, which were obtained by averaging measurements within each grid element (ranging from 5° latitude×5° longitude at the equator to 2.5° latitude×5° longitude at 60° latitude) over the period of one week.

Although the precision of the temperature measurement is greatly enhanced by reducing the random error in the averaging process, there are a number of systematic errors which must be considered. It has been attempted to reduce these errors by applying the appropriate corrections, but uncertainties still remain, causing the overall accuracy to be considerably poorer than the fraction of 1°K stated for the precision. Aside from correcting for an instrumental degradation which will be discussed later, a correction was applied to the data for an as yet unexplained deviation from

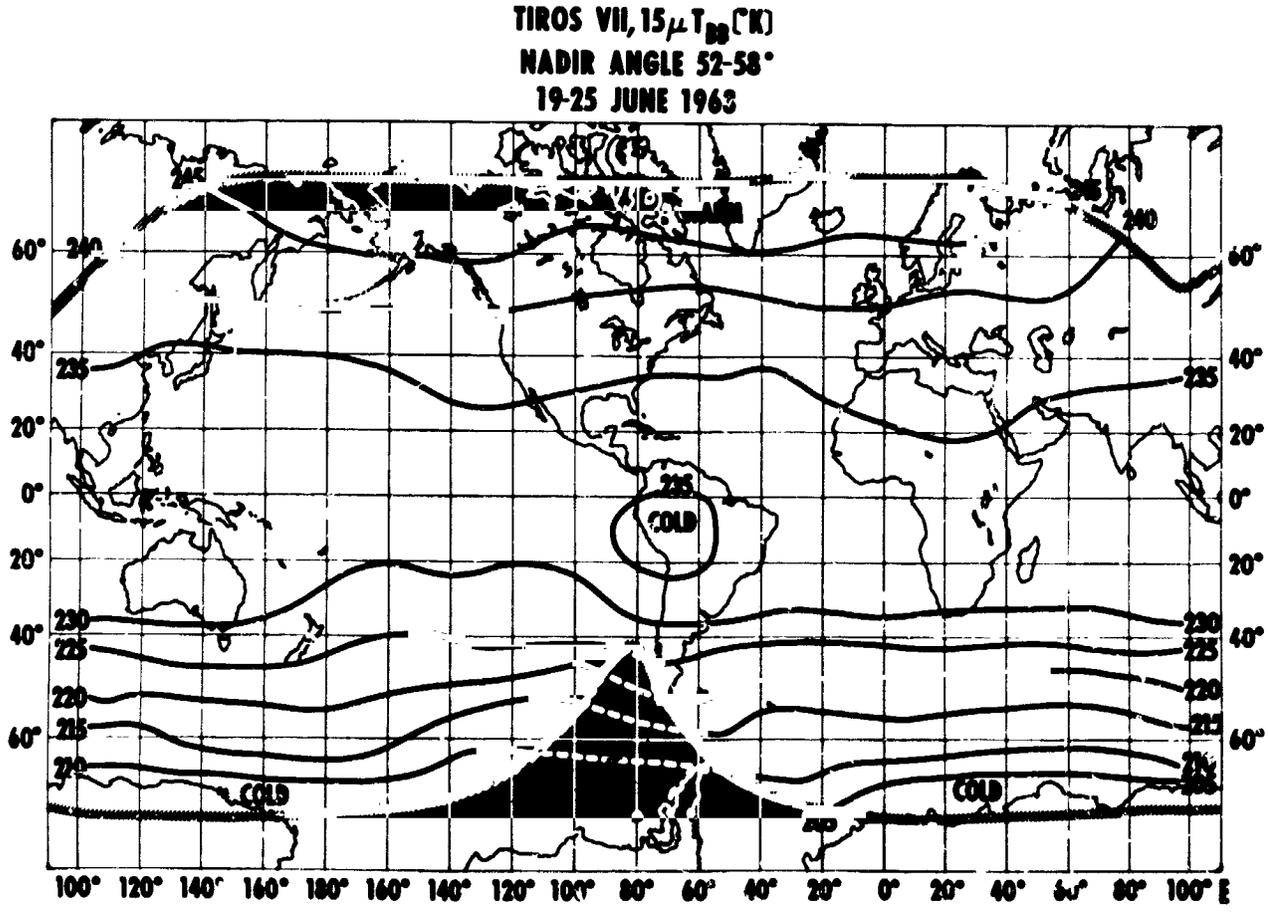


FIGURE 3(b).—Quasi-global map of 15μ equivalent blackbody temperatures averaged over the week 19–25 June 1963; nadir angle 52°–58°. Radiation data cannot be obtained outside the area enclosed by the dashed border.

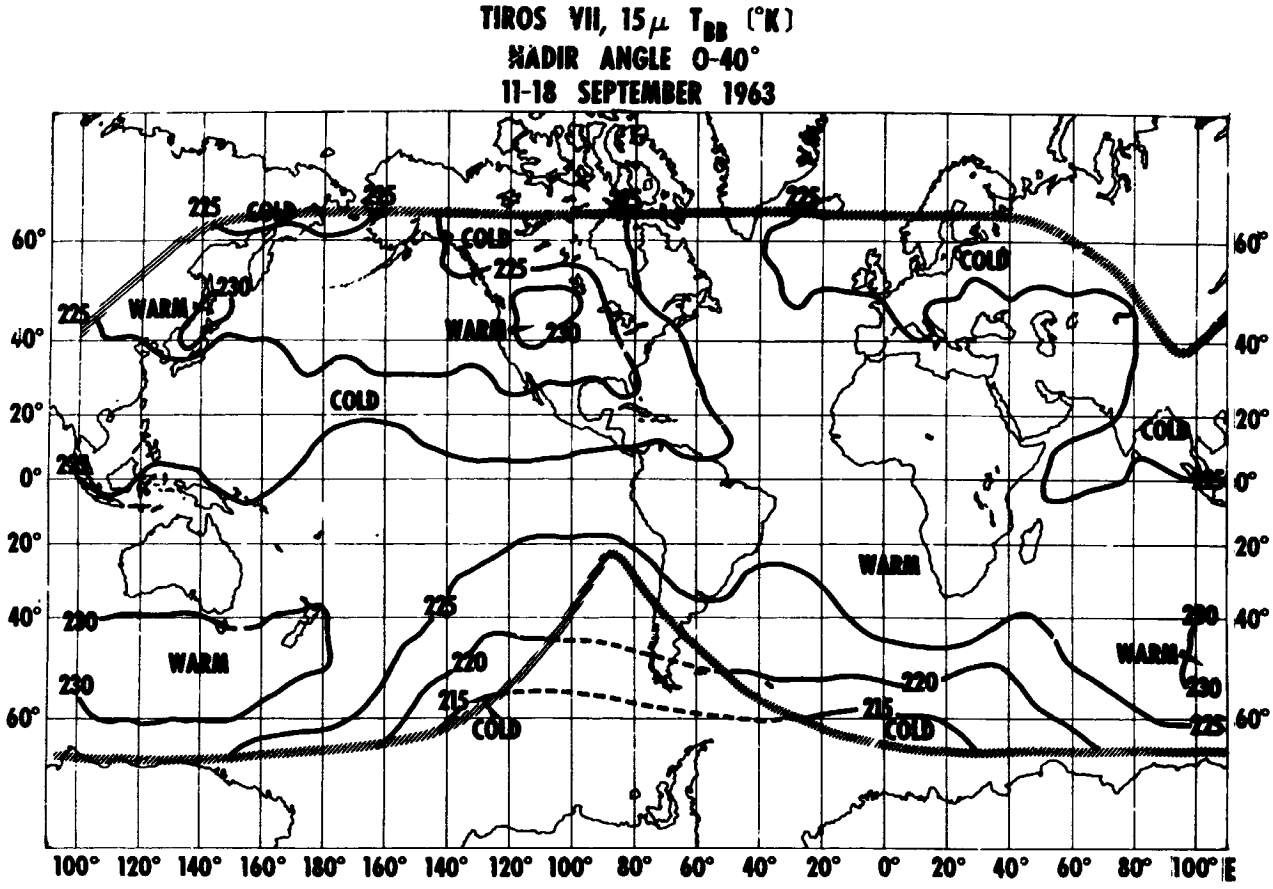


FIGURE 4.— Quasi-global map of 15μ equivalent blackbody temperatures averaged over the week 11–18 September 1963; nadir angle 0°–40°. Radiation data cannot be obtained outside the area enclosed by the dashed border.

the original preflight calibration. The radiometer, which alternatingly views the earth through one of two viewing ports showed that a consistently lower response on one of these viewing ports with respect to the other were apparently causing half of the measured temperatures to be too low and half to be too high. According to the preflight calibration, there should be no such difference. For lack of a better correction procedure, the difference between the temperatures measured through the two viewing ports over the same area at the same time were determined for all cases shown in the maps in Figures 3 through 12. Temperatures for the lower side were then corrected upward by half this difference, while those for the higher side were corrected downward by the same amount. Total differences were generally 5 to 8°K, so that the corrections were in the order of 2.5 to 4°K.

The maps were prepared from the magnetic tapes (FMRT's) containing all the TIROS VII radiation data (Reference 4) by an IBM 7094 computer and contour lines were drawn manually. This mapping procedure was the same as the one used in previously published similar maps (Reference 9).

III. ANALYSIS OF STRATOSPHERIC TEMPERATURES, JUNE 1963 TO MARCH 1964

A. Solstice June 1963

A typical temperature pattern at the start of the northern hemisphere summer and southern hemisphere winter for the quasi-globe is shown in Figure 3a. Stratospheric mean temperatures averaged over nadir angles between 0° and 40° and averaged over the week from 19 June to 25 June, the first week of TIROS VII, range from 240°K near 60° N to 215° near 60° S. The generally zonal structure of the temperature pattern, with isotherms running practically parallel with

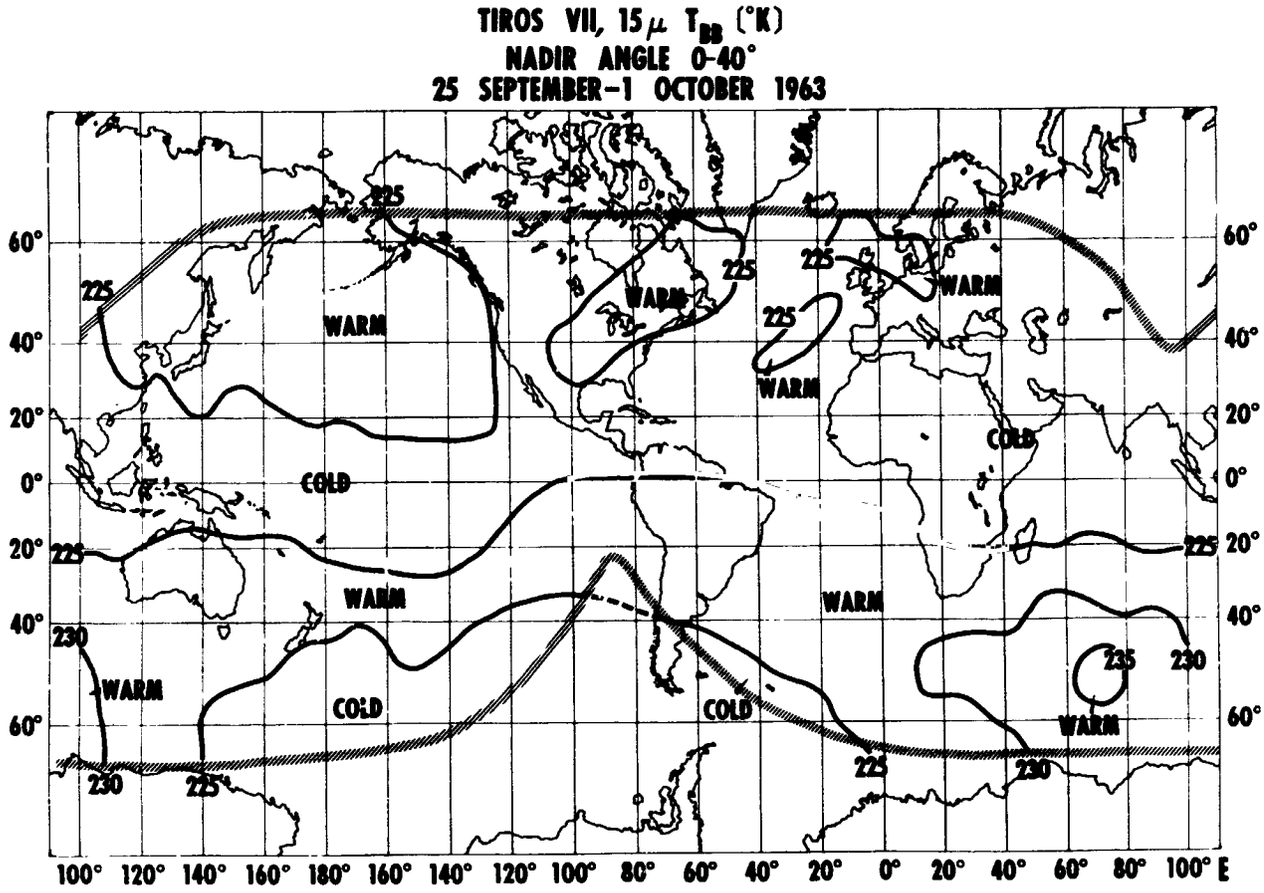


FIGURE 5.—Quasi-global map of 15μ equivalent blackbody temperatures averaged over the week 25 September–1 October 1963; nadir angle 0°–40°. Radiation data cannot be obtained outside the area enclosed by the dashed border.

latitude circles, confirms the established knowledge of the physics of the lower stratosphere; namely, the existence of a very strong cyclonic circulation system around a low pressure core centered near the winter pole. This low pressure is caused by the intense cooling of the stratosphere over the winter pole. Conversely, anticyclonic circulation prevails over the summer pole where solar heating causes a warm stratosphere resulting in a high pressure system. The predominantly zonal isotherms at high northern latitudes (summer) suggest that solar heating is primarily responsible for this pattern. At high southern latitudes (winter) however, there are slight but apparently significant disturbances in this zonal pattern. Isotherms over the South Pacific between 30° S and 60° S deviate from a perfect zonal course; they show a definite slope from the southwest to the northeast in that area. Temperatures increase rapidly from the winter pole toward the equator causing very steep temperature gradients with latitude in the winter hemisphere. A temperature of about 230°K is reached near 20° S. From there to the summer pole, the temperature increases gradually, and the gradients are not as steep as over the winter hemisphere. This is expected and it demonstrates clearly that the winter cyclone is considerably stronger than the summer anticyclone. The measured mean temperatures of 215°K near 60° S and of 240°K near 60° N agree reasonably well with the computed values of 213° and 238°K respectively (Table I) for the temperature profiles for high latitude cold winter and summer shown in Figure 2a. The measured temperatures of 230°K within the belt 40° S to 25° N are 3°K higher than those computed (Table I) for the tropical temperature profile shown in Figure 2a. This disagreement with the computation seems to be consistent with the cases at 60° N and 60° S where the measured temperatures were 2°K too high. The reason for this consistent discrepancy could be due to a number

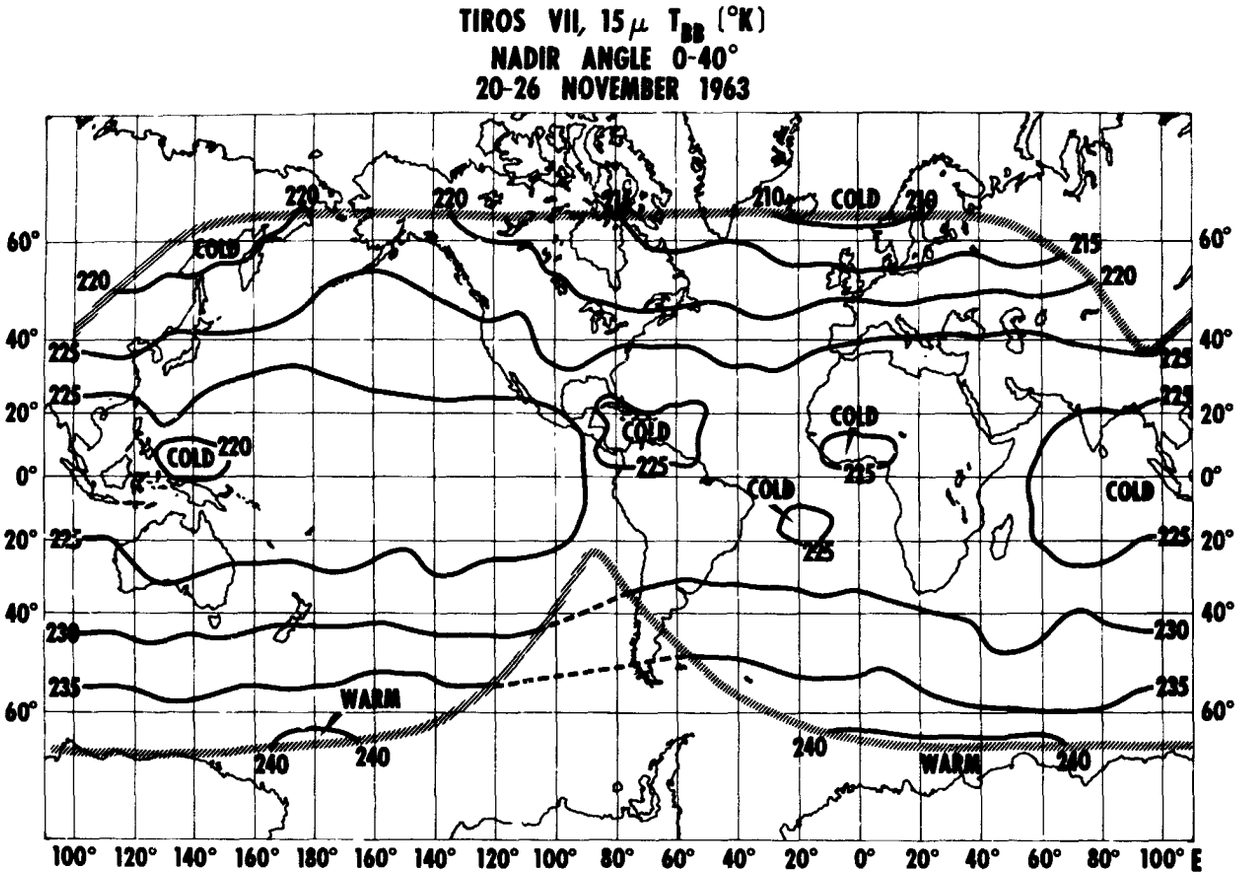


FIGURE 6.—Quasi-global map of 15μ equivalent blackbody temperatures averaged over the week 20–26 November 1963; nadir angle 0°–40°. Radiation data cannot be obtained outside the area enclosed by the dashed border.

of assumptions contained in the radiative transfer calculations which may have resulted in the computed values of T_{BB} being several degrees too low.

A coarse picture of the temperature gradient with altitude can be obtained by comparing Figures 3a and 3b. In Figure 3b, temperatures measured at nadir angles between 52 and 58 degrees were plotted. The weighting functions in Figures 2b and 2c show that all measurements shown in Figure 3b reflect temperatures at altitudes approximately 3–4 km higher than the temperatures shown in Figure 3a. Thus, by comparing Figures 3a and 3b, one may distinguish three categories in the temperature profiles:

1. Increasing temperatures with altitude between 15 and 35 km where the measured temperature differences are positive.

2. Decreasing temperatures with altitude between 15 and 35 km where the measured temperature differences are negative.

3. Isothermal regions between 15 and 35 km if the temperature differences are close to zero. It can be seen that differences of about +2°K occur in the tropics where strongly positive temperature gradients with altitude prevail. This difference agrees reasonably well with the values calculated for a tropical profile (Table 1). Over the higher northern latitudes (summer) moderately increasing temperatures with altitude and over high southern latitudes (winter) isothermal temperatures are suggested. This is in fair agreement with the calculations (Table I).

B. Seasonal Reversal, September–October 1963

As the solar illumination gradually becomes symmetrical with respect to northern and southern hemispheres during September, 1963, temperature patterns at high latitudes markedly reflect the resulting change in solar heat input. In some areas, this change takes place rapidly while other regions undergo a much slower transition. During the first two months following the

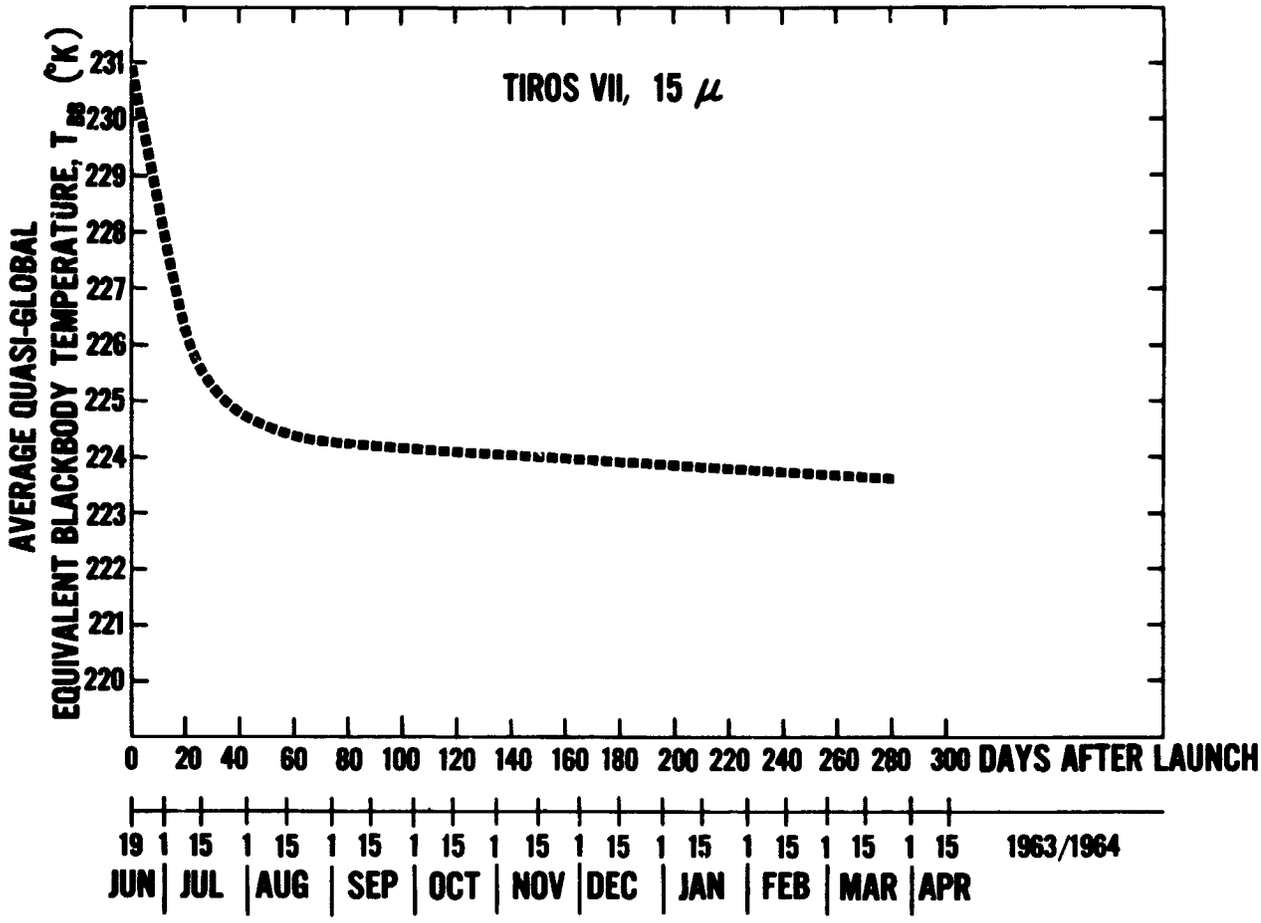


FIGURE 1(b).—Average quasi-global equivalent blackbody temperatures vs. time. The heavy dashed curve indicates the derived apparent degradation in the instrumental response.

solstice, there were practically no temperature gradients along any given latitude circle in the northern hemisphere; but strong longitudinal temperature patterns begin to develop now during the transitional season, especially after the September equinox. During July and August, the temperature patterns remain essentially the same as those during 19–25 June, shown in Figure 3a. There is one important change, however, which exhibits the same characteristics as similar changes observed on measurements in other spectral regions in previous TIROS satellites (References 7 and 8). *All* measured temperatures over the entire world seem to decline during the first three weeks of observation. It is extremely probable that a gradual degradation in the response of the radiometer is the cause for this decline. By mid-July 1963, the magnitude of this sensitivity decrease is about 7°K. After that, the degradation seems to have leveled off and continued at a much lesser rate. An approximate time history of this instrumental degradation based on the preliminary information available at this time is given in Figure 1b. A more complete version will be given in the TIROS VII Radiation Data Catalog and Users' Manual (Reference 4). An appropriate correction based on Figure 1b should, therefore, be added to all temperature measurements quoted thereafter and shown in the maps, Figure 4 to 12. Although this degradation leaves some uncertainties in the exact absolute magnitudes of the temperature measurements, it does not appreciably affect the ability to assign the appropriate temperature profile (Figure 2a) to a measurement at a given location or the implications drawn from the distribution of relative temperature patterns.

The first indication of any significant change in the post solstice temperature pattern occurs during the week 28 August to 3 September: the temperatures at high northern latitudes decrease by about 5°K to a temperature of about 228°K (+7° correction) at 60° N. At northern

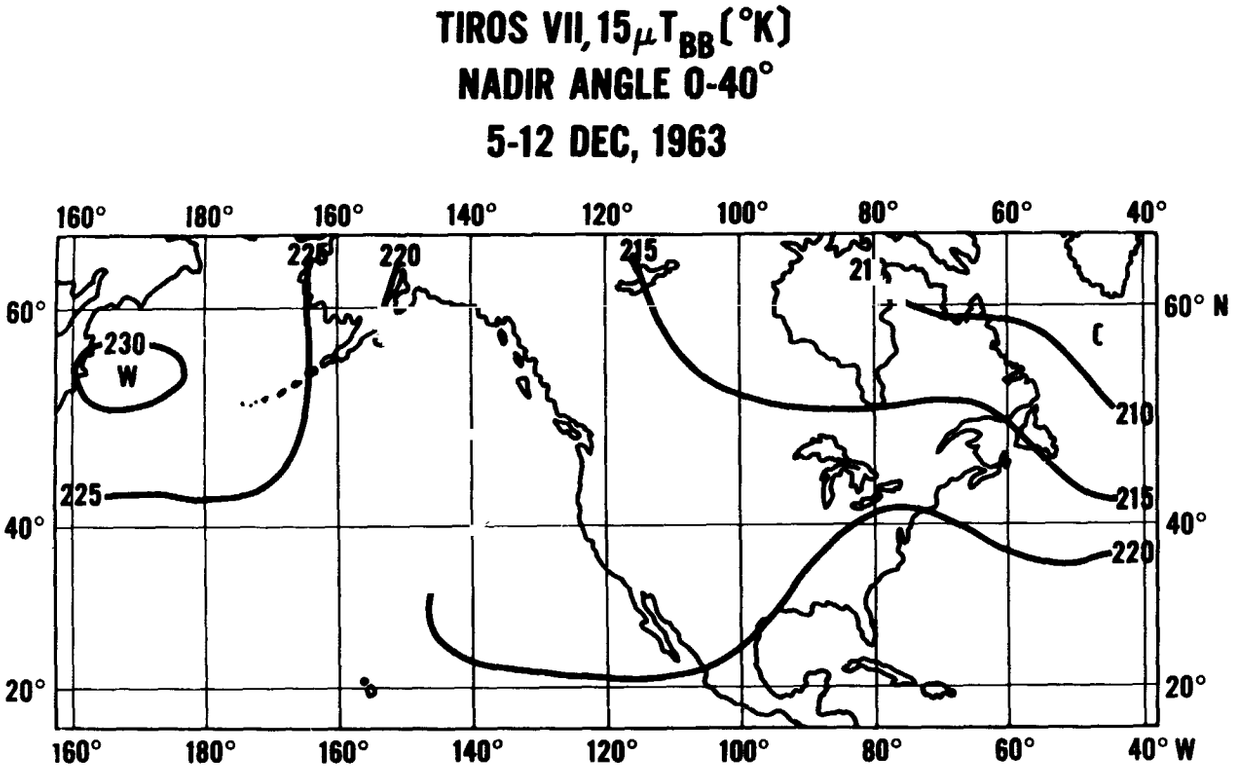


FIGURE 7(a).—Map from the North Pacific to the North Atlantic of 15μ equivalent blackbody temperatures averaged over the week 5–12 December 1963; nadir angle 0°–40°.

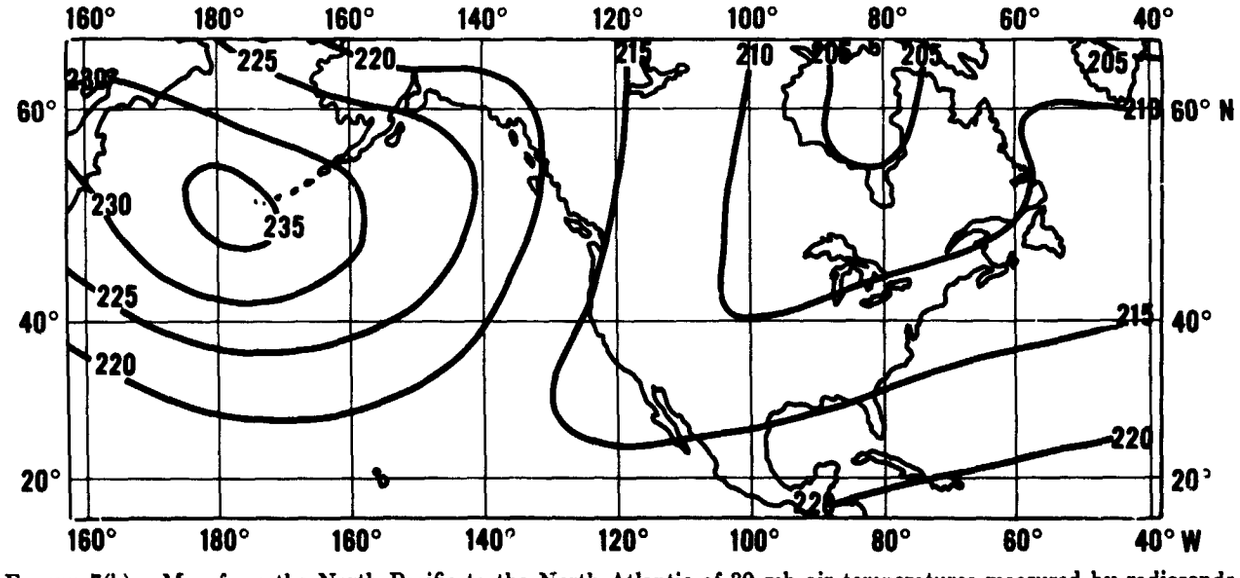


FIGURE 7(b).—Map from the North Pacific to the North Atlantic of 30 mb air temperatures measured by radiosonde balloons averaged over the week 5–12 December 1963 (Reference 16).

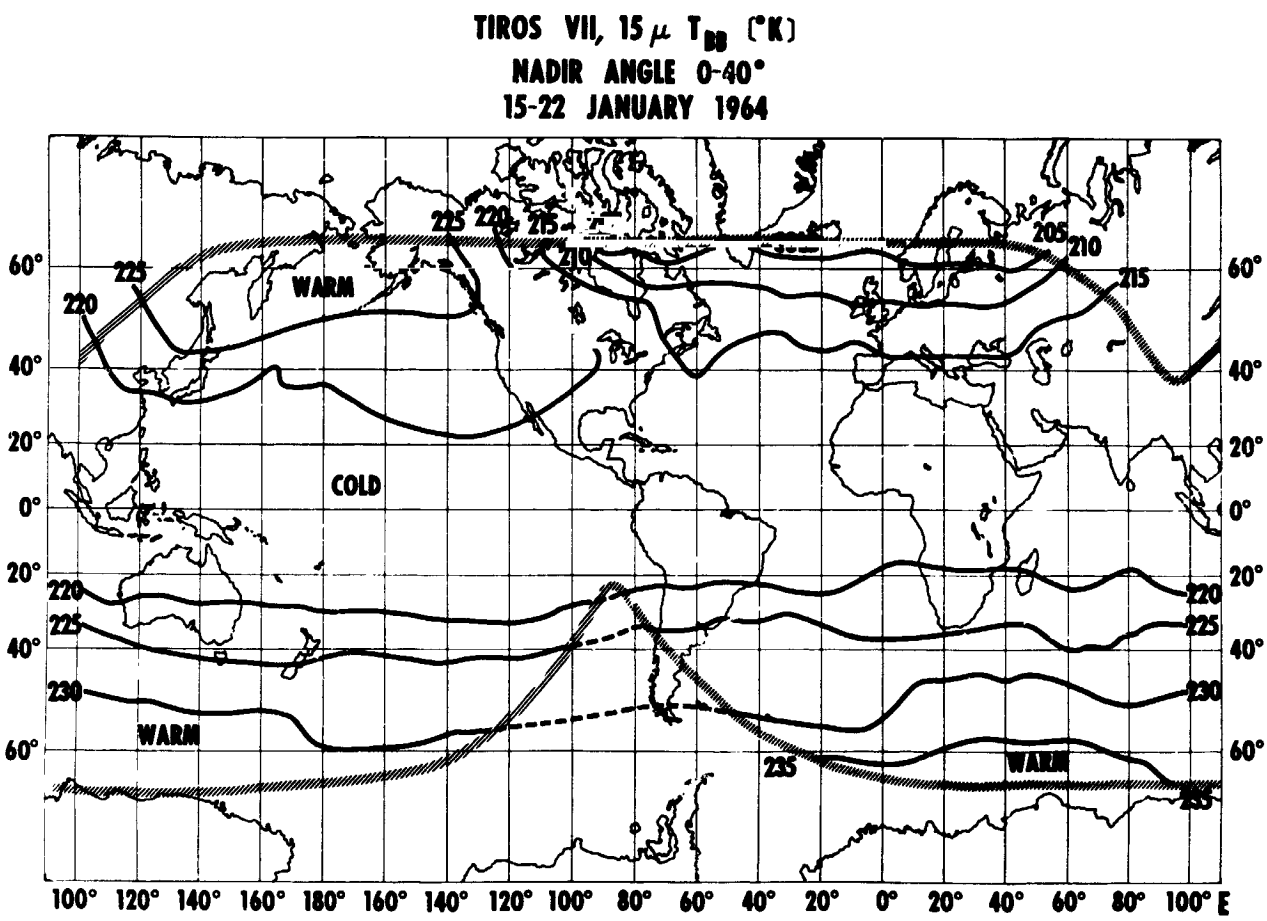


FIGURE 8(a).—Quasi-global map of 15μ equivalent blackbody temperatures averaged over the week of 15–22 January 1964; nadir angle 0°–40°. Radiation data cannot be obtained outside the area enclosed by the dashed border.

mid-latitudes (20°−50° N) there is also a slight temperature decrease from the summer (19–25 June) conditions. However, here this decrease does not take place uniformly at all longitudes. There are pocket-like regions where warmer temperatures seem to persist longer than over other regions causing some inhomogeneities in contrast to the quite uniform zonal pattern observed in June. Going southward, temperatures in general decrease very gradually from about 228+7°K at 60° N to about 223+7°K at 40° S with the latter temperatures prevailing over most of the zone from about 25° N to 40° S. From there southward temperatures decrease rapidly with latitude. Thus, the remainder of the southern hemisphere remains essentially unchanged, still reflecting typical winter conditions (205+7°K at 60° S) except for a warm region of 225+7°K, which developed during later winter over the Indian Ocean between 40° to 50° S. The disturbance noted during June at these latitudes over the western South Pacific has grown throughout July and early August, and a warm region extends now over the entire Indian Ocean showing a definite longitudinal pattern. For example, temperatures over the southern Indian Ocean at 60° S are about 216°+7°K. At the same latitude over the Atlantic and central South Pacific, temperatures are 213+7°K and 203+7°K respectively.

During September, the European and West Asian portions of the northern hemisphere and parts of North America begin to cool off rapidly. There temperatures at 60° N have cooled to about 223+7°K while over the northwest Atlantic and North Pacific somewhat warmer temperatures still prevail (228+7°K) during the week of 11–18 September (Figure 4). At the same time, the breakdown of the winter over the southern hemisphere is gaining impetus. The same, even intensified longitudinal patterns as observed in

August are apparent. Temperatures in the Indian Ocean warm area amount up to 234+7°K. This area has now spread out between the longitudes of 95° to 180° E along the 50th parallel, yet temperatures at this latitude over the South Atlantic and the eastern South Pacific still remain relatively cold (220+7°K). Temperatures at low and mid-latitudes, in general, range around 225+7°K (Figure 4).

The week following the equinox (25 September–1 October) shows that the temperature gradients between the two hemispheres (Figure 5) have reached their minimum during this transition period. With the exception of the very warm region at 50–55° S over the Indian Ocean reaching maximum temperatures of 236+7°K at 70° E temperatures over the entire quasi-globe are nearly constant at 225+7°K. It is significant to note that at most longitudes the temperature gradient along a given latitude belt can now be much greater than gradients between 60° N and 60° S: temperatures along 50° S vary from 236+7°K at 70° E to 223+7°K near 170° W, while along the 180th meridian, for instance, temperatures at 60° S are about 222+7°K, at 60° N about 228+7°K and at the equator about 223+7°K. This is a total variation of 6°K along the meridian compared to a variation of 13°K along the 50° S parallel (Figure 5). The phenomenon of very pronounced temperature differences between the Indian and Pacific Oceans at all latitudes between 40° S and 65° S persists during the entire transition period from winter to summer. The maximum temperature difference is about 15°K throughout this time, with the Indian Ocean area showing the warmest and the eastern South Pacific showing the coldest temperatures (Figure 5). This phenomenon first appears in large scale during the middle of August and lasts until about 20 October. During early and middle October, the eastern portion of the southern hemisphere (40 W to 120 E) is relatively warm (230+7°K at 60° S) with a warm center of

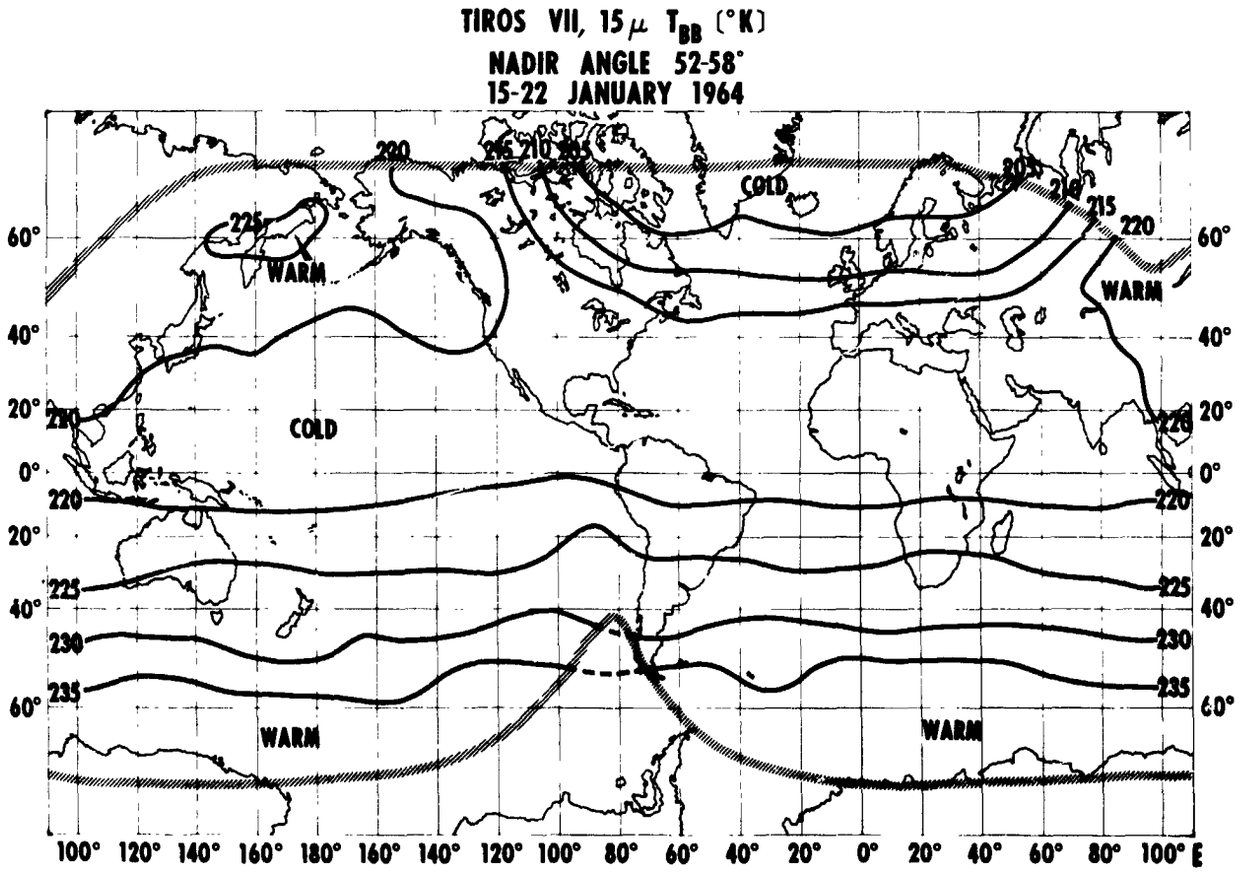


FIGURE 8(b).—Quasi-global map of 15μ equivalent blackbody temperatures averaged over the week 15–22 January, 1964; nadir angle 52°–58°. Radiation data cannot be obtained outside the area enclosed by the dashed border.

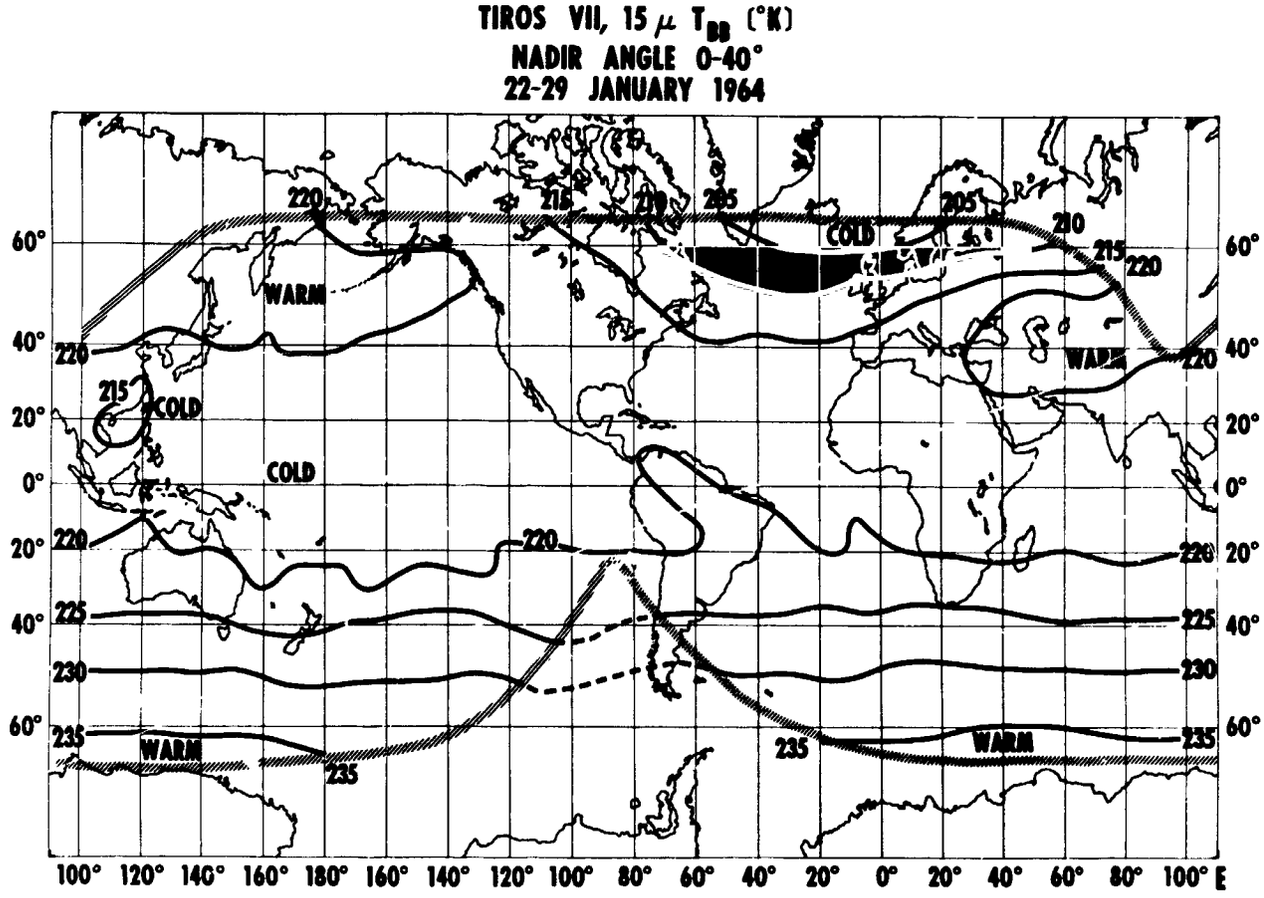


FIGURE 9(a).—Quasi global map of 15μ equivalent blackbody temperatures averaged over the week 22–29 January, 1964 nadir angle 0°–40°. Radiation data cannot be obtained outside the area enclosed by the dashed border.

234+7°K at 80° E, while the western portion remains relatively cold with the cold center at 220+7°K at 170° E. During late October, the week of 23rd to 29th, the entire southern hemisphere finally achieves a typical summer temperature pattern with wide and uniform latitudinal temperature belts ranging from about 230+7°K at 60° S to about 225+7°K at 30° S. There is a uniform belt with temperatures of almost exactly 225+7°K between 30° S and 30° N. Over the Atlantic, Europe and Western Asia, temperatures drop rapidly from 225+7°K at 30° N to 215+7°K at 60° N, while over the Pacific they remain at 225+7°K up to 60° N.

During the entire fall transition period, the north Pacific area has not cooled off as rapidly as the rest of the northern hemisphere and in general remains warmer throughout much of the winter. During the course of October, this pattern intensified whereby temperatures in all northern regions except over the north Pacific and eastern Asia continue to decline. Finally, at the end of October, as stated above, the northern hemisphere shows a typical winter pattern with the same tight gradients with latitude as observed three months previously in the southern hemisphere. At 60° N, temperatures are uniformly cold (215+7°K) over Eurasia, the North Atlantic and North America, but warm over Alaska and the Bering Sea (225+7°K). The global picture does not change appreciably between the end of October and the week 20–26 November shown in Figure 6. Although the warm area (224+7°K) over the North Pacific has somewhat diminished in size compared to the end of October, and it is now located over the Gulf of Alaska, the isotherms over the area still show a very definite departure from a zonal pattern and run northwest to southeast over North America and northeast to southwest over east Asia. While temperatures are still uniform with latitude over the Atlantic and Eurasian continent and range from 213+7°K at 60°N

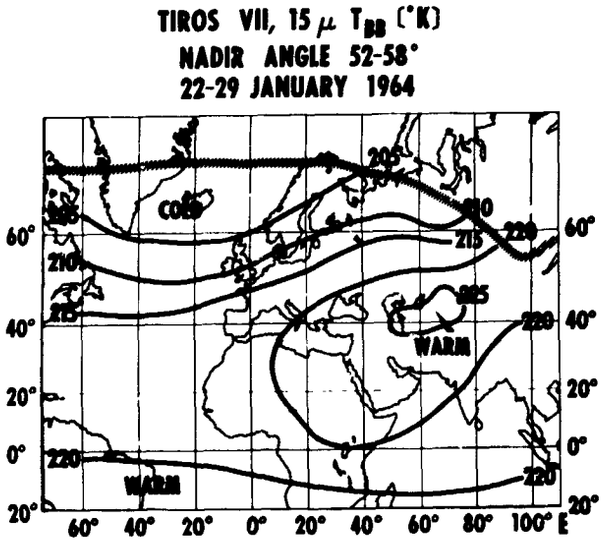


FIGURE 9(b).—Regional map of 15μ equivalent blackbody temperatures averaged over the week 22–29 January, 1964; nadir angle 52°–58°. The effects of a stratospheric warming over Eastern Europe are illustrated. No essential change in the temperatures occurred in other areas of the quasi-globe relative to Figure 8(b).

to 225+7°K at 40° N, the summer temperature pattern in the southern hemisphere is perfectly uniform with latitude and temperatures range from 235+7°K at 60° S to 225+7°K at 25° S (Figure 6). This pattern is identical to the temperature pattern during June, however, with northern and southern latitudes reversed.

C. The "Aleutian Anticyclone", December 1963

An unexpectedly warm stratosphere occurring as a regular climatological feature during winter in the general area of the Aleutian Island chain has long been recognized as the cause of a strong anticyclonic stratospheric circulation system in that area (Reference 10). Such circulation is unusual because in winter one would expect a very cold stratosphere resulting in a cyclonic circulation symmetric around the winter pole. Indeed, this cyclone exists, but in the northern hemisphere during mid-winter it is strongly disturbed and displaced by the Aleutian anticyclone. The morphology of this stratospheric warming over the North Pacific has received much attention in the past and a number of theories have been advanced attempting to explain its origin on the basis of dynamic processes (Reference 11). The satellite temperature measurements made it possible to follow the variation of mean temperatures in the lower stratosphere over this entire region throughout the winter of 1963. Figure 7a shows the temperature measured by the satellite between 40° W and 160° E and 20° N and 65° N averaged for the week from 5 to 12 December. Globally, there is practically no change from the patterns shown in Figure 6, except that the warm North Pacific region, which previously had remained at higher temperatures because it simply cooled slower than the other areas, has now actually warmed between the end of November and early December. It is centered at the end of the Aleutian Island chain and stands out clearly with temperatures of 232 +7°K. An analysis of temperatures at the 30 mb level from conventional radiosonde observations analyzed and obtained from the Freien Universität Berlin for the same area and time is shown in Figure 7b for comparison. The excellent agreement in the patterns between the two maps can serve as a measure of the validity of the satellite data. As can be seen from Figures 8, 9, 11, and 12, the Aleutian warm region is a permanent feature of the entire winter from the week of 5–12 December on. However, although temperatures within the warm center had not shown an

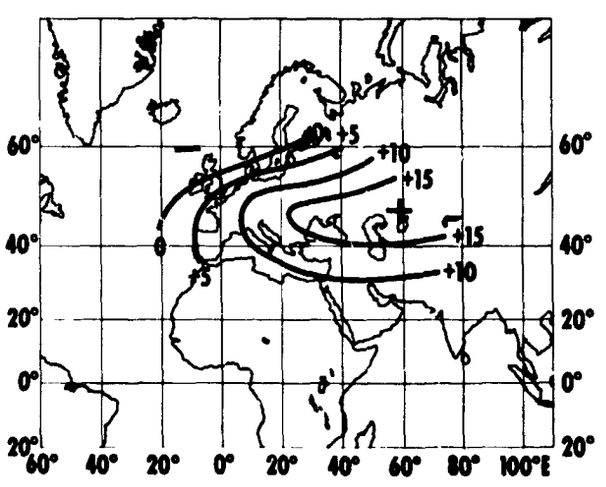


FIGURE 10.—Regional map of the differences between the 15μ equivalent blackbody temperatures measured on 27 January 1964 and those averaged over the week 9–15 January; nadir angle 0°–40°. The effects of a stratospheric warming over Eastern Europe are illustrated.

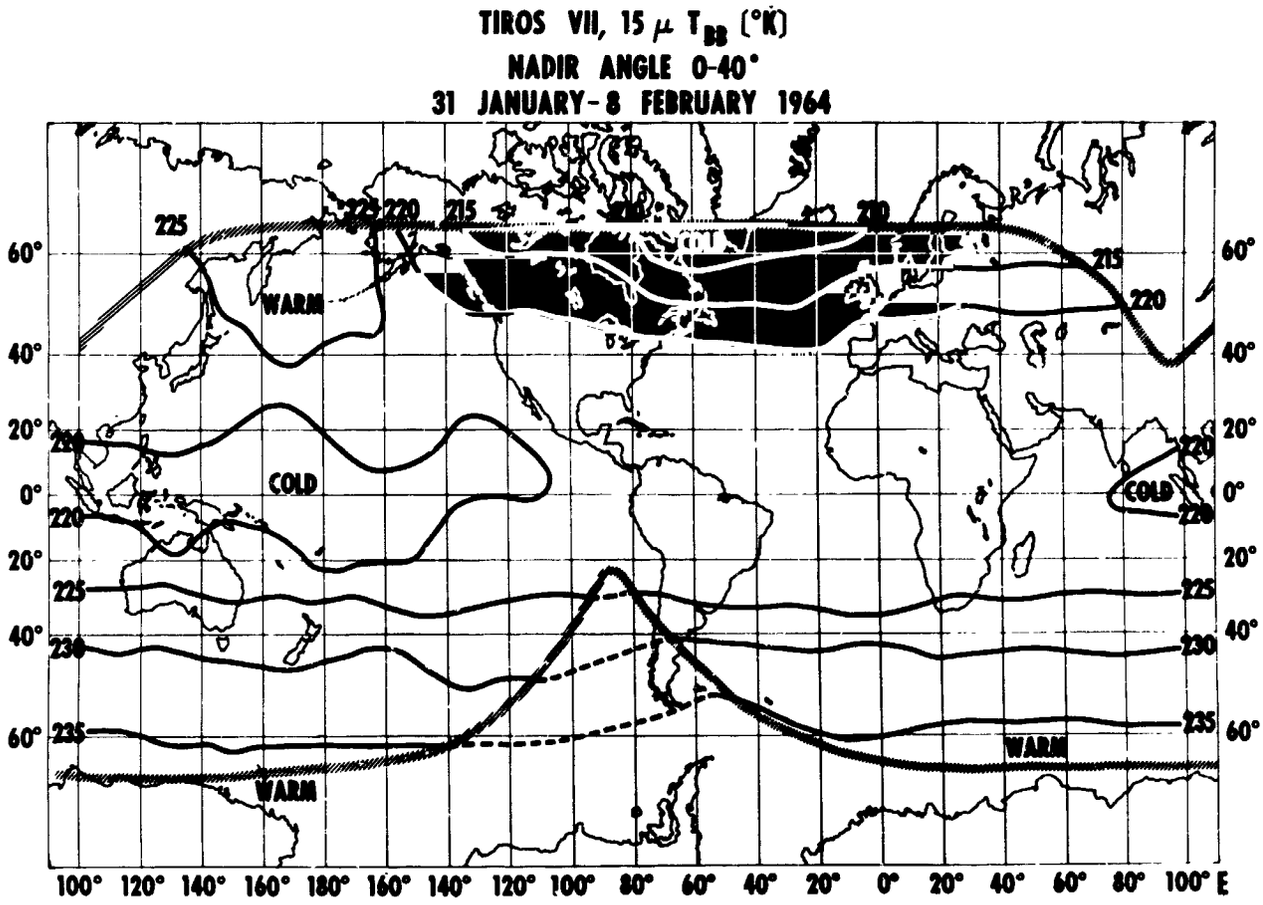


FIGURE 11(a).—Quasi-global map of 15μ equivalent blackbody temperatures averaged over the week 31 January-8 February 1964; nadir angle 0°–40°. Radiation data cannot be obtained outside the area enclosed by the dashed border.

actual increase until 5–12 December, a warm area in one form or another was present since the end of summer over the North Pacific as described above and shows in Figures 4, 5, and 6. This means that during this year the winter anomaly over the Pacific, though variable in extent during early winter, grows directly out of the summer conditions without ever reaching as low temperatures as other portions of the northern hemisphere do. Although, during early December, the system takes on a more permanent form and location and intensifies somewhat, it originated not because of a sudden heating phenomenon during winter, but simply because it refused to cool off at the end of summer. Naturally, there must be a mechanism to maintain the higher temperature during winter, while other regions at equal latitudes cool rapidly. It should be noted that those dynamic mechanisms which would have to depend on the existence of the complete winter cyclone could not explain the warm regions which already exist even before the cyclone is fully established. It might be interesting to investigate mechanisms by which the warm regions can be maintained throughout the winter by absorption of infrared radiation.

The satellite measurements show that the situation in the northern hemisphere is actually not vastly different from the southern hemisphere where a warm region over the southeastern Indian Ocean, though not quite as intense as the Aleutian one, persisted throughout the winter and then spread to make one half of the hemisphere almost 15°K warmer than the other. In the northern winter, almost a mirror image of this process occurs: the Pacific is warm and the Atlantic and Eurasia are cold, while in the south the Pacific is cold and the Atlantic and Indian Ocean areas are warm. In the north, the temperature gradients between warm and cold regions are somewhat greater (about 20°K between 180° W and 40° W at 60° N). Of course, so far satellite data exist only for the seasonal cycle 1963/64, and the ob-

servations described here must be tested by experiments in future years.

D. Solstice, December 1963

Between November 1963 and the middle of January 1964, there is very little change from the temperature patterns shown in Figure 6. As discussed above, the Aleutian warm center becomes somewhat stronger and quite stationary during early December and, in contrast, the Atlantic and Eurasian cold regimes remained unchanged with very tight latitudinal temperature gradients and isotherms aligned with latitude during the week 15–22 January 1964 (Figure 8a). It is interesting that the temperature difference of about 15°K between 60° N and the tropical belt over the eastern portion of the northern hemisphere is less than the longitudinal temperature difference between the Aleutian warm center and the North Atlantic at 60° N. A similar situation existed in the southern hemisphere at the end of winter (Figure 5); but during the solstice, the Australian warm area did not extend to such high latitudes as the Aleutian area does. As shown in Figure 8a, the mid-latitude and tropical regions have cooled considerably; and the temperature between 25° S and 40° N is now about 218+7°K, while earlier temperatures in the same zone ranged near 225+7°K. This remarkable cooling over so wide an area is particularly significant since it precedes by about one week the occurrence of a stratospheric warming over western Asia. Instrumental degradation cannot be responsible for such a temperature decrease, since Figure 12 does indicate higher temperature again in March. Also, ever since November temperatures at high southern latitudes remained unchanged. Relatively moderate and very uniform temperature zonal temperature gradients exist at all longitudes. Temperatures at 60° S are 235+7°K.

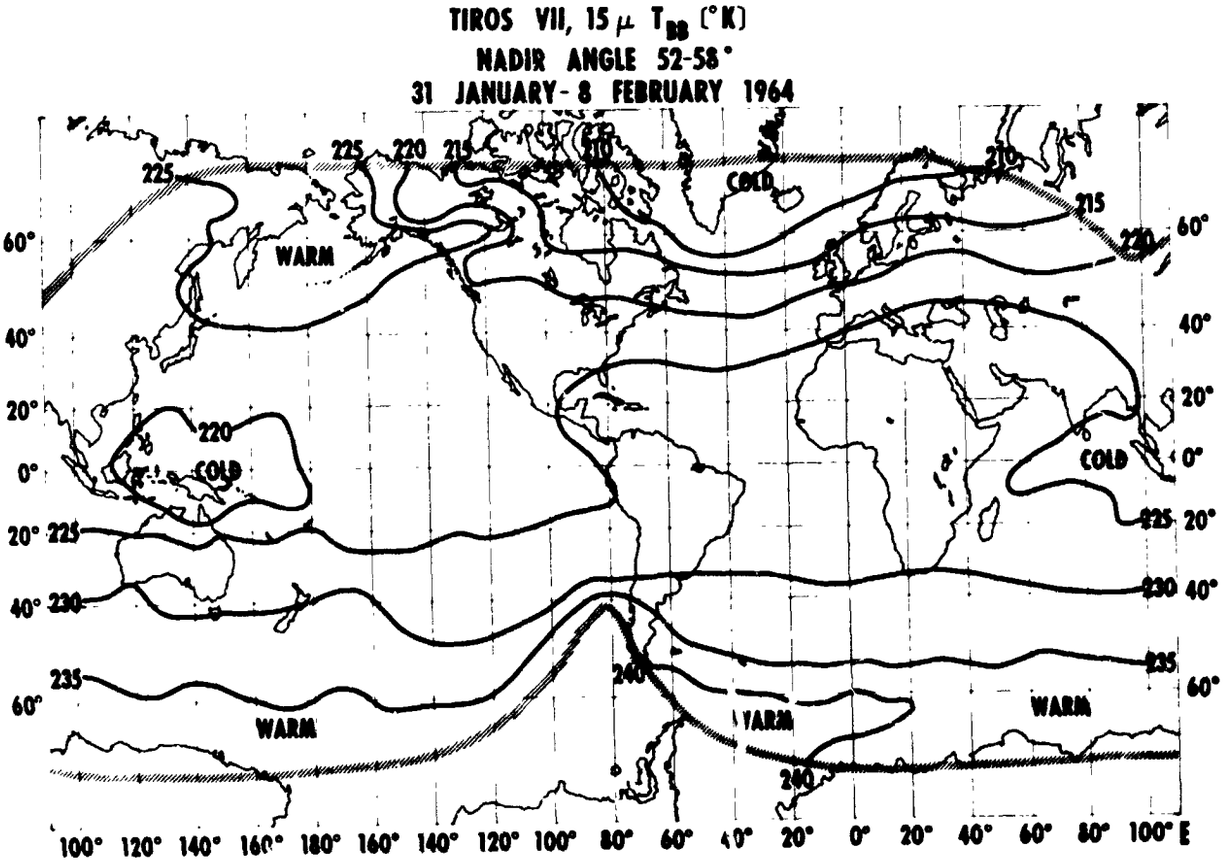


FIGURE 11(b).—Quasi-global map of 15μ equivalent blackbody temperatures averaged over the week 31 January–8 February 1964; nadir angle 52°–58°. Radiation data cannot be obtained outside the area enclosed by the dashed border.

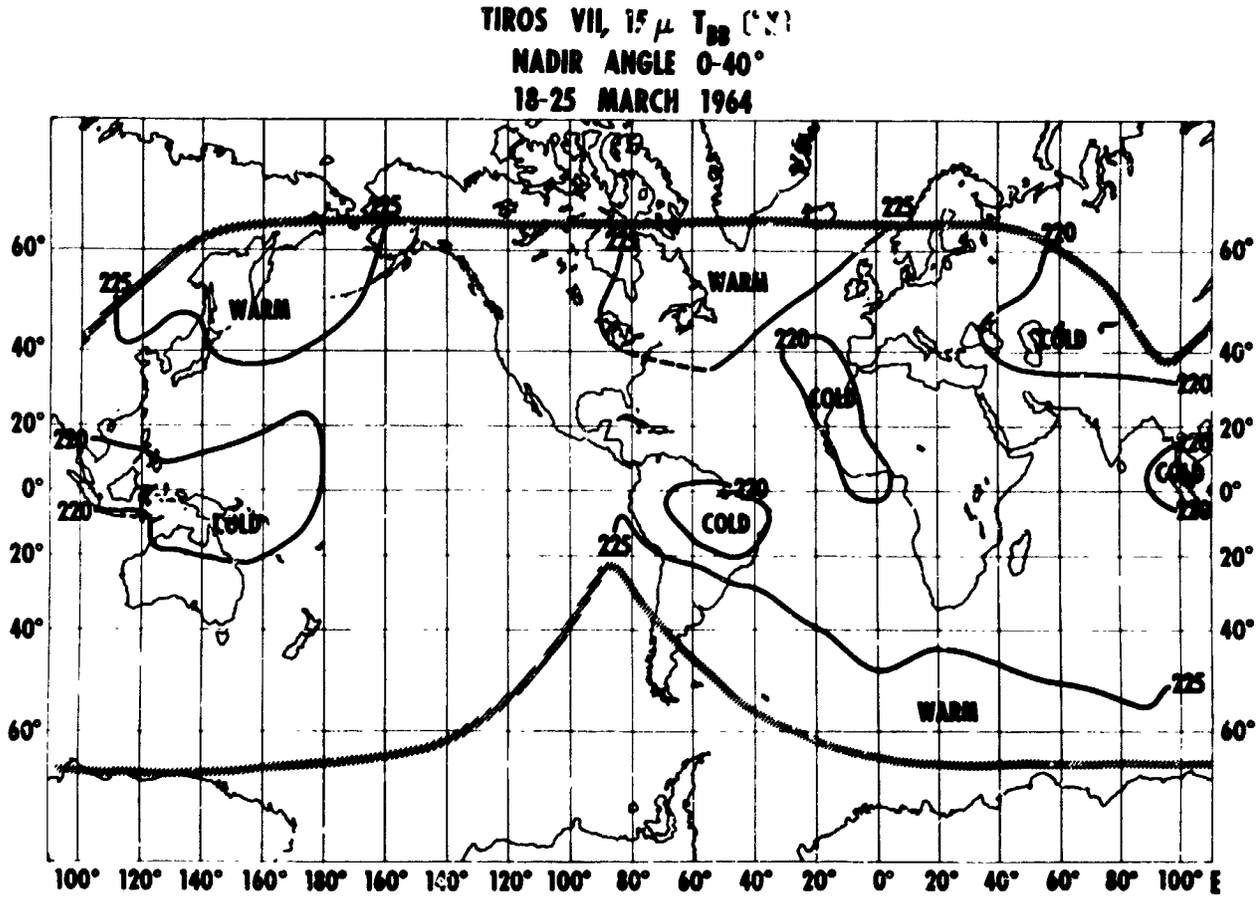


FIGURE 12.—Quasi-global map of 15μ equivalent blackbody temperatures averaged over the week 18–25 March, 1964; nadir angle 0°–40°. Radiation data cannot be obtained outside the area enclosed by the dashed border.

Temperatures of 231+7°K at 50° S and of 206+7°K at 60° N (eastern portion only) are compatible with the computed temperatures of 238 and 213°K respectively for the "summer" and "cold winter" profiles (Table I). Temperatures calculated for the "tropical" profile (227°K, Table I) are about 2°K higher than those measured in the now cold mid-latitude belt (218 +7°K). Mean temperatures of 232+7°K in the center of the Aleutian warm region measured during December suggest that the average temperature of the stratosphere in the 15 to 35 km region was about 12°K higher than that shown in the "warm winter" profile of Figure 2, which yields a mean temperature of only 229°K (Table I). Of course, such a temperature profile in this partic[illegible]ar altitude region is merely postulated since, for example, a low average stratospheric temperature of 205°K in the altitude region from 15 to 35 km, such as shown in the "cold winter" profile, Figure 2a, could be balanced by an extraordinarily high average temperature of 300°K in the region from 35 to 60 km to produce the measured radiation temperature of 239°K. But, such a temperature profile is much less probable than the one postulated. The validity of the assumption of an average stratosphere temperature increase of 12°K between 15 and 35 km is also suggested by comparing the satellite measurements at low nadir angles (0–40°, Figure 8a) to those at high nadir angles (52–58°, Figure 8b). Temperatures over the Aleutian warm area are practically the same at both nadir angle ranges indicating a nearly isothermal or at least a very slowly varying temperature gradient with altitude in the region where the weighting functions $\Psi(h)$ are at their maximum (15–35 km, Figures 2*b* and 2*c*). A temperature profile such as the one mentioned above with a mean temperature of 205°K between 15 and 35 km and 300°K between 35 and 60 km would produce a difference of 9°K between the radiation temperatures measured at 0° nadir and

at 58° nadir angle. Figures 8a and 8b, however, indicate no temperature difference at all over the Aleutian warm area. Thus, one finds that the difference in the mean temperatures between 15 to 35 km level and the 35 to 50 km level cannot be excessively large, and with that restriction, a temperature increase of 10°K for the standard "warm winter" profile from 15 to 35 km follows for the North Pacific area. Similarly, one finds that there is no temperature differences for the two nadir angles over the North Atlantic and Northern Europe (Figures 8a and 8b) which agrees well with the calculated differences of 1°K for the "cold winter" profile. The larger temperature differences over the tropics—the 220°K isotherm is about 10° further north in Figure 8b than in Figure 8a—indicate the rapid temperature increase with altitude above 15 km and agree well with the computed difference of 4°K for the "tropical" profile (Table I); while the calculated values of 238 and 240°K for "high latitude summer" (Table I) agree well with the observed temperatures (+7° correction) and temperature differences between 50 and 55° S.

This picture has changed considerably during late January (Figures 11a and 11b) where temperature differences at mid-latitudes and in the tropics over the entire eastern hemisphere are as high as or exceed the differences observed previously within the tropics. Such large differences exceeding 5°K extend particularly into North Africa, the eastern Mediterranean and the Near East (Figures 11a and 11b). This indicates that neither the "tropical" nor the "summer" temperature profiles (Figure 2a) apply in this case. In that area, temperatures at 15 km must be somewhat higher than those given by the "tropical" profile; and there must be a pronounced increase of temperature with altitude resulting in a very warm upper stratosphere, i.e. a very strong vertical temperature gradient between the lower and upper stratosphere. It is significant that this change has occurred over the area into which the sudden warming, described below, has moved in its decaying stages.

E. Sudden Warming, January 1964

The type of temperature measurement from a satellite described here was expected to be most useful for the detection of the occurrence and the observation of the behavior of stratospheric warmings. These warmings were first observed over Berlin by Scherhag (Reference 12) and have since then been the subject of intensive observation and analysis (Reference 13). The sudden warmings have now been observed almost every year during the northern hemisphere winter and, during their occurrence, the lower stratosphere over a limited area may be heated by as much as 30°K within a few days. Causes for such warmings have thus far not been explained. In contrast to the Aleutian warm center described above, the sudden warmings show no regularity in their time and location of occurrence. In order to be detectable by the TIROS radiometer, the warming must occur at altitudes between 15 and 35 km, or, if it occurs at other levels, it must be unproportionally intense. Calculations based on the "warm winter" temperature profile of Figure 2a show that a uniform temperature increase of 12°K between 15 and 35 km will increase the radiation temperature by 10°K, while a similar change of the temperature profile between 34 and 44 km will cause a rise in the radiation temperature of less than 1°K.

A number of small and weak warmings occurred over America and Europe during December 1963 and January 1964, but those did not penetrate to altitudes below 30 km (Reference 14). Therefore, they could not be detected by the satellite. During the last week of January (22–29), however, a warming of apparently moderate intensity occurred over the region of the Caspian Sea, extending between the Himalayan Mountains and the Black Sea (Figure 9a). Radiation temperatures over that area have increased by about 8°K over the previous week (Figure 8a). During that week (15–22 January), there was practically no indication of any disturbance in the perfect latitudinal temperature structure over that region, except perhaps for a very small area over Iran where a temperature increase of about 2–3 degrees could be observed; but in the high nadir angle measurements, there is an indication of increased temperatures over South Central Asia (Figure 8b) inferring that the subsequent warming might have originated in the upper stratosphere (above 30 km) in that area. Unfortunately, the warming developed in an area where no conventional meteorological data at these altitudes are

available, and the satellite observations provide the only method of analyzing the event. Furthermore, even the satellite observations are limited in coverage, since this area is located near the data gap over the central Soviet Union which exists because of the particular geometry of the data readout stations for TIROS (see above). The satellite data were, therefore, analyzed for individual days of 22, 24, 25, 26, 27, and 29 January. These daily analyses showed that the temperatures within the warm center were as high as 230+7°K on 29 January, but because they extended only over very small regions and because of the motion of the center from day to day, these high temperatures became obscured in the weekly averages (Figure 9a). The first definite indication of the warming occurs on 22 January, where a small and relatively weak warm center of 225+7°K appears over Pakistan, Afghanistan and Eastern Iran. This is already a few degrees higher than the average temperature over the same area for the preceding week of 14–22 January (Figure 8a). Two days later on 24 January, the warm region had spread out somewhat to the north, and the center had also moved in that direction but had not significantly intensified. On 25 January, temperatures of 225+7°K range over the entire middle East extending as far eastward as central China. The center is now just east of the Caspian Sea and has just barely reached a temperature of 230+7°K. On 26 January, the same situation prevails with the center becoming slightly warmer and larger. On 27 January, the center has grown considerably larger and temperatures of 230+7°K extend over an area of at least 30° longitude and 10° latitude northeast of the Caspian Sea. The 225+7°K isotherm has advanced to a latitude of 50° N, causing a tremendously tight temperature gradient over northern Europe where temperatures increase by more than 20°K from southern Norway to the Black Sea. Due to the warming, isotherms have been tilted to follow a course from southeast to northwest instead of the perfect eastwesterly direction still prevailing on 22 January. Finally, on 29 January a warm center extending over 10° great circle arc with temperatures of 230+7°K stretches from the Black Sea eastward to China. One 225° isotherm now runs east-west along the 40th parallel between 40° E and 80° E. The other runs southwest to northeast from the Black Sea to about 50° E and 55° N. The center is located near 65° E and 50° N. This situation is summarized in Figure 10 which shows that on 27 January the radiation temperatures had increased by 15°K over the averages for the week 9–15 January near the center and that the periphery of the warming extended as far west as the North Sea, and the Atlantic and as far south as North Africa and Arabia. Figure 11a shows that during the week 31 January to 8 February the warming had passed its climax. Rather homogeneous temperature patterns with latitude return over northern Europe and eastern Asia, but south of 40° N a broad warm area of about 223+7°K now covers the entire region between 40° N and 25° S latitude and 110° W and 80° E longitude (Figure 11a). This means that the mid-latitude belt which had cooled so drastically before the warming had been restored close to its normal temperature of about 223+7°K between 25° S and 40° N. Figures 8, 9, and 11 show that the Aleutian warm center had remained undisturbed during the entire period of the warming. On 30 January, the warming was detected by a radiosonde ascent over Berlin, the only one available for this analysis (Reference 15). This radiosonde temperature profile, when compared to a prewarming profile shows that the warming over Berlin occurred primarily above 30 km. For the week of 22–29 January, temperatures at high nadir angles are about 5°K higher than at low nadir angles not only over the warm center, but also extending far beyond the center, particularly along the southern and western periphery of the warming (Figures 9a and 9b). This means that over the center temperatures have increased not only in the lower stratosphere but more so in the upper stratosphere. The much greater extent of the warming in the high nadir angle temperature patterns compared to the low nadir angle patterns suggested that along the western and southern periphery the warming occurred only in the upper stratosphere. This confirms the above mentioned radiosonde observation over Berlin. Since the much larger warming (13°K) in the center would require an unreasonably large temperature if the warming were confined to altitudes above 30 km, and since such a strong vertical temperature gradient would be clearly noticeable as a much

greater temperature increase at the larger nadir angles, it must be concluded that the warming near the center penetrated to much lower altitudes where the Ψ functions are at maximum. If the warming occurred uniformly at altitudes from 15 to 35 km, a temperature increase of about 15°K would satisfy the radiation measurements. Such an increase is quite commensurate with a moderate stratospheric warming.

F. Equinox, March 1964

The global temperature structure undergoes a major change during the period from early February (Figure 11a) to the week of 18–25 March shown in Figure 12. The steep zonal winter time temperature gradients at high northern latitudes, particularly over the eastern hemisphere as well as the zonal summer pattern in the southern hemisphere have disappeared. The southern hemisphere has cooled considerably, and parts of the northern hemisphere have warmed. This results again, as during the September equinox, in an almost constant temperature over the entire quasi-globe. In fact, temperature gradients, longitudinally or latitudinally are even shallower than during September. Over the eastern Pacific, there is no temperature gradient at all between 60° S and 60° N, with temperatures of about 223+70°K prevailing at all latitudes. As the southern hemisphere cools, a temperature gradient along 60° S seems to develop similarly as during the previous winter. The South Indian and Atlantic Oceans are now about 5°K warmer than the South Pacific.

In the northern hemisphere near 60° the North Atlantic and northeast North America have warmed rapidly by about 20°K since February, while the region east of the Black Sea, where the sudden warming occurred in January, has returned to its pre-warming winter temperatures of 217+7°K. Temperatures in the Aleutian region have remained practically constant throughout the winter and are still at about 228+7°K. This produces temperature differences of about 10°K between the Atlantic and Pacific warm regions and central Asia at 50° N. That is the largest temperature difference found during this equinox (Figure 12). Since these data have been reduced only very recently, there has not yet been an adequate opportunity to analyze the January to March transition as thoroughly as the June to October period. Such an analysis will be made in the future.

IV. CONCLUSIONS AND SUMMARY

Satellite measurements of emitted radiation in the 15 micron carbon dioxide band have demonstrated that mean temperatures in the lower stratosphere can be mapped on a global scale. By comparing measurements at high and low nadir angles, a qualitative inference can be made about the vertical temperature gradient between the lower and upper stratosphere. Analysis of global temperature patterns from June 1963 to March 1964 revealed that temperature patterns in both hemispheres remain relatively undisturbed during summer. These patterns are oriented very precisely along latitude circles with the temperature maximum occurring at the summer pole. Near maximum temperatures are reached in the southern summer about one month after the equinox. After that, temperatures increase only slightly through the solstice and remain high until the next equinox. The northern summer could only be analyzed through its peak and declining phases, June to September 1963. Temperatures decline very slowly during the first two months following the solstice. Then, a rapid decline occurs just before the equinox. During both equinoxes, temperatures were relatively uniform and differences of about 10–15°K may exist both with latitude and longitude. Winter temperature patterns in both hemispheres exhibit strong temperature gradients between the pole and 40°, but a completely uniform pattern never exists in either winter hemisphere. A fairly uniform region with temperatures generally around 230°K extends between 25° latitude of the summer hemisphere and 40° latitude of the winter hemisphere during solstice. The satellite measurements not only permitted for the first time a continuous analysis of the southern hemisphere from June 1963 to March 1964 up to a latitude of 65° S, but also demonstrated the usefulness of uniform and continuous data coverage over the entire quasi-globe. A warm area over the Indian Ocean, though initially small, seems to play a similar role in the southern winter as the Aleutian warm center during the northern winter. During winter 1963, this warm region grew out of a disturbance in the isotherms

during June and finally divided the entire stratosphere of the winter hemisphere into a "warm" and "cold" portion with the warm portion located between Africa and Australia. Temperature differences between these warm and cold areas amounted to about 15°K and disappear with the warming of the South Pacific area during late October, giving way to the homogeneous summer pattern. Similarly, there is no homogeneous winter pattern in the northern hemisphere. A large region over the North Pacific does not cool off as rapidly after the equinox as the rest of the stratosphere. Just before the solstice, this region becomes well established and intensifies in the area where the Aleutian anticyclone is usually found.

The morphology of a stratospheric warming over southwestern Asia and its penetration to Europe was observed by the satellite indicating that heating of possibly 20°K penetrated down to at least 20 km over the center of the warming, but at the periphery warming of perhaps somewhat greater magnitude remained at altitudes above 30 km. The persistent climatological inhomogeneities of warm cells over both winter hemispheres lead to the speculation that during winter the stratospheric temperatures may be influenced by radiation from underlying surfaces, while during summer solar heating is dominant.

A study was also made of the effect of how high altitude clouds on the measured temperatures, since the weighting functions (Figure 2b) are not negligible at 10 km. The result indicated that very large and high cloud systems such as hurricanes do decrease the measured temperatures by as much as 10°K if the clouds reach above 10 km and by less than 4°K if the clouds, filling the field of view are at 6 km; but it was shown that such effects are only isolated and are averaged out by the weekly analyses which formed the basis for this investigation.

ACKNOWLEDGMENTS

Experiments resulting in findings such as the ones presented here always require the efforts of many hundreds of dedicated individuals. We gratefully acknowledge the contributions of those who have contributed to the success of the TIROS VII experiments. We are particularly indebted to Mr. Robert Hite and his team for the computational efforts in reducing the vast amount of radiation measurements to the maps presented here.

BIBLIOGRAPHY

1. Bandeen, W. R., Hanel, R. A., Licht, J., Stampfl, R. A. and Stroud, W. G.: "Infrared and Reflected Solar Radiation Measurements from the TIROS II Meteorological Satellite," *J. of Geophys. Res., 66*, 3169–3185, October 1961.
2. Hanel, R. A., Bandeen, W. R., and Conrath, B. J.: "The Infrared Horizon of the Planet Earth," *J. of Atmos. Sciences, 20*, 73–86, March 1963.
3. Bandeen, W. R., Conrath, B. J., and Hanel, R. A.: "Experimental Confirmation from the TIROS VII Meteorological Satellite of the Theoretically Calculated Radiance of the Earth within the 15 Micron Band of Carbon Dioxide," *J. of Atmos. Sciences, 20*, 609–614, November 1963.
4. "TIROS VII Radiation Data Catalog and Users' Manual," NASA/Goddard Space Flight Center, Greenbelt, Maryland (to be published).
5. Elsasser, W. M., with Margaret F. Culbertson: "Atmospheric Radiation Tables," *Meteorological Monographs*, 4, 23, 43 pp., August 1960, American Meteorological Society.
6. "TIROS II Radiation Data Users' Manual" NASA/Goddard Space Flight Center, Greenbelt, Maryland, 15 August 1961.
7. "TIROS III Radiation Data Users' Manual Supplement," NASA/Goddard Space Flight Center, Greenbelt, Maryland, 1 December 1963.
8. "TIROS IV Radiation Data Catalog and Users' Manual," NASA/Goddard Space Flight Center, Greenbelt, Maryland, 15 December 1963.
9. Bandeen, W. R., Conrath, B. J., Nordberg, W. and Thompson, H. P.: "A Radiation View of Hurricane Anna from the TIROS III Meteorological Satellite," NASA/Goddard Space Flight Center, Greenbelt, Maryland (to be published in Tellus, 1964).
10. Wege, K.: Druck-, Temperatur- und Strömungsverhältnisse in der Stratosphäre über der Nordhalbkugel, Meteorol. Abhandl. d. Freien Universität Berlin, Bd. V—, Heft 4 (1957).
11. Boville, B. W.: "What are the Causes of the Aleutian Anti-cyclone?" *Proceedings of the International Symposium on Stratospheric and Mesospheric Circulation 20 to 31 August, 1962, Berlin, Germany, Meteorologische Abhandlungen*, Institut für Meteorologie und Geophysik der Freien Universität Berlin.
12. Scherhag, R. A.: "Die Explosionsartigen Stratosphärenerwärmungen des Spätwinters 1951/52." *Berichte des Deutschen Wetterdienst in der U.S. Zone, 38*, 51–63, 1952.
13. Reed, R. J.: "On the Cause of the Stratospheric Sudden Warming Phenomenon." *Proceedings of*

the International Symposium on Stratospheric and Mesospheric Circulation 20 to 31 August, 1962, Berlin, Germany, Meteorologische Abhandlungen, Institut für Meteorologie und Geophysik der Freien Universität Berlin.

14. Teweles, S.: "Stratowarm Alerts 1964" dissemination by *IQSY* Stratwarm Agency, Stratosphere Analysis Project, U.S. Weather Bureau, Washington 25, D.C.
15. Stratospheric Analysis Chart for 1963–1964, *Meteorologische Abhandlungen*, Institut fur Meteorologie und Geophysik der Freien Universitat Berlin.
16. Cole, Allen E., Kantor, A. J., and Court, A.: "Supplemental Atmospheres," *Interim Notes on Atmospheric Properties—32* (*Rev.*), 6 June 1963, Air Force Cambridge Research Laboratories, Office of Aerospace Research, U.S. Air Force, Bedford, Mass.

N64-31317

CLOUD HEIGHTS AND NIGHTTIME CLOUD COVER FROM TIROS RADIATION DATA*

S. I. RASOOL†

Radiation data obtained from TIROS III have been analyzed, separately for day and for night, for the period July through September 1961. The global distribution of the average effective temperatures measured by the 8–12μ channel of the satellite radiometer shows a close correlation with the cloud cover data.

An estimate of the latitudinal distribution of cloud heights has been obtained using the TIROS radiation data for daytime and the distribution of cloud cover recently obtained from the TIROS photographs. Combining these values of the cloud heights with the nighttime radiation data determines the latitudinal distribution of nighttime cloud cover.

The results indicate that in the Southern Hemisphere the percentage cloudiness at night is considerably higher than in the day, while in the case of the Northern Hemisphere the cloudiness appears to decrease at night.

1. INTRODUCTION

In a recent paper, Arking (1964) has derived the global distribution of cloud cover by the analysis of TIROS III video pictures. In Fig. 7 (Curve 1) is plotted the latitudinal distribution of average *daytime* percentage cloud cover as obtained by Arking for the period of 12 July through 10 September 1961.

In the present study we attempt, using these data and the simultaneous radiation measurements by TIROS III, to derive the average latitudinal distribution of *cloud top heights* and thence the *nighttime* cloud cover. Curve 2 in Fig. 7 shows the resulting percentage cloud cover during the night as a function of latitude.

2. TIROS INFRARED DATA

The radiation instrumentation in TIROS and the physical significance of this experiment have already been described by several authors (Bandeen *et al.*, 1961; Nordberg *et al.*, 1962). Three of the five channels of the TIROS radiometer measure terrestrial radiation in the far infrared corresponding to wavelength intervals of 5.8–6.5 μ, 8–12 μ, and 7–30 μ. The other two channels record the solar radiation in the visible, as reflected by the earth, to obtain an estimate of the albedo of any region. The wavelengths of the infrared channels have been chosen so as to provide information on the temperature at different levels of the atmosphere and to give an estimate of the total outgoing radiation in the infrared. In this discussion we shall only be concerned with Channel 2, sensitive in the 8–12 μ region of the infrared.

The main absorber of infrared radiation in the earth's atmosphere is water vapor. The absorption coefficient, however, varies considerably with wavelength between 4 and 50 μ, where almost all the energy of the terrestrial radiation is confined. In the 8–12 μ spectral interval the absorption by water vapor and also by CO_2, the other important absorber of the infrared radiation, is at its minimum. Consequently, *in the absence of clouds*, Channel 2 of the satellite radiometer, measuring in this atmospheric "window," records energy originating from very near the ground corresponding to an effective temperature only ~5–10 deg K less than the ground temperature (Wark, Yamamoto and Lienesch, 1962; Prabhakara and Rasool, 1963).

Reasonably thick clouds, however, are practically opaque to infrared radiation of wavelengths >4 μ (Shiffrin, 1961). As there is very little water vapor above the clouds, and if we may assume that the albedo of the clouds in the far

*Published in *Journal of the Atmospheric Sciences*, 21(2):152–156, March 1964.

†Goddard Space Flight Center and New York University, New York.

infrared is negligible, the radiometer *in the presence of thick clouds* will record energy corresponding to the cloud top temperature. This may be as much as 40–50 deg K lower than the surface temperature.

If, therefore, for any given instant, both the TIROS measured temperatures in this "window" channel and the actual surface temperatures are available, a knowledge of the temperature lapse rate in the atmosphere will enable one to obtain a quick estimate of cloud top height. On the other hand, on a climatological basis, if the satellite-measured temperatures for a given region are averaged over a season, then the departure of this temperature from the mean ground temperature will be mainly dependent on three variables—cloud amount, cloud height, and water vapor distribution in the atmosphere. As mentioned earlier, the effect of water vapor is comparatively small, and from our knowledge of climatological distribution of the water vapor over the globe, its effect on the temperatures measured by the satellite can be accounted for.

The strong effect of clouds on the TIROS III measured temperatures in the "window" channel is demonstrated by Fig. 1. In this figure are plotted the regional and temporal averages of the temperatures measured by Channel 2 of TIROS III during daytime (0600–1800 hours Local Time) for the period of July 12 through 10 September 1961. Measurements for the polar regions are not available because the inclination of the orbit was only about 48°. The surface of the earth between 50N and 50S has been divided into a 10 deg latitude by 10 deg longitude grid, and all measurements taken by the satellite in each grid at nadir angle <25 deg, between local time 0600 and 1800, have been averaged for this period.

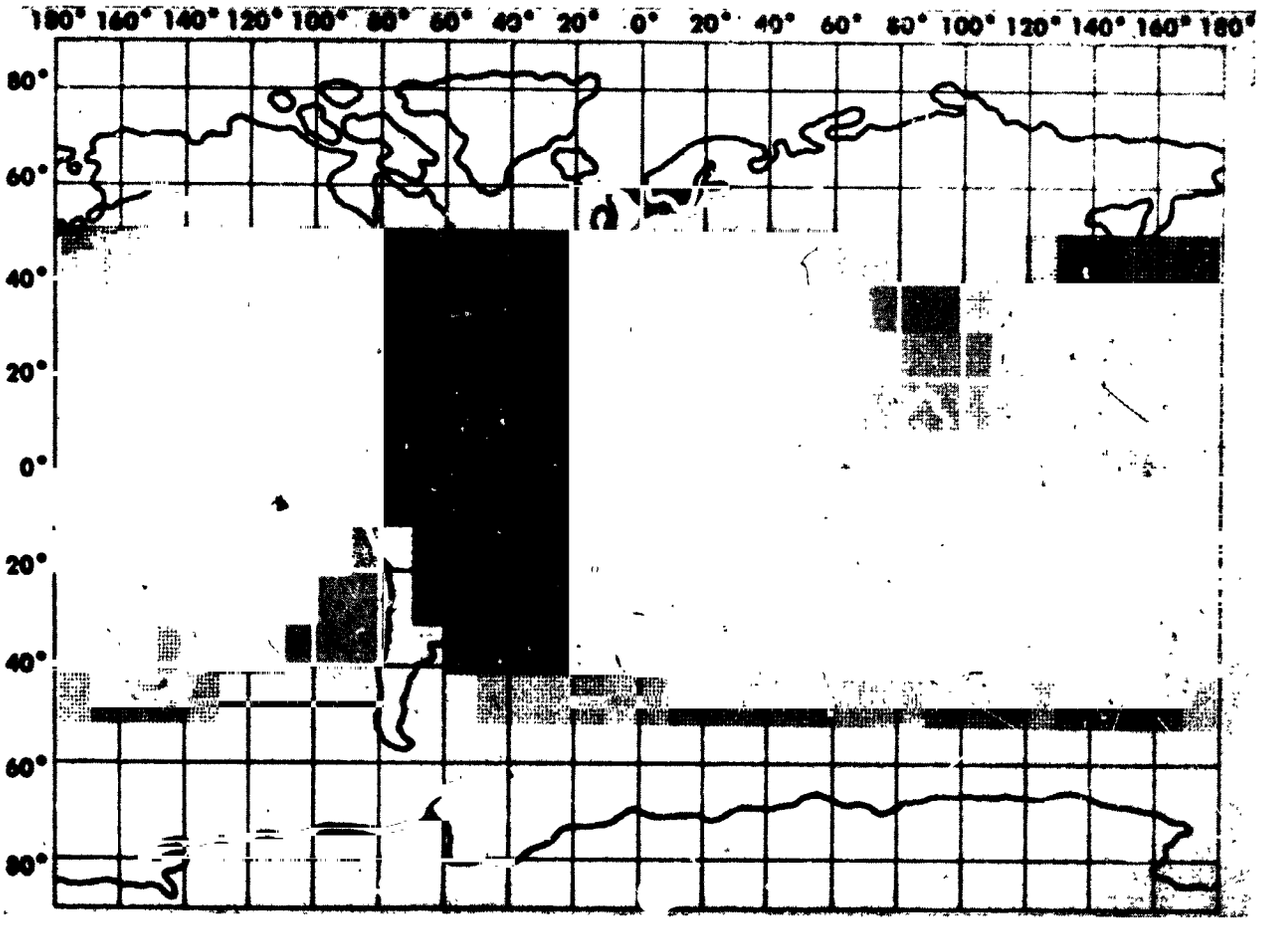


FIGURE 1.—Global distribution of average temperatures measured by Tiros radiometer channel 2 for the period 12 July through 10 September 1961. Darkest shade, $T > 295$K, lightest, $T < 255$K.

In most cases there are more than 500 observation points per grid. The observations were corrected for the instrumental degradation (Bandeen, private communication).

The very dark shades correspond to high radiation intensities with effective temperatures ranging above 290K. The lightest shades correspond to temperatures of the order of 240K. The two triangular areas, which are diagonally opposite and comprise a part of South America and Southern Siberia, have been left blank because of the unavailability of the telemetry from the TIROS for these regions.

Several interesting features are revealed by examining this figure. 1) The relatively cold temperature belt near the equator corresponds to the high cloudiness usually observed in the equatorial region. 2) The subtropical belt in each of the hemispheres show relatively higher temperatures and, therefore. probably fewer clouds than in the equatorial regions. 3) Extremely low temperatures measured over East Pakistan and India imply a heavy cloud cover over these regions which is, in all probability, the monsoon activity of this season. An almost quasipermanent cloud cover over central Africa is also indicated from this diagram.

These conclusions are supported by comparing Fig. 1 with Fig. 2 in which is plotted the global distribution of cloud cover, also in a 10×10 degree grid, taken from the climatological estimates of Haurwitz and Austin (1944). In Fig. 2, the darkest shades correspond to 20 percent cloudiness, and the lightest to 70 per cent cloudiness. The agreement in the main features of the two figures is striking, except that the TIROS III measured extremely high temperatures over the South Pacific, indicating a relatively clear area which is not noticeable in Fig. 2, undoubtedly due to lack of observational data in this region.

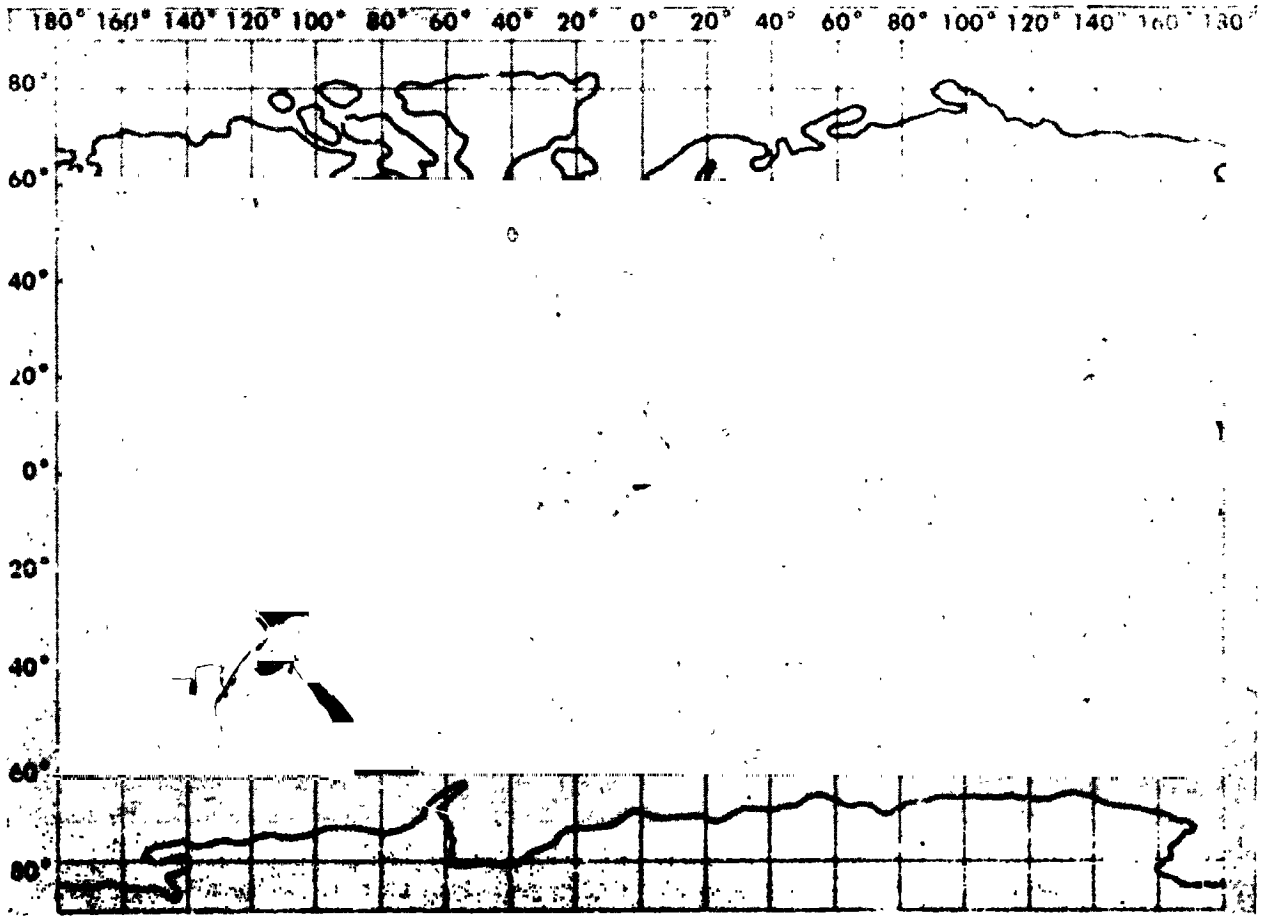


FIGURE 2.—Global distribution of average cloud cover for July (after Haurwitz and Austin, 1944). Darkest shade—cloud cover <20 percent, lightest shade—cloud cover >70 percent.

3. DEDUCTION OF CLOUD HEIGHTS AND CLOUD COVER

Although it is interesting to note the resemblance between Figs. 1 and 2, the actual derivation of cloud amounts from the TIROS radiation data is fairly involved. As mentioned earlier, several factors, viz., cloud amount, cloud height, surface temperature, vertical distribution of water vapor and of temperature, influence the 'effective' temperatures measured by the satellite. By the method described below we have, however, attempted to separate the effect of these parameters in order to obtain a reasonable estimate of cloud height from TIROS radiation data.

Using the climatological estimates of the global distribution of surface temperature (Haurwitz and Austin, 1944), the latitudinal variations in the vertical distribution of temperature (Davis, 1962, London, 1957), the total amounts of water vapor and ozone and their vertical distributions, we have constructed ten different model atmospheres which correspond to the ten latitudinal belts between 50N and 50S for the months of July and August.

The radiation flux in the 8–12 μ region, which would be emitted from such atmospheres, was then calculated following a method described in an earlier paper (Prabhakara and Rasool, 1963). The flux values thus calculated give the equivalent black body temperatures TIROS will observe above these atmospheres *if the atmosphere were completely cloudless.* Similar calculations were then repeated assuming the presence of thick clouds at various altitudes and of varying proportions. It is assumed that the clouds are completely opaque to the far infrared and the radiation is emitted only from the top of the clouds.

In Figs. 3 and 4 the results obtained for two such model atmospheres are shown. The family of five curves shows the difference (ΔT) between the surface temperature and the effective black body temperature observed by the satellite in the presence of different amounts of clouds located at altitudes of 2, 3, 4, 5, or 6 km. The intersection point of these curves with the ordinate in Fig. 3 is at 11.5 deg. This indicates that in a hot and humid atmosphere, even in complete absence of clouds, the satellite would observe an effective temperature about 12 deg lower than the surface

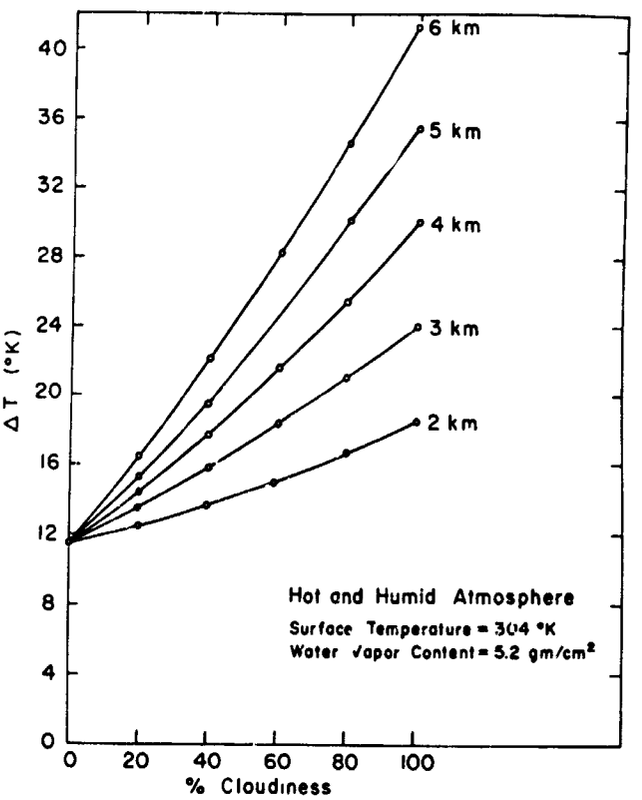


FIGURE 3.—ΔT as a function of percentage cloud cover with cloud tops at various altitudes. ΔT is the difference between the average surface temperature and the equivalent blackbody temperature measured by the Tiros III, channel 2. A case of hot and humid atmosphere.

temperature. In Fig. 4 is shown another extreme case, this time of a dry and cold atmosphere where the ΔT, in the absence of clouds, is only 3.5 deg.

If, therefore, for any given region, one has an estimate of surface temperature, of the amount of water vapor in the atmosphere and of the percentage cloud cover, then with such a family of curves, combined with the TIROS radiation data of the "window" channel, one can derive an approximate value for the effective height of the cloud tops.

In order to determine the latitudinal distribution of the effective cloud heights for the months of July and August 1961, we calculate the following parameters for each 10-deg latitudinal belt between 50N and 50S.

The satellite-observed temperatures in the "window" channel are averaged separately for day and for night for the period 12 July through 10 September 1961.

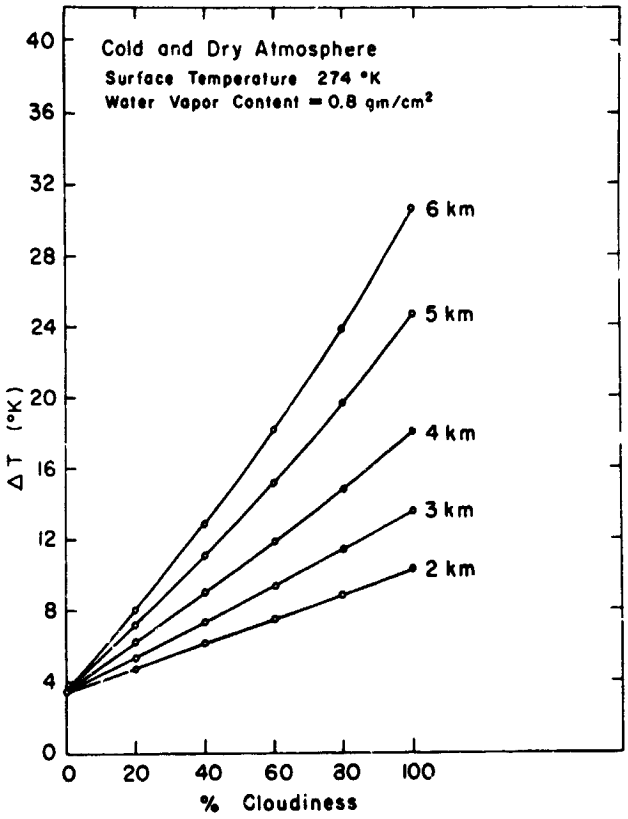


FIGURE 4.—Same as Figure 3—a case of cold and dry atmosphere.

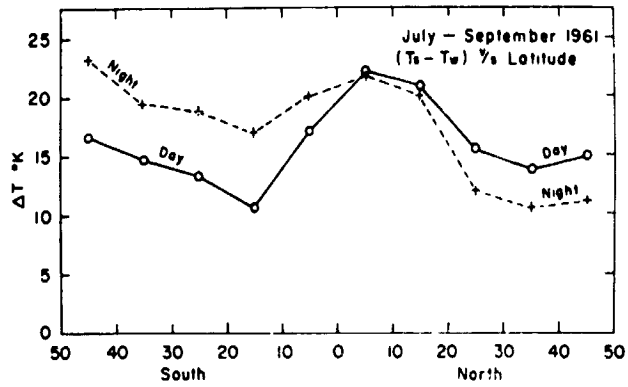


FIGURE 5.—ΔT as a function of latitude—for day and for night.

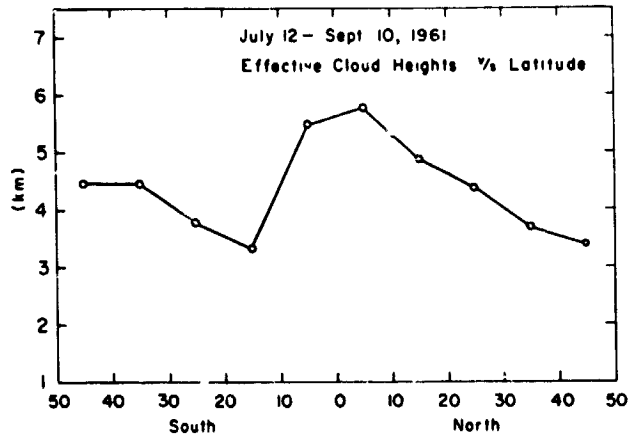


FIGURE 6.—Effective cloud top heights deduced from the daytime observation of Tiros III as a function of latitude.

The latitudinal means of day and night surface temperatures for the same period were determined from the *Monthly Climatic Data of the World* (U. S. Weather Bureau); supplementing these data were climatological estimates of Haurwitz and Austin (1944) for the oceanic regions.

The difference (ΔT) between the mean surface temperatures and the average effective temperatures measured by the TIROS channel 2 in each latitudinal belt is plotted in Fig. 5 separately for day and for night.

Latitudinal averages of daytime percentage cloud cover for the same period given by Arking are shown in Fig. 7 (Curve 1).

Now with both the ΔT and percentage cloud cover being known for daytime, we can use our nomograms (e.g., Figs. 3 and 4) to determine the average daytime cloud height for each latitudinal belt. The heights thus deduced are plotted as a function of latitude in Fig. 6.

If we assume that, on an average, cloud heights derived from the daytime measurements are also valid for the night, we can now use the ΔT values for the night (Fig. 5) and the heights plotted in Fig. 6 to determine the nighttime cloud cover as a function of latitude.

The resulting latitudinal distribution of the nighttime cloud cover is plotted as Curve 2 in Fig. 7.

A comparison of the day and night curves immediately shows that in the Southern Hemisphere, which is largely oceanic, the percentage cloudiness is much higher at night than during the day. In the Northern Hemisphere, on the other hand, the nocturnal cloud cover could be as low as ~20 percent for the latitudes of 20 to 40 deg.

These results are based on the debatable assumption that, on an average, for a given latitudinal belt the cloud-top heights do not change from day to night. In case this assumption is not correct, our results would then indicate that,

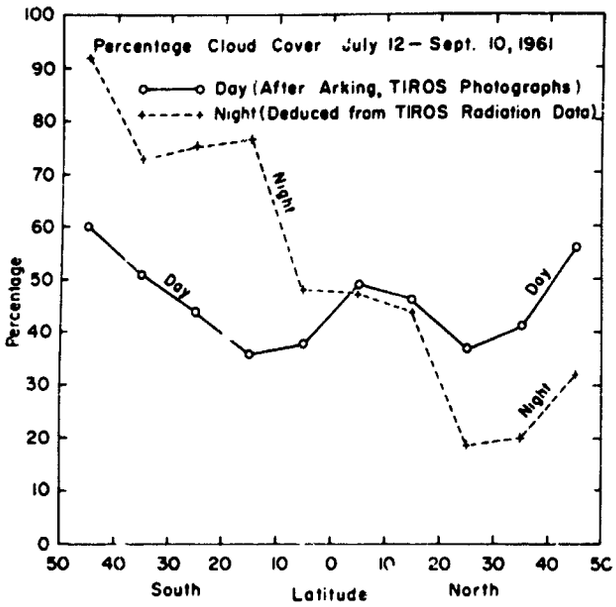


FIGURE 7.—Curve 1. Daytime average percentage cloud cover for the period 12 July through 10 September as a function of latitude (after Arking). Curve 2. Nighttime average percentage cloud cover for the same period deduced from Tiros window channel data.

in order to have the same amount of cloudiness for both day and night, the cloud heights in the Southern Hemisphere should be 2 to 4 km higher at night than in the daytime, and lower by 1 to 3 km in the Northern Hemisphere.

Similar analysis of the TIROS radiation data can also be made to determine the longitudinal as well as latitudinal distribution of nighttime cloud cover, but it requires more detailed information on the global distribution of surface temperatures, water vapor, and variations in the emissivities of land areas (Beuttner and Kern, 1963).

ACKNOWLEDGMENT

I wish to thank Professor R. M. Goody for reading over the manuscript and making valuable comments, L. Umscheid for programming and computing on an IBM 7094 and A. Liebman for his valuable assistance during the preparation of this paper.

REFERENCES

ARKING, A., 1964: The latitudinal distribution of cloud cover from TIROS photographs. *Science*, **143**, 569–572.

BANDEEN, W. R., R. A. HANEL, J. LICHT, R. A. STAMPFL and W. G. STROUD, 1961: Infrared and reflected solar radiation measurements from the TIROS II meteorological satellite. *J. Geophys. Res.*, **66**, 3169–3185.

BEUTTNER, K. J. K., and C. D. KERN, 1963: Infrared emissivity of the Sahara from TIROS data. *Science*, **142**, 671–672.

DAVIS, P., 1962: A re-examination of the heat budget of the troposphere and lower stratosphere. *Scientific Report No. 3*, Contract No. AF 19(604)-6146, Research Division, College of Engineering, New York University.

HAURWITZ, B., and J. M. AUSTIN, 1944: *Climatology*. New York, McGraw-Hill Book Co., Inc., 410 pp.

LONDON, J., 1957: A study of the atmospheric heat balance. Final Report, Contract No. AF 19(122)–165, Research Division College of Engineering, New York University.

NORDBERG, W., W. R. BANDEEN, B. J. CONRATH, V. KUNDE and I. PERSANO, 1962: Preliminary results of radiation measurements from the TIROS III meteorological satellite. *J. Atmos. Sci.*, **19**, 20–30.

PRABHAKARA, C., and S. I. RASOOL, 1963: Evaluation of TIROS infrared data. *Proc. First Internat. Symposium on Rocket and Satellite Meteor.*, Amsterdam, North-Holland Publishing Co., 234–246.

SHIFFRIN, K. S., 1961: Spectral properties of clouds. *Geofis. pura. appl.*, **48**, 129–137.

WARK, D. Q., G. YAMAMOTO and J. H. LIENESCH, 1962: Methods of estimating infrared flux and surface temperature from meteorological satellities. *J. Atmos. Sci.*, **19** 369-384.

N64-24[illegible]

GLOBAL DISTRIBUTION OF THE NET ENERGY BALANCE OF THE ATMOSPHERE FROM TIROS RADIATION DATA*

S. I. RASOOL†

The radiation data from Tiros II and Tiros III have been analyzed to obtain the monthly averages of the global distribution of the total outgoing radiation from the earth. These data have been combined with climatological estimates of the incoming radiation for the same periods to obtain monthly averages of the regional distribution of the net energy balance. The results show that these data could be useful in studies of the meridional circulation and the evolution of large-scale weather systems.

We have carried out an analysis of the radiation data acquired by the Tiros meterorological satellities, in order to determine the regional and time variations in the energy balance of the earth and its atmosphere.

The energy balance of the earth-atmosphere system is made up of the difference between the incoming solar radiation, mostly in the visible, and the outgoing terrestrial radiation in the infrared. It is well known that the latitudinal variation of the energy balance shows an excess of incoming solar radiation over outgoing radiation near the equator, and a deficiency at the poles. It is this variation of the energy balance with latitude that drives the atmospheric heat engine. Thus, through the determination of the latitudinal average of the energy balance with the aid of the Tiros data, we obtain the information which is necessary to understand the general circulation of the atmosphere.

At the same time, we have obtained some information regarding the regional variations in the energy balance of the earth-atmosphere system. Knowledge of these variations gives us a very important source of information for the understanding of large-scale weather systems.

Simpson (*1*) carried out the first extensive studies of the atmospheric energy balance, by the application of a simplified radiation theory to empirical physical models of the atmosphere. This problem has since been re-examined by Houghton (*2*), London (*3*), Budyko (*4*), and others.

The radiation instrumentation in Tiros, and the physical significance of this experiment, have already been discussed in detail by Bandeen *et al.* (*5*) and Nordberg *et al.* (*6*).

Three of the five channels of the Tiros radiometer measure terrestrial radiation in the far infrared, corresponding to wavelength intervals of 5.8 to 6.5 μ, 7 to 13 μ, and 7 to 30 μ. The other two channels record the solar radiation reflected by the earth, in the visible spectrum, which gives an estimate of the albedo of any region.

In Tiros II and Tiros III, which were active for 5 and 3 months, respectively, the channels which measure the reflected visible energy degraded very quickly, and almost no data are available for these channels. It is also possible to obtain information regarding the solar input by the statistical analysis of the Tiros cloud-cover photographs (*7*), but this analysis is still in a preliminary stage. Therefore, for this first study of the energy balance, we have used the climatological estimates of Budyko (*4*) for the incoming radiation. These give the monthly averages of solar radiation reaching the ground as a function of latitude and longitude, as derived from ground-based pyrohelimeter data.

For the outgoing radiation we have used the Tiros data in the infrared channels. Channel 4, which is sensitive in the 7- to 30-μ interval, and thus records almost 80 percent of the total outgoing radiation from the earth, also degraded after approximately 1 month in each of the

*Published in *Science*, 143(3606):567-569, February 7, 1964.

†Goddard Space Flight Center and New York University, New York.

satellites. The data from channel 2, sensitive in the 7- to 13-μ interval, however, is available from 23 November 1960 to 12 April 1961 for Tiros II and from 12 July to 10 September 1961 for Tiros III. There was some degradation in this case also, but fairly good estimates of the correction factors for this channel are now available.

Wark *et al.* (*8*) have shown that data from this channel can be used to obtain reasonably good estimates of the total outgoing radiation from the earth. Numerical factors for converting channel 2 measurements to total outgoing radiation are now available for both Tiros II and III. A comparison between the values of total outgoing radiation derived from channels 4 and 2 for the overlapping period give satisfactory results.

To obtain an estimate of the temporal as well as regional variations in the total outgoing flux, we have divided the surface of the earth into intervals of 10° latitude by 10° longitude between 50°N and 50°S. Because the inclination of the satellite orbit is 48° the data for polar regions are not available. All measurements made by the satellite in each of these intervals, at nadir angles less than 25°, have been converted to the total outgoing flux and averaged separately for each month. There are approximately 500 observation points per month in each 10° square.

The total outgoing flux estimated in this manner is, however, liable to have a small diurnal bias. This is because the orbit of the satellite precesses in right ascension at a rate of 6° per day, or it has a precession period of 9 weeks. Therefore our results for the total outgoing radiation are affected by the diurnal variation of the ground temperature. However, this effect is small for the particular case of this study which is concerned with the long-period averages over large areas. Because of the cloudiness and the presence of water vapor, the average of the channel-2 data over large regions and extended periods of time does not give the ground temperature, but has been shown (*9*) to give the effective temperature at a height of about 3 km, and therefore the effect of the diurnal variation is much reduced.

As mentioned earlier, for the incoming radiation we have used the climatological estimates of Budyko (*4*), which are based on the ground observations of solar radiation, and are given as monthly averages, also in 10° by 10° grids. These values were corrected for the absorption of solar radiation by the atmosphere, which, according to London (*3*), varies with latitude and season and is about 30 to 40 percent of the radiation which reaches the ground. The energy balance for each 10° by 10° grid, that is, the difference between these corrected values of incoming energy and the Tiros-measured values of the outgoing energy, is plotted on a month by month basis in Figs. 1 through 6.

These correspond to the months of December 1960 and January, February, March, July, and August 1961, respectively. The dark areas correspond to a positive energy balance, that is, an excess of incoming radiation over the outgoing. The light areas indicate regions of negative balance. In these figures, the darkest shade (for example, southwestern United States in July) has a value of $>+1.75\times10^5$ ergs cm^{-2} sec^{-1}, while the lightest shade (for example, 40°N to 50°N belt in December) corresponds to $<-1.65\times10^5$ ergs cm^{-2} sec^{-1}. The intermediate values of the energy balance are plotted, in steps of about 0.5×10^5 ergs cm^{-2} sec^{-1}, as seven gradations of the shading level.

The two triangular regions comprising parts of South America and Siberia have been left blank because data from Tiros are not available in those regions. Also the several other 10° by 10° grids left blank in March and July are due to nonavailability of data from Tiros for these regions.

A preliminary examination of these figures reveals several interesting features.

1) In December and January, the regions of maximum positive energy input are located in the latitudinal belt of 20°S to 50°S, while in July and August it is the 10°N to 40°N belt which has a high excess of energy. The evolution of this phenomenon is revealed by the charts for February and March.

2) The geographical distribution of the energy balance appears to be such that the desert areas of Africa, Australia, the Middle East, and southwestern United States show extreme positive energy inputs in the local summer.

3) The effect of the monsoon over India is noticeable in comparing the charts for March and August 1961. The net energy input over India is lower in August than in March, presumably

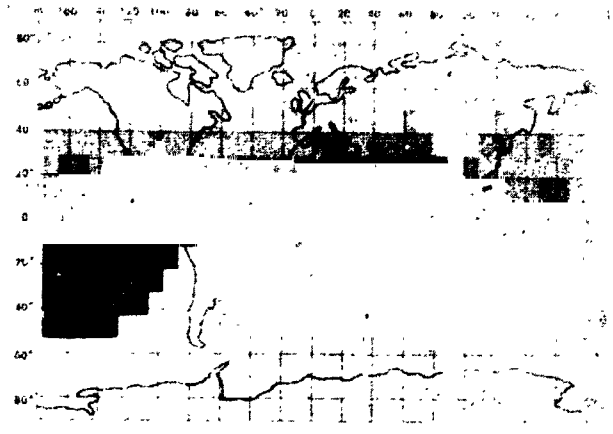


FIGURE 1.—December 1960

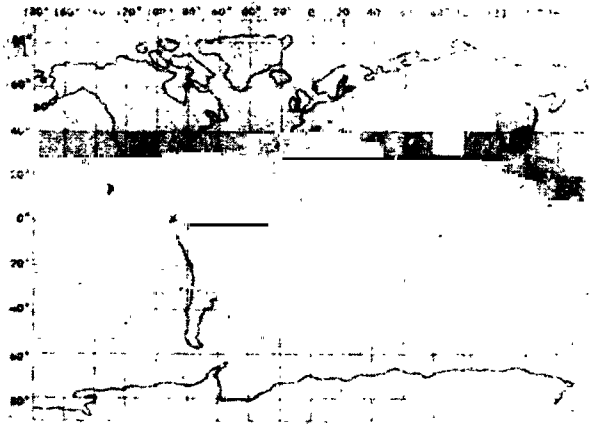


FIGURE 2.—January 1961

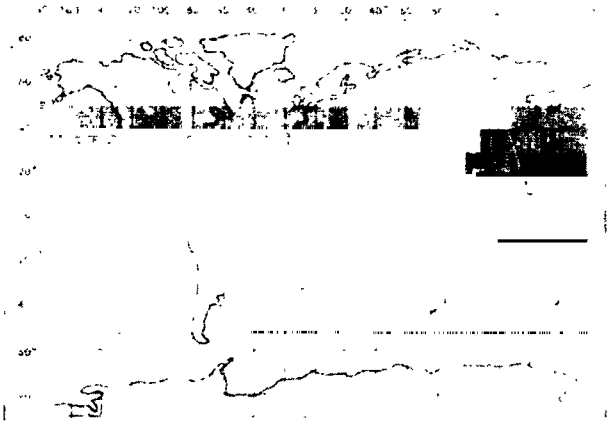


FIGURE 3.—February 1961

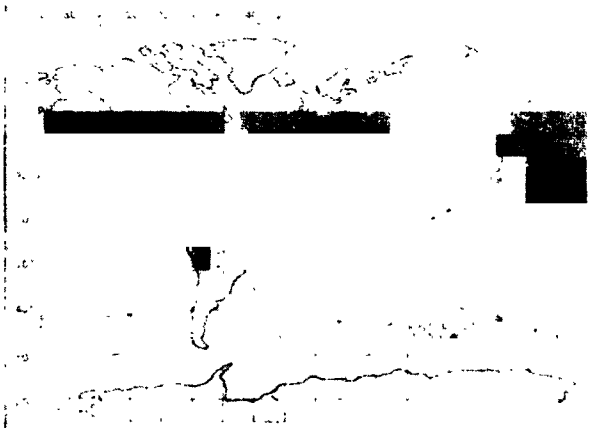


FIGURE 4.—March 1961

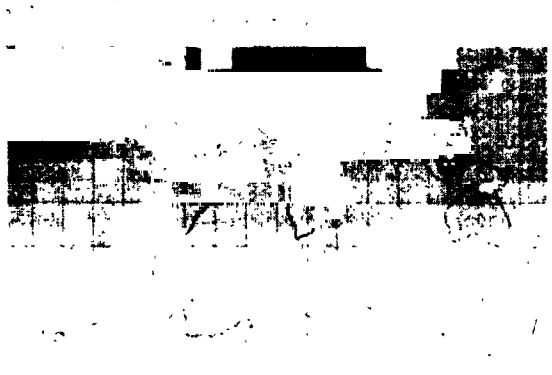


FIGURE 5.—July 1961

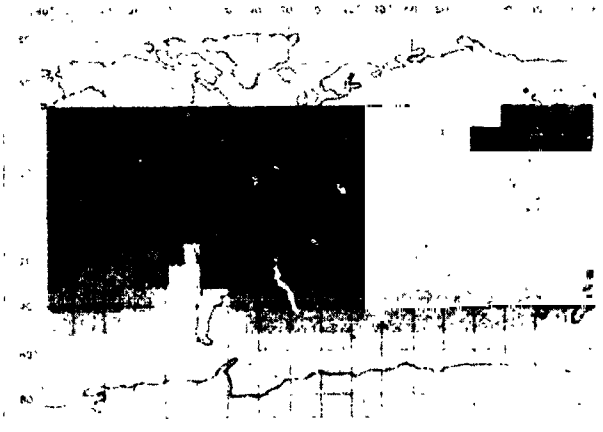


FIGURE 6.—August 1961

FIGURES 1–6.—Distribution of net energy balances for the months of December 1960 (Figure 1), January 1961 (Figure 2), February 1961 (Figure 3), March 1961 (Figure 4), July 1961 (Figure 5), and August 1961 (Figure 6). Extreme dark shade $\sim > +1.75 \times 10^5$ ergs cm^{-2} sec^{-1} and extreme light shade $< = 1.65 \times 10^5$ ergs cm^{-2} sec^{-1}.

because of the heavy monsoon cloud cover, although if it were not for the monsoon, one would expect a very high excess of energy in the summer month of August.

4) In the Northern Hemisphere, the region of the western Pacific appears to show a relative deficit in energy during all the 6 months. This area is well known for strong cyclogenic activity.

As in the case of the monsoon, the high cloudiness produced by the cyclogenesis probably accounts for the observed low input in energy.

A detailed analysis of these charts, in conjunction with the actually observed global distribution of the weather patterns for the respective months, may provide a better understanding of the role played by the energy balance of the atmosphere in the evolution of weather systems.

REFERENCES AND NOTES

1. G. C. Simpson, *Mem. Roy. Meteorol. Soc.* **3**, No. 23, 53 (1929).
2. G. H. Houghton, *J. Meteorol.* **11**, (1954).
3. J. London, "A study of the atmospheric heat balance," *Final Report*, Contract No. AF 19 (122)-165, Research Division, College of Engineering, New York University, New York, 1957.
4. F. A. Budyko, Jr., *The Heat Balance of the Earth's Surface*. Translated from the Russian edition by Nina A. Stepanova (U.S. Weather Bureau, Washington, D.C., 1958).
5. W. R. Bandeen, R. A. Hanel, J. Licht, R. A. Stampfl, W. G. Stroud, *J. Geophys. Res.* **65**, 3165 (1961).
6. W. Nordberg, W. R. Bandeen, B. J. Conrath, V. Kinde, I. Persano, *J. Atm. Sci.* **19**, 20 (1962).
7. A. Arking, *Science*, this issue.
8. D. G. Wark, G. Yamamoto, J. H. Lienesch, "Methods of estimating infrared flux and surface temperature from meteorological satellites," *J. Atmospheric Sci.* **19**, 369 (1962).
9. C. Prabhakara and S. I. Rasool, in *Proceedings of the 1st International Symposium on Rocket and Satellite Meteorology*, H. Wexler and J. E. Caskey, Jr., Eds. (North-Holland, Amsterdam, 1963).

10. I thank J. London, R. M. Goody, J. Charney, and R. Jastrow for many illuminating discussions, and Conrad Hipkins who computed and plotted these data on an IBM 7094 and SC 4020.

INTENSITIES OF 9.4μ AND 10.4μ CO_2 BANDS*

S. I. RASOOL

The intensities of the 9.4 μ and 10.4 μ CO_2 bands have been computed theoretically using, as the basis of our calculations, the line profiles at the centers of these bands published by Migeotte, Neven, and Swensson (1957). The band intensity values thus obtained are as follows:

9.4 μ band -0.0180 cm^{-2}-$atm.^{-1}$ at 265°K

10.4 μ band -0.0098 cm^{-2}-$atm.^{-1}$ at 265°K

The ratio of the two band intensities is 1.84 which is closer to the value of 1.7 given by Kostkowski (1955) than to 2.4 reported by Penner (1959).

The 9.4 μ and 10.4 μ CO_2 bands are extremely weak, and for the amount of CO_2 present in the earth's atmosphere ($\sim$0.03 percent) they do not play any significant role in the radiation processes within the terrestrial atmosphere. For this reason theoretical and experimental investigations of these bands have been rather limited. Also the two results available in the literature (Penner, 1959 and Kostkowski, 1955) differ rather significantly from each other.

For the purposes of studying radiative properties of the atmospheres of the other planets, Venus and Mars for example, where the amount of CO_2 in the atmosphere may be greater than in the earth's by several orders of magnitude, these bands do become important (Jastrow and Rasool, 1963). Therefore, in the absence of generally accepted values of the intensities of 9.4μ and 10.4μ bands, we have attempted to estimate them independently.

The method used is as follows: The solar infrared spectrum published by Migeotte *et al.* (1957) contains many lines at the center of the 9.4 μ and 10.4 μ bands due to absorption by CO_2 present in the earth's atmosphere. Measuring the total absorption within these lines and applying the Ladenburg and Reiche (1911) solution we calculate the intensities of lines in the P and R branches of the bands. The half width of the lines was assumed to be 0.09 cm^{-1} at NTP, and the amount of CO_2 in the absorption path was determined on the assumption that the total CO_2 in the earth's atmosphere is 2.40 m-atm. These line intensity values were then used to obtain a mean value of the band intensity using the classical relation (see, for example, Goody, 1963).

Internal consistency in the band intensity calculations was achieved when the effective temperature of the atmosphere above Jungfraujoch observing station was assumed to be 265°K. This is $\sim$6°K lower than the recorded surface temperatures during these observations.

In the following table we give the band intensity values determined by the method mentioned and values give by the other authors.

Intensities in cm^{-2}-atm^{-1} of CO_2 *Bands at* 9.4 μ *and* 10.4 μ

	9.4 μ (S′)	10.4 μ (S″)	T	S′/S″
This work	0.0180	0.0098	265°K	1.8
Kostkowski (1955)	0.0465	0.0281	329°K	1.7
Penner (1959)	0.0532	0.0219	297°K	2.4

The band intensity values given by Kostkowski, when reduced to 265°K, compare fairly well with our results, but those of Penner, especially of the 9.4 μ band, are significantly higher than ours.

I wish to thank Professor R. M. Goody of Harvard University for suggesting this method of calculation.

REFERENCES

R. M. Goody, *Atmospheric Radiation*, Cambridge Univ. Press, Cambridge, 1963.

R. Jastrow and S. I. Rasool, "Radiative transfer in the atmospheres of Venus and Mars," in *Space Research III*, ed. W. Priester, North Holland Publ. Co., 1963.

H. J. Kostkowski, Dissertation, the John Hopkins Univ., 1955.

R. Ladenburg and R. Reiche, *Annalen der Physik*, **421**, 181, 1911.

M. Migeotte, L. Neven and J. Swensson, The solar spectrum from 2.8 μ to 23.7 μ. *Mém. Soc. Roy. Sc., Liège*, 1957.

S. S. Penner, *Quantitative Molecular Spectroscopy and Gas Emissivity*. Addisor Wesley, Mass. 587 pp., 1959.

*Published in Mémoires Soc. R. Sc. Liége, Cinquième Série, Tome IX, Extract pages 55–57.

N66 37967

DATA FROM EXPLORER 17 ON COMPOSITION OF THE UPPER ATMOSPHERE*

CARL REBER

INTRODUCTION

The Explorer 17 aeronomy satellite carried two double-focusing magnetic mass spectrometers designed [*Meadoes*, 1960; *Hall et al.*, 1960; *Spencer and Reber*, 1962] to measure the concentrations of the major neutral-particle constituents of the earth's upper atmosphere. The ambient densities of helium, oxygen, and nitrogen were obtained between north and south latitudes of 58° and from the perigee altitude of 257 km to an altitude in excess of 700 km. Data for various local times, altitudes, and geographic locations have been combined to provide a broader altitude coverage than could have been experienced at any single

TABLE 1.—*Tabulated Mass Spectrometer Data*

The local sun time, angle of attack (α) geographic latitude and longitude are averaged over the 4-minute pass. The stations involved are: BP, Blossom Point, Md.; COL, College, Alaska; FTM, Fort Myers, Fla.; GF, Grand Forks, Minn.; QUI, Quito, Equador; MOJ, Mojave, Calif.; NFL, Newfoundland; OOM, Woomera, Australia; JOB, Johannasburg, South Africa.

Pass and Station	Date	Local Time, hours	α	Geog. Lat.	Geog. Long.
15 BP	4/ 4/63	21.15	6°	38.5°	−75.0°
50 COL	4/ 6/63	0.65	16°	57.0°	−149.0°
80 COL	4/ 8/63	0.99	9°	55.0°	−147.0°
80 FTM	4/ 8/63	4.89	63°	18.0°	−92.0°
118 BP	4/10/63	18.81	70°	37.0°	−72.0°
120 GF	4/11/63	20.32	51°	51.0°	−98.5°
138 BP	4/12/63	2.51	12°	37.0°	−84.0°
152 BP	4/13/63	2.01	14°	39.5°	−68.5°
167 BP	4/14/63	1.65	20°	39.5°	−75.0°
182 BP	4/15/63	1.54	25°	37.0°	−78.0°
183 QUI	4/15/63	3.26	23°	4.5°	−79.0°
197 BP	4/16/63	1.43	27°	34.0°	−81.5°
211 BP	4/17/63	0.53	45°	41.5°	−71.5°
226 BP	4/18/63	0.48	53°	38.5°	−74.0°
241 BP	4/19/63	24.19	62°	38.0°	−79.5°
242 MOJ	4/19/63	0.64	54°	31.0°	−121.5°
254 NFL	4/20/63	22.75	82°	49.0°	−53.0°
270 BP	4/21/63	23.30	80°	41.50°	−71.5°
271 GF	4/21/63	22.88	85°	45.0°	−101.0°
708 NFL	5/20/63	7.18	39°	49.5°	−49.5°
795 OOM	5/26/63	15.81	63°	−34.0°	137.5°
800 JOB	5/26/63	15.90	65°	−37.5°	19.0°
888 JOB	6/ 1/63	13.24	33°	−27.0°	25.0°

*Published in *Journal of Geophysical Research*, 69(21):4681–4685, November 1, 1964.

location. Data are presented for helium, atomic oxygen, and molecular nitrogen; molecular oxygen and atomic nitrogen data will be presented and discussed in a later report along with other data obtained under more varied conditions.

EXPERIMENT

The mass spectrometer measured the number densities of the various species which were in its ionizing, or sampling, region. Before reaching this sampling region, however, some of the neutral particles underwent collisions with various parts of the spectrometer structure; thus the number measured by the spectrometer was generally different from the ambient number density. The relationship between the spectrometer measurements and the ambient number densities was found to be a function of particle species, satellite velocity, and the angle between the spectrometer axis and the satellite velocity vector (angle of attack α). Using satellite velocity and angle of attack, both obtained from the tracking stations and the satellite's optical aspect system, calculations have been carried out for each species and have been used to convert the measured quantities to the ambient values presented here. The contribution of residual gas has also been taken into account.

The absolute accuracy of the number density data is ± 40 percent, reflecting laboratory vacuum calibration error and the uncertainties in the ambient number density calculations. The relative accuracy of the data obtained for similar angles of attack is about 5 percent; for data obtained when the angles of attack differ by 70° or more, the relative accuracy is about 30 percent. The mean mass data and the concentration ratios are good to ± 10 percent. Since the relative accuracy of the number density varies with the angle of attack, this quantity, the local sun time, and other pertinent information are given in Table 1.

DATA

Figure 1 shows the ambient number densities of the major constituents plotted as a function of

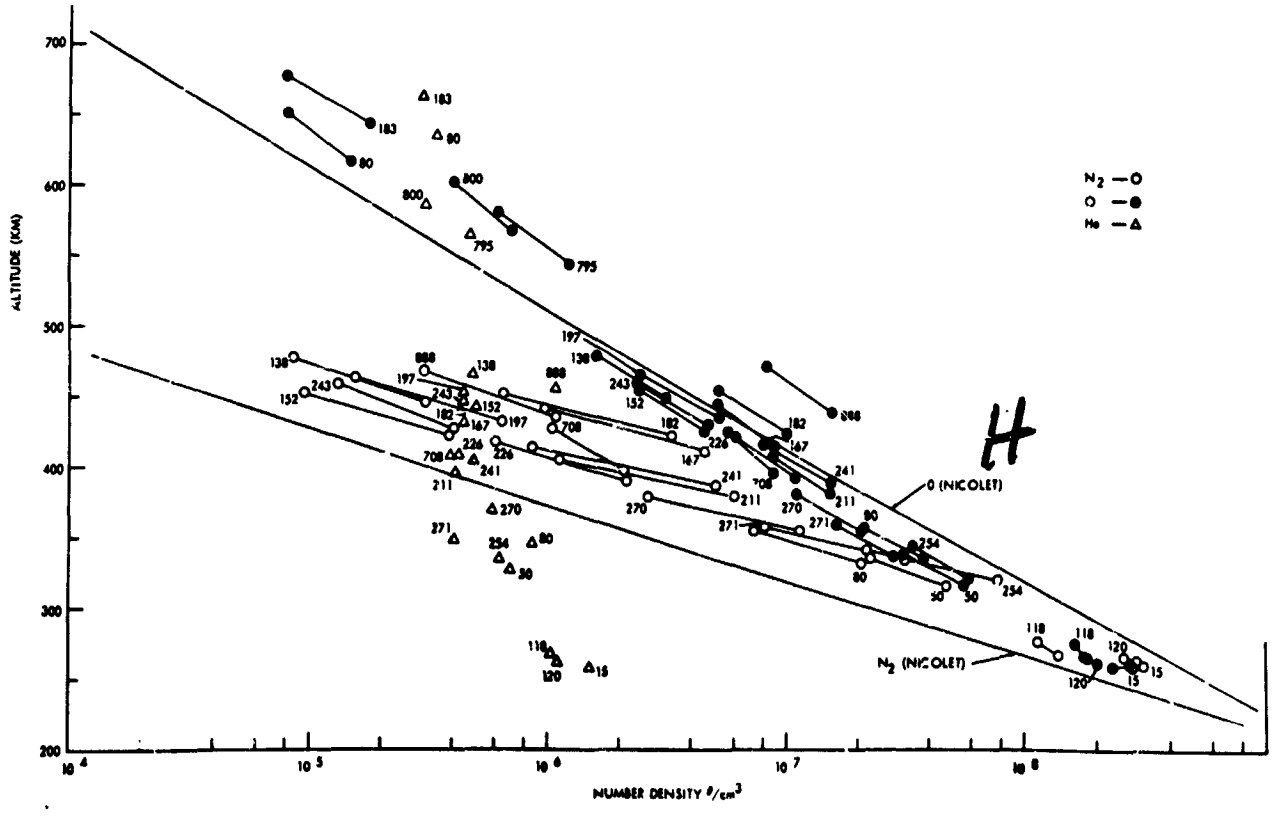


FIGURE 1.—Ambient number densities of the major constituents versus altitude. These data include many local times and geographic locations. The orbit numbers refer to Table 1. Nicolet's model for $T = 700°K$ is included for reference.

altitude for a number of different local times and locations. A single satellite interrogation (pass) is indicated by two data points joined by a straight line for molecular nitrogen and atomic oxygen and by one point for helium. The Nicolet model (private communication) is included for reference.

For the times concerned, the number densities of helium and atomic oxygen are comparable at about 600 km, with helium predominant at higher altitudes. Between 300 and about 600 km, atomic oxygen is the major constituent; in the 250- to 300-km region, molecular nitrogen and atomic oxygen have nearly equal concentrations.

There is, at present, insufficient daytime data to determine scale heights for the sunlit atmosphere. However, during the night at higher altitudes the concentration gradients are consistent with the scale heights for temperatures of 700° to 750°K.

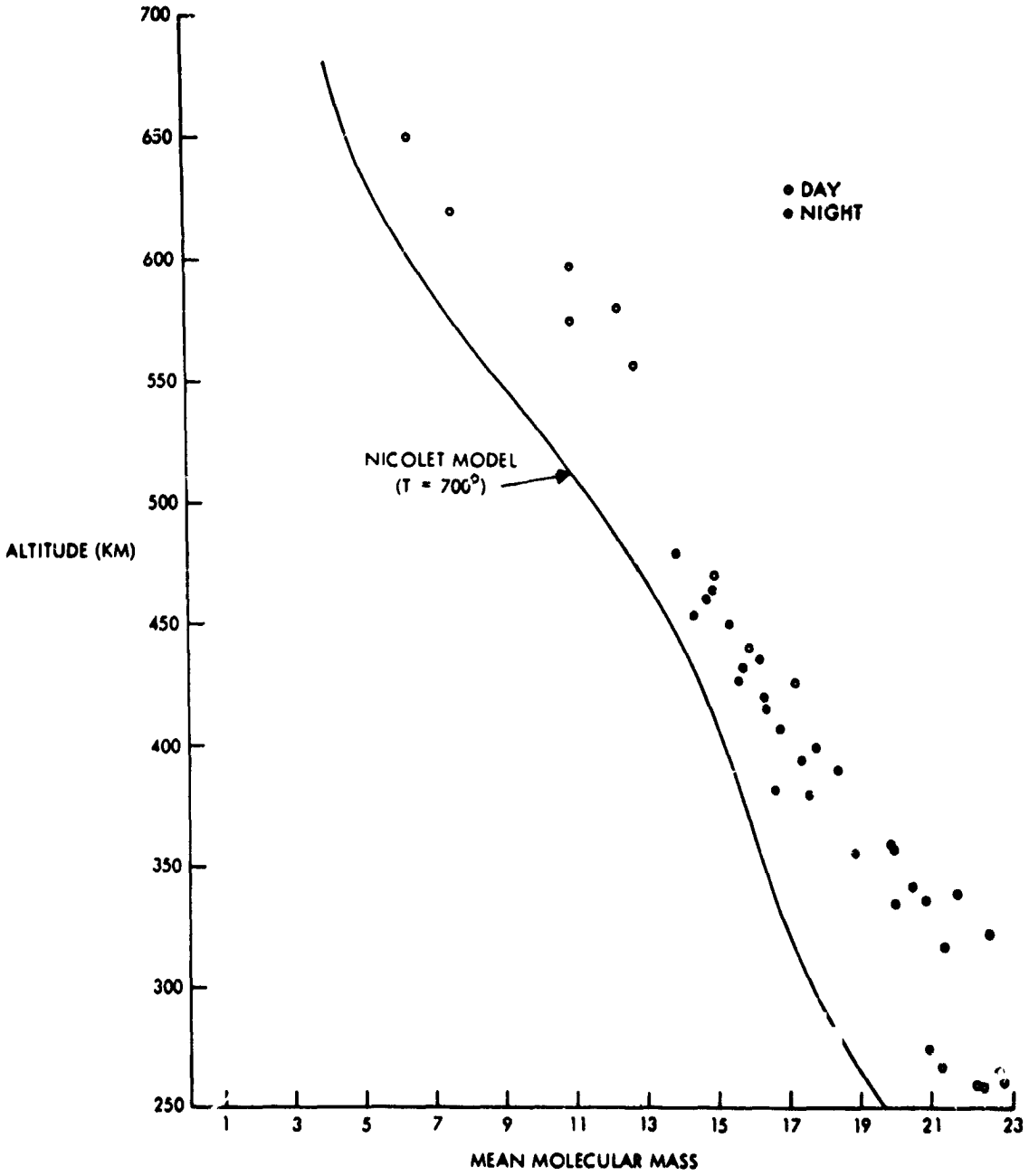


FIGURE 2.—Mean molecular mass versus altitude. The mean mass was calculated using the three major constituents measured, N_2, O, and He. Nicolet's model is included for reference.

There appears to be a deviation from nighttime diffusive equilibrium in the 300- to 350-km altitude range for N_2. This may be indicative of longer times required for N_2 to diffuse at these altitudes; however, as noted above, the data shown are representative of many time and geographic locations and may not provide an accurate instantaneous vertical distribution. It is interesting to note that the data from passes 167, 182, and 183, for which the densities are higher than for other nighttime passes, were obtained shortly after a minor magnetic disturbance which occurred on April 14. All data shown are measured values and are not averaged or smoothed in any way.

Figure 2 shows the variation of mean molecular mass with altitude. These data were calculated using the measurements of helium, atomic oxygen,

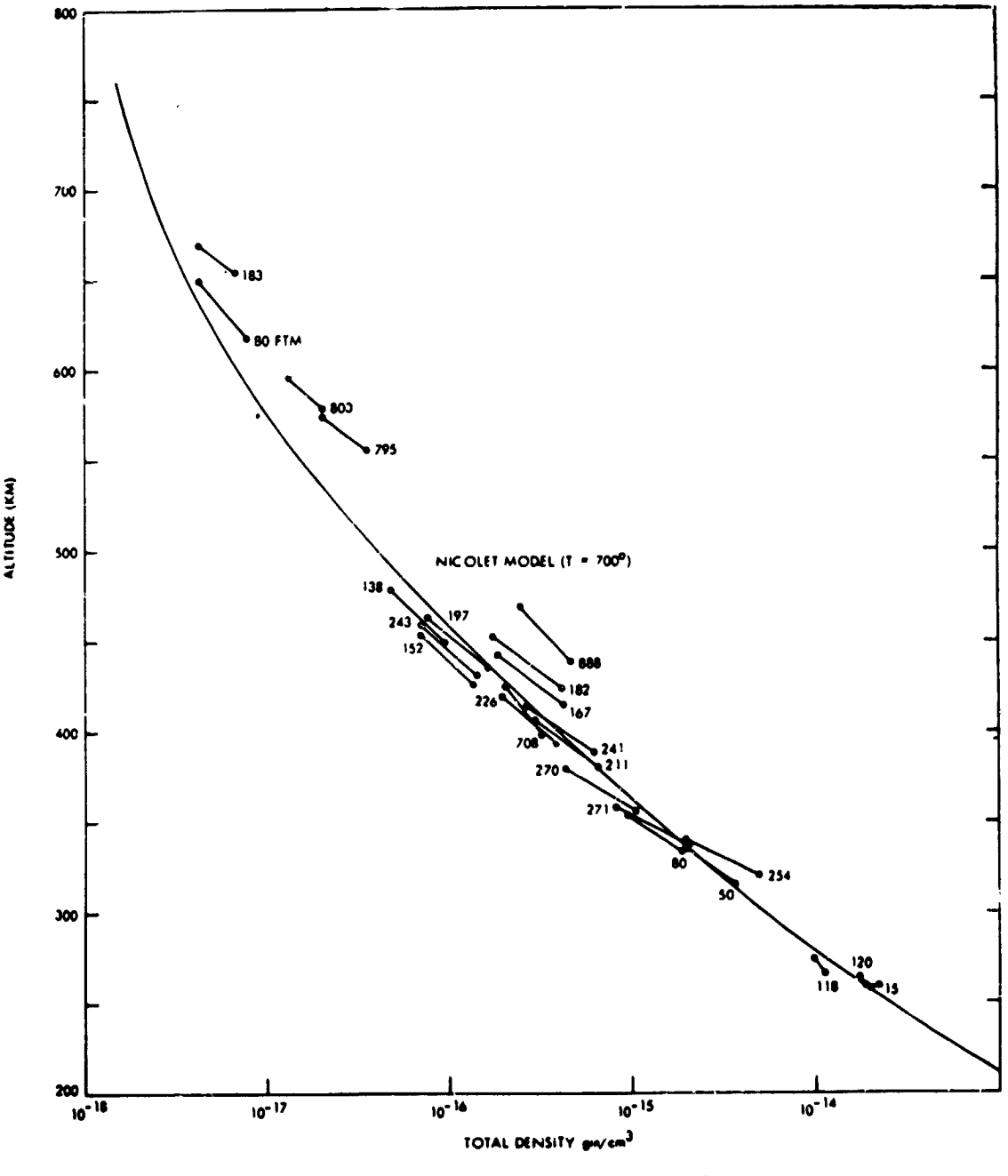


FIGURE 3. Total mass density versus altitude. The total mass density was calculated using the densities for N_2, O, and He. The orbit numbers refer to Table 1. Nicolet's model is included for reference.

and molecular nitrogen only, other measured gases providing a negligible contribution ot the mean mass. The presence of hydrogen, which the instrument was not designed to measure, would reduce the value of the mean mass somewhat at higher altitudes.

The total mass density, shown in Figure 3, was computed by summing the contributions from molecular nitrogen, atomic oxygen, and helium. The results obtained in this way agree with the densities obtained independently by the pressure gage experiments on Explorer 17. This agreement lends support to the calculations relating the concentrations in the sampling region to the ambient number densities, as the pressure gage kinetics have been studied extensively [*Horowitz,*

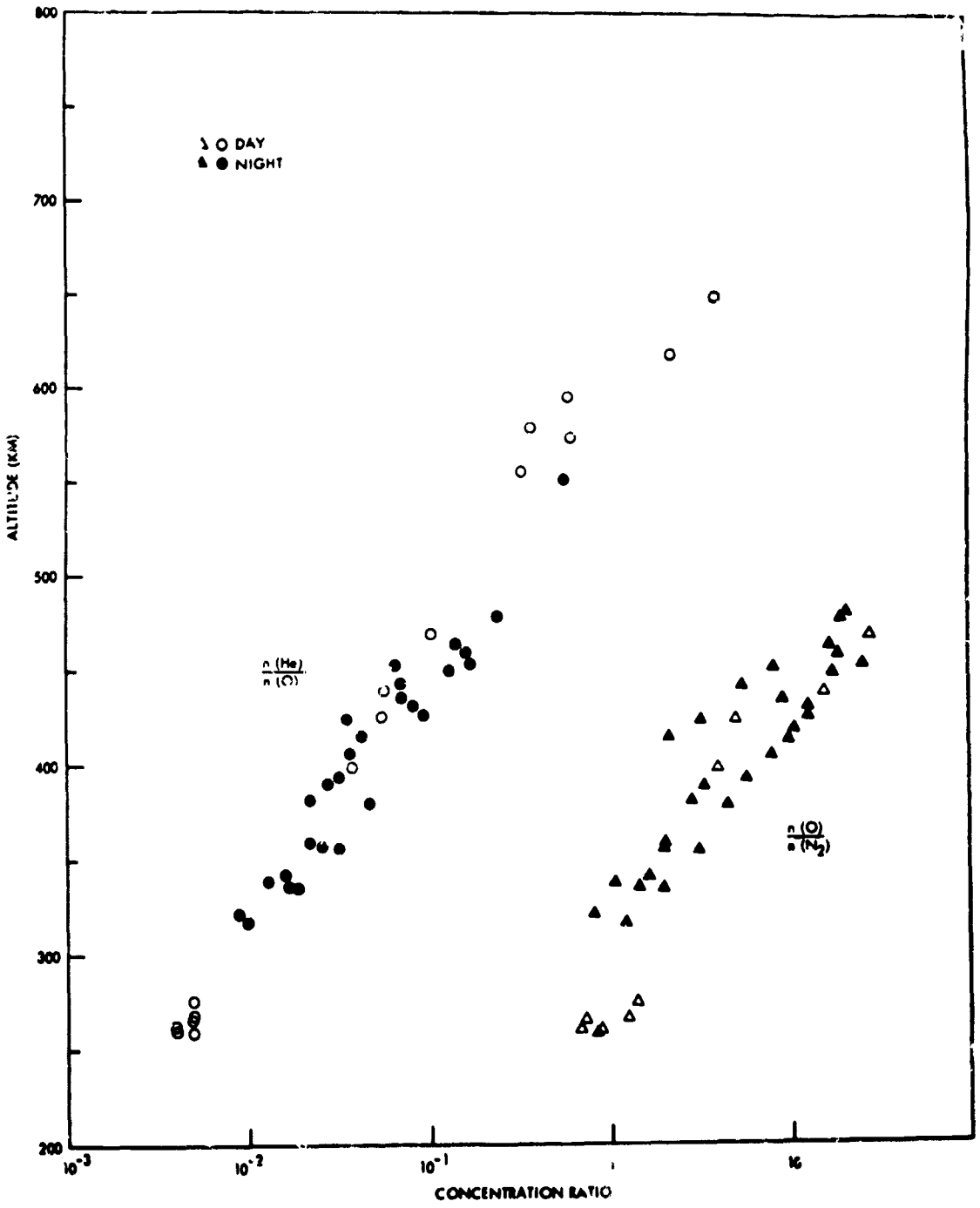


FIGURE 4.—Concentration ratios of the major constituents versus altitude.

1953]. Also, the total density measurements from the mass spectrometers agree in general within a factor of 2 with atmospheric density calculated from satellite drag observations [*Bryant*, 1964].

Figure 4 shows the variation with altitude of the concentration ratios of helium to atomic oxygen and atomic oxygen to molecular nitrogen.

ACKNOWLEDGMENT

I an indebted to Professor Marcel Nicolet for many hours of stimulating discussions about the data from Explorer 17, particularly with respect to extracting the scientifically interesting from the technically involved.

REFERENCES

BRYANT, R., Densities obtained from drag on Explorer 17, *J. Geophys. Res.*, *69*, 1423–1425, 1964.

HALL, L. G., P. F. HOWDEN, and T. F. IWASKAE, Design of a mass spectrometer for use in a satellite, paper presented at Meeting of ASTM Committee E-14 on Mass Spectroscopy, 8th Annual Meeting, Atlantic City, June 26 through July 1, 1960.

HOROWITZ, R., and D. KLEITMAN, A method for determining density in the upper atmosphere during rocket flight, *Naval Res. Lab. Rept. 4246*, Washington, D. C., 1953.

MEADOWS, E. B., Design requirements of a mass spectrometer for satellite use, paper presented at Meeting of ASTM Committee E-14 on Mass Spectroscopy, 8th Annual Meeting, Atlantic City, June 26 through July 1, 1960.

SPENCER, N. W., and C. A. REBER, A mass spectrometer for an aeronomy satellite, *Space Res.*, *3*, 1151–1155, 1964.

N64 - 33630

A NIGHT-TIME MEASUREMENT OF OZONE ABOVE 40 KM*

EDITH I. REED AND REUBEN SCOLNIK

The ozone distribution between 40 and 70 km was measured near midnight, May 27, 1960, from Wallops Island, Virginia by means of photometers sensitive to the ultraviolet airglow at wavelengths between 2400 and 2900 A. Below 60 km, the densities are within a factor of two of the daytime photochemical equilibrium, as represented by Johnson's late afternoon measurement of June 14, 1949. Above 60 km, the ozone density increased with altitude, with its maximum increase, a factor of 6 over the day time value, occurring at 63 km.

INTRODUCTION

At high altitudes, above the principal ozone maximum, ozone concentration as a function of altitude should be governed principally by the presence or absence of sunlight, and vary in a predictable manner from day to night and from season to season [*Chapman*, 1930][a,b]. The first direct measurement of the daytime profile was a result of studies of the sun conducted by the Naval Research Laboratory on a V-2 rocket in 1949. Ozone densities up to 70 km were deduced from solar spectra and were consistent with computed photoequilibrium profiles [*Johnson, et al*, 1952]. Chapman also suggested that at high altitudes, above the ozone maximum and below the atomic oxygen maximum, ozone would increase at night as a result of a reaction between atomic and molecular oxygen. Several have treated this problem numerically, including *Nicolet* [1957], *Barth* [1961], *Dütsch* [1961], *Paetzold* [1961], *Wallace* [1962], and *Hunt* [1964], but with varying results, depending on the set of reactions, reaction rates, and initial concentrations which were chosen. Of particular difficulty is the computation of the effects of minor constituents such as hydrogen, nitrogen oxides, and the hydroxyl radical.

Ground based measurements of ozone content at these altitudes have not been satisfactory. Measurements of total ozone content are not particularly helpful since variations in the ozone content below 30 km due to air movements are comparable to the expected night time increase at higher altitudes. However, the discovery of the ultraviolet airglow and a general improvement in the techniques of ultraviolet photometry made a night time measurement of ozone feasible.

In 1957, the Naval Research Laboratory flew an ultraviolet photometer with a response from 2600 to 2900 A, and did observe an ultraviolet airglow layer centered at 101 km [*Tousey*, 1958]. This had been predicted from laboratory observations which showed that the Herzberg bands of molecular oxygen, the visible end of which had been observed in the airglow [*Chamberlain*, 1955], extend to 2563 A in the ultraviolet [*Broida* and *Gaydon*, 1954]. But because an unknown amount of the observed airglow could be due to an OI line at 2972 A, where the filter transmission is still 10 percent of its maximum; the ozone density could not be determined unambiguously.

In May 1960, Goddard Space Flight Center flew a number of ultraviolet photometers, including some whose filters were centered at 2620 A and narrow enough so that the absorption cross section of ozone varied by only 50 percent over the bandwidth of the filter. These data, when interpreted with the aid of an airglow spectrum obtained by T. Stecher (of GSFA), provide an ozone density profile between 40 and 70 km.

INSTRUMENTATION

NASA Aerobee 4.05, one of a series of payloads designed for stellar photometry [*Boggess*, 1961],

*Published as *Goddard Space Flight Center Document X-613-64-267*, September 1964.

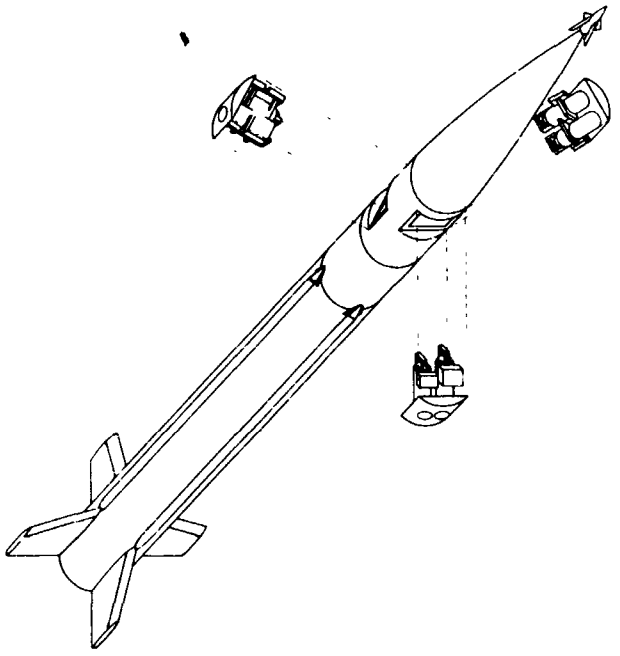


FIGURE 1.—Location of photometers on Aerobee 4.05.

contained three pairs of photoelectric photometers, each pair mounted 120 degrees apart around the rocket axis (See Figure 1) looking out at three different angles to that axis: nominally, 75°, 90°, and 105°. One photometer of each pair was sensitive to light in the spectral region centered near 2620 A, while the response of the other was centered near 2260 A, but with their optical axes parallel. Each pair was mounted on a removable door, which, when installed, became an integral part of the rocket skin. Since the principal purpose of this instrumentation was ultraviolet star spectroscopy, 2620 A photometers were preferred to the 2680 A ones (to reduce ambiguity in the interpretation of stellar data due to the strong magnesium doublet at 2800 A). However, to correlate the data from this flight with that of earlier flights, one 2680 A photometer was included. This was mounted on the door containing the 90° photometers, and also looked at 90° with respect to the rocket axis.

The optical system of each photometer was similar and is shown in Figure 2. Calcium fluoride was used for all lenses in the 2260 A units; quartz for those in the 2620 and 2680 A photometers. The field of view had a total width of between 4 to 5°.

Isolation of the 2600 A region was achieved by combining two millimeters of 0.05 percent lead-

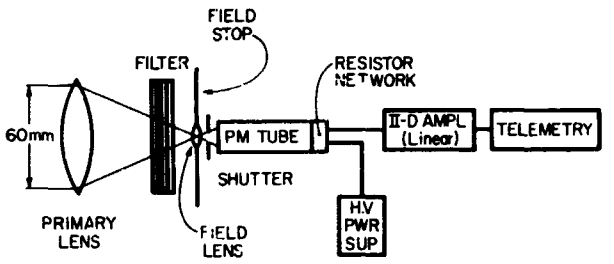


FIGURE 2.—Optical-electrical schematic of a photometer.

doped KCl:KBr (1:1) crystal with Cation-X in thin sheets of polyvinyl alcohol [*Childs*, 1961]. Three millimeters of nickel sulfate hexahydrate provided a sharp cutoff for longer wavelengths; one Corning No. 7–54 and one Corning 9–54 filter sharpened the shorter wavelength cutoff. A typical filter had a transmittance of 0.18, an effective wavelength of 2620 A and a 200 A bandwidth.

The 2700 A filter consisted of three Corning 7–54 filters, one sheet of Cation-X and 5 millimeters of nickel sulfate hexahydrate. It has an effective wavelength of 2680 A, a transmittance of 0.16, and a bandwidth of 320 A.

The relative spectral response of the filters (see Figure 4) was measured by C. Childs, formerly of this laboratory, with a recording spectrophotometer, Cary Model No. 14, with an analytical accuracy of ½ of 1 percent for relative spectral transmission, a wavelength calibration of 4A, and a resolving power of 1A. Over the wavelength regions covered by each filter, the photomultiplier was assumed to have a constant sensitivity and the lenses a constant transmissivity. The rela-

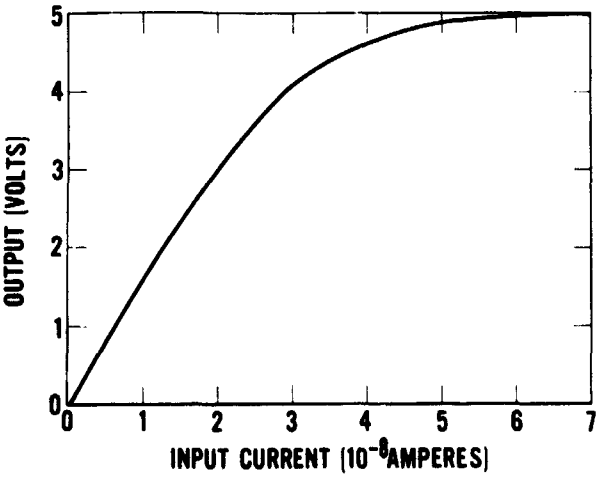


FIGURE 4.—Calibration curve for converter-amplifier.

tive response of the photometers was determined by use of the 2537 A line, to which all three types of photometers respond with an easily measured signal.

The 2680 A photometer used RCA's well-known 1P28 as a detector; all others used EMI's No. 6256B, an end-on fused silica window multiplier with cesium antimonide photocathode. A solenoid-operated shutter between the field stop and multiplier gave optical zero signals several times during the flight. The 1P28 and 6256B's were operated at 1000 and 1200 volts respectively; each detector had its own solid state DC to DC inverter power source with a resistor divider network at the base of each multiplier.

Each photomultiplier output was amplified to the zero to five volt range required by the telemetry system with what was essentially an impedance converter. Designed by G. Baker of the laboratory, the converter-amplifier had an electrometer tube (5886) input stage and ended with an emitter follower, with an overall voltage gain of 2.5 for small signals. It was purposely nonlinear in order to extend the dynamic range. A typical calibration curve is shown in Figure 3. The output voltage went directly to a pulse position modulation telemetry transmitter which relayed the data to the ground receiving equipment.

BASIC DATA

Aerobee-Hi NASA 4.05 was launched at 0030 EST on May 27, 1960 from Wallops Island, Virginia (37°50′ N, 75°29′ W). The vehicle performance was normal: propulsion ended at 52.4 sec after launch at an altitude of 36.6 km with a vertical velocity of 1.84 km/sec and a horizontal velocity of .260 km/sec at an azimuth of 111°. The rocket spun about its longitudinal axis at a rate of 2.16 rps and soon entered a precession cone of 5.7° half angle, whose axis was 15.6° from zenith at an azimuth of 96°, with a period of 75 sec. Aspect during the ascent and free-fall portions of the flight was determined from a combination of data from magnetometers, trajectory information, and horizon and star data from the photometers. Down-leg aspect could not be determined with useful accuracy below the free-fall region. A peak altitude of 215.3 ± 0.2 km was reached 249.4 seconds after launch. Telemetry ceased at 468 sec; no recovery of instrumentation was attempted.

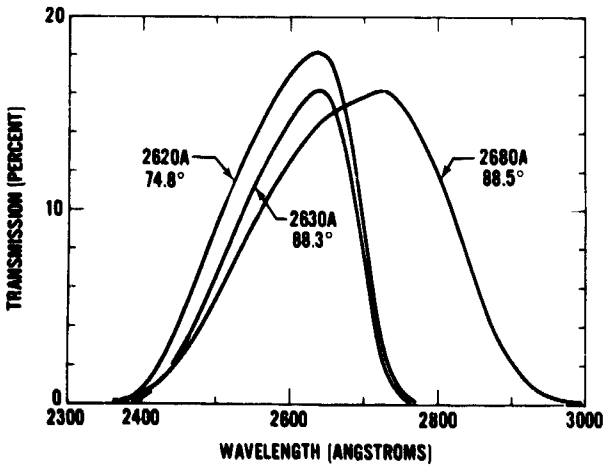


FIGURE 3.—Spectral characteristics of flight filters.

In Figure 5 is a sample of the telemetry record from two of the photometers for a period corresponding to three revolutions of the vehicle, while the airglow layer is still above the rocket. Since these photometers are nearly perpendicular to the rocket axis, the zenith angle of the photometer axis changes from a minimum (76° for the time in Figure 5) as the photometer points skywards, through 90° as it scans the horizon, and to a maximum of 100° as it points earthward. Ozone is relatively opaque around 2600 A, and the airglow in Figure 5 can be seen only when the zenith angle of the photometer is near a minimum. The ozone is relatively transparent to the light passed by the 2260 A filter, and the brightening of the airglow at the horizons can be clearly seen. (The light is probably of wavelengths longer than 2700 A, passed through the long wavelength tail of the 2260 A filter.) The southern horizon appears wider because it is merged with several bright stars in the Milky Way. As the vehicle increased in altitude, the signal from the 2620 A photometer resembled that from the 2260 A photometer, with bright horizons, a less bright sky toward zenith, and a dark earth. Above the airglow layer, the sky was dark (except for the brighter stars) and the earth appeared light. The noise in the record is due partly to the photomultiplier and partly to stars. The records were read at the midpoint between horizons, and analyzed to yield both a distribution of ozone with altitude and the volume emission of the airglow versus altitude.

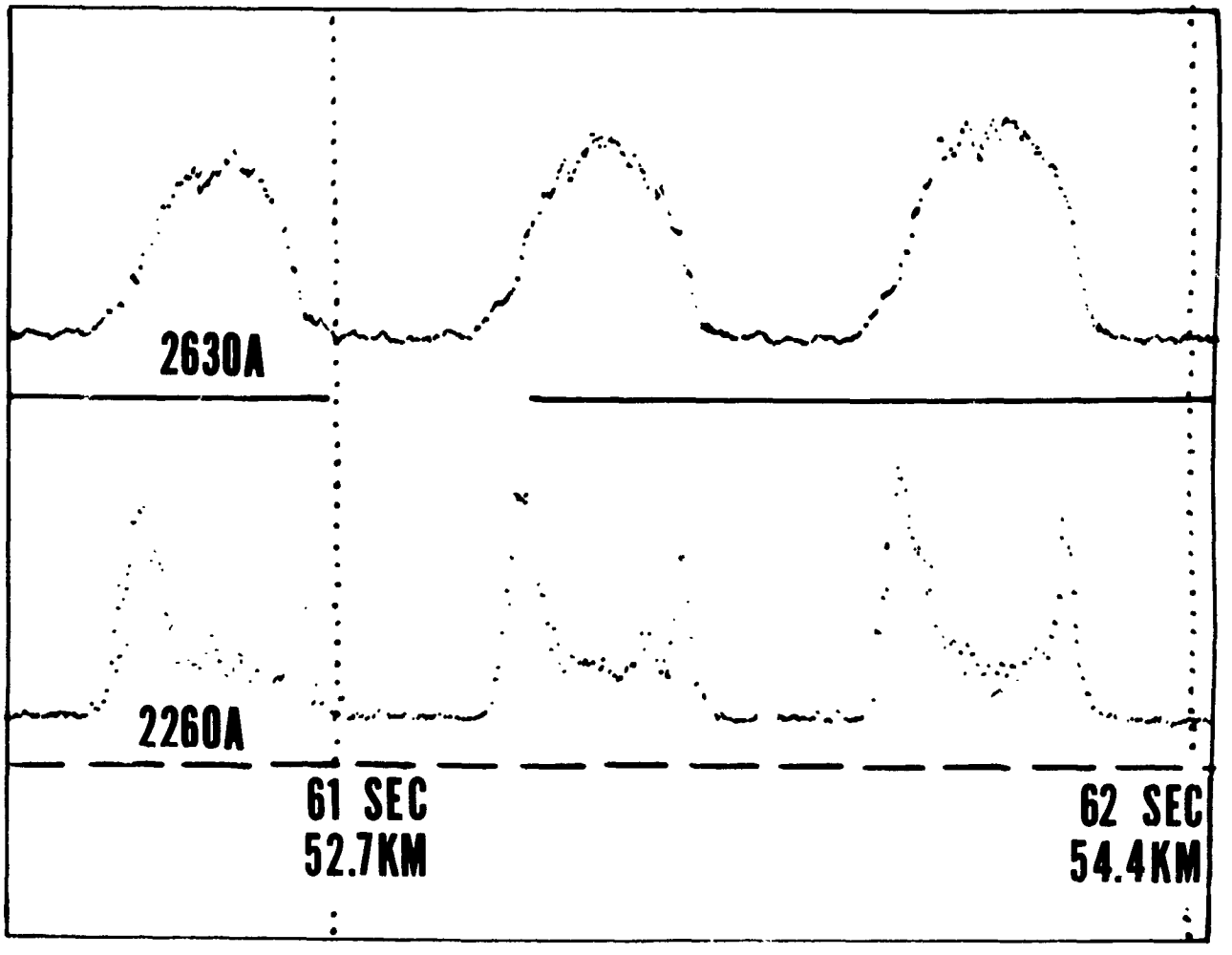


FIGURE 5.—Sample of telemetry record showing roll modulation of the airglow signal. Increasing signal represents increasing light.

The data obtained in the region of interest is shown in Figure 6. Data from 3 of the 7 photometers proved to be useful for ozone measurements. The pass band of the three 2260 A photometers was too wide to permit an accurate determination of an effective cross section for ozone; the down looking 2620 A photometer could not see the airglow while the vehicle was below the airglow layer. Data points for the photometers are shown to indicate the scatter in the raw data. As would be expected, the scatter increased rapidly as the signal rose into the non-linear portion of the amplifier response curve. In addition to the data, the angle of the rocket's longitudinal axis with respect to local zenith is given. The rocket took a spiral path with a zenith angle of 8° at thrust termination (52.4 sec) until it entered

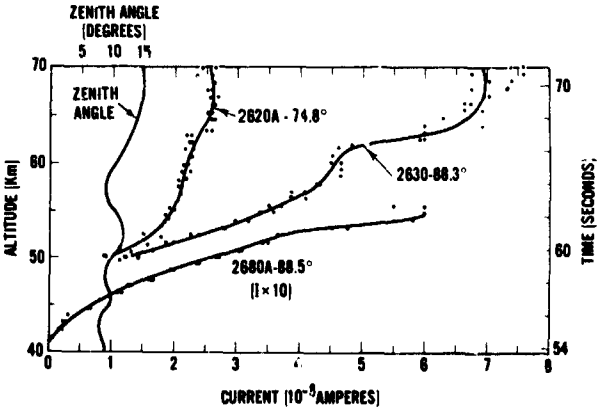


FIGURE 6.—Data from three photometers during the ascent of the rocket. The zenith angle is the angle between the longitudinal axis of the rocket and local zenith; the angle stated for each photometer is the angle between the photometer's optical axis and the rocket's longitudinal axis.

its regular precession cone of motion at about 75 seconds.

OZONE

The ozone content of the atmosphere is obtained from the rate of increase of the airglow signal as the vehicle rose. The energy observed is related to the ozone density in the following manner:

$$n(O_3)=\frac{(\log E_2-\log E_1)n\cos\gamma}{a(h_2-h_1)}$$

where $n(O_3)$ is the number of ozone molecules per cm^3

h_1 and h_2 are the lower and upper ends of the altitude interval,

E_1 and E_2 are energies observed at the corresponding altitudes,

n is Loschmidt's number, 2.687×10^{19} cm^{-3},

γ is the angel between the photometer axis and zenith, and

a is the absorption coefficient, cm^{-1}, base 10.

This was applied to the smooth curves in Figure 6 at one-second intervals.

Since the cross section of ozone varies appreciably over the wavelength interval passed by each filter, some assumption must be made concerning the spectrum of the airglow. For this purpose, the spectrum observed by T. Stecher (private communication) was used, and the absorption coefficient, a, computed for each photometer where

$$a=\frac{\Sigma E_i R_i a_i}{\Sigma E_i R_i}$$

where E_i is the energy in a particular wavelength interval, i,

R_i is the corresponding relative transmission of the filter, and

a_i is the corresponding absorption coefficient.

The absorption coefficients used were from the tabulation of *Inn* and *Tanaka* [1953, 1959] which in the region of interest here are about 10 percent higher than those of Vigroux [1953] and 1 or 2 percent lower than those of *DeMore* and *Raper* [1964]. The effective absorption coefficients calculated on this basis are: 2620 A (88.3°), 114.4 cm^{-1}; 2620 A (74.8°), 113.8 cm^{-1}; and 2630 A (88.5°), 79.48 cm^{-1}.

The resulting ozone density is shown in Figure 7 as a function of altitude. It is felt that the spectral characteristics of the photometers, the airglow and the ozone absorption coefficient are sufficiently well known such that they contribute no more than a total of about 20 percent uncertainty to the number density of ozone. This would be a systematic error which would not affect the shape of the curve.

The angle of the photometers with respect to zenith could easily contribute an uncertainty of 20 percent to both the absolute and relative values of the ozone densities. It is based upon magnetometer data and the assumption that at the end of thrust the rocket axis was aligned with the velocity vector.

The largest source of error is in the character of the data, which contains noise from the photomultiplier dark current, stray pulses, and stars. It is difficult to separate this from the possible temporal and spatial variations of the airglow itself. A temporal variation could be responsible for some of the shape of the curve, but it is unlikely that the airglow would vary sufficiently in the 16 seconds of time that the ozone curve represents, to be responsible for its major features. Spatial variation is not thought to be a major

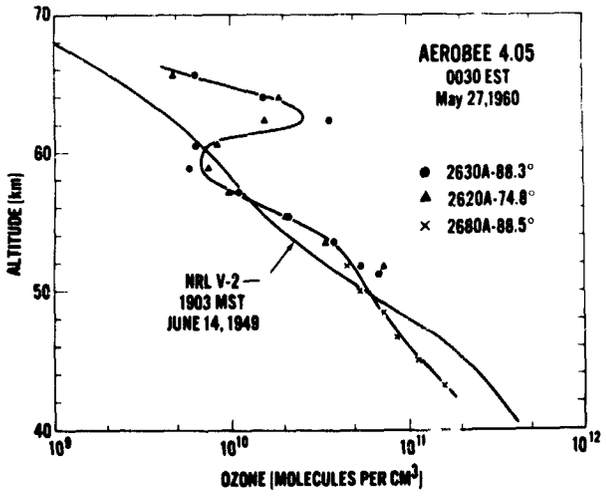


FIGURE 7.—Altitude distribution of ozone.

source of uncertainty in the data since the numbers derived from the two photometers which were looking at different portions of the sky agree reasonably well. The magnitude of the noise-like errors is apparent from the scatter of points about the curve and is on the order of 50 percent.

For comparison the daytime ozone distribution as measured by absorption of the solar spectra between 2500 and 3400 A [*Johnson, et al*, 1952] is shown. The daytime profile was computed using the ozone absorption coefficients of *Ny* and *Choong* [1933], which in the spectral region of interest are 10 to 15 percent higher than those of Inn and Tanaka used for the night time profile.

The most important feature of the profile is the factor of 6 increase in ozone density over the daytime profile above 60 km. While this may be in error by 50 percent or more, it is believed that the shape of the curve does indicate an increase of ozone density at night in this region, and that it is on the order of a half of a magnitude.

Techniques are being developed by various workers to use satellites for the measurement of ozone in this region. *Venkateswaran* [1961], observed the sunlight reflected from Echo I as it emerged from the earth's shadow, using various wavelength pairs between 4700 and 7000 A. His results above 55 km are about a factor of 15 higher than the measurements by *Johnson, et al.* However *Venkateswaran* [1963] states that this method probably gives too high values at levels above the principal ozone maximum.

The second type of observation was made at sunrise and sunset by a satellite borne radiometer with a response center at about 2630 A [*Rawcliffe, et al*, 1963]. At 60 km his data are about 20 percent lower than Johnson's, but approach Johnson's data, and above 80 km are somewhat higher than the trend of Johnson's data.

It is expected that the night time values would be higher. The magnitude of the effect, besides depending on reactions among the various oxygen species and third bodies, depends critically on such things as the initial hydrogen concentrations chosen [*Bates* and *Nicolet*, 1950, and *Wallace*, 1962] and possible reactions involving atomic nitrogen [*Barth*, 1961]. Perhaps the most recent computation of ozone densities has been done by *B. G. Hunt* [1964] using an atmosphere in which he assumes the only reactive constituent is oxygen. (The effects of atomic nitrogen and hydrogen would be to lower the calculated 0 and 0_3 concentrations.) Between 40 and 50 km, Hunt's curve for just before sunrise conditions is as much as 30 percent lower than Johnson's daytime profile, crosses it at 53 km, and reaches a maximum value of 5×10^{10} molecules of 0_3 per cm at 69 km.

AIRGLOW

The other principal result from the analysis of data from these photometers is information concerning the distribution of the airglow. Volume emission can be deduced from the data obtained as the vehicle passes through the emitting region. The energy calibration of the photometers has been used in preparing these curves, (Figure 8) so that they do rightly represent the relative energy in the portion of the spectrum passed by the different filters. Ten arbitrary units represent on the order of one photon cm^{-3} sec^{-1} per A. The airglow as measured by the 2680 A photometer is 2.3 times that sensed by the 2620 A photometers. The airglow measured by the 2260 A photometers is 0.15 times that of the 2620 A photometers; nearly all the energy measured by the 2260 A photometers has come through a long wavelength tail of the filters. This pattern is completely consistent with a spectrum of the airglow horizon obtained by T. Stecher with a spectrograph flown at 0030 local time on July 19, 1963.

The 2680 A filter was similar in construction and characteristics to those flown in March 1957 [*Tousey*, 1958] and November 1959 [*Friedman*, 1961, *Packer*, 1961]. The altitude of maximum

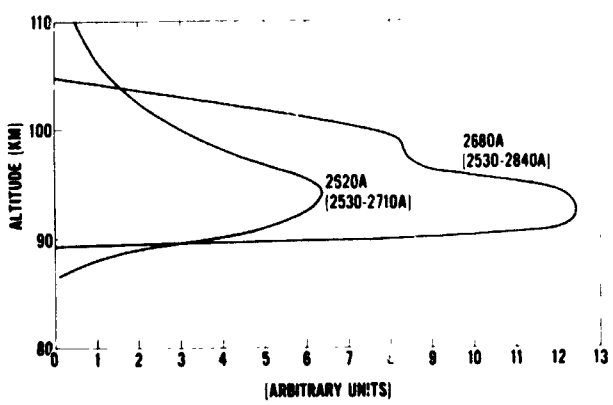


FIGURE 8.—Altitude distribution of the ultraviolet emission.

emission were 101 and 96 km respectively, compared to 92 km for this flight. The zenith intensity for the 1957 flight was 3.4 rayleighs per Angstrom (Dunkelman, private communication) and for the 1959 flight was 1.94 rayleigh per Angstrom. The airglow during the flight of Aerobee 4.05 was somewhat brighter, but within a factor of 10 of these values.

The brighter stars and the Milky Way were readily noted in the records as the photometer scanned across the sky. The signal in the absence of obvious stars indicated that less than 15 percent of the light from extended sources originated above the emission layer indicated in Figure 8.

ACKNOWLEDGMENTS

We thank the following colleagues at Goddard Space Flight Center for their help: Dr. A. Boggess III for making the data from this rocket payload available to us for this purpose; G. Baker and R. M. Windsor for the electronic instrumentation; E. Serra for engineering assistance in the payload preparation; T. Stecher for an advance copy of his airglow spectrum, wlthout which the effective absorption cross section could not have been determined with significant accuracy; L. Dunkelman for the absolute values of the night airglow measurement in March 1957; and Dr. J. E. Kupperian, Jr. for his encouragement and suggestion of analysis of this data.

REFERENCES

BARTH, C. A., Nitrogen and oxygen atomic reactions in the chemospher, *Chemical Reactions in the Lower and Upper Atmosphere*, Stanford Research Institute, Interscience Publ., 303–326, 1961.

BATES, D. R., and M. NICOLET, The photochemistry of atmospheric water vapor, *J. Geophys. Res., 55* 301–327, 1950.

BOGGESS III, A., Ultraviolet astronomical photometry from rockets, *Space Astrophysics*, W. Liller, ed., McGraw-Hill Book Co., 121–132, 1961.

BROIDA, H. P., and A. G. GAYDON, The Herzberg bands of O_2 in an oxygen afterglow and in the night-sky spectrum, *Proc. Roy. Soc., London U222*, 181–195, 1954.

CHAMBERLAIN, J. W., The ultraviolet airglow spectrum, *Astrophys. J. 121*, 277–286, 1955.

CHAPMAN, S., On ozone and atomic oxygen in the upper atmosphere, *Phil. Mag., 10*, 369–383, 1930.

CHAPMAN, S., A theory of upper-atmospheric ozone, *Mem. R. Met. Soc., 3* 103–125, 1930.

CHILDS, C. B., Broad-band ultraviolet filters, *J. Opt. Soc. Am., 51*, 895–897, 1961.

DEMORE, W. and O. RAPER, Hartley band extinction coefficients of ozone in the gas phase and in liquid nitrogen, carbon monoxide, and argon, *J. Phys. Chem., 68*, 412–414, Feb. 1964.

DÜTSCH, H. U., Current problems of the photochemical theory of atmospheric ozone, *Chemical Reactions in the Lower and Upper Atmosphere*, Stanford Research Institute, Interscience Publ., 167–180, 1961.

FRIEDMAN, H., A survey of NRL rocket research results obtained since the last COSPAR meeting, *Space Research II*, H. C. van de Hulst, ed., Interscience, N.Y. 1021–1035, 1961.

HUNT, B. G., A non-equilibrium investigation into the diurnal photochemical atomic oxygen and ozone variations in the mesosphere, *W. R. W. Report, Tech. Note PAD 82*, Feb. 1964 (Box 1424H, G. P. O. Adelaide, South Australia).

INN, E. C. Y. and Y. TANAKA, Absorption coefficient of ozone in the ultraviolet and visible regions, *J. Opt. Soc. Am., 43.*, 870–873, 1953.

INN, E. C. Y., and Y. TANAKA, Ozone absorption coefficients in the visible and ultraviolet regions, *Advan. Chem. Ser., No. 21*, American Chemical Society, Washington, D.C., 263–268, 1959.

JOHNSON, F. S., J. D. PURCELL, R. TOUSEY, and K. WATANABE, Direct measurements of the vertical distribution of atmospheric ozone to 70 kilometers altitude, *J. Geophys. Res., 57*, 157–176, 1952.

NICOLET, M. Nitrogen oxides and the airlow, *The Theshlod of Space*, M. Zelikoff, ed., Pergamon Press, 40–57, 1957,

NY, T.-Z., and S.-P. CHOONG, *Chinese J. Phys., 1*, 38. 1933.

PACKER, D. M., Altitudes of the night airglow radiations, *Ann. Geophys., 17*, 67–75, 1961.

PAETZOLD, H. K., The photochemistry of the atmospheric ozone layer, *Chemical Reactions in the Lower and Upper Atmosphere*, Sanford Research Institute, Interscience Publ., 181–195, 1961.

RAWCLIFFE, R. D., G. E. MEILOY, R. M. FRIEDMAN, and E. H. ROGERS, Measurement of vertical distribution of ozone from a polar orbiting satellite, *J. Geophys. Res. 68*, 6425–6429, Dec. 15, 1963.

TOUSEY, R., Rocket Measurements of the Night Airglow, *Ann. Geophys., 14*, 186–195, April-June 1958.

VENKATESWARAN, S. V., J. G. Moore, and A. J. Krueger, Determination of the vertical distribution of ozone by satellite photometry, *J. Geophys. Res., 66*, 1751–1771, 1961.

VENKATESWARAN, S. V., On some problems of exploration of the upper atmosphere between 50 and 100 km by means of rockets and satellites, *Rocket and Satellite Meteorology: Proc. 1st Int'l. Symp. Rocket and Satellite Meteorology*, H. Wexler and J. E. Caskey, Jr., eds., John Wiley and Sons, 199–209, 1963.

VIGROUX, E., Contribution a l'étude expérimentale de l'absorption de l'ozone, *Ann. de Phys., 8*, 709–762, 1953.

WALLACE, L., The OH nightglow emission, *J. Atmos. Sci., 19*, 1–16, 1962.

SOME ASPECTS OF WIND SHEAR IN THE UPPER ATMOSPHERE*

R. G. ROPER

The wind motions responsible for the shearing of sodium vapor trails ejected from rockets in the 70 to 140 km region of the upper atmosphere are subjected to an analysis based on generally accepted theories of hydrodynamic turbulence. The region from 80 to 100 km is of particular interest in that here the predictions of shear turbulence theory are well substantiated. The energy spectrum of the height shear is found to follow the 4/3 power law proposed by Tchen, and is associated with a vertical correlation distance of approximately 6 km. The existence of an isotropic inertial region of maximum scale 3 km, previously indicated by analysis of meteor data, is confirmed. The vertical scale of the turbulent eddies is found to be the atmospheric pressure scale height, a phenomenon which has been observed by others, but which, as yet, has no satisfactory explanation.

1. TURBULENCE THEORY

The complete development of the relationships used in the analysis of wind shear is beyond the scope of this work; the following, used in conjunction with the references quoted, should provide an adequate background for consideration of the subsequent analysis.

1.1 Energy Spectrum Analysis

If there exists in a turbulent flow field a range of scales which receive energy from larger scale motions and pass it on undiminished to smaller scale motions, then, for this so-called inertial (non-dissipative) range of scales, the only form of energy spectrum function dimensionally possible is

$$E(k)=\alpha\epsilon^{2/3}k^{-5/3} \quad \text{(Kolmogoroff, 1941)}$$

where ϵ is the rate at which the turbulent energy is received by (and leaves) the inertial range of scales, k is the wavenumber vector corresponding to the real space scale r, and α is an absolute constant of order unity.

Batchelor (1953) has shown that, for such an inertial region which also possesses the property of isotropy, the fluid velocity differences measured as a function of the separation r follow the relation

$$[u(x)-u(x+r)]^2=4.82\alpha(\epsilon r)^{2/3} \tag{1}$$

In real space, the energy spectrum function $E(r)$ is defined by

$$E(r)=[u(x)-u(x+r)]^2 \tag{2}$$

Tchen (1954) has considered an otherwise isotropic region subjected to a mean shear, and finds that equation 1 is modified, becoming

$$[u(x)-u(x+r)]^2=a\,r^{4/3} \tag{3}$$

where a involves α, ϵ, and the mean gradient. Thus, for what may be termed shear turbulence, in real space

$$E(r)\sim r^{4/3} \tag{4}$$

1.2 Correlation Analysis

An energy spectrum function equivalent to that based on velocity differences can be formulated from the lateral or longitudinal velocity correlations defined as

$$f(r)=\frac{u_f(x)u_f(x+r)}{u_f^2}$$

and

$$g(r)=\frac{u_n(x)u_n(x+r)}{u_n^2} \tag{5}$$

where $u_f(x)$, $u_f(x+r)$, $u_n(x)$, $u_n(x+r)$ are the turbulent components of the velocity at two points x and $x+r$ respectively, measured parallel (suffix f) and normal (suffix n) to the vector separation r.

*Published as *Goddard Space Flight Center Document* X-651-64-337, November 1964.

In isotropic turbulence, $u_f^2 = u_n^2 = u_0^2$, the velocity characteristic of the energy bearing eddies of the Kolmogoroff spectrum.

Introduction to the equation of continuity for incompressible fluids leads to the relation

$$g(r) = f(r) + \frac{1}{2} r \frac{\partial f}{\partial r} \tag{6}$$

(von Karman and Howarth, 1938)

between the functions f and g, or, if the turbulence is isotropic in two dimensions only,

$$g(r) = f(r) + r \frac{\partial f}{\partial r} \tag{7}$$

These functions, f and g, may be called Eulerian space correlation functions, and either may be denoted by $R(r)$. If the range of scales under observation is inertial, then $R(r)$ must depend only on ϵ, and the only form dimensionally possible is

$$u_0^2[1 - R(r)] \sim \epsilon^{2/3} r^{2/3}$$

i.e.
$$1 - R(r) \sim r^{2/3}$$

(for any given flow field, u_0 and ϵ are constant).

Introduction of the relations 6 and 7 above gives, with c a constant

$$f = 1 - cr^{2/3} \tag{8}$$

and
$$g = 1 - 4/3\ cr^{2/3} \tag{9}$$
for three dimensional isotropy

or
$$g = 1 - 5/3\ cr^{2/3} \tag{10}$$
for isotropy in two dimensions only.

Since the variations of the correlation difference function $[1 - R(r)]$ with r is the same as that of $E(r)$ in equation 1, we may suppose that the dependence of $[1 - R(r)]$ on r will be similarly modified in the presence of a mean shear

i.e.
$$[1 - R(r)] \sim r^{4/3} \tag{11}$$

If, in fact, equation 11 is pertinent, then the relationships between f, g, and r (equations 8, 9, 10) will now become

$$f = 1 - ar^{4/3}$$

and
$$g = 1 - 5/3\ ar^{4/3}$$

or
$$g = 1 - 7/3\ ar^{4/3}$$

An indication of the degree of isotropy can be obtained by considering the ratio

$$S = \frac{1 - f}{1 - g}$$

For two dimensional isotropy

$S = 0.60$ without mean shear

and $S = 0.43$ with mean shear

For three dimensional isotropy

$S = 0.75$ without mean shear

and $S = 0.60$ with mean shear

The relative importance of mean shear will be indicated by the form of either the energy spectrum function $E(r)$, or the correlation difference function $[1 - R(r)]$.

2. THE PRACTICAL APPLICATION OF TURBULENCE THEORY

2.1 The Mean Wind Profile

In applying the relations developed above to the wind vectors measured by means of sodium vapor trails ejected as a tracer into the upper atmosphere, the relative importance of the mean motion must not be overlooked. The velocities used in the energy spectrum and correlation analyses of the previous section must be the turbulent velocities, or departures from the mean motion. In normal correlation analysis, the mean value of a set of observations is usually taken as the mean value of the measured data, which, when applied to wind height shear data, would be tantamount to the assumption of a mean wind profile which is constant with height. Such a profile is the exception rather than the rule in meteorological phenomena.

Practically any attempt to prescribe a mean wind profile to the measured winds will be subjective to a certain extent. If one uses a polynomial fit, for example, it must be truncated before it can assume any of the features of the measured profile which are due to the turbulent motions present. From experience based on the measurement of winds in the 75 to 105 km region by means of radio reflections from meteor trails (Elford, 1958, 1964), a quadratic change with height of the mean wind over any given 20 km interval should best

describe the contribution of the mean motion without destroying any of the characteristics of the turbulent flow field. In the present analysis, the windspeed/azimuth data are converted to zonal and meridional components, and a polynomial profile of order $Z+1$, where

$$Z=\frac{\text{Total height range covered by data (km)}}{20},$$

is fitted to each. The profiles thus determined are subtracted from the relevant measured profiles to give zonal and meridional turbulent velocities.

2.2 The Determination of E[Δh]

The available sodium trail data lists wind speed and azimuth against height, and usually involves irregularly spaced height intervals. The sampling irregularity exists for two reasons:

a) there is an occasional difficulty in absolutely identifying the same point on the trail in consecutive photographs;

b) the wind profile between consecutive observational heights is linear. This is usually obvious from the photographs, and can easily be allowed for in subsequent analysis.

Whereas equal height interval sampling is not absolutely necessary for subsequent reduction, it does simplify the analysis, and so linear interpolation between the listed data points is used to provide a profile with data points spaced 0.2 km in height.

The energy spectrum function is computed as

$$E(\Delta h_i)=\frac{1}{N}\sum_{k=1}^{N}(u_{k+i}-u_k)^2 \qquad (12)$$

in which

$$N=5(H_2-H_1)-i$$

where $i=1, 2, 3....N$

such that $\Delta h_1=0.2$ km.

$\Delta h_2=0.4$ km

etc.

and H_1, H_2 (in km) are the lower and upper bounds respectively of the region for which $E(\Delta h)$ is being determined. Such partitioning of the height range is necessary since there is considerable variation in the characteristics of the flow over the total height range sampled (usually some 70 to 200 km).

Energy spectrum functions may be calculated using

a) zonal turbulent velocities

b) meridional turbulent velocities

and c) turbulent windspeed.

The turbulent windspeed w is defined here as

$w=$ measured windspeed $-$
((mean zonal wind)$^2+$
(mean meridional wind)$^2)^{1/2}$

2.3 Correlation Analysis

The two correlation functions f and g defined in 1.2 refer to turbulent velocity components measured parallel and normal to the separation vector. Since vertical velocities in this region of the upper atmosphere are so much less than the associated horizontal components, the magnitude of the vertical component cannot be determined from sodium trail photographs. (Most workers in this field consider the upper limit for mean plus random vertical motions to be some 10 metres/sec). However, we may redefine f and g in terms of the orthogonal zonal and meridional flow fields, with a view to investigating possible isotropy. This has been done by meteorologists in the past, with at least partial success (see, for example, Hutchings (1955)).

The normalizing factors $\bar{u}_f^2$ and $\bar{u}_n^2$ are best estimated by the standard correlation function definition. The correlation functions f and g then become

$$f(\Delta h_i)=\frac{\sum_{k=1}^{N} u_{k+i}u_k}{\left[\sum_{k=1}^{N} u_{k+1}^2 \sum_{k=1}^{N} u_k^2\right]^{1/2}} \qquad (14)$$

where u are the zonal turbulent wind velocities and

$$g(\Delta h_i)=\frac{\sum_{k=1}^{N} v_{k+i}v_k}{\left[\sum_{k=1}^{N} v_{k+i}^2 \sum_{k=1}^{N} v_k^2\right]^{1/2}}, \qquad (15)$$

where v are meridional turbulent wind velocities, and N and i are as previously defined for equation 12.

3. PRELIMINARY RESULTS

The analysis of Section 2 has been applied to data obtained from a sodium trail release over the Eglin Air Force Base, Florida (29.6° N, 86.6° W) at 1910 CST on May 21, 1963 (Edwards et al., 1963). In this experiment, wind speed and azimuth were obtained over a height range of 69 to 140 km.

3.1 The Mean Wind Profile

Since the data covers the height range from 69 to 140 km, polynomials of order 4 are fitted to the zonal and meridional measured profiles. These yield mean zonal and meridional profiles

$$u_{\text{mean}} = 38.6 - 126h - 244h^2 + 174h^3 + 205h^4$$

$$v_{\text{mean}} = -30.4 + 19.8h + 148h^2 - 16.8h^3 - 125h^4$$

where u, v are in metres/sec and h is the normalized height given by

$$h = (2z - z_{\min} - z_{\max})\,(z_{\min} - z_{\max})$$

where z is the height variable

$z_{\max}$ the maximum and

$z_{\min}$ the minimum heights of the available data, all heights being in kilometers.

Normalization of the height range stabilizes the least squares fitting process, and makes the relative importance of the individual terms of the fitted polynomials more obvious than is the case when the mean velocities are expressed as power series in the actual height z. Results are plotted in Figs. 1 and 2.

The meteorological significance of these profiles, in particular the reversal of both the zonal and meridional components above 110 km, cannot be evaluated from consideration of this single firing.

3.2 The Turbulent Wind Profile

As mentioned in Section 2, the 70 to 200 km height range covers a number of characteristically different regions. The wind motions observed below approximately 105 km indicate the presence of small-scale structure, while those above 110 km do not appear to be at all turbulent, even though vertical shear is present. In the results presented here, discussion is confined to the consideration of

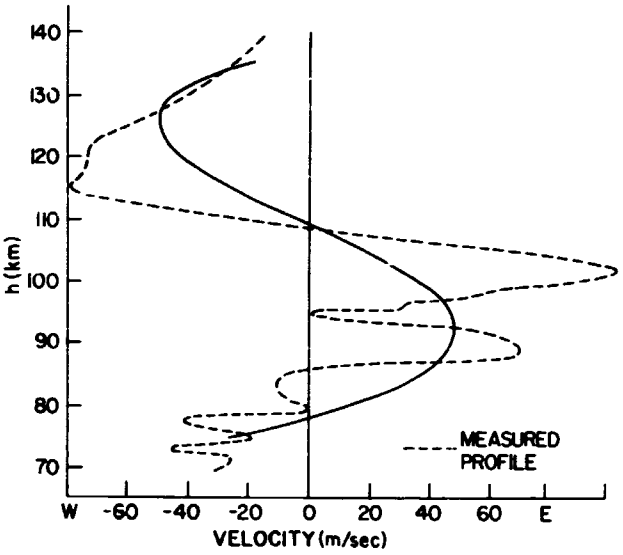


FIGURE 1.—Mean zonal profile.

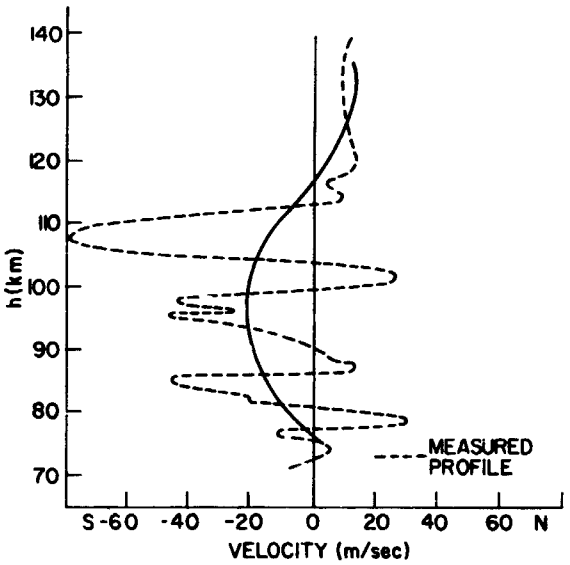


FIGURE 2.—Mean meridional profile.

the region from 80 to 100 km, which has been found to be representative of a turbulent region which can be adequately described by available statistical theories of hydrodynamic turbulence (Lamont and Jager, 1961; Zimmerman, 1962; Roper, 1962).

The measured zonal and meridional profiles, and the deviations from the mean wind for the height range 80 to 100 km are plotted in Figures 3 to 6. Whereas the immediately obvious wave-like nature of the turbulent profile would suggest a wave theory approach as likely to be the most profitable for consideration of the wind motions

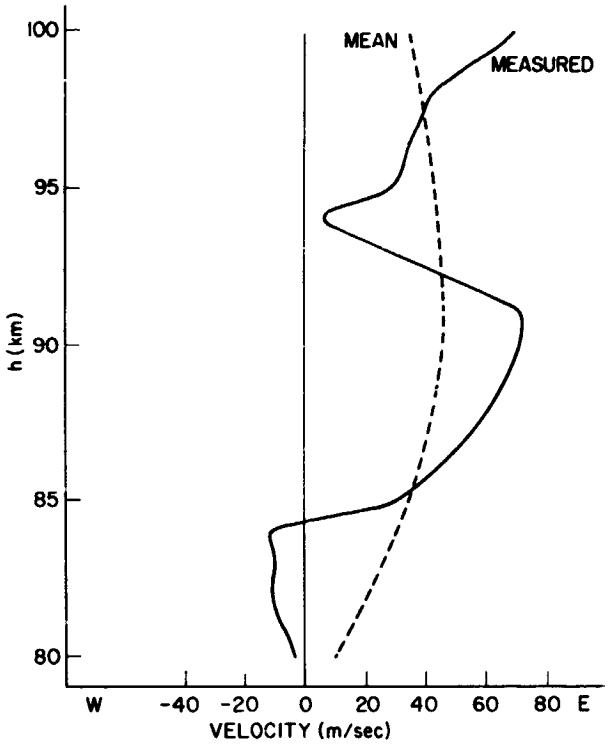


FIGURE 3.—Zonal profile, 80-100 km.

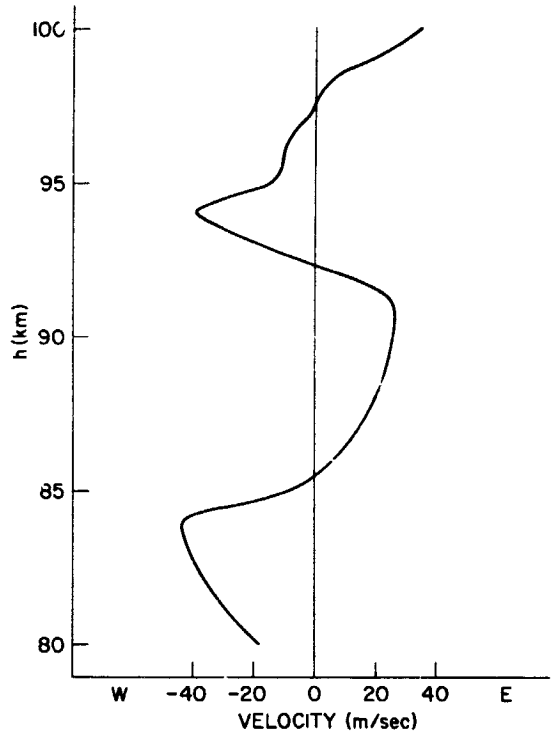


FIGURE 5.—Turbulent zonal profile.

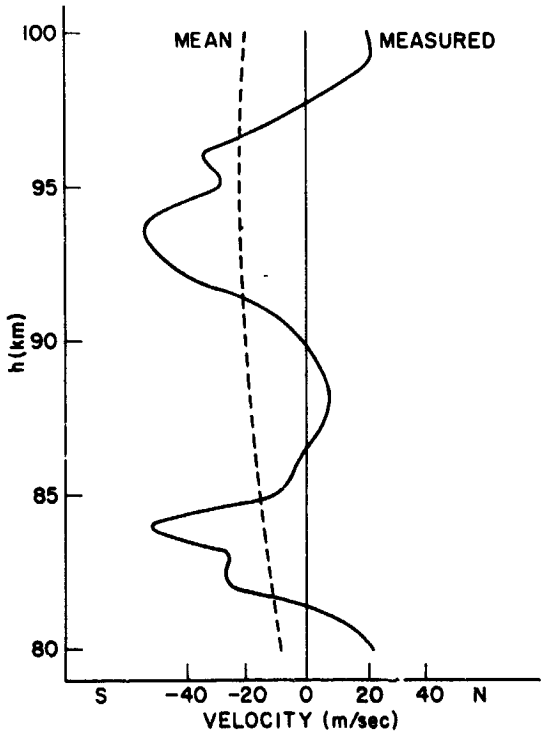


FIGURE 4.—Meridional profile, 80-100 km.

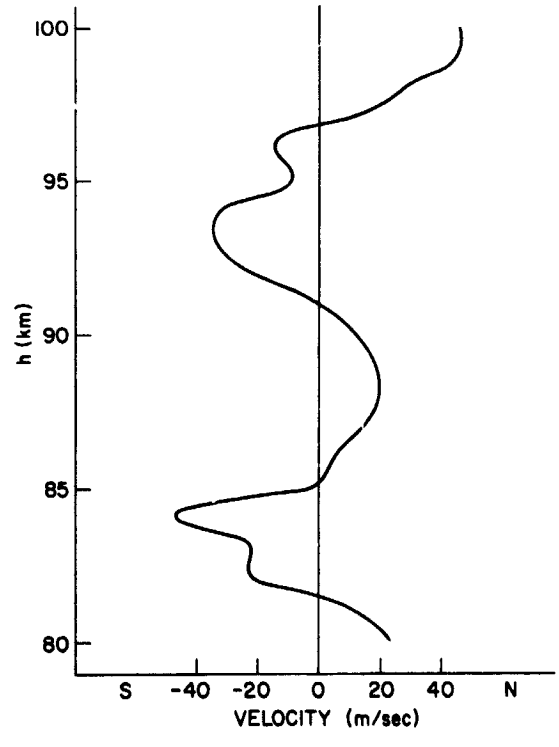


FIGURE 6.—Turbulent meridional profile.

in this region, applications of available wave theories (e.g., Hines, 1959) have not produced consistent results. The possibility of generation of turbulence in the 80 to 100 km region by vertically propagating gravity waves has been proposed by Hines (1963). The purpose of the present work, however, is not to determine the source of the turbulent energy, but rather to substantiate the evidence that the observed shears characterize a region of hydrodynamic turbulence.

3.3 The Energy Spectrum Functions

The energy spectrum functions for the zonal and meridional turbulent wind profiles for the 80 to 100 km region have been computed using the methods described in Sections 2.2 and 2.3. As can be seen from Figures 7 and 8, the velocity difference spectrum $E(\Delta h)$ and the correlation difference function $[1-f(\Delta h)]$ for the zonal turbulent velocities are completely equivalent. This is not surprising, since both functions are solutions of the same equation

$$F = a\Delta h^m$$

(see Sections 1.1 and 1.2).

Similarly for the meridional spectra of Figures 9 and 10. The slope m of the log log plots of both the zonal and meridional $E(\Delta h)$ (or $[1-R(\Delta h)]$) spectrum functions against height difference Δh is constant at 4/3 for separations Δh up to 3 km, indicative of an isotropic region subject to a mean wind shear (Tchen, 1954).

3.4 Isotropy

Since the energy spectrum functions and the correlation difference functions all follow an established (shear) law at small scales in the height range from 80 to 100 km, it is possible to determine the nature of the isotropy at these scales in this region.

Figure 13, curve A, is a plot of the variation of

$$S = \frac{1-f}{1-g}$$

against Δh.

In the region up to a scale of 3.5 km, S has a value of approximately 0.7, indicating an isotropy lying somewhere between two and three dimensional. The increase in S at scales greater than approximately 3 km is due to the breakdown

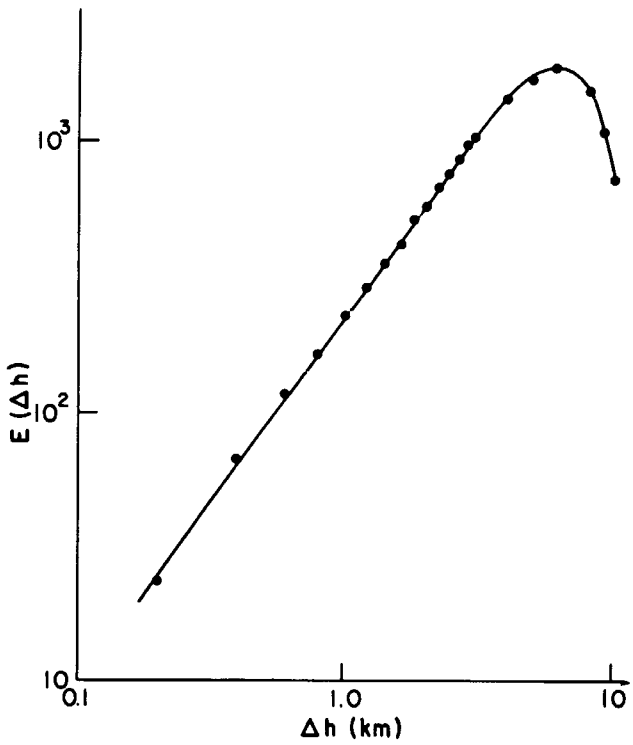


FIGURE 7.—The zonal energy spectrum function.

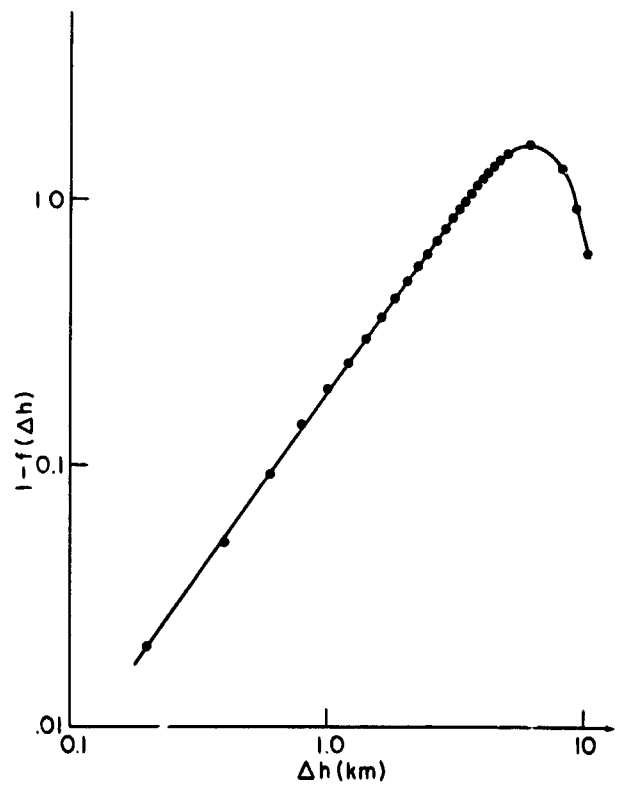


FIGURE 8.—The zonal correlation difference function.

at this scale of the 4/3 power law in both zonal and meridional spectrum functions.

The function S has been found to be quite sensitive to variation in the parameters specifying the mean wind profiles. For example, if, instead of the fourth order polynomial fits to the total height range of the zonal and meridional data, linear mean wind profiles are fitted over the height range 80 to 100 km only, subsequent spectrum analysis yields the S function of curve B of Figure 13. This curve would indicate the existence of three dimensional isotropy for scales up to 3 km. It is possible that the 4th order polynomial fits are, in fact, attributing a small fraction of the random wind variations to the mean motion. However, in order to correlate spectra obtained for different height strata (as is done in the next section), a continuous mean profile is required, and has therefore been used throughout the analysis.

3.5 Vertical Scale

The vertical scale associated with the turbulent wind structure is conveniently defined by the Δh corresponding to the maximum value of either $E(\Delta h)$ or $[1-R(\Delta h)]$. As can be seen from Figures 7, 9 and 11, which are plotted using zonal, meridional and windspeed turbulent velocities respectively, the vertical correlation distance as defined above is not the same for each component, i.e., the wind motions are not isotropic at the maximum of $E(\Delta h)$. However, for the purpose of comparison with stratospheric and lower mesospheric data, where meridional winds are, for the most part, negligible, consideration need only be given to the zonal energy spectrum.

The variation of vertical scale with height in the stratosphere has been determined by Webb (1964) from a series of Robin soundings at Eglin. His results indicate an exponential increase of the vertical scale with height, from approximately 800 metres at 35 km to 2 km at 55 km. He has extrapolated this exponential to a height of 90 km, and finds a vertical scale of 6 km which is the value determined for this height by Greenhow (1959) from a correlation analysis of wind shears determined by means of radio reflections from meteor trails at Jodrell Bank (53°N).

In order to investigate the change with height of the vertical correlation distance in the 80 to 130 km region, the sodium trail data was sub-

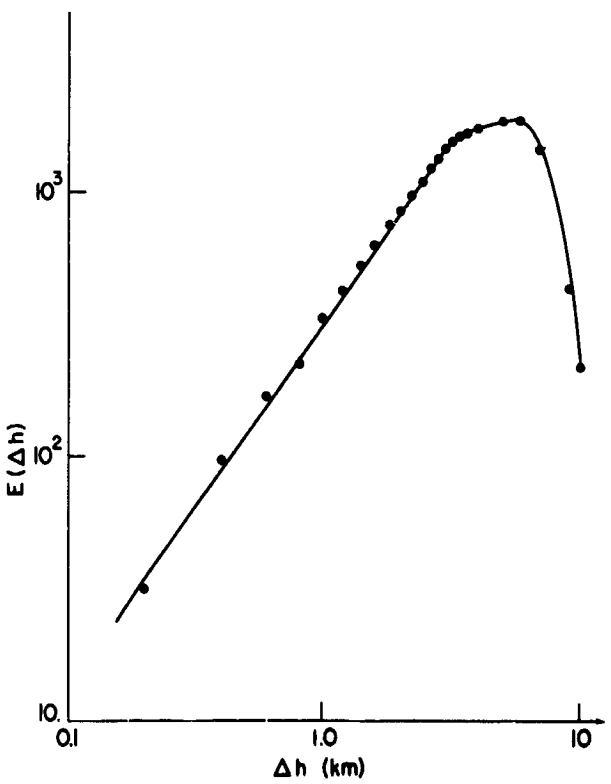


FIGURE 9.—The meridional energy spectrum function.

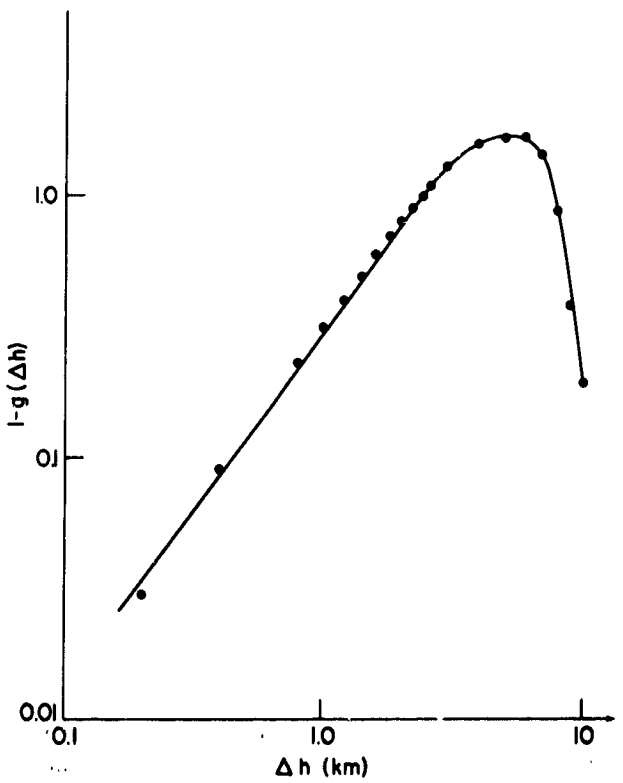


FIGURE 10.—The meridional correlation difference function.

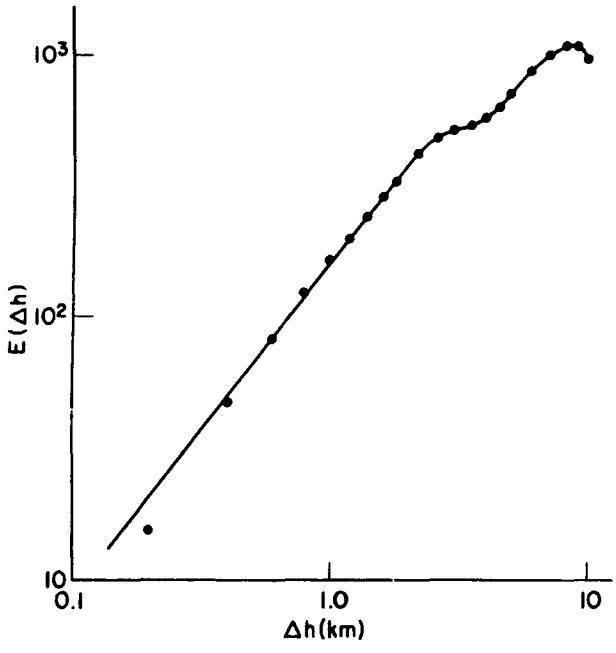


FIGURE 11.—The windspeed energy spectrum function.

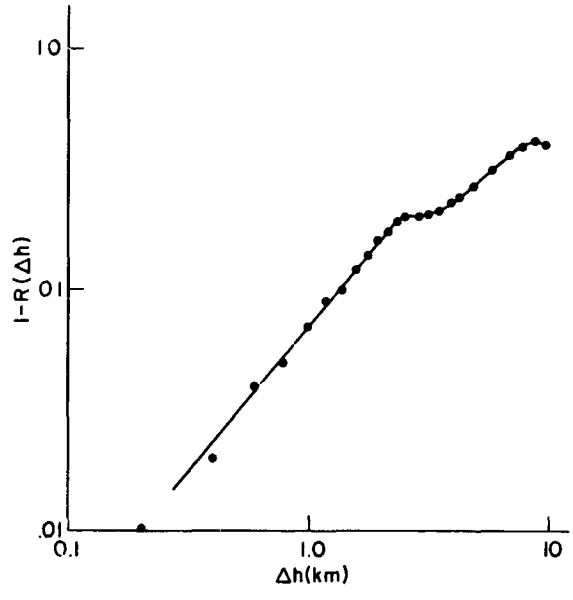


FIGURE 12.—The windspeed correlation difference function

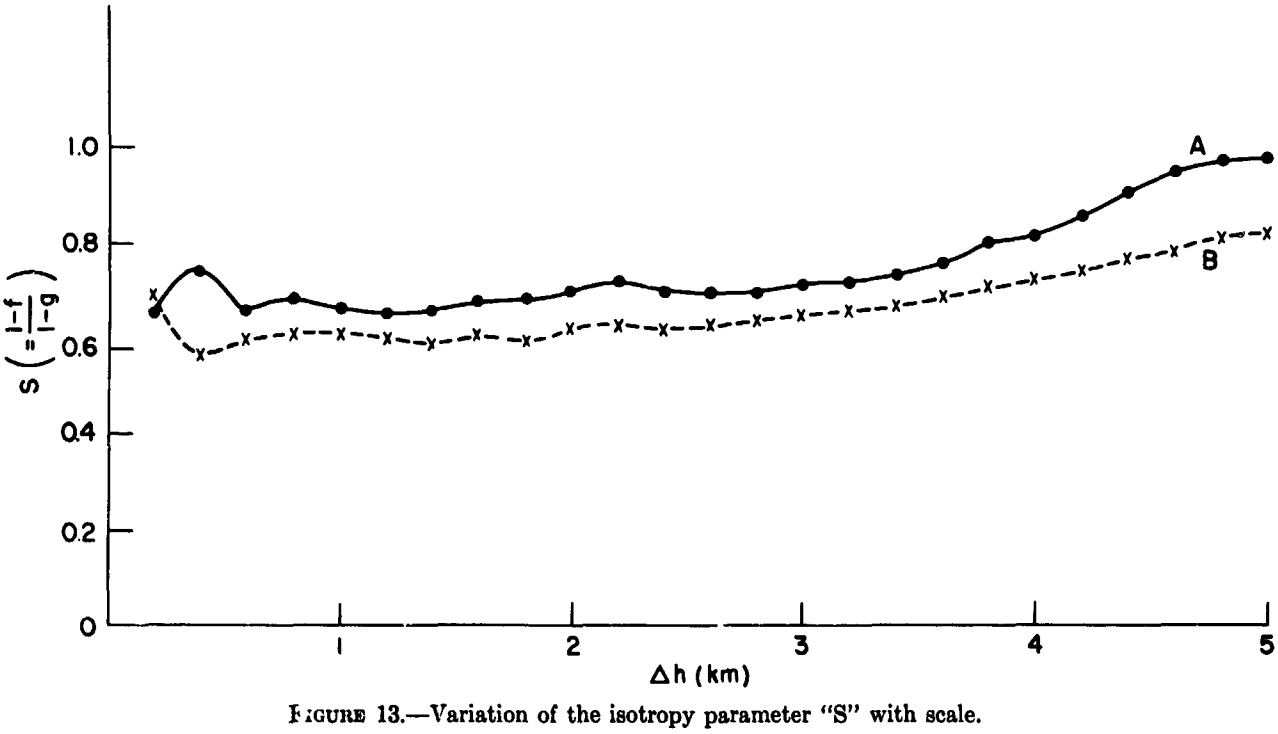


FIGURE 13.—Variation of the isotropy parameter "S" with scale.

jected to a stepwide analysis, starting with the 75 to 95 km height range, and proceeding via the 80 to 100, 85 to 105, etc. ranges to 115 to 135 km. The maxima of the resultant $E(\Delta h)$ curves were then plotted as the vertical scales at the midpoints of the respective height ranges. The results are shown in Figure 14, together with Greenhow's determination, and the 7.8 km at 94 km (October, 1961) determined from radio meteor trail shears at Adelaids (35°S) (Roper, 1962). The dashed line is Webb's extrapolated variation. While there is excellent agreement with Webb's proposed exponential increase, there is also good agreement between the vertical scales measured

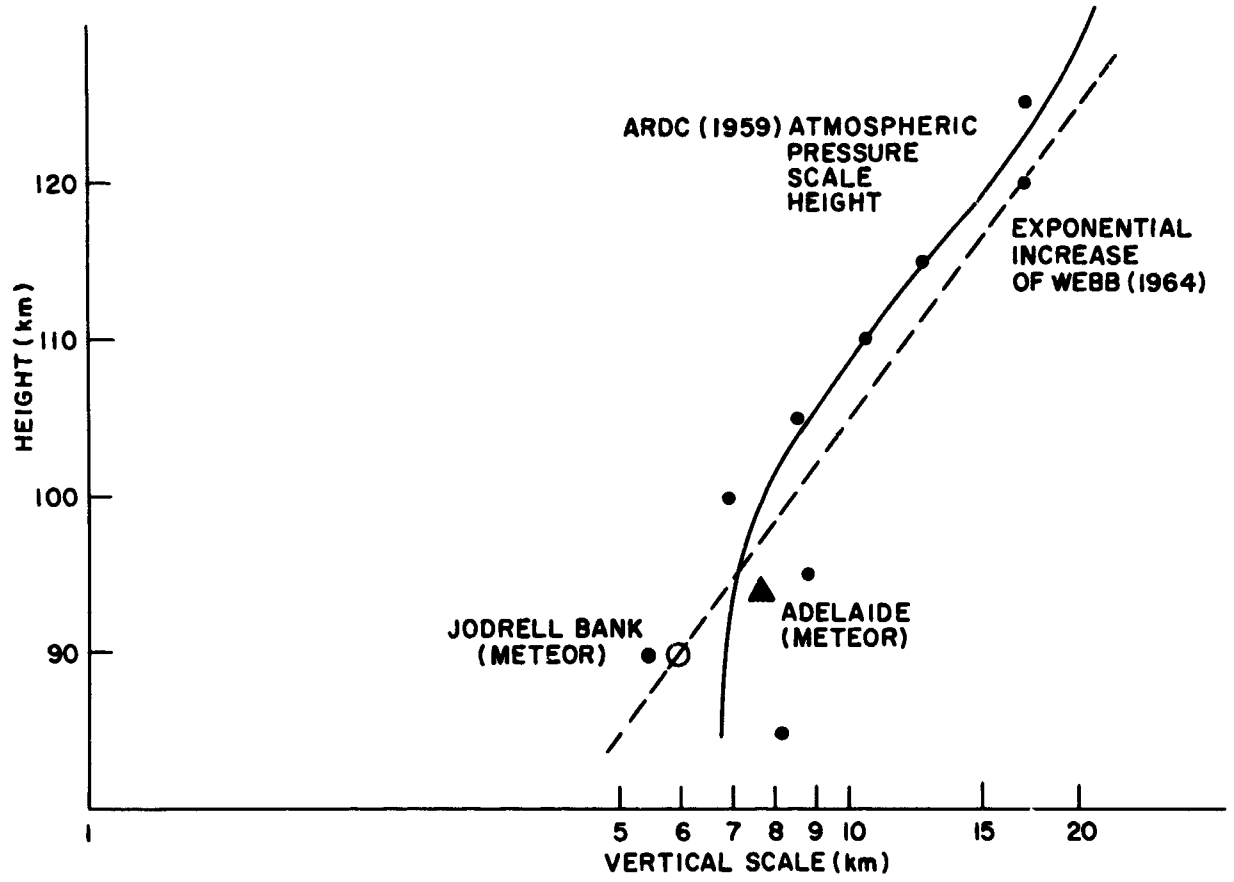


FIGURE 14.—Variation of the vertical scale of the zonal turbulent wind component with height.

at various altitudes for the Eglin firing, and the atmospheric pressure scale height (also shown in Figure 14). This phenomenon has also been observed by others (e.g., Zimmerman (1964)). As yet, no satisfactory explanation for such a dependence has been proposed.

Above 125 km, the magnitudes of the deviations of the measured winds from the mean wind values becomes insignificant. Wind motions at these heights for this particular firing are characteristically nonturbulent.

CONCLUSIONS

The techniques of spectrum analysis based on hydrodynamic turbulence theory can be profitably applied to wind data obtained from sodium trail rocket firings, at least in the height range from 80 to 100 km. The analysis of further firings should indicate whether or not the mean profiles determined as a prelude to spectrum analysis have any meteorological significance.

REFERENCES

GENERAL

KOLMOGORFF, A., C. R. AKAD. Sci. U.R.S.S., *30*, 301, 1941.
BATCHELOR, G. K., The Theory of Homogeneous Turbulence, Cambridge University Press, 1953
CORRSIN, S., Jour. Geophys. Res., *64*, 2134, 1959
TCHEN, C. M., Phys. Rev., *94*, 4, 1954
KARMAN, T., VON and HOWARTH, L., Proc. Roy. Soc., London *164*, 192, 1938
ELFORD, W. G., Plan. and Space Sci., *1*, 94, 1958
ELFORD, W. G., 1964 (unpublished)
HUTCHINGS, J. W., Jour. Met., *12*, 263, 1955
EDWARDS, H. D. et al., Jour. Geophys. Res. *68*, 6062, 1963
BLAMONT, J. E., and DE JAGAR, C., Ann. de Geophys., *17*, 134, 1961
ZIMMERMAN, S. P., Ann de Geophys. *18*, 116, 1962.
ROPER, R. G., Ph.D. Thesis, University of Adelaide, 1962
HINES, C. O., Jour. Geophys. Res., *64*, 2210, 1959
HINES, C. O., Quart. Jour. Roy. Met. Soc., *89*, 1, 1963
WEBB, W. L., U.S. Army E.R.D.A.—130, White Sands, New Mexico. April 1964
GREENHOW, J. S., and NEUFELD, E. L., Jour. Geophys. Res., *64*, 2129, 1959
ZIMMERMAN, S. P., Jour. Geophys. Res., *65*, 784, 1964.

APPENDIX I

RESUME OF THE COMPUTER PROGRAM

The computer program has been written in FORTRAN IV for execution on an IBM 7094, and consists of the following routines:

1. The main program SHEAR which reads and does some processing of the input data (which is punched on cards).

SHEAR calls subroutines

a) FIT, which fits polynomial profiles to the zonal and meridional wind velocity/height profiles by the method of least squares, utilizing

b) MATS, to solve the set of independent linear equations

c) The total wind speed profile is calculated by TRADE.

d) OUTPUT performs the energy spectrum and correlation analyses, and printout of same.

e) PAGE is a utility routine which turns and numbers the output pages, and prints an appropriate heading on each.

Input Data

The input data is on punched cards, and is read as follows:

1. Header Card; the information punched on this card is printed at the top of each page of output. The full 72 columns can be used for any type of alphanumeric heading.

2. The parameters specifying the mean zonal and meridional profiles. These determine the order of the polynomial fitted to

a) the zonal (punched in columns 2, 3) and

b) the meridional (punched in columns 5, 6) measured data to determine the mean profile. e.g., a 5 punched in column 3 will fit a polynomial of the form

$$a_1+a_2h+a_3h^2+a_4h^3+a_5h^4$$

to the zonal wind data.

The maximum allowable value of the subscript i in the a_i above is 10.

3. The height range (must be integral, and in km) over which the energy spectrum and correlation analyses are to be performed. This can be any fraction of the height interval covered by the input data, but must not be less than 10 km unless an appropriate spectral range is punched in column 13. The minimum of the height range interval is punched in columns 2 to 5, the maximum in columns 6 to 9. If OUTPUT is required to produce spectra over height differences to other than a maximum of 10 km, the required maximum height difference can be punched in columns 10 to 13. This entry must not exceed the height range specified in columns 2 to 9. If this last field is left blank, spectra to 10 km are output.

4. The trial data, which must not exceed a maximum to minimum height range of 150 km. The data is punched as year, month, day, hour (local time, 24 hour clock), height (km), windspeed (metres/sec), and zaimuth (degrees) as follows

YR	MO	DAY	HOUR	HEIGHT	WINDSPEED	AZIMUTH
3	3	9	14	19	24	29

all integer

must be right adjusted if integer; anywhere in appropriate field if decimal punch is included.

5. After all data of 4. a blank card to flag "end of data."

6. Further height ranges as for 3 as required.

7. a) If a change in the mean wind profile is required a card with a 2 punched in column 1 may be substituted for any of the cards 6, and followed by a card of format 2 above which specified the new profile. This card is then followed by further height ranges (as for 6.) as required.

7. b) If a completely new set of header plus data cards is to be processed after either stages 5, 6, or 7a, a card with a 1 punched in column 1 will return control to the start of the program, which will then read in the new set of data cards sequenced as from 1 above.

APPENDIX II

Listing of the FORTRAN IV Program

```
                SHEAR                                                    08/
          EXTERNAL FORMULA NUMBER — SOURCE STATEMENT — INTERNAL
C       SODIUM TRAIL HEIGHT SHEAR ANALYSIS PROGRAM
C
C       CALCULATES THE ENERGY SPECTRUM FOR TOTAL SHEAR, AND FOR ZONAL AND
C       MERIDIONAL SHEAR COMPONENTS
C
C       READS INPUT DATA AS FOLLOWS
C       A HEADER CARD, PUNCHED WITH DETAILS OF FIRING TIME, ETC.
C       FORMAT 12A6
C       THE PARAMETERS SPECIFYING THE MEAN ZONAL AND MERIDIONAL PROFILES,
C       FORMAT 213
C       THE HEIGHT RANGE OVER WHICH ANALYSIS IS TO BE PERFORMED, ZMIN,
C       ZMAX, AND THE MAXIMUM OF THE OUTPUT SPECTRAL RANGE DESIRED.
C       FORMAT  1X2F4.0, 14.
C       SPECTRAL RANGE MUST NOT EXCEED ZMAX-ZMIN.
C       IF NO SPECTRAL RANGE IS SPECIFIED, ZMAX-ZMIN MUST BE GREATER THAN
C       10KM, AND SPECTRA WILL BE OUTPUT TO 10KM.
C       THE TRAIL DATA, YEAR MONTH DAY LOCAL TIME (HOURS AND MINUTES,
C       24 HOUR CLOCK) HEIGHT(KM) WIND SPEED(METRES/SEC) WIND AZIMUTH
C       (DEGREES).
C       FORMAT 313, 15, F5.1, 2F5.0
C       A BLANK CARD
C       FURTHER ZMIN, ZMAX AS REQUIRED.  IF A 2 APPEARS IN COLUMN 1,
C       PROGRAM WILL READ NEXT CARD AS NEW PROFILE SPECIFICATION, FOLLOWED
C       BY FURTHER ZMIN, ZMAX CARDS.     IF A 1 APPEARS IN COLUMN 1, THE
C       PROGRAM WILL READ NEXT CARD AS HEADER CARD OF A COMPLETELY NEW SET
C       OF DATA.
C
        DIMENSION BZ(10), BM(10)
        DIMENSION RESULT(12), ZI(750), WINDI(750), AZRADI(750), ZONAL(750),
       1ERID(750),              TZONAL(750,) TMERID(750)
        DIMENSION TSPEED(750), AW(5)
        COMMON ZI, WINDI, AZRADI, ZONAL, ERID, TZONAL, TMERID, TSPEED
        C=0.05
        CS=0.000001
        PI=3.1415926
        TWOPI=2.0*PI
  999   NGO=0
        NI=0
        I=0
        NSUM=0
        ZERO=0.0
        READ      (2,1)RESULT
    1   FORMAT(12A6)
   17   READ      (2,4)NP,NQ
    4   FORMAT (213)
        NFIT=-1
    2   READ(2,3)NGO, ZMIN, ZMAX, NDIFF
    3   FORMAT(II,2F4.0,I4)
        IF(NGO)19, 19, 18
   18   IF(NGO-2)999, 17, 17
   19   CONTINUE
        IF(NDIFF)20, 20, 21
   20   NDIFF=10
   21   MIN=ZMIN+C
        MAX=ZMAX+C
```

SHEAR 08/
EXTERNAL FORMULA NUMBER — SOURCE STATEMENT — INTERNAL

```
      IF(NI)5, 5, 101
C     TRAIL DATA INPUT
    5 READ      (2,6)MYEAR, MONTH, JOUR, LTIME, Z, WIND, AZDEG

    6 FORMAT(3I3, I5, F5.1, 2F5.C)
      IF(MYEAR)100,100, 7
    7 I=I+1
      IF(I-1)11, 11, 8
    8 ZDIFF=Z-ZI(I-1)
      MZ=10.0*Z+C
      LZ=10.0*ZI(I-1)+C
      LDIFF=MZ-LZ
      IF(LDIFF-2)12, 11, 9
   12 WRITE      (3,6)MYEAR, MONTH, JOUR, LTIME, Z, WIND, AZDEG
      WRITE(3,13)
   13 FORMAT(1X/////1X26HDATA CARDS OUT OF SEQUENCE//////1X20HEXECUTION
     1TERMINATED//////1X)
      PRINT 13
      CALL EXIT
C     INTERPOLATION ROUTINE
    9 NPOINT=LDIFF/2-1
      NSUM=NSUM+NPOINT
      GRAD=(WIND - WINDI(I-1) )/ZDIFF
      AZDIFF=AZDEG/57.3-AZRADI(I-1)
      IF(ABS(AZDIFF)-PI)94,94,91
   91 IF(AZDIFF)92,94,93
   92 AZDIFF=AZDIFF+TWOPI
      GOTO94
   93 AZDIFF=AZDIFF-TWOPI
   94 AZGRAD=AZDIFF/ZDIFF
      DO10J=1,NPOINT
      ZI(I)=ZI(I-1)+0.2
      WINDI(I)=WINDI(I-1)+C.2*GRAD
      AZRADI(I)=AZRADI(I-1)+C.2*AZGRAD
      IF(AZRADI(I))95,96,96
   95 AZRADI(I)=AZRADI(I)+TWOPI
   96 IF(AZRADI(I)-TWOPI)98,98,97
   97 AZRADI(I)=AZRADI(I)-TWOPI
   98 CONTINUE
      I=I+1
   10 CONTINUE
   11 ZI(I)=Z
      WINDI(I)=WIND
      AZRADI(I)=AZDEG/57.3
      GO TO 5
  100 NI=I
C     ALL DATA IN.  COMMENCE COMPUTATION
      HDIFF=ZI(NI)-ZI(1)
      HPLUS=ZI(NI)+ZI(1)
C     CALCULATE ZONAL AND MERIDIONAL WIND COMPONENTS
      DO201J=1,NI
      ZONAL(J)=WINDI(J)*SIN(AZRADI(J))
  201 ERID(J)=WINDI(J)*COS(AZRADI(J))
  101 NZ=0
      LMIN=10*MIN
      LMAX=10*MAX
```

```
                SHEAR                                                    08/
      EXTERNAL FORMULA NUMBER  —  SOURCE STATEMENT  —  INTERNAL
      CALL PAGE (RESULT, ZERO)
      DO104J=1,NI
      LZ=ZI(J)*10.0+C
      IF(LZ-LMIN)104,102,102
  102 IF(LZ-LMAX)103,103,104
  103 NZ=NZ+1
  104 CONTINUE
      WRITE        (3,105)NI, NSUM, ZMIN, ZMAX, NZ, NP, NQ
  105 FORMAT(1X/////1X33HTOTAL NUMBER OF INPUT DATA POINTS I8///1X26HINC
     1LUDING INTERPOLATION OF 15,7H POINTS ////1X45HNUMBER OF DATA POINT
     2S WITHIN THE HEIGHT RANGE F5.0,6H KM TO,F5.0,3H KM I8////1X31HEAST
     3 WEST PROFILE SPECIFICATION I5 ///1X33HNORTH SOUTH PROFILE SPECIFI
     4CATION I3 ///1X)
      START=1
      CALL PAGE (RESULT, START)
      WRITE        (3,106)ZMIN, ZMAX
  106 FORMAT(1X50HCALCULATED ZONAL AND MERIDIONAL MEAN WIND PROFILES
     110X12HHEIGHT RANGE F5.0,6H KM TO F5.0,3H KM/1HO
     2/10X 6HHEIGHT 10X5HZONAL 12X10HMERIDIONAL 10X10HWIND SPEED /1X/
     31X,18X,3(4X12HDATA      MEAN,4X) /1X,)
      NHITE=(ZMAX-ZMIN)/20.0+C
      NHITE=NHITE+1
      INK=5*NHITE-1
      NFIT=NFIT+1
      IF(NFIT)202,202,203
C     CALCULATE COEFFICIENTS OF MEAN PROFILE POLYNOMIALS
  202 CALL NORMAL (ZONAL, ERID, ZI, NI, BZ, BM, NP, NQ)
  203 CONTINUE
C     DETERMINE AND PRINT CUT MEAN WIND PROFILES
      SUBS=(ZMAX+ZMIN-ZI(1))*5.0+1.0+C
      DO108L=MIN, MAX, NHITE
      IZ=SUBS-FLOAT(L*5)
      J=MAX+MIN-L
      H=J
      S=(2.0*H-HPLUS)/HDIFF+CS
      AZONAL=0.0
      AMERID=0.0
      DO204M=1,10
      AZONAL=AZONAL+BZ(M)*S**(M-1)
      AMERID=AMERID+BM(M)*S**(M-1)
  204 CONTINUE
      MZONAL=AZONAL
      MMERID=AMERID
      MSPEED=SQRT(AZONAL**2+AMERID**2)
      WRITE(3,107)J, ZONAL(IZ), MZONAL, ERID(IZ), MMERID, WINDI(IZ), MSPEED
  107 FORMAT(1XI13,4X,3(  F8.0,4XI14,4X),/1X)
  108 CONTINUE
      CALL PAGE (RESULT, START)
      WRITE(3,109)ZMIN, ZMAX
  109 FORMAT(1X,22HTURBULENT WIND PROFILE, 10X12HHEIGHT RANGE F5.0,
     1 6HKM TO F5.0, 3H KM/1X/1X, 10X6HHEIGHT, 10X6HTZONAL, 10X11HTMERIDIONA
     2L, 10X6HTSPEED/1X)
C     DETERMINE AND PRINT CUT TURBULENT WIND PROFILES
      NH=(MAX-MIN)*5+1
      K=(ZMAX-ZI(1))*5.0+2.0+C
      NK=INK
```

SHEAR 05/
EXTERNAL FORMULA NUMBER — SOURCE STATEMENT — INTERNAL

```
      RMSZ=0.0
      RMSM=0.0
      RMST=0.0
      SUMSQT=0.0
      DO112J=1,NP
      TZONAL(J)=0.0
      TMERID(J)=0.0
      K=K-1
      H=ZMAX-0.2*FLOAT(J-1)
      IH=H+0
      S=(H*H-HPLUS)/HDIFF+CS
      NK=NK+1
      DO205M=1,10
      TZONAL(J)=TZONAL(J)+BZ(M)*S**(M-1)
      TMERID(J)=TMERID(J)+BM(M)*S**(M-1)
  205 CONTINUE
      TZONAL(J)=ZONAL(K)-TZONAL(J)
      TMERID(J)=ERID(K)-TMERID(J)
      CALL TRADE(J,K)
C     CALCULATE RMS TURBULENT VELOCITIES FOR THE HEIGHT RANGE ZMIN/ZMAX
      RMSZ=TZONAL(J)**2+RMSZ
      RMSM=TMERID(J)**2+RMSM
      RMST=TSPEED(J)**2+RMST
      SUMSQT=SUMSQT+1.0
      IF(NK-INK)112,112,110
  110 WRITE(3,111)IH, TZONAL(J), TMERID(J), TSPEED(J)
  111 FORMAT(1X, 10X14, 12XF5.0, 11XF8.0, 13XF5.0/1X)
      NK=0
  112 CONTINUE
      CALL PAGE(RESULT, START)
      RMSZ=SQRT(RMSZ/SUMSQT)
      RMSM=SQRT(RMSM/SUMSQT)
      RMST=SQRT(RMST/SUMSQT)
      WRITE (3,206)ZMIN, ZMAX, (BZ(I), I=1, NP)
  206 FORMAT(1X10HNORMALIZED
     1          1X12HHEIGHT RANGE 10XF7.0,7H KM. TO F7.0, 4H KM./1H0/
     2          1X42HCOEFFICIENTS GIVING BEST FIT TO ZONAL DATA/1X/10E11.3)
      WRITE(3,207)(BM(I),I=1,NQ)
  207 FORMAT(1X/1X/
     11X/1X47HCOEFFICIENTS GIVING BEST FIT TO MERIDIONAL DATA/1X/1X10E11
     2.3/1X/)
      WRITE(3,208)RMSZ, RMSM, RMST
  208 FORMAT(1X/1X
     1       1X22HRMS TURBULENT VELOCITY/1X/1X10X5HZONALF11.0,11HMETRES/S
     2E0./1X/1X10X10HMERIDIONAL F6.0,11HMETRES/SEC./1X/1X10X10HWIND SPEE
     3D F6.0,11HMETRES/SEC.)
      CALL PACE(RESULT, START)
      WRITE      (3,113)ZMIN,ZMAX
  113 FORMAT(1X46HENERGY SPECTRUM OF ZONAL WIND HEIGHT VARIATION,
     110X12HHEIGHT RANGE F5.0,6H KM TO F5.0,3H KM/1X)
C     CALCULATE AND OUTPUT ENERGY SPECTRUM FUNCTIONS
      CALL OUTPUT(TZONAL, NH, NDIFF)
      CALL PAGE (RESULT, START)
      WRITE      (3,114) ZMIN, ZMAX
  114 FORMAT(1X51HENERGY SPECTRUM OF MERIDIONAL WIND HEIGHT VARIATION,
```

```
                SHEAR                                                    08/
      EXTERNAL FORMULA NUMBER -- SOURCE STATEMENT -- INTERNAL
      CALL PAGE(RESULT, ZERO)
      DO104J=1,NI
      LZ=ZI(J)*10.0+C
      IF(LZ-LMIN)104,102,102
  102 IF(LZ-LMAX)103,103,104
  103 NZ=NZ+1
  104 CONTINUE
      WRITE      (3,105)NI, NSUM, ZMIN, ZMAX, NZ, NP, NQ
  105 FORMAT(1X/////1X33HTOTAL NUMBER OF INPUT DATA POINTS I8///1X26HINC
     1LUDING INTERPOLATION OF 15,7H POINTS ////1X45HNUMBER OF DATA POINT
     2S WITHIN THE HEIGHT RANGE F5.0,6H KM TO,F5.0,3H KM I8////1X31HEAST
     3 WEST PROFILE SPECIFICATION I5 ///1X33HNORTH SOUTH PROFILE SPECIFI
     4CATION 13 ///1X)
      START=1
      CALL PACE(RESULT,START)
      WRITE      (3,106)ZMIN, ZMAX
  106 FORMAT(1X50HCALCULATED ZONAL AND MERIDIONAL MEAN WIND PROFILES
     110X12HHEIGHT RANGE F5.0,6H KM TO F5.0,3H KM/1HO
     2/10X 6HHEIGHT 10X5HZONAL 13X10HMERIDIONAL 10X10HWIND SPEED /1X/
     31X,18X,3(4X12HDATA      MEAN,4X),/1X/)
      NHITE=(ZMAX-ZMIN)/20.0-C
      NHITE=NHITE+1
      INK=5*NHITE-1
      NFIT=NFIT+1
      IF(NFIT)202,202,203
C     CALCULATE COEFFICIENTS OF MEAN PROFILE POLYNOMIALS
  202 CALL NORMAL (ZONAL, ERID, ZI, NI, BZ, BM, NP, NQ)
  203 CONTINUE
C     DETERMINE AND PRINT OUT MEAN WIND PROFILES
      SUBS=(ZMAX+ZMIN-ZI(1))*5.0+1.0+C
      DO108L=MIN, MAX, NHITE
      IZ=SUBS-FLOAT(L*5)
      J=MAX+MIN-L
      H=J
      S=(2.0*H-HPLUS)/HDIFF+CS
      AZONAL=0.0
      AMERID=0.0
      DO204M=1,10
      AZONAL=AZONAL+BZ(M)*S**(M-1)
      AMERID=AMERID+BM(M)*S**(M-1)
  204 CONTINUE
      MZONAL=AZONAL
      MMERID=AMERID
      MSPEED=SQRT(AZONAL**2+AMERID**2)
      WRITE(3,107)J, ZONAL(IZ), MZONAL, ERID(IZ), MMERID, WINDI(IZ),MSPEED
  107 FORMAT(1XI13,4X,3(  F8.C,4XI4,4X),/1X)
  108 CONTINUE
      CALL PAGE (RESULT, START)
      WRITE(3,109)ZMIN, ZMAX
  109 FORMAT(1X,22HTURBULENT WIND PROFILE, 10X12HHEIGHT RANGE F5.0,
     1 6HKM TO F5.0,3H KM/1X/1X,10X6HHEIGHT, 10X6HTZONAL, 10X11HTMERIDIONA
     2L,10X6HTSPEED/1X)
C     DETERMINE AND PRINT OUT TURBULENT WIND PROFILES
      NH=(MAX-MIN)*5+1
      K=(ZMAX-ZI(1))*5.0+2.0+C
      NK=INK
```

SHEAR 08/
EXTERNAL FORMULA NUMBER — SOURCE STATEMENT — INTERNAL

```
      RMSZ=0.0
      RMSM=0.0
      RMST=0.0
      SUMSQT=0.0
      DO112J=1,NH
      TZONAL(J)=0.0
      TMERID(J)=0.0
      K=K-1
      H=ZMAX-0.2*FLOAT(J-1)
      IH=H+C
      S=(2.0*H-HPLUS)/HDIFF+CS
      NK=NK+1
      DO205M=1,10
      TZONAL(J)=TZONAL(J)+BZ(M)*S**(M-1)
      TMERID(J)=TMERID(J)+BM(M)*S**(M-1)
  205 CONTINUE
      TZONAL(J)=ZONAL(K)-TZONAL(J)
      TMERID(J)=ERID(K)-TMERID(J)
      CALL TRADE(J,K)
C     CALCULATE RMS TURBULENT VELOCITIES FOR THE HEIGHT RANGE ZMIN/ZMAX
      RMSZ=TZONAL(J)**2+RMSZ
      RMSM=TMERID(J)**2+RMSM
      RMST=TSPEED(J)**2+RMST
      SUMSQT=SUMSQT+1.0
      IF(NK-INK)112,112,110
  110 WRITE(3,111)IH,TZONAL(J),TMERID(J),TSPEED(J)
  111 FORMAT(1X,10XI4,12XF5.0,11XF8.0,13XF5.0/1X)
      NK=0
  112 CONTINUE
      CALL PAGE (RESULT, START)
      RMSZ=SQRT(RMSZ/SUMSQT)
      RMSM=SQRT(RMSM/SUMSQT)
      RMST=SQRT(RMST/SUMSQT)
      WRITE(3,206)ZMIN,ZMAX,(BZ(I),I=1,NP)

  206 FORMAT(1X10HNORMALIZED
     1          1X12HHEIGHT RANGE 10XF7.0,7H KM.  TO F7.0,4H KM./1H0/
     2          1X42HCOEFFICIENTS GIVING BEST FIT TO ZONAL DATA/1X/10E11.3)
      WRITE(3,207)(BM(I),I=1,NQ)
  207 FORMAT(1X/1X/
     11X/1X47HCOEFFICIENTS GIVING BEST FIT TO MERIDIONAL DATA/1X/1X10E11
     2.3/1X/)
      WRITE(3,208)RMSZ, RMSM, RMST
  208 FORMAT(1X/1X
     1       1X22HRMS TURBULENT VELOCITY/1X/1X10X5HZONALF11.0,11HMETRES/S
     2EC./1X/1X10X10HMERIDIONAL F6.0,11HMETRES/SEC./1X/1X10X10HWIND SPEE
     3D F6.0,11HMETRES/SEC.)
      CALL PAGE (RESULT, START)
      WRITE      (3,113)ZMIN, ZMAX
  113 FORMAT(1X46HENERGY SPECTRUM OF ZONAL WIND HEIGHT VARIATION,
     110X12HHEIGHT RANGE F5.0,6H KM TO F5.0,3H KM/1X)
C     CALCULATE AND OUTPUT ENERGY SPECTRUM FUNCTIONS
      CALL OUTPUT(TZONAL, NH, NDIFF)
      CALL PAGE (RESULT, START)
      WRITE      (3,114) ZMIN, ZMAX
  114 FORMAT(1X51HENERGY SPECTRUM OF MERIDIONAL WIND HEIGHT VARIATION,
```

```
      SHEAR                                                          08/
EXTERNAL FORMULA NUMBER — SOURCE STATEMENT — INTERNAL
     110X12HHEIGHT RANGE F5.0,6H KM TO F5.0, 3H KM/1X)
      CALL OUTPUT(TMERID, NH, NDIFF)
      CALL PAGE (RESULT, START)
      WRITE      (3,115)ZMIN,ZMAX
  115 FORMAT(1X46HENERGY SPECTRUM OF WIND SPEED HEIGHT VARIATION,
     110X12HHEIGHT RANGE F5.0,6H KM TO F5.0,3H KM/1X)
      CALL OUTPUT(TSPEED, NH, NDIFF)
      GOTO2
      END
```

```
/$ID JO2T          RGR     BLDG 11
 $PAUSE
 $EXECUTE          IBJOB
 IBJOB VERSION 2,  7090 - PP - 929
1


0
 $IBJOB            GO
 $IBFTC OUT        M94,XR7
1                 OUT
              EXTERNAL FORMULA NUMBER  —  SOURCE STATEMENT  —  INTERN
0
      SUBROUTINE OUTPUT(TSPEED,NH,NDIFF)
C
C     PRODUCES HEIGHT SPECTRUM AND PERFORMS CORRELATION ANALYSIS.
C
      DIMENSION TSPEED(750)
      WRITE(3,1)
    1 FORMAT(1X,10X7HDELTA H,5X10HE(DELTA H),
     15X11HCORRELATION,3X13HSIGNIF      1-G   /1X)
      NEND=5*NDIFF
      LINE=0
      DO125K=1,NEND
      LINE=LINE+1
      DELTAH=FLOAT(K)*0.2
      SUM=0.0
      ENERGY=0.0
      ERGX=0.0
      GJPKXJ=0.0
      SQJ=0.0
      SQJPK=0.0
      DO123J=1,NH
      IF(NH-J-K)124,122,122
  122 JPK=J+K
      ENERGY=ENERGY+(TSPEED(JPK)-TSPEED(J))**2
      SUM=SUM+1.0
      GJPKXJ=GJPKXJ+TSPEED(JPK)*TSPEED(J)
      SQJ=SQJ+TSPEED(J)**2
      SQJPK=SQJPK+TSPEED(JPK)**2
  123 CONTINUE
  124 IF(SUM)117,117,118
  117 G=0.0
      SIGNIF=9.99
      GOTO119
  118 ENERGY=ENERGY/SUM
      G=GJPKXJ/SQRT(SQJ*SQJPK)
      SIGNIF=(1.0-G**2)/SQRT(SUM-1.0)
  119 DIFF=1.0-G
      IF(LINE-50)125,125,116
  116 WRITE(3,2)
    2 FORMAT(1H1/1X)
      WRITE(3,1)
      LINE=0
  125 WRITE(3,115)DELTAH,ENERGY,      G,SIGNIF,DIFF
  115 FORMAT(1X,6XF8.1,  7XF8.1 ,8XF8.3,F11.3,F8.2)
      RETURN
      END
```

```
$1BJOB          GO
$IBFTC NORMAL   M94,XR7



            NORMAL
       EXTERNAL FORMULA NUMBER  —  SOURCE STATEMENT  —  INTE

      SUBROUTINE NORMAL (ZONAL,ERID,ZI,NI,BZ,BM,NP,NQ)
C
C     DETERMINES BEST FIT PROFILE TO THE NORMALIZED TOTAL HEIGHT RANG
C
      DIMENSION ZONAL(750),ERID(750),ZI(750),BZ(1C),BM(1C),ZINORM(750
      DO1J=1,NI
      ZINORM(J)=(2.0*ZI(J)-ZI(NI)-ZI(1))/(ZI(NI)-ZI(1))
    1 CONTINUE
      CALL FIT (ZONAL,NP,BZ,NI,ZINORM)
      CALL FIT (ERID, NQ,BM,NI,ZINORM)
      RETURN
      END



$IBJOB          GO
$IBFTC TRADE    M94,XR7



              TRADE                                              08/
       EXTERNAL FORMULA NUMBER  —  SOURCE STATEMENT  —  INTERNAL

      SUBROUTINE TRADE(J,K)
C
C     CALCULATES TURBULENT WINDSPEED PROFILE
C
      DIMENSION ZI(750)
      DIMENSION WINDI(750),AZRADI(750),ZONAL(750),ERID(750),TZONAL(750),
     1TMERID(750),TSPEED(750)
      COMMON ZI, WINDI,AZRADI,ZONAL,ERID,TZONAL,TMERID,TSPEED
      TSPEED(J)=WINDI(K)-SQRT((ZONAL(K)-TZONAL(J))**2+(ERID(K)-TMERID(J)
     1)**2)
      RETURN
      END



                                                                 08/17
$IBFTC FIT      M94,XR7
```

FIT 08/17
EXTERNAL FORMULA NUMBER -- SOURCE STATEMENT -- INTERNAL FO

```
      SUBROUTINE FIT (BLOW,NSPEC,A,N,X)
C
C     FITS A HEIGHT PROFILE SPECIFIED BY NSPEC TO THE MEASURED DATA
C     BLOW, DETERMINING THE COEFFICIENTS A BY THE METHOD OF LEAST
C     SQUARES.  PROFILE FITTED OVER TOTAL HEIGHT RANGE OF INPUT DATA.
C
      DIMENSION BLOW(750), A(10), X(750), S(10,11)
      DO3J=1,10
    3 A(J)=0.0
      IF(NSPEC)100,100,4
    4 NADD=NSPEC+1
      DO10J=1,NSPEC
      DO10K=1,NADD
   10 S(J,K)=0.0
      DO5J=1,NSPEC
      DO5K=1,NSPEC
      DO5I=1,N
    5 S(J,K)=S(J,K)+X(I)**(K+J-2)
      DU6J=1,NSPEC
      K=0
      DO6I=1,N
      K=K+1
    6 S(J,NADD)=S(J,NADD)+X(I)**(J-1)*BLOW(K)
      CALL MATS(S,A,NSPEC,MISS)
      IF(MISS)100,100,7
    7 WRITE(3,8)
    8 FORMAT(1X75HERROR IN INPUT DATA HAS RESULTED IN MATRIX S BEING UNS
     1UITABLE FOR INVERSION ////1X20HEXECUTION TERMINATED ///////)
      PRINT 8
      CALL EXIT
  100 RETURN
      END
```

```
                MATS                                                    08/17
      EXTERNAL FORMULA NUMBER — SOURCE STATEMENT — INTERNAL FO

      SUBROUTINE MATS (S,A,NSPEC,MISS)
C
C     REDUCES THE AUGMENTED MATRIX S TO PRODUCE THE COEFFICIENTS A
C
      DIMENSION S(10,11), A(10)
      MISS=-1
      MM=NSPEC+1
      N=NSPEC
      DO15I=2,N
   70 II=I-1
    7 DO15J=1,II
    8 IF(S(I,J))9,15,9
    9 IF(ABS(S(J,J))-ABS(S(I,J)))11,10,10
   10 R=S(I,J)/S(J,J)
      GOTO 130
   11 R=S(J,J)/S(I,J)
      DO12K=1,MM
      B=S(J,K)
      S(J,K)=S(I,K)
   12 S(I,K)=B
  130 JJ=J+1
   13 DO14K=JJ,MM
   14 S(I,K)=S(I,K)-R*S(J,K)
   15 CONTINUE
      IF(ABS(S(N,N))-1.0E-10)16,16,17
   16 MISS=+1
      GOTO29
   17 A(N)=S(N,MM)/S(N,N)
      DO28I=2,N
      JJ=N-I+1
      B=0.0
      II=N-I+2
      DO25K=II,N
   25 B=B+S(JJ,K)*A(K)
      IF(ABS(S(JJ,JJ))-1.0E-10)16,16,28
   28 A(JJ)=(S(JJ,MM)-B)/S(JJ,JJ)
   29 RETURN
      END
```

08/17

```
$IBFTC MATS      M94,XR7
```

08/17

```
$IBFTC PAGE      M94,XR7
```

```
      PAGE                                                    08/17
     EXTERNAL FORMULA NUMBER — SOURCE STATEMENT — INTERNAL FO

      SUBROUTINE PAGE (RESULT, START)
C
C     TURNS PAGE, NUMBERS IT, AND WRITES HEADING AS APPEARING
C     ON RESULT CARD.
C
      DIMENSION RESULT (12)
      IF(START)1,1,2
    1 NOP=0
    2 NOP=NOP+1
      WRITE  (3,3) RESULT, NOP
    3 FORMAT (1H1/1X12A6,30X4HPAGE15//)
      RETURN
      END
```

APPENDIX III

Sample Data And Computer Output

PRELIMINARY PAPER EGLIN DATA, MAY 21, 1963. 16/9/64. PROFILE,5,5.

5	5					
80	100					
63	5	21	1910	69.0	34	255
63	5	21	1910	70.6	31	256
63	5	21	1910	71.0	27	270
63	5	21	1910	72.4	43	272
63	5	21	1910	72.6	20	290
63	5	21	1910	77.0	41	256
63	5	21	1910	80.4	25	358
63	5	21	1910	82.0	08	310
63	5	21	1910	82.4	28	202
63	5	21	1910	83.6	27	200
63	5	21	1910	84.0	54	190
63	5	21	1910	85.0	54	194
63	5	21	1910	85.4	30	108
63	5	21	1910	88.0	60	080
63	5	21	1910	91.6	76	095
63	5	21	1910	95.0	50	178
63	5	21	1910	96.0	40	120
63	5	21	1910	96.6	54	150
63	5	21	1910	97.6	40	115
63	5	21	1910	99.0	42	084
63	5	21	1910	99.6	64	078
63	5	21	1910	100.0	58	070
63	5	21	1910	100.8	65	069
63	5	21	1910	102.0	80	079
63	5	21	1910	103.0	93	094

	63	5	21	1910	104.0	115	104
	63	5	21	1910	104.2	111	120
	63	5	21	1910	105.0	129	122
	63	5	21	1910	105.6	140	130
	63	5	21	1910	105.8	130	128
	63	5	21	1910	107.0	125	132
	63	5	21	1910	108.0	92	140
	63	5	21	1910	109.0	50	160
	63	5	21	1910	110.0	42	240
	63	5	21	1910	111.0	56	250
	63	5	21	1910	113.0	60	262
	63	5	21	1910	114.0	84	274
	63	5	21	1910	115.0	68	270
	63	5	21	1910	117.2	72	282
	63	5	21	1910	125.0	45	286
	63	5	21	1910	130.0	30	296
	63	5	21	1910	135.0	14	310
	63	5	21	1910	140.0	15	316
							← BLANK CARD
7 8							← EOF

TRIAL ANALYSIS. EGLIN DATA, MAY 21, 1963. 15/9/64. PROFILE,5 5.

TOTAL NUMBER OF INPUT DATA POINTS 356

INCLUDING INTERPOLATION OF 313 POINTS

NUMBER OF DATA POINTS WITHIN THE HEIGHT RANGE 80. KM TO 100. KM 101

EAST WEST PROFILE SPECIFICATION 5

NORTH SOUTH PROFILE SPECIFICATION 5

TRIAL ANALYSIS. EGLIN DATA, MAY 21, 1963. CONTINUOUS PROFILE,5,5.

CALCULATED ZONAL AND MERIDIONAL MEAN WIND PROFILES HEIGHT RANGE

HEIGHT	ZONAL		MERIDIONAL		WIND SPEED	
	DATA	MEAN	DATA	MEAN	DATA	MEAN
100	70.	34	20.	−21	72.	40
99	56.	37	21.	−21	60.	43
98	42.	39	4.	−22	42.	45
97	39.	41	−11.	−22	41.	47
96	34.	43	−35.	−22	48.	49
95	31.	45	−28.	−22	42.	50
94	6	46	−51.	−22	52.	51
93	31.	46	−51.	−21	59.	51
92	55.	47	−38.	−21	67.	51
91	73.	47	−13.	−21	74.	51
90	72.	46	−1.	−20	72.	50
89	67.	45	4.	−19	67.	49
88	62.	43	8.	−18	63.	47
87	55.	41	5.	−18	55.	45
86	44.	39	−4.	−16	44.	43
85	31.	36	−9.	−15	32.	39
84	−12.	33	−53.	−14	54.	36
83	−9.	29	−25.	−13	27.	32
82	−10.	24	−26.	−11	28.	27
81	−8.	19	10.	−10.	12.	22
80	−3.	14	23.	−8	23.	16

TRIAL ANALYSIS. EGLIN DATA, MAY 21, 1963. CONTINUOUS PROFILE,5,5.

TURBULENT WIND PROFILE HEIGHT RANGE 80. KM TO 100. KM

HEIGHT	TZONAL	TMERIDIONAL	TSPEED
100	35.	41.	42.
99	19.	42.	26.
98	2.	26.	16.
97	−3.	11.	31.
96	−10.	−12.	32.
95	−14.	−6.	29.
94	−40.	−29.	8.
93	−16.	−29.	31.
92	8.	−16.	58.
91	26.	8.	62.
90	25.	19.	52.
89	22.	24.	45.
88	18.	27.	38.
87	13.	24.	33.
86	4.	13.	30.
85	−6.	7.	27.
84	−45.	−38.	18.
83	−39.	−12.	−5.
82	−35.	−14.	0.
81	−28.	20.	3.
80	−18.	31.	10.

TRIAL ANALYSIS. EGLIN DATA, MAY 21, 1963. 15/9/64. PROFILE,5,5.

NORMALIZED HEIGHT RANGE 80. KM. TO 100. KM.

COEFFICIENTS GIVING BEST FIT TO ZONAL DATA

0.386E 02 −0.126E 03 −0.244E 03 0.174E 03 0.205E 03

COEFFICIENTS GIVING BEST FIT TO MERIDIONAL DATA

−0.304E 02 0.198E 02 0.148E 03 −0.168E 02 −0.125E 03

RMS TURBULENT VELOCITY

ZONAL	22. METRES/SEC.
MERIDIONAL	23. METRES/SEC.
WIND SPEED	17. METRES/SEC.

TRIAL ANALYSIS. EGLIN DATA, MAY 21, 1963. CONTINUOUS PROFILE,5,5.

ENERGY SPECTRUM OF ZONAL WIND HEIGHT VARIATION HEIGHT RANGE 80.

DELTA H	DELTA V**2	CORRELATION	SIGNIF	I-G
0.2	24.3	0 979	0.004	0.02
0.4	64.9	0.943	0.011	0 06
0.6	113.7	0.900	0.019	0.10
0.8	169.4	0.851	0.028	0.15
1.0	230.7	0.797	p 037	0.20
1.2	297.1	0 739	0.047	0.26
1.4	369.4	0.676	0.056	0.32
1.6	450.4	0.604	0.066	0.40
1.8	534.8	0.525	0.076	0.47
2.0	619.9	0.449	0.084	0.55
2.2	703.2	0.374	0.091	0.63
2.4	789.0	0 297	0.097	0.70
2.6	877.1	0.217	0.102	0.78
2.8	964.9	0.136	0.106	0.86
3.0	1054.0	0.053	0.108	0.95
3.2	1133.4	−0.023	0.109	1.02
3.4	1209.9	−0.099	0.109	1.10
3.6	1285.2	−0.176	0.107	1.18
3.8	1360.8	- 0.257	0.104	1.26
4.0	1436.0	−0.342	0.099	1 34
4.2	1507.0	−0.429	0.092	1.43
4.4	1567.1	−0.514	0.083	1.51
4.6	1614.3	−0.597	0.073	1.60
4.8	1681.4	−0.651	0.066	1.65
5.0	1745.7	−0 696	0.060	1.70
5.2	1804.0	−0.735	0.053	1 74
5.4	1852.3	−0.766	0.048	1.77
5.6	1885.5	−0.788	0.045	1.79
5.8	1897.3	−0.800	0.043	1.80
6.0	1880.3	−0.801	0.043	1.80
6.2	1863.6	−0.795	0.044	1.80
6.4	1847.4	−0.783	0.047	1.78
6.6	1827.7	−0.762	0.051	1.76
6.8	1803.8	−0.730	0.057	1.73
7.0	1775.0	−0.690	0.065	1.69
7.2	1740.9	−0.642	0.073	1.64
7.4	1701.2	−0.588	0.082	1.59
7.6	1656.5	−0.528	0.092	1.53
7.8	1608.4	−0.466	0.100	1.47
8.0	1555.6	−0.401	0.108	1.40
8.2	1497.1	−0.333	0.116	1.33
8.4	1432.6	−0.263	0.122	1.26
8.6	1361.9	−0.190	0.128	1.19
8.8	1285.1	−0.116	1.132	1.12
9.0	1202.6	−0.040	0.135	1.04
9.2	1114.7	0.038	0.136	0.96
9.4	1020.5	0.120	0.135	0.88
9.6	920.9	0.207	0.133	0.79
9.8	819.6	0.296	0.128	0.70
10.0	733.6	0.372	0.122	0.63

TRIAL ANALYSIS. EGLIN DATA, MAY 21, 1963. CONTINUOUS PROFILE,5,5.

ENERGY SPECTRUM OF MERIDIONAL WIND HEIGHT VARIATION HEIGHT RANGE

DELTA H	DELTA V**2	CORRELATION	SIGNIF	I-G
0.2	31.3	0.973	0.005	0.03
0.4	94.1	0.917	0.916	0.08
0.6	164.4	0.853	0.028	0.15
0.8	241.7	0.781	0.040	0.22
1.0	326.3	0.701	0.052	0.30
1.2	418.3	0.612	0.064	0.39
1.4	515.2	0.518	0.076	0.48
1.6	616.0	0.420	0.086	0.58
1.8	723.6	0.316	0.094	0.68
2.0	825.3	0.219	0.100	0.78
2.2	935.1	0.116	0.105	0.88
2.4	1052.2	0.008	0.107	0.99
2.6	1175.9	−0.104	0.106	1.10
2.8	1305.7	−0.219	0.103	1.22
3.0	1429.3	−0.325	0.097	1.33
3.2	1534.6	−0.411	0.091	1.41
3.4	1614.7	−0.479	0.085	1.48
3.6	1667.7	−0.535	0 079	1.54
3.8	1711.0	−0.582	0.074	1.58
4.0	1741.1	−0.617	0.069	1.62
4.2	1756.8	−0.641	0.066	1.64
4.4	1758.3	−0.656	0.065	1.66
4.6	1745.6	−0.662	0.064	1.66
4.8	1776.6	−0.680	0.062	1.68
5.0	1836.6	−0.715	0.056	1.72
5.2	1878.5	−0.733	0.054	1.73
5.4	1900.7	−0.735	0.054	1.74
5.6	1902.8	−0.723	0.056	1.72
5.8	1885.6	−0.699	0.061	1.70
6.0	1853.4	−0.666	0.066	1.67
6.2	1809.8	−0.625	0.073	1.63
6.4	1751.9	−0.573	0.081	1.57
6.6	1679.1	−0.510	0.090	1.51
6.8	1592.6	−0.435	0.100	1.44
7.0	1494.3	−0.350	0.109	1.35
7.2	1385.9	−0.257	0.117	1.26
7.4	1269.8	−0.156	0.123	1.16
7.6	1149.0	−0.051	0.127	1.05
7.8	1028.4	0.057	0.128	0.94
8.0	912.5	0.164	0.126	0.84
8.2	802.4	0.266	0.121	0.73
8.4	697.1	0.366	0.114	0.63
8.6	598.6	0.459	0.105	0.54
8.8	508.4	0.544	0.094	0.46
9.0	427.7	0.620	0.083	0.38
9.2	357.9	0.684	0.072	0.32
9.4	302.8	0.735	0.063	0.27
9.6	260.1	0.774	0.056	0.23
9.8	231.1	0.801	0.050	0.20
10.0	219.4	0.812	0.048	0.19

TRIAL ANALYSIS. EGLIN DATA, MAY 21, 1963. CONTINUOUS PROFILE,5,5.

ENERGY SPECTRUM OF WIND SPEED HEIGHT VARIATION HEIGHT RANGE 80.

DELTA H	DELTA V**2	CORRELATION	SIGNIF	I-G
0.2	15.1	0.993	0.001	0.01
0.4	47.1	0.979	0.004	0.02
0.6	83.2	0.964	0.007	0.04
0.8	122.0	0.947	0 011	0.05
1.0	161.5	0.930	0.014	0.07
1.2	200.6	0.914	0.017	0.09
1.4	240.7	0.898	0.020	0 10
1.6	286.9	0.879	0.024	0.12
1.8	337.0	0.859	0.028	0.14
2.0	382.2	0.841	0.031	0.16
2.2	421.0	0.827	0.034	0.17
2.4	455.3	0.815	0 036	0.19
2.6	483.4	0.805	0.038	0.20
2.8	502.9	0.799	0.039	0.20
3.0	512.5	0.797	0.040	0.20
3.2	517.8	0.796	0.040	0.20
3.4	528.9	0.793	0.041	0.21
3.6	542.8	0.789	0.042	0.21
3.8	561.7	0.783	0.043	0.22
4.0	584.0	0.775	0.045	0.22
4.2	606.6	0.768	0.046	0 23
4.4	628.3	0.761	0.048	0.24
4.6	650.9	0.755	0.049	0.25
4.8	679.5	0.746	0.051	0.25
5.0	712.5	0.734	0.053	0 27
5.2	748.8	0.722	0.056	0.28
5.4	785.1	0.710	0.058	0.29
5.6	818.7	0.700	0.060	0.30
5.8	847.5	0.692	0.062	0.31
6.0	870.6	0.686	0.063	0.31
6.2	894.8	0.680	0.065	0.32
6.4	919.4	0.674	0.066	0.33
6.6	943.4	0.668	0.068	0.33
6.8	968.1	0.661	0.069	0.34
7.0	996.2	0.653	0.071	0.35
7.2	1023.8	0.645	0.073	0.35
7.4	1048.3	0.638	0.075	0.36
7.6	1070.2	0.630	0.077	0.37
7.8	1088.9	0.623	0.078	0.38
8.0	1103.4	0.616	0.080	0.38
8.2	1114.0	0.608	0.082	0.39
8.4	1121.1	0.600	0.084	0.40
8.6	1120.2	0.593	0.086	0.41
8.8	1114.6	0.588	0.087	0.41
9.0	1103.8	0.585	0.089	0.42
9.2	1087.7	0.584	0.090	0.42
9.4	1066.1	0.586	0.090	0.41
9.6	1039.0	0.590	0.090	0.41
9.8	1007.8	0.596	0.090	0.40
10.0	973.4	0.603	0.090	0.40

N64-31451

TEMPERATURE, PRESSURE, DENSITY, AND WIND MEASUREMENTS WITH THE ROCKET GRENADE EXPERIMENT, 1960-1963*

W. SMITH, L. KATCHEN, P. SACHER, P. SWARTZ, AND J. THEON

Complete data from 28 rocket grenade experiments at Wallops Island, Virginia, and Fort Churchill, Canada, are presented. Pressures, temperatures, densities, and winds have been derived directly from the recorded times of explosions and sound arrivals; but no attempt has been made to analyze the meteorological significance of these measurements. Error analyses on 16 of the Wallops experiments are also included.

INTRODUCTION

During the period 1960–1963 a total of 28 rocket grenade experiments were carried out by Goddard Space Flight Center at Wallops Island, Virginia, and Fort Churchill, Canada. It is the purpose of this report to present the complete data from these experiments. The data have been reduced but not analyzed, i.e., the physical measurements (pressure, density, temperature and winds) have been derived directly from recordings of the times of grenade explosions and sound arrivals, and from the rocket trajectories; but no attempt has been made to analyze the measurements for their meteorological significance. Rather, this report should be considered as a record of the unsmoothed, raw measurements which may serve as the basis of further investigation and interpretation of the structure of the atmosphere. This is the first such report since the program started in 1960. As the program continues, future results will be documented in similar fashion.

The grenade experiment is one of the several techniques employed in NASA's Meteorological Rocket Sounding Program, the objective of which is to obtain as representative as possible a sample of the synoptic structure of the mesosphere and lower ionosphere. In order to extend the findings over the widest possible geographic area, and perhaps eventually to attain a global network of soundings, the NASA launchings are coordinated whenever possible with soundings in other parts of the world. The instrumental details and methods of data reduction have been thoroughly described elsewhere.†

EXPERIMENTAL METHOD

One and two pound explosive charges (grenades) are carried aloft in the nosecone of a Nike-Cajun sounding rocket. The grenades are ejected and exploded at 4–6 km intervals. Either a precise radar such as the FPS–16 or a doppler tracking system, or both, are used to determine the rocket position and hence the exact position of the explosion. The time of the explosion is detected by small rocket borne infrared photocells and telemetered to the ground. A ground based array of hot wire microphones with frequency response peaked at 4 cycles per second is used to detect and record the sound waves from each exploding grenade. The measured experimental parameters are the times of the grenade explosions, the positions of the grenade explosions, and the times of arrival of the sound waves at the ground based microphones.

Elevation and azimuth angles are computed for each arriving sound wave front by applying a least-squares fit to the arrival times at the various microphones. Each wave is then analytically traced up through the atmosphere by means of Snell's law. Data from radiosonde ascents or

*Published as *NASA Technical Report* R-211, October 1964.

†Nordberg, W., and Smith W., "The Rocket Grenade Experiment," NASA Technical Note D–2107, March 1964.

from small meteorological rockets obtained at the time of the grenade soundings are used for this tracing up to the first explosion; above this, the results of the experiment itself are used for each successive explosion. This *apparent* position of the *wave* is then compared with the *known* position of the *wave source*, the grenade explosion. The amount by which the sound wave has been displaced horizontally from one explosion to the next is a measure of the average wind velocity vector in the layer between any two adjacent explosions. The average speed of the sound, and hence average temperature between two adjacent explosions, may also be determined. This analysis yields an altitude profile of temperatures and winds as a direct measurement. The temperature profile consists of discrete points, each representing an average value for an altitude layer between grenade explosions. It may be used to derive density and pressure profiles, if the pressure or density is known at a given level at the bottom of the temperature profile; and the latter data are available from the accompanying radiosonde. Pressure is calculated by use of the hydrostatic equation, and the ideal gas law; an integration is performed over the temperature profile starting at the known pressure. The density at any level is then computed from the pressure, temperature, and universal gas constant.

During the period between July 8, 1960 and December 7, 1963, 23 successful rocket grenade soundings were conducted from Wallops Island, Virginia (37° 50′N, 75° 29′W). In addition to offering excellent launch and tracking support, the geographical location of Wallops Island makes it a good middle-latitude observation site. Five soundings were conducted from a sub-arctic range, Fort Churchill, Canada, (58° 47′N, 94° 17′W) in the period between December 1962 and March 1963. Originally, it was planned to distribute

Table 1

Dates, Times, and Locations of GSFC Grenade Experiments 1960-1963

Date	Local Time	Location
8 July 1960	2259	Wallops Island
14 February 1961	1850	Wallops Island
16 February 1961	2126	Wallops Island
5 April 1961	0757	Wallops Island
5 May 1961	1800	Wallops Island
5 May 1961	2354	Wallops Island
13 July 1961	1707	Wallops Island
14 July 1961	1102	Wallops Island
20 July 1961	0530	Wallops Island
16 September 1961	1855	Wallops Island
1 March 1962	1901	Wallops Island
2 March 1962	0615	Wallops Island
23 March 1962	1854	Wallops Island
27 March 1962	1904	Wallops Island
17 April 1962	0428	Wallops Island
6 June 1962	2005	Wallops Island
7 June 1962	2053	Wallops Island
1 December 1962	1625	Wallops Island
4 December 1962	0105	Fort Churchill
6 December 1962	0032	Wallops Island
5 December 1962	2343	Fort Churchill
20 February 1963	1734	Fort Churchill
20 February 1963	1847	Wallops Island
28 February 1963	1547	Fort Churchill
28 February 1963	1711	Wallops Island
8 March 1963	1901	Wallops Island
8 March 1963	1801	Fort Churchill
7 December 1963	0812	Wallops Island

the number of firings about equally between the two sites, but a fire early in 1961 destroyed part of the Fort Churchill launch facility, and it was not operational again until November 1962. Table 1 lists the soundings conducted. Four of the recent firings from each site were conducted nearly simultaneously.

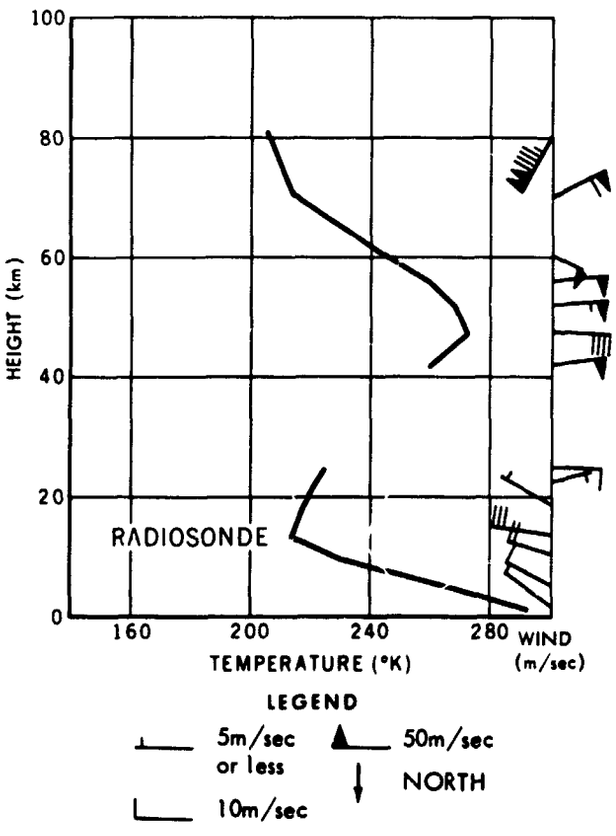
The results of the 28 soundings are presented in Figures 1 through 43. In Figures 1 through 28, the *directly measured* parameters of temperatures and winds are tabulated (actual computer printouts) in the columns on the left side of the page immediately above the graph. Tables of pressure and density *derived* from the temperature profiles are also shown. The temperatures were linearly interpolated between each of the measured points.

Winds and temperatures below 30 km obtained from radiosondes, and below 50 km from smaller meteorological sounding rockets (Arcas* and Haspt whenever available) launched nearly

*"Data Report of the Meteorological Rocket Network, Spring 1961 Firings," U.S. Army Signal Missile Agency, White Sands Missile Range, New Mexico, September 1961.

†Private Communication from M. J. Parker, Naval Ordnance Laboratory, White Oak, Maryland.

WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
82.256477	52.356449	259.43228	41659.187
93.878967	41.380489	272.74658	47575.437
85.262176	57.176998	268.03988	52210.906
84.192276	47.764659	259.17248	56140.277
116.43370	48.733200	244.54368	60321.136
63.638649	58.033428	213.90589	70569.476
208.64318	145.25718	205.61149	81196.117



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2543300	05	.2482000	04	.3835052	01-	.2254700	03
.2643300	05	.2137334	04	.3272117	01-	.2275630	03
.2743300	05	.1843137	04	.2796004	01-	.2296561	03
.2843300	05	.1591645	04	.2392689	01-	.2317491	03
.2943300	05	.1376346	04	.2050514	01-	.2338422	03
.3043300	05	.1191765	04	.1759771	01-	.2359352	03
.3143300	05	.1033998	04	.1512361	01-	.2380283	03
.3243300	05	.8970617	03	.1301517	01-	.2401213	03
.3343300	05	.7797774	03	.1121576	01-	.2422144	03
.3443300	05	.6786743	03	.9677938	02-	.2443075	03
.3543300	05	.5914055	03	.8361840	02-	.2464005	03
.3643300	05	.5159808	03	.7233967	02-	.2484936	03
.3743300	05	.4507097	03	.6266098	02-	.2505866	03
.3843300	05	.3941552	03	.5434443	02-	.2526797	03
.3943300	05	.3450925	03	.4718900	02-	.2547727	03
.4043300	05	.3024784	03	.4102479	02-	.2568658	03
.4143300	05	.2654210	03	.3570777	02-	.2589588	03
.4243300	05	.2331652	03	.3110230	02-	.2611737	03
.4343300	05	.2050637	03	.2712010	02-	.2634241	03
.4443300	05	.1805533	03	.2367629	02-	.2656746	03
.4543300	05	.1591495	03	.2069427	02-	.2679251	03
.4643300	05	.1404367	03	.1810893	02-	.2701755	03
.4743300	05	.1240576	03	.1586474	02-	.2724260	03
.4843300	05	.1096353	03	.1404877	02-	.2718758	03
.4943300	05	.9685901	02	.1245812	02-	.2708605	03
.5043300	05	.8553499	02	.1104301	02-	.2698451	03
.5143300	05	.7550246	02	.9784574	03-	.2688297	03
.5243300	05	.6661352	02	.8674291	03-	.2675386	03
.5343300	05	.5872438	02	.7712033	03-	.2652820	03
.5443300	05	.5171588	02	.6849906	03-	.2630252	03
.5543300	05	.4549578	02	.6078186	03-	.2607686	03
.5643300	05	.3997722	02	.5395126	03-	.2581482	03
.5743300	05	.3507529	02	.4798628	03-	.2546492	03
.5843300	05	.3072005	02	.4261343	03-	.2511502	03
.5943300	05	.2685671	02	.3778074	03-	.2476512	03
.6043300	05	.2343572	02	.3343293	03-	.2440092	03
.6143300	05	.2041450	02	.2948384	03-	.2403197	03
.6243300	05	.1775295	02	.2596163	03-	.2382302	03
.6343300	05	.1541136	02	.2282449	03-	.2352406	03
.6443300	05	.1335591	02	.2003429	03-	.2322511	03
.6543300	05	.1155330	02	.1755630	03-	.2292616	03
.6643300	05	.9975435	01	.1535887	03-	.2262720	03
.6743300	05	.8596653	01	.1341322	03-	.2232825	03
.6843300	05	.7393951	01	.1169322	03-	.2202930	03
.6943300	05	.6346729	01	.1017517	03-	.2173034	03
.7043300	05	.5436574	01	.8837577	04-	.2143139	03
.7143300	05	.4650316	01	.7597811	04-	.2132319	03
.7243300	05	.3975242	01	.6518719	04-	.2124513	03
.7343300	05	.3396371	01	.5590007	04-	.2116708	03
.7443300	05	.2900249	01	.4791119	04-	.2108903	03
.7543300	05	.2475274	01	.4104263	04-	.2101098	03
.7643300	05	.2111429	01	.3514024	04-	.2093292	03
.7743300	05	.1800085	01	.3007071	04-	.2085487	03
.7843300	05	.1533808	01	.2571876	04-	.2077682	03
.7943300	05	.1306198	01	.2198481	04-	.2069876	03
.8043300	05	.1111745	01	.1878277	04-	.2062071	03
.8143300	05	.9453360	00	.1611055	04-	.2044250	03

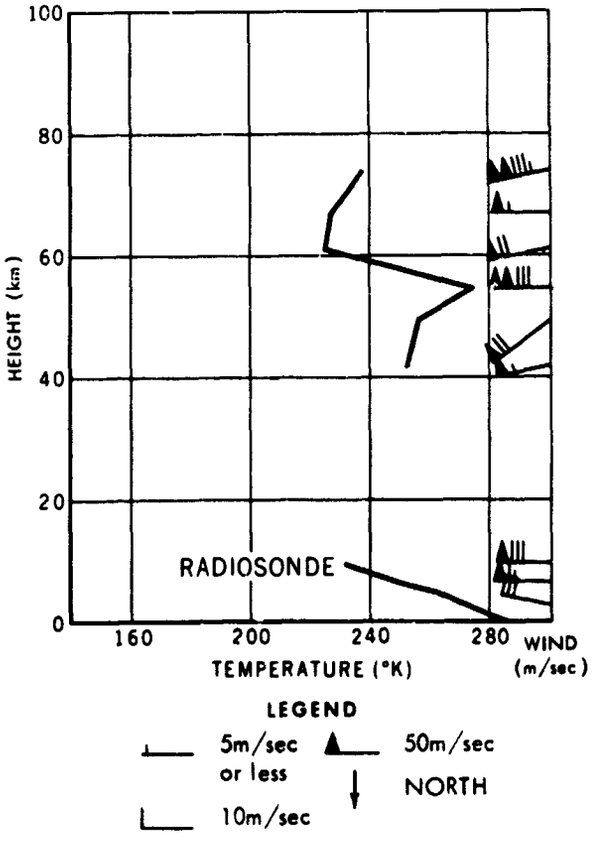
FIGURE 1

simultaneously with the grenade soundings are shown graphically. Winds and temperatures above 40 km from the grenade soundings are shown on the graph.

RESULTS

In the results from these soundings all data points, regardless of our degree of confidence, are presented in tabular form. Some individual data points may be subject to later revision, as there may be occasional errors in some of the records which have not yet been detected. In the preparation of these graphs some data points, where such errors seem obvious, were omitted and the graphs note the omissions. For the most part, these occur at the upper altitude limit of the experiment where the signal-to-noise ratio of detected acoustic waves is poor. Random errors contained in either the measurements or in the data reductions can usually be recognized by comparing two adjacent data points. Since these errors will generally be contained in the parameters (time and space coordinates of the explosion) associated with one individual grenade explosion, the temperature data points both below and above this explosion will contain errors of approximately equal magnitude but opposite sign. Therefore, if large excursions of this nature occur between two adjacent data points, it is very likely that the recordings from the explosion between these two data points were in error.

WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
258.25717	53.823459	252.57649	42001.500
233.48279	69.739135	256.26120	49508.047
271.57638	129.50918	274.10839	54864.257
259.95138	72.340553	225.56340	61136.257
272.36947	55.853248	227.78839	67134.976
260.74188	137.33760	237.91288	73796.414



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.3045000	05	.1172100	04	.1770008	01-	.2307000	03
.3145000	05	.1012915	04	.1517165	01-	.2325938	03
.3245000	05	.8764247	03	.1302125	01-	.2344876	03
.3345000	05	.7592442	03	.1118990	01-	.2363814	03
.3445000	05	.6585141	03	.9628182	02-	.2382753	03
.3545000	05	.5718173	03	.8294655	02-	.2401691	03
.3645000	05	.4971072	03	.7154511	02-	.2420629	03
.3745000	05	.4326487	03	.6178466	02-	.2439567	03
.3845000	05	.3769690	03	.5341862	02-	.2458506	03
.3945000	05	.3288167	03	.4623898	02-	.2477444	03
.4045000	05	.2871259	03	.4007002	02-	.2496382	03
.4145000	05	.2509887	03	.3476315	02-	.2515320	03
.4245000	05	.2195938	03	.3026265	02-	.2527966	03
.4345000	05	.1902229	03	.2643927	02-	.2532875	03
.4445000	05	.1683143	03	.2310599	02-	.2537784	03
.4545000	05	.1474232	03	.2019901	02-	.2542692	03
.4645000	05	.1291637	03	.1766311	02-	.2547601	03
.4745000	05	.1131990	03	.1545018	02-	.2552509	03
.4845000	05	.9903686	02	.1351853	02-	.2557418	03
.4945000	05	.8702238	02	.1183190	02-	.2562327	03
.5045000	05	.7638526	02	.1025884	02-	.2593998	03
.5145000	05	.6715994	02	.8905447	03-	.2627319	03
.5245000	05	.5914699	02	.7744703	03-	.2660639	03
.5345000	05	.5217455	02	.6747233	03-	.2693960	03
.5445000	05	.4609684	02	.5888430	03-	.2727280	03
.5545000	05	.4073033	02	.5263770	03-	.2695747	03
.5645000	05	.3589853	02	.4776475	03-	.2618348	03
.5745000	05	.3152153	02	.4321850	03-	.2540949	03
.5845000	05	.2756824	02	.3898576	03-	.2463549	03
.5945000	05	.2400886	02	.3505355	03-	.2386149	03
.6045000	05	.2081485	02	.3140903	03-	.2308750	03
.6145000	05	.1797386	02	.2774642	03-	.2256798	03
.6245000	05	.1549716	02	.2388385	03-	.2260507	03
.6345000	05	.1336562	02	.2056503	03-	.2264216	03
.6445000	05	.1153058	02	.1771253	03-	.2267925	03
.6545000	05	.9950316	01	.1526008	03-	.2271634	03
.6645000	05	.8589099	01	.1315100	03-	.2275343	03
.6745000	05	.7417070	01	.1132002	03-	.2283672	03
.6845000	05	.6409877	01	.9718122	04-	.2297870	03
.6945000	05	.5545036	01	.8351683	04-	.2313069	03
.7045000	05	.4801653	01	.7184823	04-	.2328268	03
.7145000	05	.4162011	01	.6187322	04-	.2343466	03
.7245000	05	.3611072	01	.5333694	04-	.2358665	03
.7345000	05	.3136058	01	.4602422	04-	.2373864	03

FIGURE 2

ERROR ANALYSIS

An error analysis was run with the aid of a digital computer on 16 of the 28 experiments reported. The firings considered were the first 16 conducted from Wallops Island. The following nomenclature is used in the error analysis:

Δw = Difference in west coordinate between successive grenade explosions.

Δn = Difference in north coordinate between successive grenade explosions.

h = Difference in altitude between successive grenade explosions.

ω = Wind direction in the layer.

W = Wind speed in the layer.

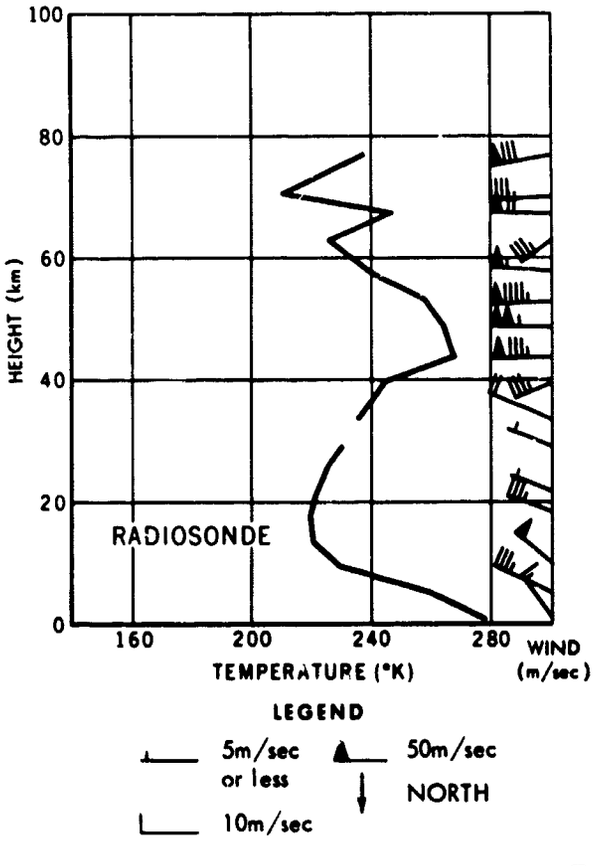
c = Speed of sound in the layer $= k(T)^{1/2}$, where k is a proportionality constant depending on molecular mass, the ratio of specific heats, and the universal gas constant.

ϕ_μ = Azimuth angle of sound wave for upper explosion.

τ = Travel time of sound wave in layer.

$K = c_0 \sec\theta_u + W_0 \cos(\phi_u - \omega_0) - W \cos(\phi_u - \omega)$, where c_0 is the mean speed of sound over

WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
292.86309	18.130565	234.89309	33683.750
248.69609	36.597328	243.78259	39602.218
269.17739	83.686485	267.01477	43709.046
272.88568	106.25949	264.04278	48449.386
267.75299	96.220848	258.29199	52904.457
275.55267	53.734848	238.25239	57711.847
235.30198	37.225559	224.72909	63107.707
271.42547	71.667457	246.47889	67362.398
265.22039	40.814918	210.58119	70596.187
259.91388	81.029823	237.40589	77047.812



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.3045000	05	.1172100	04	.1770008	01-	.2307000	03
.3145000	05	.1012724	04	.1520784	01-	.2319966	03
.3245000	05	.8757722	03	.1307817	01-	.2332933	03
.3345000	05	.7579851	03	.1125665	01-	.2345900	03
.3445000	05	.6566228	03	.9691282	02-	.2360440	03
.3545000	05	.5693507	03	.8350074	02-	.2375460	03
.3645000	05	.4941439	03	.7201560	02-	.2390480	03
.3745000	05	.4292708	03	.6217047	02-	.2405500	03
.3845000	05	.3732577	03	.5372276	02-	.2420519	03
.3945000	05	.3248481	03	.4646686	02-	.2435539	03
.4045000	05	.2832492	03	.3969752	02-	.2485784	03
.4145000	05	.2477074	03	.3394386	02-	.2542354	03
.4245000	05	.2172744	03	.2912549	02-	.2598923	03
.4345000	05	.1911267	03	.2507462	02-	.2655493	03
.4445000	05	.1684033	03	.2201049	02-	.2665502	03
.4545000	05	.1484003	03	.1944181	02-	.2659232	03
.4645000	05	.1307393	03	.1716853	02-	.2652963	03
.4745000	05	.1151502	03	.1515720	02-	.2646693	03
.4845000	05	.1013933	03	.1337810	02-	.2640420	03
.4945000	05	.8924222	02	.1183270	02-	.2627511	03
.5045000	05	.7850110	02	.1045991	02-	.2614603	03
.5145000	05	.6901172	02	.9241122	03-	.2601694	03
.5245000	05	.6063301	02	.8159643	03-	.2588786	03
.5345000	05	.5321813	02	.7241816	03-	.2560179	03
.5445000	05	.4662762	02	.6450012	03-	.2518494	03
.5545000	05	.4076493	02	.5733930	03-	.2476809	03
.5645000	05	.3555969	02	.5087391	03-	.2435124	03
.5745000	05	.3094737	02	.4504634	03-	.2393439	03
.5845000	05	.2687866	02	.3961082	03-	.2364024	03
.5945000	05	.2330782	02	.3471657	03-	.2338962	03
.6045000	05	.2018126	02	.3038519	03-	.2313899	03
.6145000	05	.1744751	02	.2655686	03-	.2288837	03
.6245000	05	.1506060	02	.2317754	03-	.2263774	03
.6345000	05	.1299073	02	.1998315	03-	.2264769	03
.6445000	05	.1122462	02	.1688329	03-	.2315908	03
.6545000	05	.9730046	01	.1432088	03-	.2367048	03
.6645000	05	.8460633	01	.1218929	03-	.2418187	03
.6745000	05	.7375770	01	.1046653	03-	.2455054	03
.6845000	05	.6416538	01	.9536543	04-	.2544056	03
.6945000	05	.5544697	01	.8650435	04-	.2633048	03
.7045000	05	.4755990	01	.7808105	04-	.2722040	03
.7145000	05	.4066359	01	.6615826	04-	.2141312	03
.7245000	05	.3484554	01	.5661265	04-	.2182890	03
.7345000	05	.2994846	01	.4690363	04-	.2224468	03
.7445000	05	.2581307	01	.3968524	04-	.2266046	03
.7545000	05	.2230991	01	.3368146	04-	.2307625	03
.7645000	05	1933332	01	.2867109	04-	.2349203	03
.7745000	05	1679607	01	.2448927	04-	.2389405	03

FIGURE 3

the microphone array; θ_u is the elevation angle of sound wave for upper explosion; W_0 is the wind speed over the microphone array; and ω_0 is the wind direction over the microphone array.

t = Travel time of sound wave from grenade explosion to ground.

T = Mean temperature in layer.

Δt = Time difference between microphones introduced by changing the time of arrival at one microphone.

ϕ'_u = Elevation angle from ground to apparent position of upper grenade explosion as obtained by ray tracing.

In order to determine the error function, deliberate errors were introduced in the following experimental parameters:

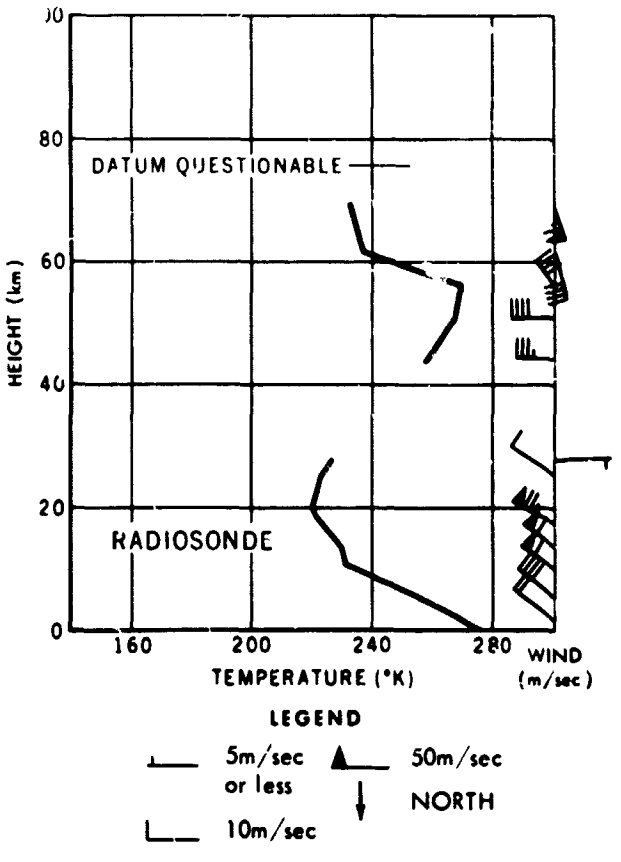
A. Position of grenade explosions

1. North: 200 meters
2. West: 200 meters
3. Up: 50 meters

B. Travel time of sound from explosions to ground: 0.3 second.

C. Time of arrival at one of the 6 microphones relative to the other five microphones: 0.02 second.

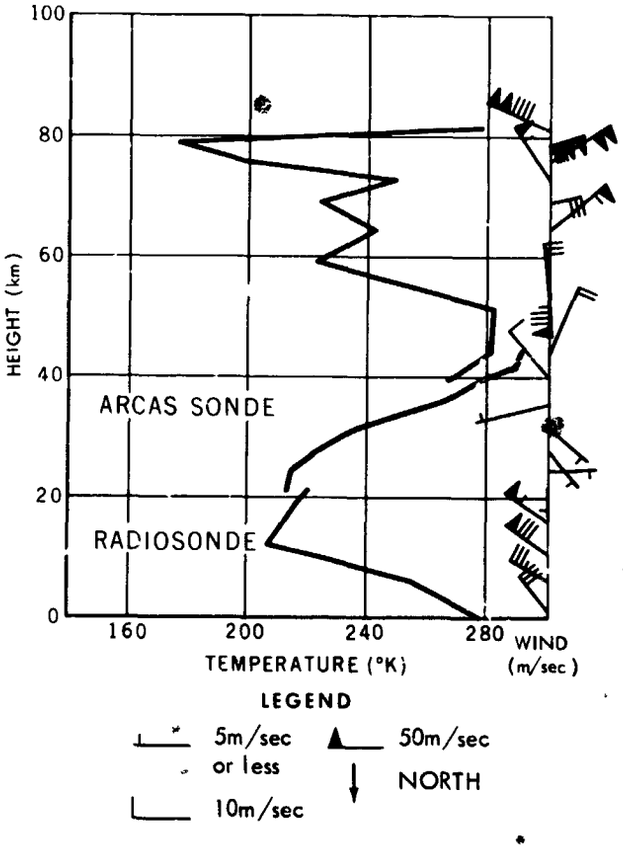
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
278.04568	32.738597	258.56097	44334.699
267.40359	40.801197	267.20568	51244.250
335.20639	41.921638	269.77447	56138.988
177.28808	39.535289	238.58029	61859.648
160.35488	66.775573	233.18949	69577.523
94.937965	25.101959	49.960929	75558.476



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2877900	05	.1516800	04	.2323785	01-	.2274000	03
.2977900	05	.1307683	04	.1986374	01-	.2294032	03
.3077900	05	.1109419	04	.1700349	01-	.2314063	03
.3177900	05	.9765115	03	.1457528	01-	.2334095	03
.3277900	05	.8453935	03	.1251036	01-	.2354107	03
.3377900	05	.7328085	03	.1075323	01-	.2374159	03
.3477900	05	.6360084	03	.925 .697	02-	.2394191	03
.3577900	05	.5526712	03	.7975310	02-	.2414223	03
.3677900	05	.4808323	03	.6881542	02-	.2434255	03
.3777900	05	.4188070	03	.5945215	02-	.2454287	03
.3877900	05	.3652430	03	.5142620	02-	.2474318	03
.3977900	05	.3188796	03	.4453766	02-	.2494350	03
.4077900	05	.2787158	03	.3861788	02-	.2514382	03
.4177900	05	.2438811	03	.3352423	02-	.2534414	03
.4277900	05	.2136335	03	.2913607	02-	.2554446	03
.4377900	05	.1873388	03	.2535110	02-	.2574478	03
.4477900	05	.1644407	03	.2210014	02-	.2594168	03
.4577900	05	.1444530	03	.1932845	02-	.2603670	03
.4677900	05	.1266788	03	.1697 08	02-	.2616101	03
.4777900	05	.1110215	03	.1490257	02-	.2636702	03
.4877900	05	.9730851	02	.1307719	02-	.2641213	03
.4977900	05	.8658475	02	.1136704	02-	.2655724	03
.5077900	05	.7630761	02	.1 770741	02-	.2666236	03
.5177900	05	.6726865	02	.5763659	02-	.2674863	03
.5277900	05	.5935381	02	.7715321	02-	.2680111	03
.5377900	05	.5237114	02	.6704348	02-	.2685359	03
.5477900	05	.4622301	02	.5 885007	02-	.2690607	03
.5577900	05	.4080211	02	.5273611	02-	.2695855	03
.5677900	05	.3600563	02	.4710670	02-	.2660845	03
.5777900	05	.3170350	02	.4234530	02-	.2608316	03
.5877900	05	.2784159	02	.3798110	02-	.2553787	03
.5977900	05	.2438266	02	.3398800	02-	.2499258	03
.6077900	05	.2109195	02	.3034191	02-	.2444729	03
.6177900	05	.1853705	02	.2702970	02-	.2390200	03
.6277900	05	.1610878	02	.2358614	02-	.2379381	03
.6377900	05	.1399185	02	.2054689	02-	.2372992	03
.6477900	05	.1214860	02	.1789078	02-	.2365410	03
.6577900	05	.1054423	02	.1557581	02-	.2358427	03
.6677900	05	.9148293	01	.1355389	02-	.2351442	03
.6777900	05	.7934166	01	.1179069	02-	.2344457	03
.6877900	05	.6878553	01	.1025200	02-	.2337472	03
.6977900	05	.5963888	01	.8930785	02-	.2277478	03
.7077900	05	.5080049	01	.7910776	02-	.2163818	03
.7177900	05	.4222664	01	.6870656	02-	.2057465	03
.7277900	05	.3379725	01	.5874618	02-	.1951111	03
.7377900	05	.2554009	01	.4916874	02-	.1844758	03
.7477900	05	.1750306	01	.3958055	02-	.1738404	03
.7577900	05	.1060159	01	.2963440	02-	.1631538	03

FIGURE 4

WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
322.41619	11.569830	266.27038	39412.546
23.538419	22.253669	280.97079	44079.898
355.26779	28.470070	282.27557	51248.187
178.72689	85.764000	222.41799	59150.648
48.711238	56.003578	242.82579	64569.699
75.535553	30.775518	223.67649	69209.593
327.44299	56.990917	249.57080	72776.789
57.703037	101.52199	198.17269	75793.546
70.997947	147.21499	175.96778	78695.750
292.62289	142.18638	278.37927	81482.250



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2216240	05	.3930000	04	.6212110	01-	.2204000	03
.2316240	05	.3372820	04	.5267826	01-	.2230591	03
.2416240	05	.2900022	04	.4476028	01-	.2257182	03
.2516240	05	.2498031	04	.3810684	01-	.2283774	03
.2616240	05	.2155583	04	.3250441	01-	.2310365	03
.2716240	05	.1863305	04	.2777740	01-	.2336956	03
.2816240	05	.1613388	04	.2378114	01-	.2363548	03
.2916240	05	.1399306	04	.2039613	01-	.2390139	03
.3016240	05	.1215597	04	.1752345	01-	.2416730	03
.3116240	05	.1057680	04	.1508107	01-	.2443321	03
.3216240	05	.9217052	03	.1300076	01-	.2469913	03
.3316240	05	.8044297	03	.1122571	01-	.2496504	03
.3416240	05	.7031186	03	.9[illegible]8521	02-	.2523095	03
.3516240	05	.6154603	03	.8409523	02-	.2549687	03
.3616240	05	.5394973	03	.7295494	02-	.2576278	03
.3716240	05	.4735695	03	.6338545	02-	.2602869	03
.3816240	05	.4162660	03	.5515215	02-	.2629461	03
.3916240	05	.3663862	03	.4805745	02-	.2656052	03
.4016240	05	.3229344	03	.4188076	02-	.2686321	03
.4116240	05	.2850577	03	.3654019	02-	.2717817	03
.4216240	05	.2519952	03	.3193201	02-	.2749313	03
.4316240	05	.2230891	03	.2794893	02-	.2780810	03
.4416240	05	.1977669	03	.2452039	02-	.2809858	03
.4516240	05	.1754424	03	.2173837	02-	.2811678	03
.4616240	05	.1556557	03	.1927421	02-	.2813498	03
.4716240	05	.1381161	03	.1709129	02-	.2815318	03
.4816240	05	.1226672	03	.1515738	02-	.2817138	03
.4916240	05	.1087815	03	.1344388	02-	.2818959	03
.5016240	05	.9655710	02	.1192541	02-	.2820779	03
.5116240	05	.8571648	02	.1057970	02-	.2822599	03
.5216240	05	.7598558	02	.9613975	03-	.2753508	03
.5316240	05	.6714575	02	.8735839	03-	.2677763	03
.5416240	05	.5912643	02	.7916435	03-	.2602017	03
.5516240	05	.5187170	02	.7153335	03-	.2526272	03
.5616240	05	.4532802	02	.6444149	03-	.2450526	03
.5716240	05	.3944418	02	.5786521	03-	.2374781	03
.5816240	05	.3417126	02	.5178136	03-	.2299035	03
.5916240	05	.2946389	02	.4614153	03-	.2224622	03
.6016240	05	.2537562	02	.3907762	03-	.2262282	03
.6116240	05	.2190958	02	.3318757	03-	.2299941	03
.6216240	05	.1896299	02	.2826147	03-	.2337600	03
.6316240	05	.1645134	02	.2412951	03-	.2375260	03
.6416240	05	.1430491	02	.2065384	03-	.2412919	03
.6516240	05	.1244940	02	.1804301	03-	.2403796	03
.6616240	05	.1081911	02	.1595413	03-	.2362525	03
.6716240	05	.9379494	01	.1407715	03-	.2321254	03
.6816240	05	.8111001	01	.1239370	03-	.2279983	03
.6916240	05	.6995789	01	.1088671	03-	.2238712	03
.7016240	05	.6039248	01	.9104204	04-	.2305929	03
.7116240	05	.5236649	01	.7670169	04-	.2378519	03
.7216240	05	.4560417	01	.6481365	04-	.2451109	03
.7316240	05	.3977512	01	.5702452	04-	.2430009	03
.7416240	05	.3449736	01	.5318705	04-	.2259634	03
.7516240	05	.2958911	01	.4933984	04-	.2089259	03
.7616240	05	.2508780	01	.4474105	04-	.1953506	03
.7716240	05	.2107967	01	.3912540	04-	.1876935	03
.7816240	05	.1758501	01	.3402604	04-	.1800485	03
.7916240	05	.1470850	01	.2653400	04-	.1931184	03
.8016240	05	.1256051	01	.1903634	04-	.2298711	03
.8116240	05	.1098134	01	.1434877	04-	.2666238	03

FIGURE 5

It was determined that an error in one of the foregoing parameters for any given grenade will cause wind and temperature errors in the two layers adjacent to the grenade. The temperature errors, in general, will be of opposite sign and approximately equal in magnitude. There are slight errors introduced in the remaining layers above the grenade; however, these are negligibly small. The errors in the temperatures and winds due to errors in the coordinates of the explosions are given by:

$$\frac{\partial W}{\partial \Delta w} = \frac{\sin \omega}{\tau}; \tag{1}$$

$$\frac{\partial W}{\partial \Delta n} = \frac{\cos \omega}{\tau}; \tag{2}$$

$$\frac{\partial W}{\partial h} = \frac{\cot \theta'_u}{\tau} \cos (\omega - \phi_u); \tag{3}$$

$$\frac{\partial c}{\partial \Delta w} = \frac{\sin \phi_u}{c\tau} \left[\frac{h^2 K}{\tau^2 (K^2 - 2c^2)} - \frac{c^2}{K} \right]; \tag{4}$$

$$\frac{\partial c}{\partial \Delta n} = \frac{\cos \phi_u}{c\tau} \left[\frac{h^2 K}{\tau^2 (K^2 - 2c^2)} - \frac{c^2}{K} \right]; \tag{5}$$

$$\frac{\partial c}{\partial h} = \frac{1}{c\tau} \left[\frac{h^2 K \cot \theta'_u - hK^2 \tau}{\tau^2 (K^2 - 2c^2)} + \frac{c^2 \cot \theta'_u}{K} \right]. \tag{6}$$

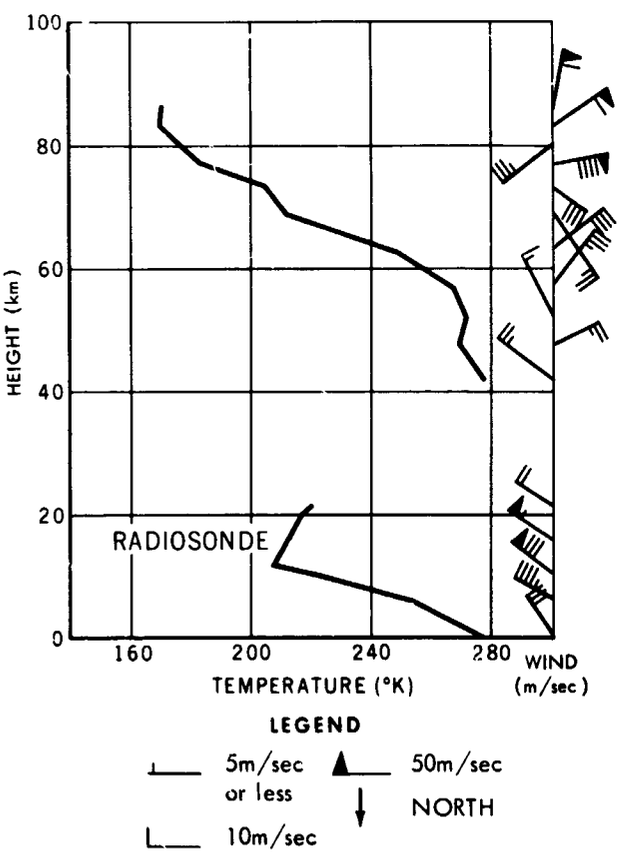
In the error analysis, the north and west coordinates of alternate grenades were changed by −200m, and the up coordinates were changed by −50m. This means that $d\Delta w$ and $d\Delta n$ for adjacent layers were alternatingly +200 and −200, and dh for adjacent layers were alternatingly +50 and −50. The azimuth angle changes slowly with altitude. Generally, $\sin\phi_u$ and $\cos\phi_u$ do not change sign during an experiment; thus the signs of $(\partial c/\partial \Delta w)d\Delta w$ and $(\partial c/\partial \Delta n)d\Delta n$ alternate regularly from layer to layer *as the coordinates of alternate grenade explosions are changed.* This accounts for the regular alternation of the temperature error, since all other signs in the equation for $\partial c/\partial \Delta w$, $\partial c/\partial \Delta n$ remain constant during an experiment. Similarly, it can be seen from the expression for $\partial c/\partial h$ that all the quantities retain the same sign throughout an experiment. Thus, the same regular alternation for the signs of the $(\partial c/\partial h)dh$ will occur for successive grenades.

The situation is different, however, for $\partial W/\partial \Delta w$, $\partial W/\partial \Delta n$, $\partial W/\partial h$. The sign of $\partial W/\partial \Delta w$ depends on the sign of $\sin \omega$ which in turn depends on the value of ω. Since ω may vary from 0 to 360°, $\sin \omega$ may be either positive or negative. The same is true for the sign of $\partial W/\partial \Delta n$, which depends on $\cos \omega$. Hence, we expect no regular alternation from layer to layer in the sign of the wind error. The sign of $\partial W/\partial h$ depends on the sign of $\cos (\omega - \phi_u)$ which in turn depends on the value of $\omega - \phi_u$. Although ϕ_u changes slowly and in general remains in the same quadrant, ω can vary from 0 to 360°, and hence $\omega - \phi_u$ can be positive or negative. Thus, $(\partial W/\partial h)dh$ exhibits no *regular* alternation in sign. In summary, we can see that the signs of $\partial c/\partial \Delta w$, $\partial c/\partial \Delta n$, $\partial c/\partial h$ are constant and hence an error in the coordinates of one individual grenade exposion will cause a positive error in the measured temperatures in the layer below the explosion if a negative error is caused in the layer above the explosion, and vice versa. In general, the magnitudes of the temperature errors caused by an error of the explosion coordinates between two adjacent layers are approximately equal.

However, this does not hold for wind errors. The error functions for the wind speed are proportional to $\sin \omega$, $\cos \omega$ and $\cos(\omega - \phi_u)$, which may vary considerably from layer to layer. The error functions of c, however, are proportional to ϕ_u and θ'_u which do not vary greatly from layer to layer.

The two other parameters of importance to the errors in addition to the grenade explosion coordinates were investigated in a similar manner. These parameters were: the time differences, Δt, between arrival of the sound wave at any two microphones within the array of six; and the travel time, t, of the sound from the explosion to the ground. The parameter Δt is important because the direction of the arriving sound wave front is derived from those differences, while t enters primarily in the calculation of the speed of sound in the layer. An error of .02 second was introduced into the arrival time at one microphone without altering the arrival times at the remaining five microphones, while an error of

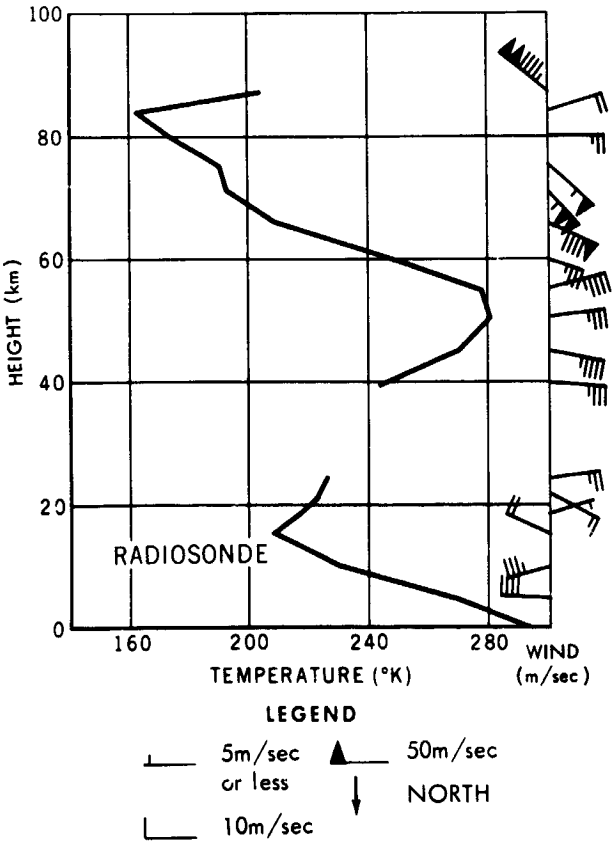
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
305.12338	23.382379	277.63378	42586.648
62.641258	23.147178	268.78369	47412.750
333.34677	14.322359	271.82907	52066.699
36.589668	35.042530	267.24627	57085.047
49.854858	39.259758	248.50579	62930.449
139.23439	25.119869	212.77249	68475.398
124.98700	42.308529	205.17138	73241.234
80.707336	89.750587	183.60189	76913.093
234.36670	35.243290	177.44699	79971.343
55.719726	59.271869	170.75879	82893.593
8.8894186	57.672039	171.12059	86514.093



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2216240	05	.3930000	04	.6212110	01-	.2204000	03
.2316240	05	.3372986	04	.5264708	01-	.2232022	03
.2416240	05	.2900583	04	.4471224	01-	.2260045	03
.2516240	05	.2499100	04	.3805162	01-	.2288067	03
.2616240	05	.2157196	04	.3244834	01-	.2316090	03
.2716240	05	.1865451	04	.2772450	01-	.2344112	03
.2816240	05	.1616022	04	.2373374	01-	.2372135	03
.2916240	05	.1402369	04	.2035545	01-	.2400157	03
.3016240	05	.1219020	04	.1748994	01-	.2428179	03
.3116240	05	.1061394	04	.1505465	01-	.2456202	03
.3216240	05	.9256422	03	.1298107	01-	.2484224	03
.3316240	05	.8085272	03	.1121219	01-	.2512247	03
.3416240	05	.7073200	03	.9700505	02-	.2540269	03
.3516240	05	.6197154	03	.8406323	02-	.2568292	03
.3616240	05	.5437630	03	.7296432	02-	.2596314	03
.3716240	05	.4778086	03	.6342969	02-	.2624337	03
.3816240	05	.4204478	03	.5522528	02-	.2652359	03
.3916240	05	.3704854	03	.4815403	02-	.2680382	03
.4016240	05	.3269027	03	.4204973	02-	.2708404	03
.4116240	05	.2888299	03	.3677194	02-	.2736426	03
.4216240	05	.2555232	03	.3220178	02-	.2764449	03
.4316240	05	.2262136	03	.2824439	02-	.2765779	03
.4416240	05	.2001977	03	.2538569	02-	.2747441	03
.4516240	05	.1770357	03	.2259951	02-	.2729103	03
.4616240	05	.1564297	03	.2010414	02-	.2710765	03
.4716240	05	.1381115	03	.1787080	02-	.2692427	03
.4816240	05	.1218960	03	.1577076	02-	.2692742	03
.4916240	05	.1076025	03	.1388775	02-	.2699286	03
.5016240	05	.9501750	02	.1223380	02-	.2705829	03
.5116240	05	.8393294	02	.1078056	02-	.2712373	03
.5216240	05	.7416382	02	.9508106	03-	.2717416	03
.5316240	05	.6552806	02	.8429293	03-	.2708284	03
.5416240	05	.5787592	02	.7470138	03-	.2699152	03
.5516240	05	.5109785	02	.6617671	03-	.2690020	03
.5616240	05	.4509622	02	.5860296	03-	.2680888	03
.5716240	05	.3978245	02	.5190883	03-	.2669982	03
.5816240	05	.3506054	02	.4630360	03-	.2637922	03
.5916240	05	.3085262	02	.4124761	03-	.2605862	03
.6016240	05	.2710788	02	.3669261	03-	.2573802	03
.6116240	05	.2378002	02	.3259410	03-	.2541742	03
.6216240	05	.2082691	02	.2891109	03-	.2509682	03
.6316240	05	.1820652	02	.2567844	03-	.2470110	03
.6416240	05	.1587082	02	.2298381	03-	.2405667	03
.6516240	05	.1378389	02	.2051100	03-	.2341224	03
.6616240	05	.1192489	02	.1824700	03-	.2276781	03
.6716240	05	.1027427	02	.1617922	03-	.2212338	03
.6816240	05	.8813632	01	.1429552	03-	.2147895	03
.6916240	05	.7535142	01	.1240156	03-	.2116768	03
.7016240	05	.6431175	01	.1066498	03-	.2100819	03
.7116240	05	.5482596	01	.9161479	04-	.2084869	03
.7216240	05	.4668439	01	.7861150	04-	.2068920	03
.7316240	05	.3970442	01	.6737739	04-	.2052971	03
.7416240	05	.3367331	01	.5872664	04-	.1997601	03
.7516240	05	.2842369	01	.5107313	04-	.1938859	03
.7616240	05	.2386899	01	.4422905	04-	.1880116	03
.7716240	05	.1994385	01	.3794710	04-	.1831001	03
.7816240	05	.1660844	01	.3195203	04-	.1810876	03
.7916240	05	.1380337	01	.2685397	04-	.1790750	03
.8016240	05	.1144845	01	.2253242	04-	.1770097	03
.8116240	05	.9473875	00	.1889038	04-	.1747210	03
.8216240	05	.7820785	00	.1580120	04-	.1724323	03
.8316240	05	.6442357	00	.1314170	04-	.1707856	03
.8416240	05	.5302585	00	.1081037	04-	.1708856	03
.8516240	05	.4365199	00	.8894120	05-	.1709855	03
.8616240	05	.3594167	00	.7318861	05-	.1710854	03
.8716240	05	.2959742	00	.6025731	05-	.1711206	03

FIGURE 6

WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
95.363586	36.568847	244.35820	39675.539
99.608467	45.432167	270.66247	45171.437
83.930687	33.290619	281.09097	50468.437
76.533165	46.691719	278.32437	54984.148
109.12829	34.365276	249.50698	60400.296
114.66099	88.817054	210.09239	66114.382
135.61579	106.67810	194.53720	71025.437
131.85289	54.567546	191.89569	75674.046
90.188316	24.938308	174.47338	80105.984
74.234520	20.360889	163.27510	83904.882
307.88009	143.44909	205.45279	87459.437



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2578400	05	.2413200	04	.3700353	01-	.2272000	03
.2678400	05	.2079899	04	.3172032	01-	.2284351	03
.2778400	05	.1794152	04	.2721527	01-	.2296703	03
.2878400	05	.1548961	04	.2337031	01-	.2309054	03
.2978400	05	.1338387	04	.2008579	01-	.2321406	03
.3078400	05	.1157391	04	.1727756	01-	.2333757	03
.3178400	05	.1001684	04	.1487445	01-	.2346109	03
.3278400	05	.8676225	03	.1281623	01-	.2358461	03
.3378400	05	.7521005	03	.1105190	01-	.2370812	03
.3478400	05	.6524728	03	.9538205	02-	.2383164	03
.3578400	05	.5664835	03	.8238468	02-	.2395515	03
.3678400	05	.4922059	03	.7121517	02-	.2407867	03
.3778400	05	.4279940	03	.6160859	02-	.2420218	03
.3878400	05	.3724401	03	.5333954	02-	.2432570	03
.3978400	05	.3243745	03	.4614837	02-	.2448773	03
.4078400	05	.2830292	03	.3949431	02-	.2496635	03
.4178400	05	.2476045	03	.3390118	02-	.2544496	03
.4278400	05	.2171629	03	.2918427	02-	.2592358	03
.4378400	05	.1909293	03	.2519363	02-	.2640219	03
.4478400	05	.1682600	03	.2180705	02-	.2688081	03
.4578400	05	.1485607	03	.1903721	02-	.2718684	03
.4678400	05	.1313234	03	.1670737	02-	.2738372	03
.4778400	05	.1161932	03	.1467694	02-	.2758059	03
.4878400	05	.1028997	03	.1290564	02-	.2777747	03
.4978400	05	.9120872	02	.1135886	02-	.2797434	03
.5078400	05	.8090337	02	.1003406	02-	.2808976	03
.5178400	05	.7177329	02	.8921157	03-	.2802850	03
.5278400	05	.6365913	02	.7929931	03-	.2796723	03
.5378400	05	.5644965	02	.7047294	03-	.2790596	03
.5478400	05	.5004524	02	.6261500	03-	.2784470	03
.5578400	05	.4432065	02	.5633871	03-	.2740686	03
.5678400	05	.3916759	02	.5077382	03-	.2687480	03
.5778400	05	.3452941	02	.4566533	03-	.2634273	03
.5878400	05	.3036351	02	.4008368	03-	.2581067	03
.5978400	05	.2662989	02	.3670071	03-	.2527861	03
.6078400	05	.2328744	02	.3286463	03-	.2468603	03
.6178400	05	.2029351	02	.2946266	03-	.2399624	03
.6278400	05	.1761442	02	.2632995	03-	.2330647	03
.6378400	05	.1522481	02	.2345206	03-	.2261668	03
.6478400	05	.1310069	02	.2081494	03-	.2192690	03
.6578400	05	.1121945	02	.1840493	03-	.2123712	03
.6678400	05	.9568949	01	.1602946	03-	.2079714	03
.6778400	05	.8137912	01	.1384308	03-	.2048040	03
.6878400	05	.6903786	01	.1192823	03-	.2016367	03
.6978400	05	.5841890	01	.1025459	03-	.1984693	03
.7078400	05	.4930323	01	.8794822	04-	.1953019	03
.7178400	05	.4153328	01	.7454443	04-	.1941061	03
.7278400	05	.3496239	01	.6293515	04-	.1935379	03
.7378400	05	.2941776	01	.5311029	04-	.1929697	03
.7478400	05	.2474116	01	.4479917	04-	.1924014	03
.7578400	05	.2079498	01	.3783824	04-	.1914634	03
.7678400	05	.1744012	01	.3239898	04-	.1875324	03
.7778400	05	.1457289	01	.2765210	04-	.1836013	03
.7878400	05	.1213048	01	.2352124	04-	.1796702	03
.7978400	05	.1005711	01	.1993715	04-	.1757391	03
.8078400	05	.8306407	00	.1677822	04-	.1724747	03
.8178400	05	.6837085	00	.1405046	04-	.1695269	03
.8278400	05	.5608830	00	.1173032	04-	.1665792	03
.8378400	05	.4585274	00	.9762405	05-	.1636314	03
.8478400	05	.3764528	00	.7550095	05-	.1737066	03
.8578400	05	.3128284	00	.5872879	05-	.1855724	03
.8678400	05	.2629726	00	.4640209	05-	.1974382	03
.8778400	05	.2230478	00	.3768877	05-	.2061788	03

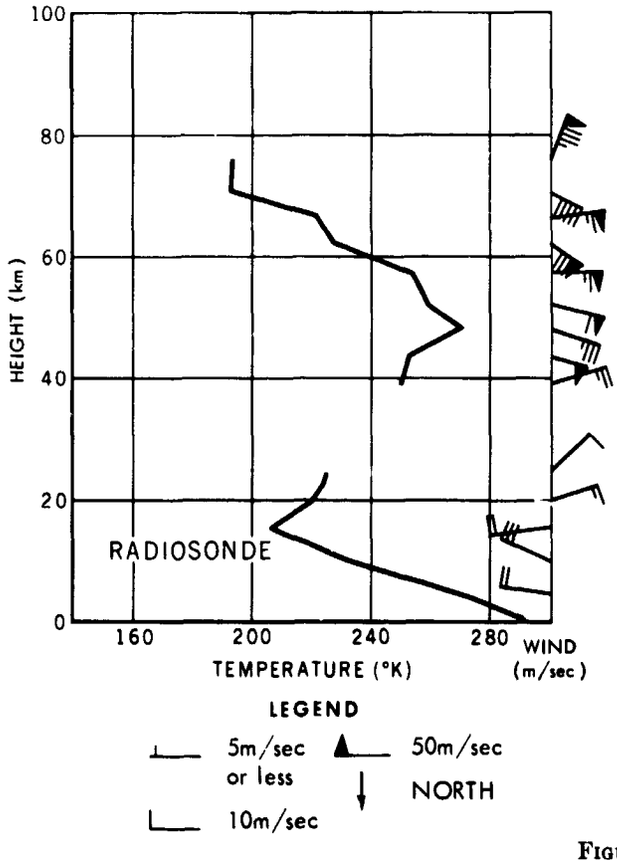
FIGURE 7

.3 second was introduced into t. As in the case of the coordinates of the grenade explosions, these errors were introduced only for alternate grenades.

A typical result of the error analysis is presented in Figures 29 through 38. From a study of these figures, the following conclusions can be reached:

1. Relative errors in the North and West coordinates of 200 meters will significantly affect only winds by about 10 m/sec but not the temperatures.
2. Relative errors in the up coordinates of 50 meters will cause significant errors of about 5°K in temperature, but not in winds.
3. An error of .3 second in t will give rise to large temperature errors—about 10°K—but in general only very small wind errors.
4. An error of .02 in the relative arrival times between microphones will cause the largest errors in both winds and temperatures: 10–40 m/sec and 10–25°K. The error analysis clearly shows that an error introduced into the input parameters of a grenade significantly affects only the layers adjacent to the particular grenade. No significant error is introduced into the calculated winds and temperature for suc-

WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
72.464332	25.952299	250.40889	38917.296
104.10459	49.421737	253.16690	43811.046
107.37399	34.035408	270.17877	48008.199
103.76009	62.433387	259.47160	52001.148
87.769149	66.201034	254.37869	57239.000
121.38459	81.264473	228.30349	62182.796
81.096214	62.609947	222.24749	66433.984
113.49089	38.723297	192.81658	70453.382
20.21242	85.848434	194.16389	75561.789



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2520500	05	.2620000	04	.4044156	01-	.2257000	03
.2620500	05	.2256273	04	.3455133	01-	.2275019	03
.2720500	05	.1945424	04	.2955705	01-	.2293039	03
.2820500	05	.1679425	04	.2531675	01-	.2311058	03
.2920500	05	.1451519	04	.2171186	01-	.2329078	03
.3020500	05	.1256009	04	.1864317	01-	.2347097	03
.3120500	05	.1088085	04	.1602760	01-	.2365117	03
.3220500	05	.9436808	03	.1379540	01-	.2383136	03
.3320500	05	.8193558	03	.1188804	01-	.2401156	03
.3420500	05	.7121926	03	.1025624	01-	.2419175	03
.3520500	05	.6197164	03	.8858514	02-	.2437195	03
.3620500	05	.5398240	03	.7659861	02-	.2455214	03
.3720500	05	.4707262	03	.6630730	02-	.2473234	03
.3820500	05	.4108987	03	.5746124	02-	.2491253	03
.3920500	05	.3590073	03	.4991495	02-	.2505710	03
.4020500	05	.3138523	03	.4353885	02-	.2511346	03
.4120500	05	.2744709	03	.3799046	02-	.2516982	03
.4220500	05	.2401129	03	.3316060	02-	.2522618	03
.4320500	05	.2101275	03	.2895479	02-	.2528253	03
.4420500	05	.1840148	03	.2516364	02-	.2547636	03
.4520500	05	.1614030	03	.2172588	02-	.2588168	03
.4620500	05	.1418642	03	.1880140	02-	.2628700	03
.4720500	05	.1249421	03	.1630725	02-	.2669232	03
.4820500	05	.1102198	03	.1424018	02-	.2696510	03
.4920500	05	.9723699	02	.1268902	02-	.2669695	03
.5020500	05	.8567838	02	.1129411	02-	.2642880	03
.5120500	05	.7539936	02	.1004101	02-	.2616065	03
.5220500	05	.6627465	02	.8905282	03-	.2592734	03
.5320500	05	.5820860	02	.7850895	03-	.2583010	03
.5420500	05	.5110126	02	.6918334	03-	.2573287	03
.5520500	05	.4484146	02	.6093878	03-	.2563564	03
.5620500	05	.3933052	02	.5365301	03-	.2553841	03
.5720500	05	.3448105	02	.4721734	03-	.2544117	03
.5820500	05	.3018251	02	.4218128	03-	.2492837	03
.5920500	05	.2634687	02	.3761671	03-	.2440093	03
.6020500	05	.2293145	02	.3346367	03-	.2387350	03
.6120500	05	.1989785	02	.2969275	03-	.2334607	03
.6220500	05	.1721090	02	.2626694	03-	.2282719	03
.6320500	05	.1485621	02	.2281564	03-	.2268473	03
.6420500	05	.1281238	02	.1980114	03-	.2254228	03
.6520500	05	.1103987	02	.1717029	03-	.2239982	03
.6620500	05	.9503990	01	.1487614	03-	.2225737	03
.6720500	05	.8161484	01	.1312699	03-	.2166019	03
.6820500	05	.6975679	01	.1161228	03-	.2092797	03
.6920500	05	.5929215	01	.1022811	03-	.2019574	03
.7020500	05	.5009836	01	.8967269	04-	.1946352	03
.7120500	05	.4216897	01	.7611329	04-	.1930148	03
.7220500	05	.3547518	01	.6394390	04-	.1932786	03
.7320500	05	.2985268	01	.5373604	04-	.1935423	03
.7420500	05	.2512848	01	.4517073	04-	.1938060	03
.7520500	05	.2115796	01	.3798167	04-	.1940698	03
.7620500	05	.1779962	01	.3233403	04-	.1917825	03

FIGURE 8

ceeding grenades, and hence only the layers adjacent to the grenade in question need be considered.

Table 2 gives the average error functions for the 16 firings tabulated by layers. Figures 39 and 40 are graphs of the two functions which are most strongly altitude dependent, $\partial T/\partial \Delta t$ and $\partial W/\partial \Delta t$, showing quite clearly the increase of the error function with altitude.

$\partial W/\partial \Delta h$, etc., by the value of the actual errors in Δw, Δn, h, Δt, and t. The maximum error is calculated by summing all the partial errors:

$$\left(\frac{\partial W}{\partial \Delta w}\, d\Delta w + \frac{\partial W}{\partial \Delta n}\, d\Delta n + \frac{\partial W}{\partial \Delta h}\, d\Delta h + \ldots \text{etc.}\right)$$

The actual errors in coordinates and times are estimated in the following manner:

The position of the grenade explosion is determined by two totally independent tracking meth-

Table 2

Absolute Average Error Functions

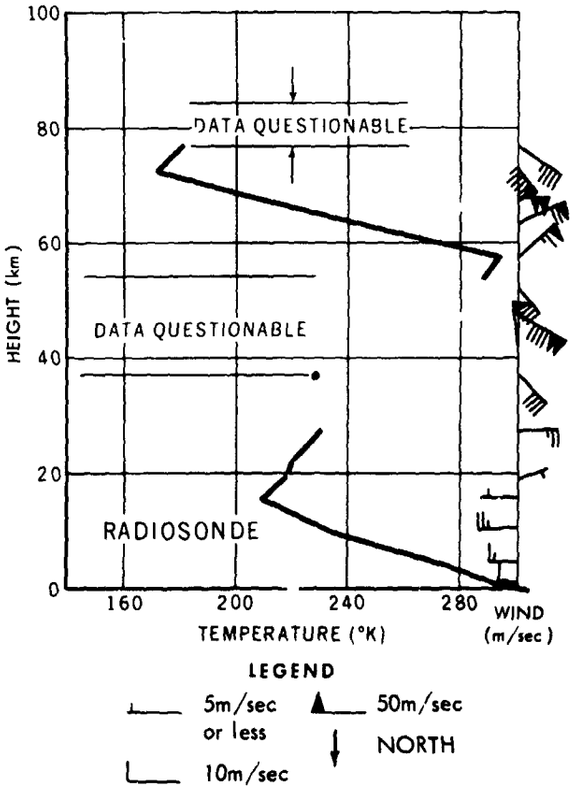
Layer	t + .3 sec		North-200m		West-200m		Up-50m		Δ t + .02 sec	
	Temp. (°K)	Wind (m/sec)	Temp. (°K)	Wind (m/sec)	Temp. (°K)	Wind (m/sec)	Temp. (°K)	Wind (m/sec)	Temp. (°K)	Wind (m/sec)
1	10.1	.5	.9	3.9	.7	9.2	5.4	.8	11.7	12.2
2	10.2	.7	.9	5.3	.7	10.6	5.6	.6	15.2	18.5
3	13.0	1.0	.9	5.2	.8	11.9	6.7	.8	16.3	22.0
4	10.5	1.1	.6	4.6	.6	10.7	5.8	.6	17.8	22.3
5	8.4	.9	.4	2.5	.4	9.5	4.6	.5	12.0	16.9
6	7.7	1.0	.4	3.7	.5	8.7	4.5	.7	15.1	26.4
7	10.0	1.3	.4	6.0	.5	10.6	5.6	.8	16.3	25.3
8	9.2	1.2	.4	6.3	.7	10.3	5.6	.7	20.9	31.9
9	11.7	2.3	.5	5.8	.8	12.1	6.4	.8	21.4	30.5
10	8.3	2.4	.5	8.2	.8	9.3	5.4	.7	22.5	41.1
11	11.5	5.2	.4	9.8	.7	11.4	6.4	.5	22.4	35.2

Figure 41 shows the dependence of $\partial T/\partial \Delta t$ upon zenith angle. As the zenith angle increases, the error function increases nearly linearly. Thus, to make the experiments as accurate as possible, it is desirable that the sound propagation be as nearly vertical as possible. A zenith angle of 0° would, of course, be ideal; however, range safety considerations make this impossible, and at Wallops Island the zenith angles usually lie in the range from 15°—30°. The zenith angles at Fort Churchill are generally less than 10°, and the probable error in the Fort Churchill data is therefore correspondingly less.

The magnitudes of the errors assumed in this analysis were chosen arbitrarily large for convenience in calculation. The actual resulting errors will be much smaller, as will be shown below. The magnitude of the actual errors in the rocket grenade experiment is determined by multiplying the error functions $\partial W/\partial \Delta w$, $\partial W/\partial \Delta n$,

ods, DOVAP and radar. The agreement of these two systems with regard to the coordinates Δn, Δw and h in general is better than 20 meters. The result is a maximum temperature error of 1°K and maximum wind error of 1–2 meters/sec, according to the error functions. The time of the grenade explosion, determined primarily by rocket borne infrared photocells, is telemetered to the ground equipment. The time of grenade explosions may also be determined by ground based flash detectors or radar signal strength records. It has been shown that this time can generally be determined within $\pm$.001 second. The errors in determining the arrival time of the sound at the ground range from $\pm$.001 to $\pm$.010 second. Thus, the maximum error expected in determining the travel time (t) of the sound between explosion and ground is approximately .01 second, which results in a temperature error of 0.3°K and virtually no errors in winds. By far the largest contribution

WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
136.33490	42.079189	228.33669	37367.250
356.12039	53.882389	294.37728	42735.046
118.10519	139.32998	202.16909	47919.597
138.40469	31.059989	285.11227	52338.449
48.023109	66.104408	293.30438	57617.648
68.479278	68.716514	242.52729	63174.636
87.361854	102.28640	205.54239	67965.382
133.97369	95.964477	172.91769	72516.437
123.23388	43.562007	181.56900	76840.343
141.17219	171.22239	125.93209	80544.648
352.19537	129.04379	163.16378	81 035.734



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2846400	05	.1586000	04	.2408618	01-	.2294000	03
.2946400	05	.1368461	04	.2079310	01-	.2292828	03
.3046400	05	.1180723	04	.1794969	01-	.2291656	03
.3146400	05	.1018716	04	.1549473	01-	.2290484	03
.3246400	05	.8789046	03	.1337503	01-	.2289312	03
.3346400	05	.7582574	03	.1154495	01-	.2288141	03
.3446400	05	.6541552	03	.9965033	02-	.2286969	03
.3546400	05	.5643290	03	.8601077	02-	.2285797	03
.3646400	05	.4868221	03	.7423581	02-	.2284625	03
.3746400	05	.4201145	03	.6376097	02-	.2295466	03
.3846400	05	.3640720	03	.5244529	02-	.2418460	03
.3946400	05	.3177682	03	.4355985	02-	.2541454	03
.4046400	05	.2791539	03	.3650016	02-	.2664448	03
.4146400	05	.2466797	03	.3083087	02-	.2787441	03
.4246400	05	.2191586	03	.2623365	02-	.2910436	03
.4346400	05	.1948238	03	.2411885	02-	.2814127	03
.4446400	05	.1721699	03	.2275227	02-	.2636275	03
.4546400	05	.1508478	03	.2137670	02-	.2458423	03
.4646400	05	.1308659	03	.1999130	02-	.2280571	03
.4746400	05	.1122333	03	.1859510	02-	.2102720	03
.4846400	05	.9572001	02	.1570115	02-	.2123877	03
.4946400	05	.8224701	02	.1239565	02-	.2311580	03
.5046400	05	.7051595	02	.9968864	03-	.2499283	03
.5146400	05	.6282062	02	.8145073	03-	.2686986	03
.5246400	05	.5564427	02	.6794634	03-	.2853071	03
.5346400	05	.4948141	02	.6009412	03-	.2868588	03
.5446400	05	.4403056	02	.5318647	03-	.2884106	03
.5546400	05	.3920617	02	.4710542	03-	.2899624	03
.5646400	05	.3493325	02	.4174818	03-	.2915141	03
.5746400	05	.3114622	02	.3702527	03-	.2930659	03
.5846400	05	.2773782	02	.3383894	03-	.2855708	03
.5946400	05	.2461865	02	.3102645	03-	.2764333	03
.6046400	05	.2176362	02	.2836594	03-	.2672958	03
.6146400	05	.1915813	02	.2585385	03-	.2581582	03
.6246400	05	.1678792	02	.2348657	03-	.2490207	03
.6346400	05	.1464081	02	.2122666	03-	.2402934	03
.6446400	05	.1270827	02	.1903640	03-	.2325733	03
.6546400	05	.1097873	02	.1701027	03-	.2248532	03
.6646400	05	.9436656	01	.1514085	03-	.2171331	03
.6746400	05	.8067246	01	.1342084	03-	.2094130	03
.6846400	05	.6857159	01	.1182824	03-	.2019679	03
.6946400	05	.5794104	01	.1036232	03-	.1947993	03
.7046400	05	.4865287	01	.9033638	04-	.1876308	03
.7146400	05	.4057879	01	.7833776	04-	.1804622	03
.7246400	05	.3359853	01	.6754543	04-	.1732936	03
.7346400	05	.2773628	01	.5827531	04-	.1748136	03
.7446400	05	.2294220	01	.4520387	04-	.1768144	03
.7546400	05	.1901843	01	.3705343	04-	.1788152	03
.7646400	05	.1579958	01	.3044156	04-	.1808160	03
.7746400	05	.1308029	01	.2646293	04-	.1722019	03
.7846400	05	.1068286	01	.2367784	04-	.1571824	03
.7946400	05	.8549683	00	.2095183	04-	.1421629	03
.8046400	05	.6674734	00	.1828935	04-	.1271434	03
.8146400	05	.5180825	00	.1329718	04-	.1357368	03
.8246400	05	.4091609	00	.9736582	05-	.1464016	03
.8346400	05	.3285713	00	.7287936	05-	.1570664	03
.8446400	05	.2668864	00	.5708503	05-	.1628780	03

FIGURE 9

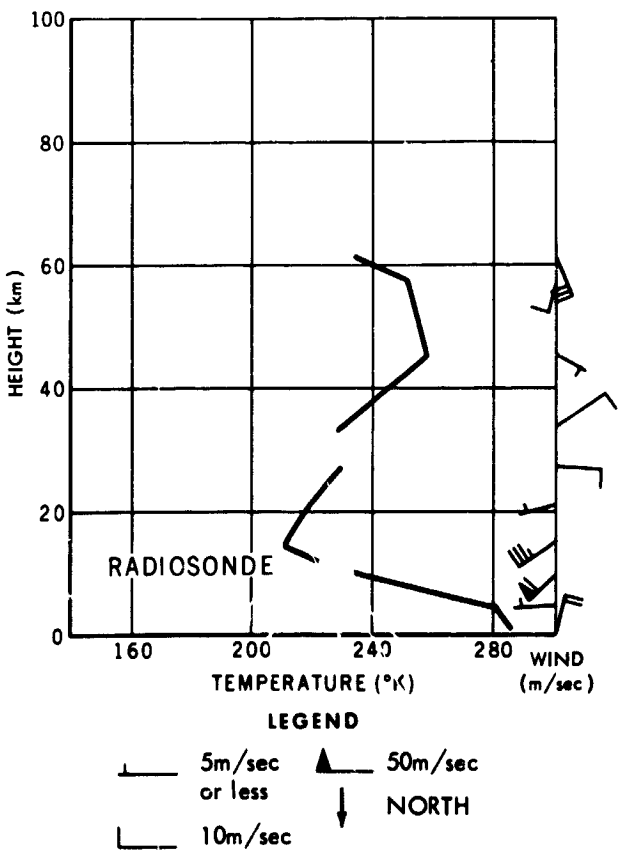
to the total error comes from errors in Δt, which also range from .001 to .010 second, depending upon the background noise level and the altitude of the explosion. The error in Δt is obtained by determining typical discrepancies among a number of independent Δt readings for each explosion. These errors result in temperature errors ranging from 1°K at 40 km to 15°K at 90 km, and wind errors from 1 to 16 m/sec over the same altitude range.

The maximum errors in the 16 soundings considered in this error analysis are presented in Figures 42 and 43, which clearly show the increase of the maximum errors with altitude. This increase is to be expected, since the amplitude of the sound waves will decrease rapidly with increasing altitudes, making it difficult to distinguish the grenade explosion from the background noise level. Since the determination of the relative arrival times between microphones is by far the most critical measured parameter, the largest part of the maximum error is due to the attenuation of the sound wave with altitude.

ACKNOWLEDGMENTS

The authors greatly appreciate the contributions of: Superior Engineering Company in designing and constructing the payload instrumentation for the grenade experiment; Texas Western College in conducting the sound ranging portion of this experiment; and New Mexico State University in operating the DOVAP tracking system.

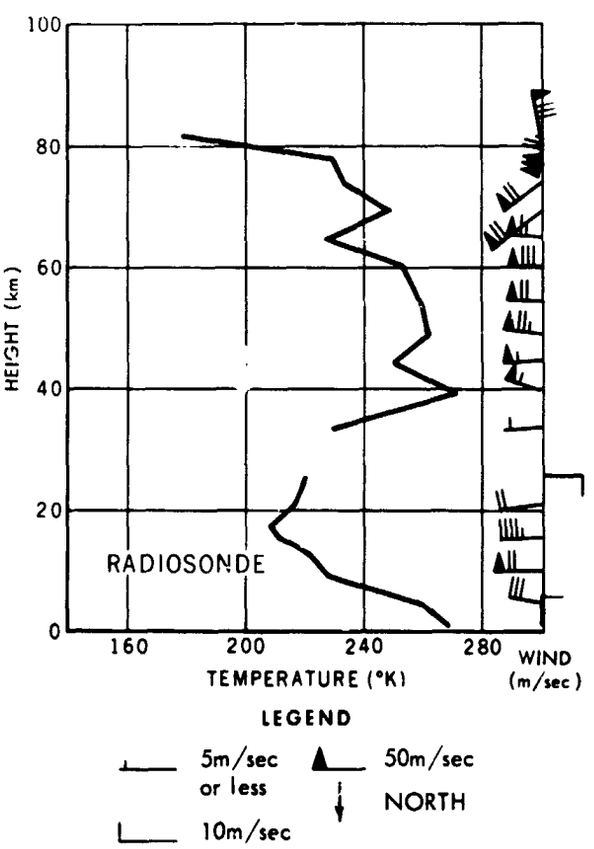
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
54.791648	7.2243499	228.665	33187.847
119.66009	4.9561348	257.62	45214.898
200.23429	12.105758	251.69139	57547.699
154.18908	29.009319	234.19510	61416.547



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2842900	05	.1585800	04	.2421628	01-	.2281200	03
.2942900	05	.1367266	04	.2087035	01-	.2282346	03
.3042900	05	.1178998	04	.1798754	01-	.2283492	03
.3142900	05	.1016779	04	.1550485	01-	.2284637	03
.3242900	05	.8769845	03	.1336642	01-	.2285783	03
.3342900	05	.7566287	03	.1149845	01-	.2292459	03
.3442900	05	.6534624	03	.9827423	02-	.2316536	03
.3542900	05	.5652446	03	.8413269	02-	.2340614	03
.3642900	05	.4896843	03	.7214396	02-	.2364691	03
.3742900	05	.4248608	03	.6196276	02-	.2388769	03
.3842900	05	.3691600	03	.5330196	02-	.2412846	03
.3942900	05	.3212236	03	.4592231	02-	.2436923	03
.4042900	05	.2799064	03	.3962408	02-	.2461001	03
.4142900	05	.2442411	03	.3424025	02-	.2485078	03
.4242900	05	.2134095	03	.2963086	02-	.2509156	03
.4342900	05	.1867181	03	.2567850	02-	.2533233	03
.4442900	05	.1635784	03	.2228439	02-	.2557310	03
.4542900	05	.1434675	03	.1940888	02-	.2575203	03
.4642900	05	.1258761	03	.1706689	02-	.2570393	03
.4742900	05	.1104189	03	.1499393	02-	.2565583	03
.4842900	05	.9684000	02	.1317473	02-	.2560773	03
.4942900	05	.8491356	02	.1157392	02-	.2555963	03
.5042900	05	.7444053	02	.1016555	02-	.2551154	03
.5142900	05	.6524572	02	.8926745	03-	.2546344	03
.5242900	05	.5717470	02	.7837294	03-	.2541534	03
.5342900	05	.5009164	02	.6879393	03-	.2536724	03
.5442900	05	.4387682	02	.6037321	03-	.2531914	03
.5542900	05	.3842498	02	.5297227	03-	.2527104	03
.5642900	05	.3364340	02	.4646889	03-	.2522294	03
.5742900	05	.2945059	02	.4075542	03-	.2517485	03
.5842900	05	.2575032	02	.3621634	03-	.2477058	03
.5942900	05	.2246321	02	.3218074	03-	.2431835	03
.6042900	05	.1954639	02	.2853272	03-	.2386611	03
.6142900	05	.1696355	02	.2524733	03-	.2340774	03

FIGURE 10

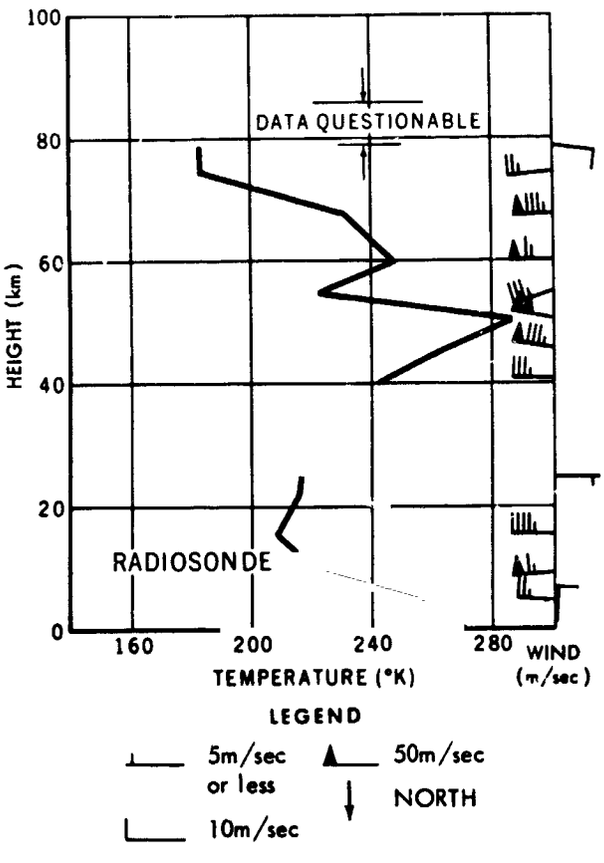
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
262.31097	5.3755097	229.40469	33886.347
299.30877	52.652637	269.77697	39334.250
252.84739	55.973369	249.59349	44584.398
281.27478	73.580688	262.48669	49082.500
273.17938	71.772773	259.51049	54465.949
269.71099	82.195953	253.38369	60122.699
277.13287	65.050956	227.26159	64997.046
242.14248	68.743583	248.01939	69636.343
234.13198	70.978942	234.93168	74051.289
356.45077	80.171730	230.69918	77846.445
201.99719	175.48089	179.71659	81413.250



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2630580	05	.2000000	04	.3164251	01-	.2202000	03
.2730580	05	.1715696	04	2699560	01-	.2214142	03
.2830580	05	.1473111	04	.2305224	01-	.2226285	03
.2930580	05	.1265935	04	.1970274	01-	.2238427	03
.3030580	05	.1088840	04	.1685504	01-	.2250570	03
.3130580	05	.9373223	03	.1443171	01-	.2262712	03
.3230580	05	.8075743	03	.1236764	01-	.2274855	03
.3330579	05	.6963709	03	.1060799	01-	.2286997	03
.3430579	05	.6014760	03	.9012165	02-	.2325130	03
.3530579	05	.5213426	03	.7570215	02-	.2399237	03
.3630579	05	.4538746	03	.6393074	02-	.2473343	03
.3730579	05	.3967746	03	.5426209	02-	.2547449	03
.3830579	05	.3482122	03	.4627466	02-	.2621555	03
.3930579	05	.3067199	03	.3964011	02-	.2695661	03
.4030579	05	.2704328	03	.3541338	02-	.2660420	03
.4130579	05	.2380297	03	.3162719	02-	.2621976	03
.4230579	05	.2091230	03	.2819980	02-	.2583532	03
.4330579	05	.1833780	03	.2510166	02-	.2545089	03
.4430579	05	.1604879	03	.2230527	02-	.2506645	03
.4530579	05	.1403553	03	.1942990	02-	.2516613	03
.4630579	05	.1228787	03	.1681899	02-	.2545276	03
.4730579	05	.1077430	03	.1458307	02-	.2573940	03
.4830579	05	.9461312	02	.1266489	02-	.2602604	03
.4930579	05	.8318938	02	.1104646	02-	.2623632	03
.5030579	05	.7317564	02	.9737283	03-	.2618104	03
.5130579	05	.6435235	02	.8581313	03-	.2612575	03
.5230579	05	.5657984	02	.7560857	03-	.2607047	03
.5330579	05	.4973458	02	.6660237	03-	.2601518	03
.5430579	05	.4370722	02	.5865544	03-	.2595990	03
.5530579	05	.3839701	02	.5172799	03-	.2586008	03
.5630579	05	.3371573	02	.4561247	03-	.2575177	03
.5730579	05	.2959015	02	.4020024	03-	.2564346	03
.5830579	05	.2595612	02	.3541273	03-	.2553515	03
.5930579	05	.2275664	02	.3117983	03-	.2542685	03
.6030579	05	.1993708	02	.2751859	03-	.2524024	03
.6130579	05	.1743413	02	.2458586	03-	.2470433	03
.6230579	05	.1520127	02	.2191239	03-	.2416843	03
.6330579	05	.1321429	02	.1948014	03-	.2363252	03
.6430579	05	.1145068	02	.1727195	03-	.2309661	03
.6530579	05	.9899216	01	.1508346	03-	.2286430	03
.6630579	05	.8564128	01	.1279872	03-	.2331174	03
.6730579	05	.7429866	01	.1089451	03-	.2375917	03
.6830579	05	.6463218	01	.9301930	04-	.2420661	03
.6930579	05	.5636944	01	.7965514	04-	.2465404	03
.7030579	05	.4921940	01	.6969441	04-	.2460348	03
.7130580	05	.4293678	01	.6153972	04-	.2430704	03
.7230580	05	.3739498	01	.5425858	04-	.2401060	03
.7330580	05	.3251400	01	.4776622	04-	.2371416	03
.7430579	05	.2822566	01	.4190692	04-	.2346478	03
.7530579	05	.2447732	01	.3651528	04-	.2335326	03
.7630579	05	.2121322	01	.3179773	04-	.2324174	03
.7730580	05	.1837257	01	.2767250	04-	.2313021	03
.7830580	05	.1587128	01	.2466969	04-	.2241334	03
.7930580	05	.1361172	01	.2259870	04-	.2098397	03
.8030580	05	.1154859	01	.2057491	04-	.1955461	03
.8130579	05	.9677157	00	.1860039	04-	.1812525	03

FIGURE 11

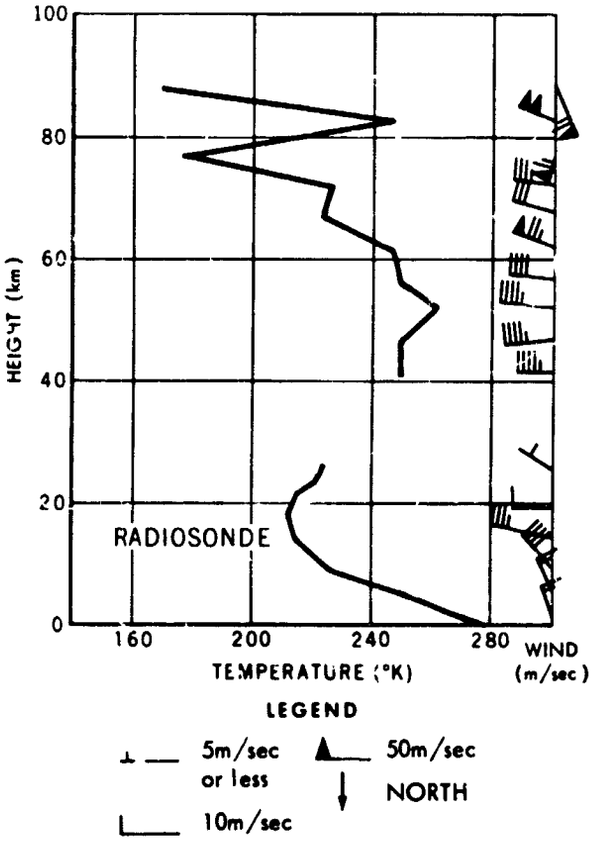
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
274.23077	34.899269	242.51269	40513.449
286.38287	85.033699	263.57437	45697.347
281.50439	101.26220	285.79770	50676.898
243.78489	33.179687	222.30879	54934.796
268.60199	62.581386	247.68548	60029.648
260.13058	85.637725	231.54458	67638.132
259.93988	24.920858	183.24919	74377.484
106.51109	10.605539	183.39129	78539.039
296.23410	234.90919	391.75579	82104.632
150.42019	103.22889	152.16539	85453.297



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2478430	05	.2600000	04	.4192929	01-	.2160300	03
.2578430	05	.2224041	04	.3558897	01-	.2177136	03
.2678430	05	.1904829	04	.3024704	01-	.2193973	03
.2778430	05	.1633445	04	.2574016	01-	.2210810	03
.2878430	05	.1402426	04	.2193269	01-	.2227647	03
.2978430	05	.1205521	04	.1871185	01-	.2244483	03
.3078430	05	.1037483	04	.1598369	01-	.2261320	03
.3178430	05	.8939038	03	.1366990	01-	.2278157	03
.3278429	05	.7710758	03	.1170506	01-	.2294993	03
.3378429	05	.6658746	03	.1003448	01-	.2311830	03
.3478429	05	.5756651	03	.8612332	02-	.2328667	03
.3578429	05	.4982214	03	.7400217	02-	.2345503	03
.3678429	05	.4316613	03	.6365886	02-	.2362340	03
.3778429	05	.3743909	03	.5482224	02-	.2379177	03
.3878429	05	.3250594	03	.4726412	02-	.2396014	03
.3978429	05	.2825198	03	.4079215	02-	.2412850	03
.4078429	05	.2458427	03	.3515725	02-	.2436131	03
.4178429	05	.2143231	03	.3014694	02-	.2476760	03
.4278429	05	.1872701	03	.2591650	02-	.2517389	03
.4378429	05	.1639926	03	.2233464	02-	.2558018	03
.4478429	05	.1437150	03	.1929377	02-	.2598647	03
.4578429	05	.1265579	03	.1670341	02-	.2639624	03
.4678429	05	.1115287	03	.1447509	02-	.2684253	03
.4778429	05	.9849314	02	.1257417	02-	.2728882	03
.4878429	05	.8716010	02	.1094828	02-	.2773511	03
.4978429	05	.7723464	02	.9554075	03-	.2818141	03
.5078429	05	.6863049	02	.8413115	03-	.2841963	03
.5178429	05	.6078233	02	.7863622	03-	.2692854	03
.5278429	05	.5346258	02	.7322078	03-	.2543746	03
.5378429	05	.4666310	02	.6788785	03-	.2394637	03
.5478429	05	.4037547	02	.6264078	03-	.2245529	03
.5578429	05	.3479338	02	.5350767	03-	.2265400	03
.5678429	05	.3005305	02	.4522272	03-	.2315209	03
.5778429	05	.2604042	02	.3835940	03-	.2365017	03
.5878429	05	.2263203	02	.3265095	03-	.2414826	03
.5978429	05	.1972702	02	.2788477	03-	.2464634	03
.6078429	05	.1721774	02	.2437531	03-	.2460845	03
.6178429	05	.1501787	02	.2144581	03-	.2439631	03
.6278429	05	.1308400	02	.1884810	03-	.2418417	03
.6378429	05	.1139582	02	.1654604	03-	.2397202	03
.6478429	05	.9896233	01	.1451055	03-	.2375988	03
.6578429	05	.8591097	01	.1271035	03-	.2354774	03
.6678429	05	.7448874	01	.1112064	03-	.2333559	03
.6778429	05	.6448914	01	.9747186	04-	.2304971	03
.6878430	05	.5565639	01	.8682091	04-	.2233309	03
.6978430	05	.4780546	01	.7704613	04-	.2161648	03
.7078429	05	.4085400	01	.6810036	04-	.2089986	03
.7178429	05	.3472421	01	.5993764	04-	.2018324	03
.7278429	05	.2934271	01	.5251311	04-	.1946662	03
.7378429	05	.2464043	01	.4578308	04-	.1875001	03
.7478430	05	.2058228	01	.3912701	04-	.1832631	03
.7578430	05	.1715962	01	.3261254	04-	.1832972	03
.7678430	05	.1430568	01	.2718503	04-	.1833314	03
.7778429	05	.1192818	01	.2266286	04-	.1833655	03
.7878429	05	.1001379	01	.1764406	04-	.1977235	03
.7978429	05	.8639777	00	.1175026	04-	.2561610	03
.8078430	05	.7685141	00	.8510464	05-	.3145986	03
.8178430	05	.6974343	00	.6513443	05-	.3730361	03
.8278430	05	.6355071	00	.6452434	05-	.3431271	03
.8378430	05	.5700333	00	.7312438	05-	.2715790	03
.8478430	05	.4945087	00	.8612610	05-	.2000310	03

FIGURE 12

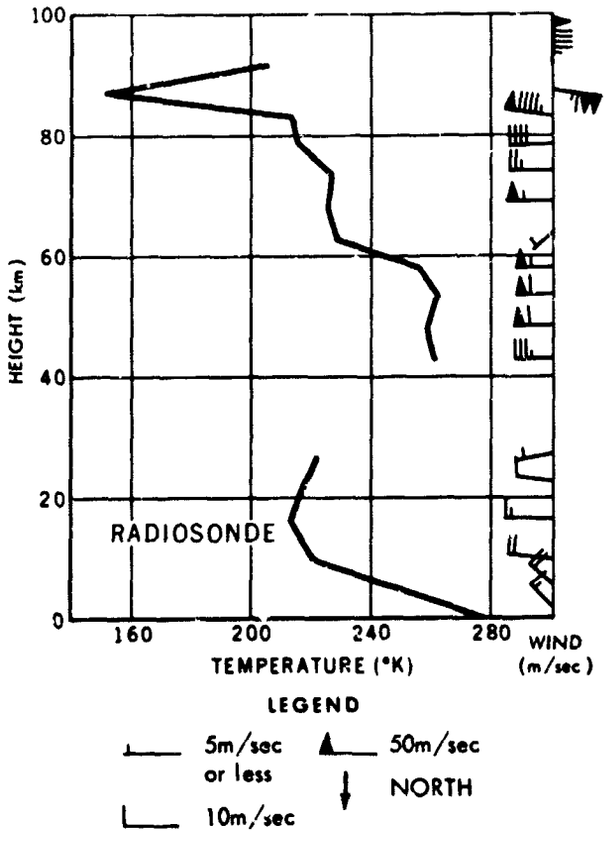
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
268.55438	43.623466	248.86549	41601.949
254.56788	43.215938	249.10910	46958.699
281.43969	43.811149	261.28778	52146.179
281.62848	38.183448	248.15359	56592.386
297.96038	76.103866	246.32100	61921.949
291.15948	27.702018	222.11869	67541.500
284.13769	34.293148	226.46260	72383.437
206.20140	66.652763	177.04888	77000.234
301.35809	97.686676	246.25959	82226.312
165.18818	65.556114	170.25909	87941.312



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2742410	05	.1700000	04	.2655842	01-	.2230000	03
.2842410	05	.1461522	04	.2264750	01-	.2248243	03
.2942410	05	.1258093	04	.1933827	01-	.2266457	03
.3042410	05	.1084331	04	.1653428	01-	.2284731	03
.3142410	05	.9357176	03	.1415513	01-	.2302974	03
.3242410	05	.8084485	03	.1213374	01-	.2321218	03
.3342410	05	.6993215	03	.1041404	01-	.2339461	03
.3442410	05	.6056336	03	.8949085	02-	.2357705	03
.3542410	05	.5251022	03	.7699543	02-	.2375949	03
.3642410	05	.4557963	03	.6632388	02-	.2394192	03
.3742410	05	.3960806	03	.5719868	02-	.2412436	03
.3842410	05	.3445678	03	.4938616	02-	.2430679	03
.3942410	05	.3000799	03	.4268940	02-	.2448923	03
.4042410	05	.2616154	03	.3694223	02-	.2467166	03
.4142410	05	.2283217	03	.3200422	02-	.2485410	03
.4242410	05	.1993932	03	.2780[illegible]63	02-	.2489029	03
.4342410	05	.1741614	03	.2[illegible]25	02-	.2489483	03
.4442410	05	.1521290	03	.2126539	02-	.2489938	03
.4542410	05	.1328954	03	.1859089	02-	.2490393	03
.4642410	05	.1161008	03	.1623852	02-	.2490848	03
.4742410	05	.1014626	03	.1412778	02-	.2502017	03
.4842410	05	.8875600	02	.1224361	02-	.2525494	03
.4942410	05	.7774005	02	.1062522	02-	.2548971	03
[illegible]	05	.6817692	02	.9233129	03-	.2572448	03
[illegible]	05	.5986396	02	.8033993	03-	.2595925	03
[illegible]	05	.5260920	02	.7036675	03-	.2604668	03
.5342410	05	.4621137	02	.6251846	03-	.2571127	03
.5442410	05	.4053256	02	.5547204	03-	.2545587	03
.5542410	05	.3549869	02	.4915319	03-	.2516047	03
.5642410	05	.3104264	02	.4349378	03-	.2486507	03
.5742410	05	.2711973	02	.3811744	03-	.2478676	03
.5842410	05	.2368622	02	.3333782	03-	.2475237	03
.5942410	05	.2068438	02	.2915329	03-	.2471799	03
.6042410	05	.1806033	02	.2549032	03-	.2468360	03
.6142410	05	.1576685	02	.2228435	03-	.2454922	03
.6242410	05	.1375504	02	.1962675	03-	.2441583	03
.6342410	05	.1197796	02	.1739797	03-	.2398515	03
.6442410	05	.1040481	02	.1538931	03-	.2355447	03
.6542410	05	.9015226	01	.1358238	03-	.2312379	03
.6642409	05	.7790551	01	.1196003	03-	.2269311	03
.6742409	05	.6713744	01	.1050631	03-	.2226243	03
.6842409	05	.5778324	01	.9030869	04-	.2229105	03
.6942409	05	.4975430	01	.7744866	04-	.2238076	03
.7042409	05	.4286865	01	.6646388	04-	.2247048	03
.7142409	05	.3695956	01	.5707451	04-	.2256019	03
.7242409	05	.3188029	01	.4[illegible]920	04-	.2260274	03
.7342409	05	.2740499	01	.4433988	04-	.2153244	03
.7442409	05	.2337805	01	.3980297	04-	.2046214	03
.7542409	05	.1977426	01	.3552548	04-	.1939184	03
.7642409	05	.1656868	01	.3150534	04-	.1832153	03
.7742409	05	.1380882	01	.2633699	04-	.1826622	03
.7842409	05	.1157901	01	.2059127	04-	.1959095	03
.7942409	05	.9822312	00	.1636125	04-	.2091689	03
.8042409	05	.8417155	00	.1318572	04-	.2223922	03
.8142409	05	.7278067	00	.1076052	04-	.2356355	03
.8242409	05	.6334245	00	.9057814	05-	.2436294	03
.8342409	05	.5504364	00	.8325551	05-	.2303310	03
.8442409	05	.4743636	00	.7614558	05-	.2170325	03
.8542409	05	.4049968	00	.6925812	05-	.2037341	03
.8642409	05	.3421191	00	.6258744	05-	.1904357	03
.8742409	05	.2855101	00	.5612258	05-	.1771373	03

FIGURE 13

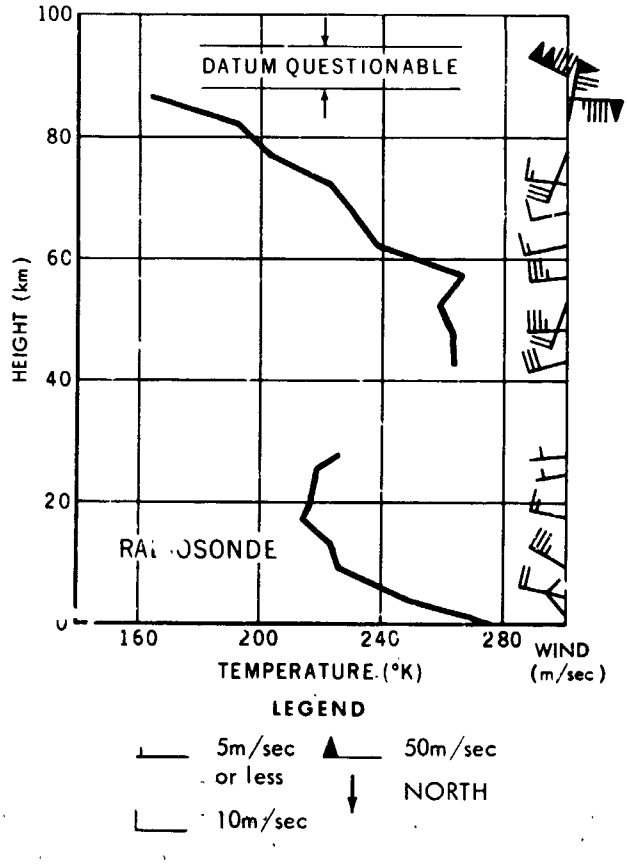
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
262.26339	37.318786	261.81979	42956.867
258.63888	59.576107	258.50137	48370.109
264.73590	57.798050	262.25830	53607.328
264.37527	55.014717	256.49829	58100.378
222.00199	1.9646849	228.56479	63469.269
267.01077	53.975399	225.95019	69130.968
266.96429	27.363178	227.11749	74019.578
259.86737	37.858348	215.65579	78665.953
291.07809	93.457389	213.82579	83089.789
115.27018	117 01439	151.20039	87297.484
355.37219	95.003875	205.23109	91614.125



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2692060	05	.1800000	04	.2833670	01-	.2213000	03
.2792060	05	.1546038	04	.2406303	01-	.2238267	03
.2892060	05	.1330240	04	.2047392	01-	.2263535	03
.2992060	05	.1146528	04	.1745157	01-	.2288803	03
.3092060	05	.9898465	03	.1490217	01-	.2314070	03
.3192060	05	.8529803	03	.1274762	01-	.2339338	03
.3292060	05	.7414078	03	.1092337	01-	.2364605	03
.3392060	05	.6431811	03	.9375978	02-	.2389873	03
.3492060	05	.5588272	03	.8061081	02-	.2415140	03
.3592060	05	.4862684	03	.6941795	02-	.2440408	03
.3692060	05	.4237558	03	.5987394	02-	.2465676	03
.3792060	05	.3698137	03	.5172223	02-	.2490943	03
.3892060	05	.3231954	03	.4474827	02-	.2516211	03
.3992060	05	.2828461	03	.3877233	02-	.2541478	03
.4092060	05	.2478712	03	.3364351	02-	.2566746	03
.4192060	05	.2175109	03	.2923492	02-	.2592014	03
.4292060	05	.1911191	03	.2543969	02-	.2617281	03
.4392060	05	.1680204	03	.2240778	02-	.2642290	03
.4492060	05	.1476788	03	.1974128	02-	.2666159	03
.4592060	05	.1297656	03	.1738760	02-	.2690029	03
.4692060	05	.1139952	03	.1531058	02-	.2593899	03
.4792060	05	.1001146	03	.1347814	02-	.2587769	03
.4892060	05	.8791744	02	.1183062	02-	.2588962	03
.4992060	05	.7722527	02	.1036311	02-	.2596136	03
.5092060	05	.6786043	02	.9081320	03-	.2603309	03
.5192060	05	.5965486	02	.7961284	03-	.2610483	03
.5292060	05	.5246212	02	.6982185	03-	.2617656	03
.5392060	05	.4614768	02	.6139661	03-	.2618567	03
.5492060	05	.4058291	02	.5425867	03-	.2605747	03
.5592060	05	.3566801	02	.4799330	03-	.2592927	03
.5692060	05	.3132953	02	.4250331	03-	.2580107	03
.5792060	05	.2750211	02	.3732070	03-	.2561287	03
.5892060	05	.2410731	02	.3290132	03-	.2522002	03
.5992060	05	.2107799	02	.2913449	03-	.2469601	03
.6092060	05	.1837712	02	.2648640	03-	.2417200	03
.6192060	05	.1597495	02	.2353440	03-	.2364800	03
.6292060	05	.1384386	02	.2085705	03-	.2312399	03
.6392060	05	.1196664	02	.1827121	03-	.2281723	03
.6492060	05	.1033289	02	.1580628	03-	.2271458	03
.6592059	05	.8920137	01	.1367078	03-	.2273193	03
.6692059	05	.7698772	01	.1182112	03-	.2268929	03
.6792059	05	.6643095	01	.1021939	03-	.2264664	03
.6892059	05	.5730858	01	.8832686	04-	.2260399	03
.6992059	05	.4943632	01	.7616044	04-	.2261387	03
.7092059	05	.4265180	01	.6563905	04-	.2263775	03
.7192059	05	.3680600	01	.5658296	04-	.2266163	03
.7292059	05	.3176766	01	.4878597	04-	.2268551	03
.7392059	05	.2742458	01	.4207197	04-	.2270938	03
.7492059	05	.2366110	01	.3665335	04-	.2248948	03
.7592059	05	.2038362	01	.3192640	04-	.2226280	03
.7692059	05	.1753177	01	.2776757	04-	.2199612	03
.7792059	05	.1505403	01	.2411364	04-	.2174944	03
.7892059	05	.1290703	01	.2086102	04-	.2155[illegible]4	03
.7992059	05	.1105743	01	.1790597	04-	.2151538	03
.8092059	05	.9470517	00	.1536573	04-	.2147231	03
.8192059	05	.8109328	00	.1318262	04-	.2143094	03
.8292059	05	.6942042	00	.1130439	04-	.2138958	03
.8392059	05	.5914240	00	.1022754	04-	.2014605	03
.8492059	05	.4982278	00	.9303110	05-	.1865769	03
.8592059	05	.4137934	00	.8396304	05-	.1716934	03
.8692059	05	.3579475	00	.7508170	05-	.1568098	03
.8792059	05	.2738231	00	.5999732	05-	.1584997	03
.8892059	05	.2239461	00	.4548786	05-	.1715166	03
.8992059	05	.1857187	00	.3516878	05-	.1840334	03
.9092059	05	.1560318	00	.2765654	05-	.1965502	03

FIGURE 14

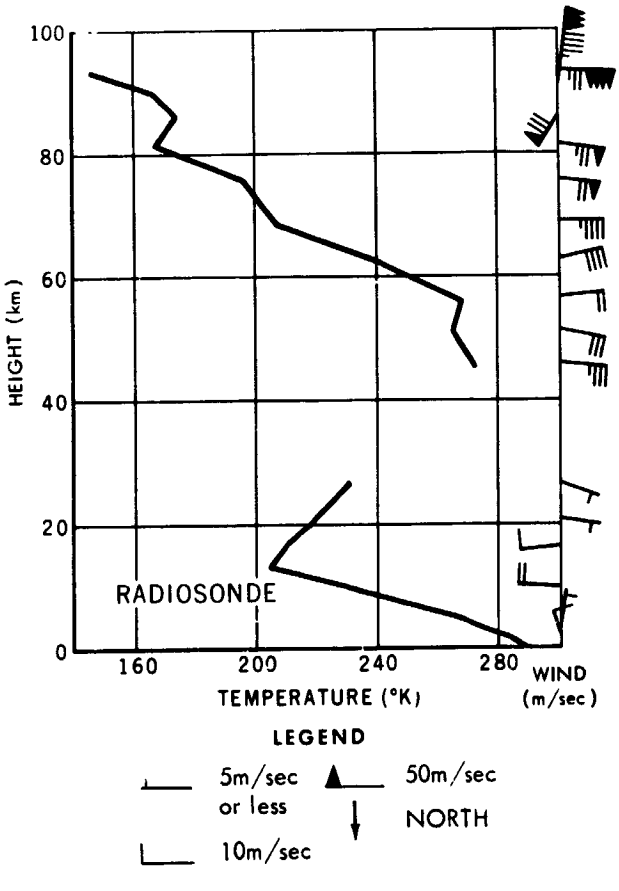
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
245.42630	2[illegible].887029	263.00979	43306.347
265.42379	3[illegible].420920	262.76327	48514.547
201.77038	22.565608	258.15268	52972.699
250.94839	35.262706	265.20490	57276.546
245.89270	17.411869	238.27420	62432.097
240.15739	11.140070	230.01030	67869.835
285.46209	13.573949	222.47969	72556.539
198.87809	30.647968	202.69108	77430.750
3.9694819	74.199607	193.52198	82064.093
92.676124	94.189346	165.26409	86408.437
297.49017	138.35170	287.35989	90139.250



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2911030	05	.1300000	04	.2015304	01-	.2247300	03
.3011030	05	.1119341	04	.1714666	01-	.2274265	03
.3111030	05	.9655342	03	.1461725	01-	.2301230	03
.3211030	05	.8343349	03	.1248473	01-	.2328195	03
.3311029	05	.7222092	03	.1068318	01-	.2355160	03
.3411029	05	.6262080	03	.9158243	02-	.2382125	03
.3511029	05	.5438643	03	.7864944	02-	.2409090	03
.3611029	05	.4731108	03	.6766029	02-	.2436055	03
.3711029	05	.4102117	03	.5830562	02-	.2463020	03
.3811029	05	.3597061	03	.5032793	02-	.2489986	03
.3911029	05	.3143625	03	.4351252	02-	.2516951	03
.4011029	05	.2751411	03	.3768001	02-	.2543916	03
.4111029	05	.2411617	03	.3268019	02-	.2570881	03
.4211029	05	.2116761	03	.2838708	02-	.2597846	03
.4311029	05	.1860570	03	.2469483	02-	.2624811	03
.4411029	05	.1636719	03	.2168320	02-	.2629717	03
.4511029	05	.1439983	03	.1908028	02-	.2629244	03
.4611029	05	.1266919	03	.1679014	02-	.2628771	03
.4711029	05	.1114675	03	.1477515	02-	.2628297	03
.4811029	05	.9807446	02	.1300223	02-	.2627824	03
.4911029	05	.8628077	02	.1146640	02-	.2621471	03
.5011029	05	.7587737	02	.1012376	02-	.2611129	03
.5111029	05	.6669698	02	.8934279	03-	.2600787	03
.5211029	05	.5859961	02	.7880948	03-	.2590445	03
.5311029	05	.5146551	02	.6939348	03-	.2583781	03
.5411029	05	.4521277	02	.6057844	03-	.2600167	03
.5511029	05	.3975365	02	.5293045	03-	.2616553	03
.5611029	05	.3498315	02	.4628885	03-	.2632939	03
.5711029	05	.3081077	02	.4051591	03-	.2649324	03
.5811029	05	.2712100	02	.3622210	03-	.2608496	03
.5911029	05	.2381021	02	.3245239	03-	.2556260	03
.6011029	05	.2086426	02	.2902837	03-	.2504024	03
.6111029	05	.1822569	02	.2594758	03-	.2451788	03
.6211029	05	.1587519	02	.2304873	03-	.2399551	03
.6311029	05	.1379677	02	.2026009	03-	.2372435	03
.6411029	05	.1197596	02	.1769967	03-	.2357238	03
.6511029	05	.1038639	02	.1545000	03-	.2342040	03
.6611029	05	.8999864	01	.1347495	03-	.2326843	03
.6711029	05	.7791455	01	.1174236	03-	.2311646	03
.6811030	05	.6739148	01	.1022459	03-	.2296239	03
.6911030	05	.5823417	01	.8897514	04-	.2280171	03
.7011029	05	.5027149	01	.7735417	04-	.2264103	03
.7111029	05	.4335417	01	.6718712	04-	.2248035	03
.7211029	05	.3735071	01	.5830010	04-	.2231967	03
.7311029	05	.3213063	01	.5082741	04-	.2202315	03
.7411030	05	.2757445	01	.4413920	04-	.2161716	03
.7511030	05	.2359697	01	.3715695	04-	.2121118	03
.7611030	05	.2013351	01	.3371367	04-	.2080519	03
.7711029	05	.1712561	01	.2924766	04-	.2039921	03
.7811029	05	.1452893	01	.2513901	04-	.2013463	03
.7911029	05	.1230324	01	.2149927	04-	.1993673	03
.8011030	05	.1040178	01	.1835879	04-	.1973884	03
.8111030	05	.8779786	00	.1565296	04-	.1954095	03
.8211030	05	.7397650	00	.1333818	04-	.1932214	03
.8311030	05	.6208957	00	.1158492	04-	.1867169	03
.8411030	05	.5179291	00	.1001253	04-	.1802124	03
.8511030	05	.4291923	00	.8607770	05-	.1737079	03
.8611030	05	.3531382	00	.7357968	05-	.1672033	03
.8711030	05	.2928459	00	.5420018	05-	.1882334	03
.8811030	05	.2488675	00	.3923858	05-	.2209597	03
.8911030	05	.2163191	00	.2970685	05-	.2536860	03
.9011030	05	.1912646	00	.2326490	05-	.2864123	03

FIGURE 15

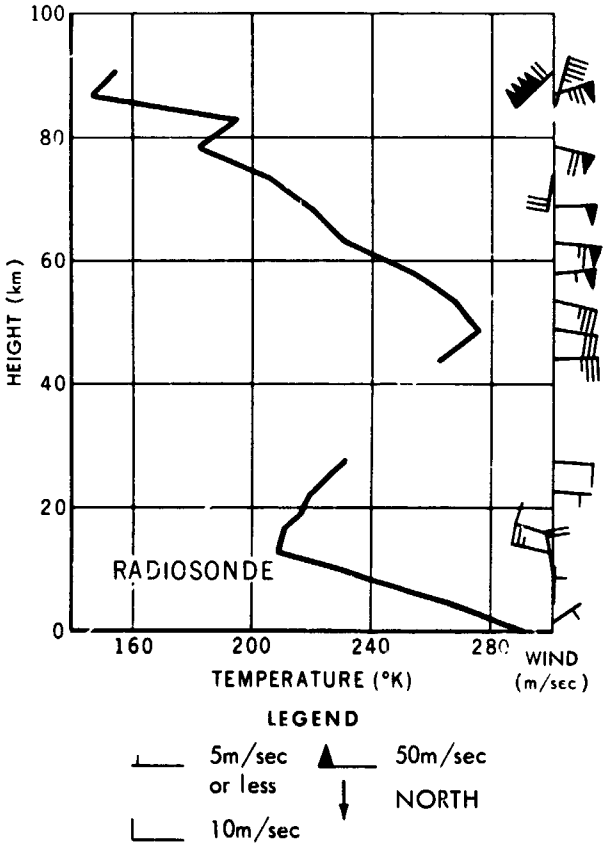
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
100.23349	35.039177	271.12729	45967.500
112.12339	31.985488	265.44738	51428.398
85.151390	20.540390	268.92038	56624.398
76.249366	37.992378	241.96479	62684.597
86.465759	47.108570	208.31379	69435.039
102.19909	72.066656	197.25399	75832.382
100.33849	77.084068	168.90759	81279.049
219.76649	81.912193	174.49778	85927.898
5.6612997	193.16529	167.00509	89940.898
97.556236	223.62478	145.78129	93665.695



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2782370	05	.1700000	04	.2561204	01-	.2312400	03
.2882370	05	.1469582	04	.2193208	01-	.2334384	03
.2982370	05	.1272189	04	.1880904	01-	.2356368	03
.3082370	05	.1102834	04	.1615445	01-	.2376352	03
.3182370	05	.9573240	03	.1389457	01-	.2400336	03
.3282370	05	.8321220	03	.1196778	01-	.2422320	03
.3382370	05	.7242419	03	.1032254	01-	.2444304	03
.3482370	05	.6311593	03	.8915658	02-	.2466288	03
.3582370	05	.5507353	03	.7710869	02-	.2488272	03
.3682370	05	.4811559	03	.6677687	02-	.2510256	03
.3782370	05	.4208802	03	.5790443	02-	.2532240	03
.3882370	05	.3685967	03	.5027484	02-	.2554224	03
.3982370	05	.3231886	03	.4370523	02-	.2576207	03
.4082370	05	.2837027	03	.3804088	02-	.2598191	03
.4182370	05	.2493247	03	.3315073	02-	.2620175	03
.4282370	05	.2193578	03	.2892359	02-	.2642159	03
.4382370	05	.1932052	03	.2526500	02-	.2664143	03
.4482370	05	.1703549	03	.2209460	02-	.2686127	03
.4582370	05	.1503671	03	.1934393	02-	.2708111	03
.4682370	05	.1327794	03	.1711766	02-	.2702367	03
.4782370	05	.1172096	03	.1516882	02-	.2691966	03
.4882370	05	.1034197	03	.1343609	02-	.2681565	03
.4982370	05	.9121133	02	.1189615	02-	.2671164	03
.5082370	05	.8040792	02	.1052812	02-	.2660763	03
.5182370	05	.7086315	02	.9291121	03-	.2657116	03
.5282370	05	.6245859	02	.8168621	03-	.2663800	03
.5382370	05	.5507040	02	.7184333	03-	.2670483	03
.5482370	05	.4857337	02	.6320927	03-	.2677168	03
.5582370	05	.4285786	02	.5563270	03-	.2683852	03
.5682370	05	.3781911	02	.4915637	03-	.2680339	03
.5782370	05	.3333633	02	.4406094	03-	.2635859	03
.5882370	05	.2932304	02	.3942177	03-	.2591379	03
.5982370	05	.2573672	02	.3520461	03-	.2546900	03
.6082370	05	.2253809	02	.3137727	03-	.2502420	03
.6182370	05	.1969090	02	.2790953	03-	.2457940	03
.6282370	05	.1716138	02	.2478020	03-	.2412713	03
.6382370	05	.1491656	02	.2199319	03-	.2362863	03
.6482370	05	.1292725	02	.1947090	03-	.2313013	03
.6582369	05	.1116886	02	.1719297	03-	.2263163	03
.6682369	05	.9618718	01	.1514023	03-	.2213313	03
.6782369	05	.8255966	01	.1329464	03-	.2163463	03
.6882369	05	.7061436	01	.1163927	03-	.2113613	03
.6982369	05	.6020476	01	.1010123	03-	.2076419	03
.7082369	05	.5122469	01	.8666698	04-	.2059131	03
.7182369	05	.4352695	01	.7426672	04-	.2041842	03
.7282369	05	.3693669	01	.6356042	04-	.2024554	03
.7382369	05	.3130174	01	.5432774	04-	.2007266	03
.7482369	05	.2648987	01	.4637561	04-	.1989978	03
.7582369	05	.2238624	01	.3953489	04-	.1972690	03
.7682369	05	.1886279	01	.3420963	04-	.1920948	03
.7782369	05	.1582039	01	.2949091	04-	.1868905	03
.7882369	05	.1320369	01	.2531813	04-	.1816861	03
.7982369	05	.1096266	01	.2164086	04-	.1764818	03
.8082369	05	.9052001	00	.1841207	04-	.1712774	03
.8182369	05	.7445886	00	.1529835	04-	.1695625	03
.8282369	05	.6123329	00	.1249243	04-	.1707650	03
.8382369	05	.5042907	00	.1021628	04-	.1719675	03
.8482369	05	.4158986	00	.8367063	05-	.1731700	03
.8582369	05	.3434779	00	.6862448	05-	.1743725	03
.8682369	05	.2836304	00	.5717470	05-	.1728252	03
.8782369	05	.2337198	00	.4764040	05-	.1709581	03
.8882369	05	.1922939	00	.3961897	05-	.1690910	03
.8982369	05	.1578369	00	.3288277	05-	.1672239	03
.9082369	05	.1290085	00	.2774782	05-	.1619749	03
.9182369	05	.1047229	00	.2334559	05-	.1562769	03
.9282369	05	.8435857	01-	.1951745	05-	.1505790	03
.9382369	05	.6706524	01-	.1684508	05-	.1387020	03

FIGURE 16

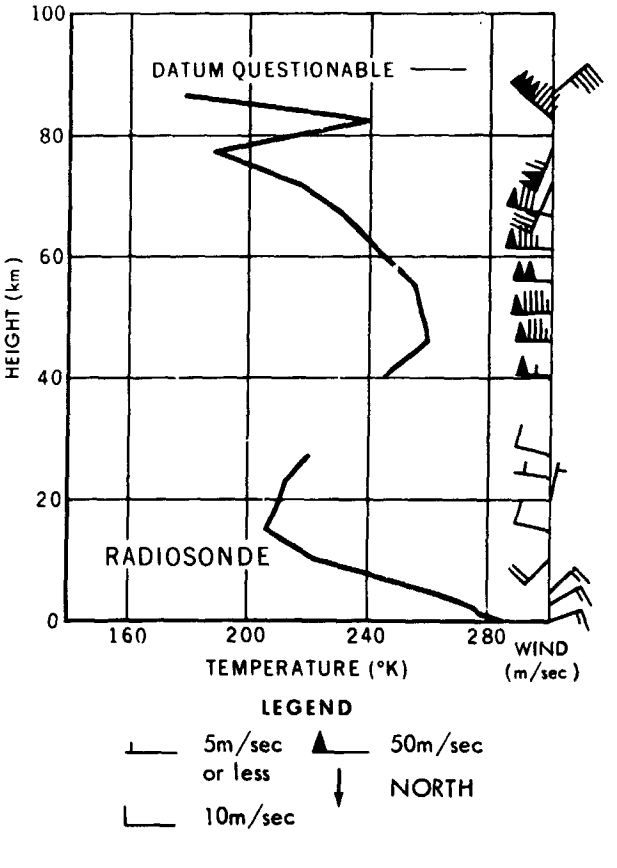
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
88.135818	35.642078	262.44598	43731.648
118.24260	29.022520	275.65139	49039.000
121.08769	36.761146	266.55227	53575.898
80.334419	54.959678	254.40489	57950.200
96.692169	64.717727	231.56219	63171.789
80.894264	51.982868	220.42698	68672.593
192.69558	28.052829	206.80468	73406.648
121.72178	70.112663	182.82728	78311.234
31.833500	44.285537	194.84429	82984.882
70.778137	72.619316	147.24949	86953.234
224.92109	217.80018	154.34379	90628.484



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.3128600	05	.1000000	04	.1498425	01-	.2325000	03
.3228600	05	.8653305	03	.1283352	01-	.2349061	03
.3328600	05	.7499351	03	.1100935	01-	.2373123	03
.3428600	05	.6508961	03	.9459506	02-	.2397184	03
.3528600	05	.5657608	03	.8140521	02-	.2421245	03
.3628600	05	.4924644	03	.7016162	02-	.2445307	03
.3728600	05	.4290652	03	.6056169	02-	.2469368	03
.3828600	05	.3746909	03	.5235211	02-	.2493430	03
.3928600	05	.3274960	03	.4532067	02-	.2517491	03
.4028600	05	.2866244	03	.3928911	02-	.2541552	03
.4128600	05	.2511792	03	.3410755	02-	.2565614	03
.4228600	05	.2203976	03	.2964967	02-	.2589675	03
.4328600	05	.1936301	03	.2580888	02-	.2613737	03
.4428600	05	.1703242	03	.2249149	02-	.2638252	03
.4528600	05	.1500085	03	.1962371	02-	.2663134	03
.4628600	05	.1322774	03	.1714400	02-	.2688015	03
.4728600	05	.1167819	03	.1499686	02-	.2712897	03
.4828600	05	.1032229	03	.1313517	02-	.2737778	03
.4928600	05	.9132106	02	.1156246	02-	.2751560	03
.5028600	05	.8078318	02	.1030332	02-	.2731504	03
.5128600	05	.7139951	02	.9173859	03-	.2711448	03
.5228600	05	.6305047	02	.8161490	03-	.2691392	03
.5328600	05	.5562814	02	.7254777	03-	.2671336	03
.5428600	05	.4902878	02	.6455825	03-	.2645803	03
.5528600	05	.4315882	02	.5743182	03-	.2618033	03
.5628600	05	.3794151	02	.5103039	03-	.2590063	03
.5728600	05	.3330998	02	.4528661	03-	.2562493	03
.5828600	05	.2919949	02	.4021824	03-	.2529359	03
5928600	05	.2554564	02	.3580484	03-	.2485612	03
.6028600	05	.2229696	02	.3181136	03-	.2441865	03
.6128600	05	.1941446	02	.2820414	03-	.2398119	03
.6228600	05	.1686231	02	.2495171	03-	.2354372	03
.6328600	05	.1460886	02	.2200091	03-	.2313310	03
.6428600	05	.1263302	02	.1919326	03-	.2293067	03
.6528600	05	.1091084	02	.1672441	03-	.2272824	03
.6628600	05	.9411522	01	.1455585	03-	.2252581	03
.6728600	05	.8107781	01	.1265319	03-	.2232338	03
.6828600	05	.6975484	01	.1098572	03-	.2212095	03
.6928600	05	.5992222	01	.9547134	04-	.2186619	03
.7028600	05	.5138047	01	.8295381	04-	.2157844	03
.7128600	05	.4396759	01	.7194509	04-	.2129069	03
.7228600	05	.3754634	01	.6227959	04-	.2100293	03
.7328600	05	.3199472	01	.5380811	04-	.2071518	03
.7428600	05	.2718522	01	.4676853	04-	.2023057	03
.7528600	05	.2301037	01	.4056556	04-	.1976169	03
.7628600	05	.1939649	01	.3506196	04-	.1907282	03
.7728600	05	.1627946	01	.3019336	04-	.1878394	03
.7828600	05	.1360111	01	.2589994	04-	.1809506	03
.7928600	05	.1135026	01	.2133584	04-	.1853336	03
.8028600	05	.9495193	00	.1760451	04-	.1879048	03
.8128600	05	.7963040	00	.1456454	04-	.1904761	03
.8228600	05	.6694253	00	.1208083	04-	.1930473	03
.8328600	05	.5629857	00	.1025636	04-	.1912328	03
.8428600	05	.4704215	00	.9143498	05-	.1792392	03
.8528600	05	.3882380	00	.8087262	05-	.1672456	03
.8628600	05	.3158847	00	.7088422	05-	.1552520	03
.8728600	05	.2536854	00	.5975985	05-	.1478918	03
.8828600	05	.2029426	00	.4719060	05-	.1498221	03
.8928600	05	.1628251	00	.3738041	05-	.1517524	03
.9028600	05	.1310111	00	.2969896	05-	.1536827	03

FIGURE 17

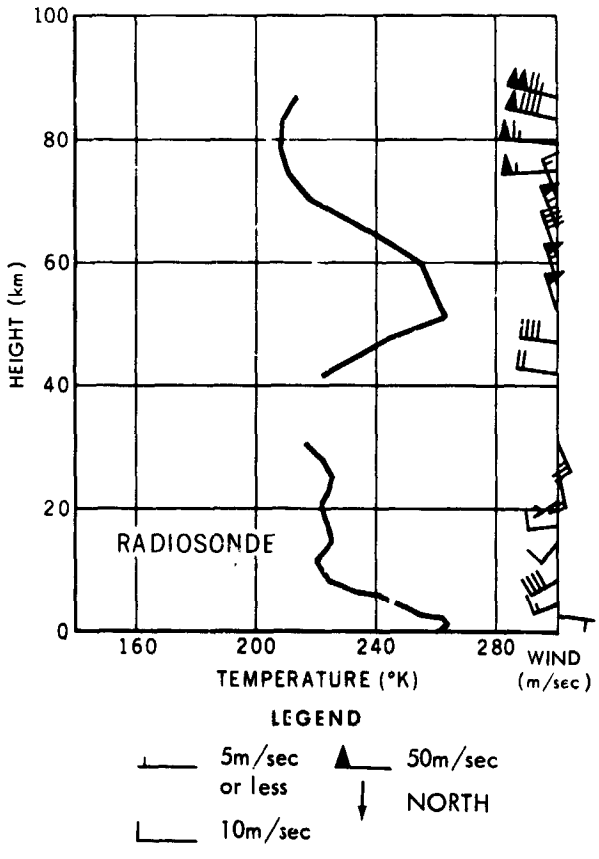
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
273.83029	53.767238	245.67549	40618.347
267.66479	83.497436	259.56347	46206.437
265.29300	94.068039	256.97729	50987.546
283.99978	99.636138	255.52859	55623.097
276.75988	85.044586	241.85888	61199.898
295.78497	80.256599	230.72470	67064.695
216.34069	36.389019	216.65219	72110.546
203.29139	115.02049	188.27578	77416.687
329.80078	196.25229	239.28830	82479.836
46.157417	45.832988	179.11309	86794.796
346.41769	389.29647	271.31540	90863.945



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2889260	05	.1300000	04	.2061350	01-	.2197100	03
.2989260	05	.1115347	04	.1750907	01-	.2219244	03
.3089260	05	.9584243	03	.1489701	01-	.2241388	03
.3189260	05	.8248469	03	.1269536	01-	.2263532	03
.3289260	05	.7109576	03	.1083645	01-	.2285676	03
.3389260	05	.6136999	03	.9264295	02-	.2307820	03
.3489260	05	.5305159	03	.7932451	02-	.2329964	03
.3589260	05	.4592600	03	.6802359	02-	.2352108	03
.3689260	05	.3981302	03	.5841933	02-	.2374252	03
.3789260	05	.3456103	03	.5024425	02-	.2396396	03
.3889260	05	.3004224	03	.4327502	02-	.2418540	03
.3989260	05	.2614878	03	.3732486	02-	.2440684	03
.4089260	05	.2278992	03	.3222819	02-	.2463571	03
.4189260	05	.1988968	03	.2784593	02-	.2488423	03
.4289260	05	.1738276	03	.2409554	02-	.2513276	03
.4389260	05	.1521260	03	.2088084	02-	.2538129	03
.4489260	05	.1333123	03	.1812104	02-	.2562982	03
.4589260	05	.1169791	03	.1574816	02-	.2587835	03
.4689260	05	.1027264	03	.1380759	02-	.2591923	03
.4789260	05	.9021051	02	.1215068	02-	.2586514	03
.4889260	05	.7920117	02	.1069016	02-	.2581105	03
.4989260	05	.6951930	02	.9403056	03-	.2575695	03
.5089260	05	.6100673	02	.8269028	03-	.2570286	03
.5189260	05	.5352674	02	.7264614	03-	.2566944	03
.5289260	05	.4695803	02	.6380881	03-	.2563819	03
.5389260	05	.4119050	02	.5603992	03-	.2560694	03
.5489260	05	.3612704	02	.4921112	03-	.2557569	03
.5589260	05	.3167762	02	.4330075	03-	.2548680	03
.5689260	05	.2775331	02	.3830493	03-	.2524168	03
.5789260	05	.2428479	02	.3384638	03-	.2499656	03
.5889260	05	.2122270	02	.2987159	03-	.2475145	03
.5989260	05	.1852264	02	.2633194	03-	.2450633	03
.6089260	05	.1618759	02	.2318330	03-	.2426121	03
.6189260	05	.1415455	02	.2035546	03-	.2405438	03
.6289260	05	.1222150	02	.1784144	03-	.2386453	03
.6389260	05	.1061614	02	.1562215	03-	.2367468	03
.6489260	05	.9211609	01	.1366489	03-	.2348484	03
.6589259	05	.7984050	01	.1194040	03-	.2329499	03
.6689259	05	.6912293	01	.1042249	03-	.2310514	03
.6789259	05	.5976166	01	.9114961	04-	.2284157	03
.6889259	05	.5158077	01	.7964442	04-	.2256268	03
.6989259	05	.4444041	01	.6947799	04-	.2228379	03
.7089259	05	.3821849	01	.6050795	04-	.2200490	03
.7189259	05	.3286607	01	.5260566	04-	.2172600	03
.7289259	05	.2808563	01	.4605162	04-	.2124699	03
.7389259	05	.2395546	01	.4029362	04-	.2071221	03
.7489259	05	.2034884	01	.3513436	04-	.2017742	03
.7589259	05	.1721051	01	.3052474	04-	.1964264	03
.7689259	05	.1448981	01	.2641856	04-	.1910785	03
.7789259	05	.1218167	01	.2198107	04-	.1930707	03
.7889259	05	.1029564	01	.1765646	04-	.2031459	03
.7989259	05	.8773210	00	.1433463	04-	.2132211	03
.8089259	05	.7531704	00	.1175086	04-	.2232964	03
.8189259	05	.6509872	00	.9718136	05-	.2333717	03
.8289259	05	.5645094	00	.8421372	05-	.2335321	03
.8389259	05	.4873984	00	.7732802	05-	.2195864	03
.8489259	05	.4168033	00	.7061230	05-	.2056407	03
.8589259	05	.3525550	00	.6407291	05-	.1916949	03
.8689259	05	.2949986	00	.5667752	05-	.1813290	03
.8789259	05	.2482211	00	.4239283	05-	.2039870	03
.8889259	05	.2127082	00	.3669587	05-	.2266468	03
.8989259	05	.1849879	00	.2585053	05-	.2493057	03
.9089259	05	.1628287	00	.2090840	05-	.2713116	03

FIGURE 18

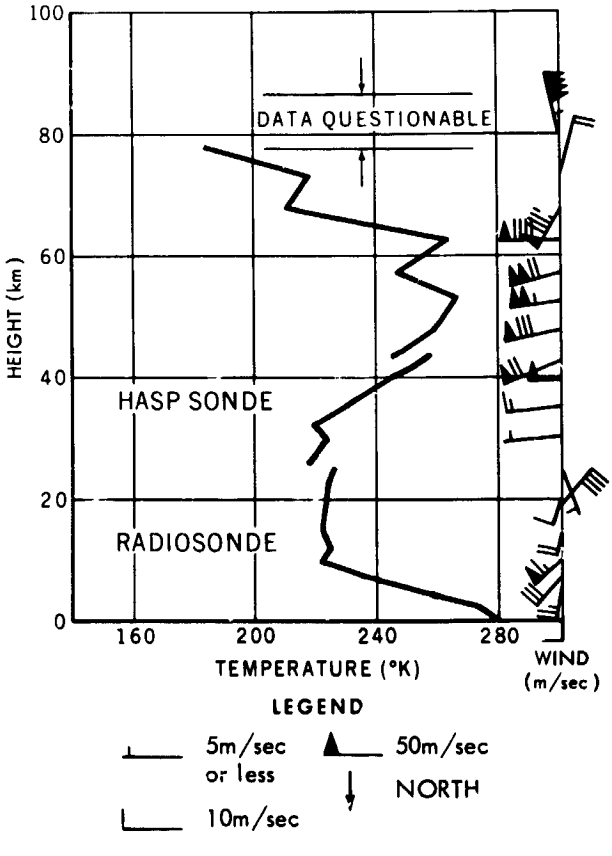
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
290.36468	20.766498	221.75138	41866.148
288.81070	40.355209	242.00459	47108.296
330.36560	48.105667	262.88278	51592.347
334.51089	54.224269	258.55648	55932.187
339.43869	39.816047	254.45228	60114.148
328.10589	55.908588	238.15919	65105.547
333.09668	13.910879	218.05169	70373.898
263.82138	54.381446	210.59829	74878.593
279.66458	64.996559	208.21719	79168.898
294.59338	91.706909	208.93289	83243.593
299.68167	125.85369	212.67819	86702.132



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.3064000	05	.1000000	04	.1603978	01-	.2172000	03
.3164000	05	.8557217	03	.1370001	01-	.2176054	03
.3264000	05	.7325066	03	.1170554	01-	.2180108	03
.3364000	05	.6272446	03	.1000484	01-	.2184163	03
.3464000	05	.5372898	03	.8554141	02-	.2188217	03
.3564000	05	.4603895	03	.7316263	02-	.2192271	03
.3664000	05	.3946274	03	.6259631	02-	.2196325	03
.3764000	05	.3383709	03	.5357393	02-	.2200380	03
.3864000	05	.2902301	03	.4586734	02-	.2204434	03
.3964000	05	.2490202	03	.3928237	02-	.2208488	03
.4064000	05	.2137321	03	.3365396	02-	.2212543	03
.4164000	05	.1835045	03	.2884153	02-	.2216597	03
.4264000	05	.1577466	03	.2445319	02-	.2247412	03
.4364000	05	.1359248	03	.2071437	02-	.2286047	03
.4464000	05	.1174197	03	.1759688	02-	.2324682	03
.4564000	05	.1016835	03	.1498048	02-	.2363318	03
.4664000	05	.8826577	02	.1280224	02-	.2401953	03
.4764000	05	.7680436	02	.1094461	02-	.2440802	03
.4864000	05	.6700262	02	.9369423	03-	.2491363	03
.4964000	05	.5860218	02	.8044393	03-	.2537924	03
.5064000	05	.5138202	02	.6926204	03-	.2584485	03
.5164000	05	.4515614	02	.5985374	03-	.2628352	03
.5264000	05	.3971945	02	.5284795	03-	.2618384	03
.5364000	05	.3492161	02	.4664186	03-	.2608415	03
.5464000	05	.3068940	02	.4114649	03-	.2598446	03
.5564000	05	.2695780	02	.3628259	03-	.2588477	3
.5664000	05	.2366912	02	.3197815	03-	.2578618	03
.5764000	05	.2077214	02	.2817140	03-	.2568804	03
.5864000	05	.1822136	02	.2480678	03-	.2558990	03
.5964000	05	.1597643	02	.2183424	03-	.2549176	03
.6064000	05	.1399718	02	.1929444	03-	.2527358	03
.6164000	05	.1224605	02	.1710147	03-	.2494715	03
.6264000	05	.1069561	02	.1513431	03-	.2462073	03
.6364000	05	.9324993	01	.1337218	03-	.2429431	03
.6464000	05	.8115300	01	.1179595	03-	.2396788	03
.6564000	05	.7048782	01	.1040017	03-	.2361194	03
.6664000	05	.6109125	01	.9161846	04-	.2323027	03
.6764000	05	.5282436	01	.8054392	04-	.2284860	03
.6864000	05	.4556648	01	.7065775	04-	.2246694	03
.6964000	05	.3920823	01	.6184899	04-	.2208527	03
.7064000	05	.3365729	01	.5388349	04-	.2176114	03
.7164000	05	.2884390	01	.4653130	04-	.2159568	03
.7264000	05	.2469075	01	.4013393	04-	.2143022	03
.7364000	05	.2111117	01	.3458677	04-	.2126476	03
.7464000	05	.1802936	01	.2976941	04-	.2109930	03
.7564000	05	.1538304	01	.2550018	04-	.2101757	03
.7664000	05	.1312049	01	.2180589	04-	.2096207	03
.7764000	05	.1118589	01	.1864000	04-	.2090657	03
.7864000	05	.9532987	00	.1592791	04-	.2085107	03
.7964000	05	.8122319	00	.1358467	04-	.2082999	03
.8064000	05	.6920720	00	.1156523	04-	.2084756	03
.8164000	05	.5897942	00	.9847768	05-	.2086512	03
.8264000	05	.5027252	00	.8386932	05-	.2088269	03
.8364000	05	.4286470	00	.7132757	05-	.2093621	03
.8464000	05	.3657261	00	.6054459	05-	.2104451	03
.8564000	05	.3123110	00	.5143724	05-	.2115280	03
.8664000	05	.2669252	00	.4373834	05-	.2126109	03

FIGURE 19

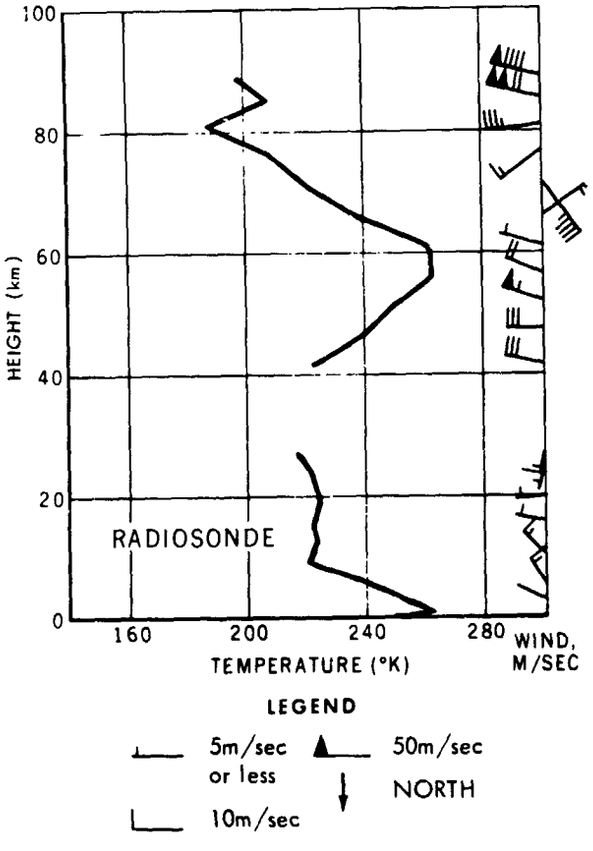
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
248.97470	71.465309	244.66299	43146.937
259.58279	82.276358	259.48867	48478.148
257.88967	103.08649	265.55178	53019.250
250.40109	122.64289	247.70709	57394.699
271.84619	91.785987	262.91629	62650.347
212.86879	87.717720	210.57238	68162.734
18.724060	20.152278	217.48979	72900.132
348.64187	202.72158	184.07159	77840.836
132.55758	339.43707	93.948997	82529.546
272.95788	385.39788	327.72747	86532.296



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2604670	05	.2000000	04	.3237469	01-	.2152200	03
.2704670	05	.1709929	04	.2745953	01-	.2169418	03
.2804670	05	.1463812	04	.2332207	01-	.2186635	03
.2904670	05	.1254708	04	.1983436	01-	.2203853	03
.3004670	05	.1076817	04	.1689032	01-	.2221071	03
.3104670	05	.9252830	03	.1440179	01-	.2238289	03
.3204670	05	.7960349	03	.1229550	01-	.2255507	03
.3304670	05	.6856570	03	.1051037	01-	.2272725	03
.3404670	05	.5912771	03	.8995484	02-	.2289943	03
.3504670	05	.5104778	03	.7708274	02-	.2307161	03
.3604670	05	.4412218	03	.6613148	02-	.2324379	03
.3704670	05	.3817895	03	.5680284	02-	.2341597	03
.3804670	05	.3307277	03	.4884667	02-	.2358814	03
.3904670	05	.2868072	03	.4205289	02-	.2376032	03
.4004670	05	.2489863	03	.3624479	02-	.2393250	03
.4104670	05	.2163815	03	.3127353	02-	.2410468	03
.4204670	05	.1882424	03	.2701365	02-	.2427686	03
.4304670	05	.1639311	03	.2335919	02-	.2444904	03
.4404670	05	.1429421	03	.2014796	02-	.2471651	03
.4504670	05	.1248333	03	.1739972	02-	.2499461	03
.4604670	05	.1091867	03	.1505138	02-	.2527270	03
.4704670	05	.9564529	02	.1304119	02-	.2555079	03
.4804670	05	.8390687	02	.1131749	02-	.2582888	03
.4904670	05	.7370027	02	.9865980	03-	.2602477	03
.5004670	05	.6479087	02	.8629041	03-	.2615829	03
.5104670	05	.5699818	02	.7552638	03-	.2629181	03
.5204670	05	.5017729	02	.6615231	03-	.2642532	03
.5304670	05	.4420119	02	.5801310	03-	.2654398	03
.5404670	05	.3891117	02	.5186698	03-	.2613614	03
.5504670	05	.3418705	02	.4629228	03-	.2572831	03
.5604670	05	.2997563	02	.4124342	03-	.2532047	03
.5704670	05	.2622805	02	.3667791	03-	.2491263	03
.5804670	05	.2292788	02	.3200282	03-	.2495939	03
.5904670	05	.2006179	02	.2768137	03-	.2524877	03
.6004670	05	.1758143	02	.2398406	03-	.2553816	03
.6104670	05	.1543129	02	.2081503	03-	.2582755	03
.6204670	05	.1356434	02	.1809400	03-	.2611694	03
.6304670	05	.1192634	02	.1603281	03-	.2591526	03
.6404670	05	.1045592	02	.1459073	03-	.2496570	03
.6504670	05	.9100528	01	.1323047	03-	.2401613	03
.6604669	05	.7912265	01	.1195027	03-	.2306656	03
.6704669	05	.6823545	01	.1074836	03-	.2211699	03
.6804669	05	.5846780	01	.9622920	04-	.2116742	03
.6904669	05	.4993262	01	.8210832	04-	.2118631	03
.7004669	05	.4267132	01	.6968768	04-	.2133233	03
.7104669	05	.3650684	01	.5921498	04-	.2147834	03
.7204669	05	.3126745	01	.5037411	04-	.2162436	03
.7304669	05	.2679761	01	.4312238	04-	.2164984	03
.7404669	05	.2291380	01	.3806145	04-	.2097346	03
.7504669	05	.1949340	01	.3345894	04-	.2029707	03
.7604669	05	.1649379	01	.2928630	04-	.1962069	03
.7704669	05	.1387493	01	.2551587	04-	.1894430	03
.7804669	05	.1158476	01	.2240764	04-	.1801147	03
.7904669	05	.9526619	00	.2062806	04-	.1608935	03
.8004669	05	.7641867	00	.1879198	04-	.1416723	03
.8104669	05	.5936266	00	.1688918	04-	.1224512	03
.8204669	05	.4416710	00	.1490566	04-	.1032300	03
.8304669	05	.3293313	00	.9241337	05-	.1241528	03
.8404669	05	.2644035	00	.5045758	05-	.1825572	03
.8504669	05	.2257515	00	.3263930	05-	.2409617	03
.8604669	05	.1995179	00	.2321867	05-	.2993661	03

FIGURE 20

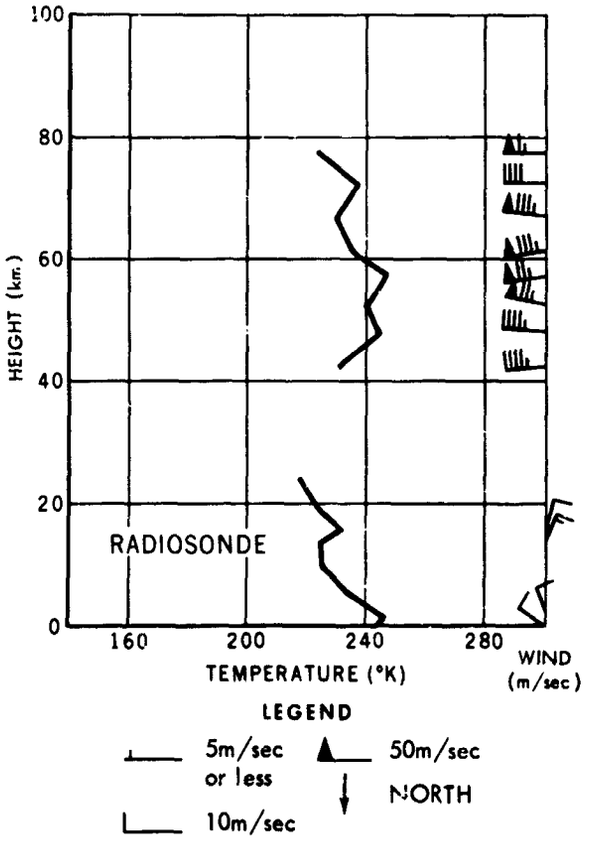
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
284.19079	32.374157	222.58299	41977.500
271.21618	29.083198	240.94349	47418.187
302.20727	56.672359	250.83938	52084.347
311.63467	21.642078	262.77108	56597.250
302.11697	3.6421175	261.85049	60956.847
55.247146	20.369548	238.49319	66180.343
127.13758	42.825447	220.66958	71686.445
235.76329	15.062839	209.03039	76435.148
257.93728	34.519538	187.90699	80973.046
300.67657	119.71708	208.01059	85273.843
301.21899	88.249855	198.10629	88936.648



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2760000	05	.1300000	04	.2085171	01-	.2172000	03
.2860000	05	.1112262	04	.1780573	01-	.2175744	03
.2960000	05	.9519342	03	.1521635	01-	.2179488	03
.3060000	05	.8145743	03	.1300476	01-	.2183232	03
.3160000	05	.6979395	03	.1111814	01-	.2186976	03
.3260000	05	.5979000	03	.9508234	02-	.2190720	03
.3360000	05	.5123596	03	.8134008	02-	.2194464	03
.3460000	05	.4391932	03	.6960573	02-	.2198208	03
.3560000	05	.3765916	03	.5958280	02-	.2201952	03
.3660000	05	.3230126	03	.5101904	02-	.2205696	03
.3760000	05	.2771425	03	.4369975	02-	.2209440	03
.3860000	05	.2378525	03	.3744217	02-	.2213184	03
.3960000	05	.2040864	03	.3209046	02-	.2216928	03
.4060000	05	.1753683	03	.2751216	02-	.2220672	03
.4160000	05	.1506487	03	.2356432	02-	.2224416	03
.4260000	05	.1295342	03	.2008496	02-	.2246837	03
.4360000	05	.1115930	03	.1704704	02-	.2260584	03
.4460000	05	.9635183	02	.1450417	02-	.2314330	03
.4560000	05	.8337308	02	.1237006	02-	.2348077	03
.4660000	05	.7209490	02	.1057441	02-	.2381824	03
.4760000	05	.6281316	02	.9067743	03-	.2413291	03
.4860000	05	.5466085	02	.7822130	03-	.2434498	03
.4960000	05	.4762602	02	.6756566	03-	.2455706	03
.5060000	05	.4154754	02	.5843763	03-	.2476914	03
.5160000	05	.3628858	02	.5060746	03-	.2498122	03
.5260000	05	.3173516	02	.4383786	03-	.2522027	03
.5360000	05	.2779123	02	.3799156	03-	.2548466	03
.5460000	05	.2437178	02	.3297495	03-	.2574905	03
.5560000	05	.2140260	02	.2866334	03-	.2601344	03
.5660000	05	.1882057	02	.2495252	03-	.2627705	03
.5760000	05	.1656052	02	.2197378	03-	.2625593	03
.5860000	05	.1457086	02	.1934930	03-	.2623482	03
.5960000	05	.1281947	02	.1703727	03-	.2621370	03
.6060000	05	.1127752	02	.1500061	03-	.2619258	03
.6160000	05	.9914414	01	.1333731	03-	.2589746	03
.6260000	05	.8699873	01	.1190908	03-	.2545030	03
.6360000	05	.7616766	01	.1061290	03-	.2500314	03
.6460000	05	.6652800	01	.9438551	04-	.2455598	03
.6560000	05	.5796647	01	.8376431	04-	.2410882	03
.6660000	05	.5038653	01	.7402469	04-	.2371352	03
.6760000	05	.4370616	01	.6509864	04-	.2338993	03
.6860000	05	.3783815	01	.5714909	04-	.2306634	03
.6960000	05	.3269279	01	.5008033	04-	.2274275	03
.7060000	05	.2818931	01	.4380495	04-	.2241917	03
.7160000	05	.2425500	01	.3824300	04-	.2209558	03
.7260000	05	.2082973	01	.3322162	04-	.2184344	03
.7360000	05	.1785787	01	.2880511	04-	.2159823	03
.7460000	05	.1528386	01	.2493629	04-	.2135303	03
.7560000	05	.1305801	01	.2155221	04-	.2110782	03
.7660000	05	.1113483	01	.1862643	04-	.2082630	03
.7760000	05	.9467931	00	.1620010	04-	.2036081	03
.7860000	05	.8020842	00	.1404517	04-	.1989532	03
.7960000	05	.6768652	00	.1213644	04-	.1942984	03
.8060000	05	.5688791	00	.1045058	04-	.1896435	03
.8160000	05	.4773957	00	.8715107	05-	.1908376	03
.8260000	05	.4017105	00	.7158104	05-	.1955100	03
.8360000	05	.3394235	00	.5906981	05-	.2001864	03
.8460000	05	.2879266	00	.4896448	05-	.2048608	03
.8560000	05	.2449343	00	.4119720	05-	.2071286	03
.8660000	05	.2083360	00	.3550499	05-	.2044246	03
.8760000	05	.1768338	00	.3054029	05-	.2017206	03
.8860000	05	.1497709	00	.2621780	05-	.1990166	03

FIGURE 21

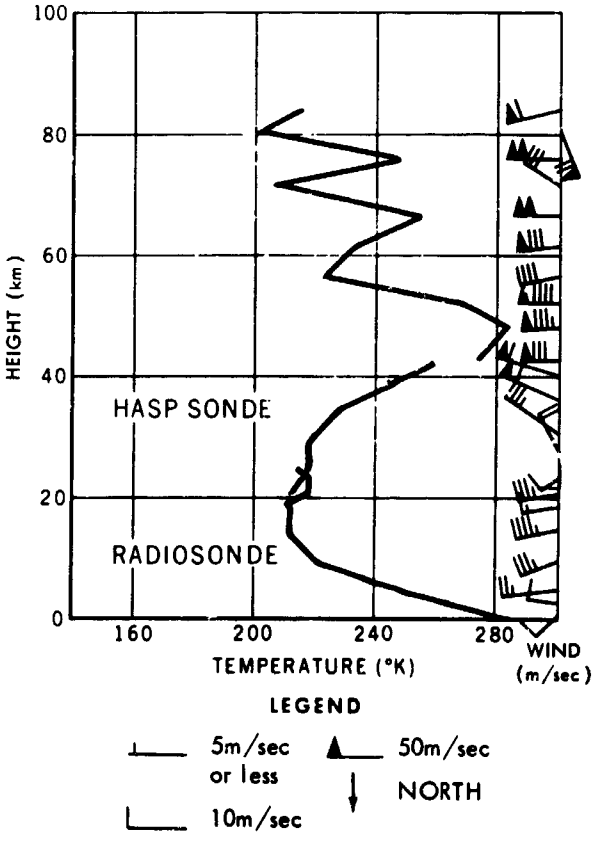
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
262.60537	43.649078	231.49218	42507.648
275.74258	43.370437	244.15739	48028.449
286.31298	76.299484	240.30319	52744.500
254.90879	75.013313	246.50650	57296.250
257.16519	83.688209	235.64189	61691.097
283.97580	82.753585	229.45658	66965.843
268.59268	41.507049	236.89219	72513.195
273.20635	64.814689	223.62619	77263.539



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2907000	05	.1230000	04	.2007083	01-	.2135000	03
.3007000	05	.1049980	04	.1702653	01-	.2148389	03
.3107000	05	.8972328	03	.1445946	01-	.2161779	03
.3207000	05	.7674885	03	.1229242	01-	.2175168	03
.3307000	05	.6571673	03	.1046107	01-	.2188557	03
.3407000	05	.5632642	03	.8911760	02-	.2201947	03
.3507000	05	.4832532	03	.7599644	02-	.2215336	03
.3607000	05	.4150106	03	.6487252	02-	.2228725	03
.3707000	05	.3567466	03	.5543195	02-	.2242115	03
.3807000	05	.3069531	03	.4741181	02-	.2255504	03
.3907000	05	.2643569	03	.4059146	02-	.2268894	03
.4007000	05	.2278827	03	.3478564	02-	.2282283	03
.4107000	05	.1966206	03	.2983853	02-	.2295672	03
.4207000	05	.1698006	03	.2561898	02-	.2309062	03
.4307000	05	.1467947	03	.2196943	02-	.2327822	03
.4407000	05	.1270763	03	.1883275	02-	.2350763	03
.4507000	05	.1101657	03	.1616881	02-	.2373704	03
.4607000	05	.9564090	02	.1390266	02-	.2396645	03
.4707000	05	.8314666	02	.1197187	02-	.2419586	03
.4807000	05	.7238067	02	.1032931	02-	.2441234	03
.4907000	05	.6303553	02	.9025898	03-	.2433062	03
.5007000	05	.5487370	02	.7883706	03-	.2424889	03
.5107000	05	.4774843	02	.6883218	03-	.2416717	03
.5207000	05	.4153059	02	.6007494	03-	.2408544	03
.5307000	05	.3611416	02	.5226069	03-	.2407468	03
.5407000	05	.3141710	02	.4520768	03-	.2421096	03
.5507000	05	.2735351	02	.3914005	03-	.2434725	03
.5607000	05	.2383494	02	.3391550	03-	.2448353	03
.5707000	05	.2078575	02	.2941298	03-	.2461981	03
.5807000	05	.1812616	02	.2581778	03-	.2445937	03
.5907000	05	.1578943	02	.2271910	03-	.2421216	03
.6007000	05	.1373506	02	.1996698	03-	.2396494	03
.6107000	05	.1195125	02	.1752553	03-	.2371773	03
.6207000	05	.1035106	02	.1533240	03-	.2351976	03
.6307000	05	.8971970	01	.1335623	03-	.2340249	03
.6407000	05	.7771385	01	.1162722	03-	.2328523	03
.6507000	05	.6726877	01	.1011542	03-	.2316797	03
.6607000	05	.5818753	01	.8794353	04-	.2305071	03
.6707000	05	.5030152	01	.7632637	04-	.2295962	03
.6807000	05	.4349217	01	.6561099	04-	.2309366	03
.6907000	05	.3763795	01	.5645184	04-	.2322769	03
.7007000	05	.3260031	01	.4861551	04-	.2336173	03
.7107000	05	.2826141	01	.4190465	04-	.2349577	03
.7207000	05	.2452096	01	.3615225	04-	.2362981	03
.7307000	05	.2127892	01	.3150047	04-	.2353372	03
.7407000	05	.1844533	01	.2763365	04-	.2325446	03
.7507000	05	.1596219	01	.2420425	04-	.2297519	03
.7607000	05	.1378956	01	.2116707	04-	.2269593	03
.7707000	05	.1189164	01	.1848115	04-	.2241667	03
.7807000	05	.1024067	01	.1602619	04-	.2226160	03

FIGURE 22

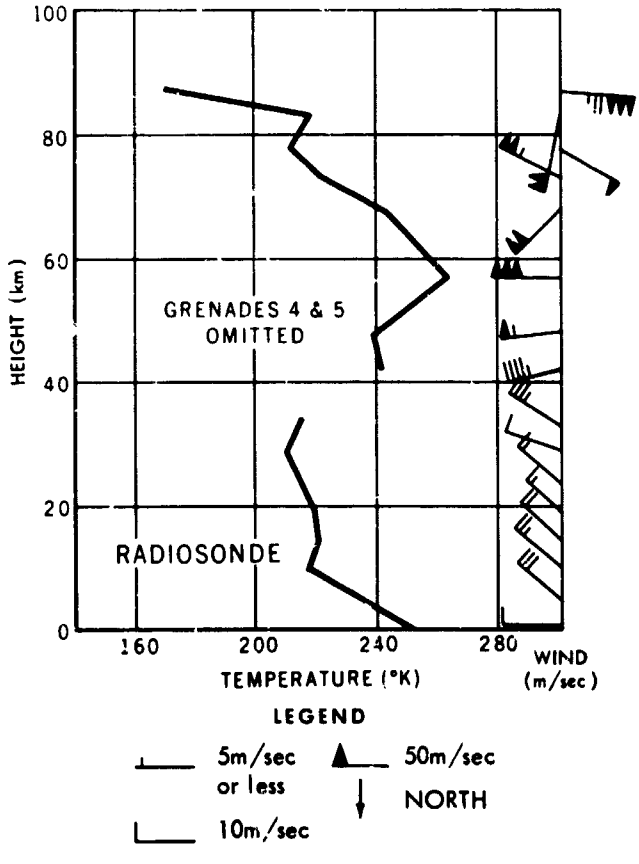
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
263.11258	80.682609	273.56659	42865.187
265.66339	84.451606	282.37439	47969.898
269.63528	91.402267	267.79379	52329.097
240.13299	37.619449	222.63978	56531.949
246.00149	78.274139	233.32418	61522.949
267.05758	102.42919	254.66459	66767.148
314.19027	30.184509	205.86940	71267.734
272.62970	97.967979	247.11109	75924.093
159.61169	68.393989	201.14839	80321.648
252.52330	60.476368	216.76379	84072.484



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2705400	05	.1700000	04	.2732550	01-	.2167400	03
.2805400	05	.1456068	04	.2302281	01-	.2203340	03
.2905400	05	.1250327	04	.1945241	01-	.2239281	03
.3005400	05	.1076315	04	.1648063	01-	.2275222	03
.3105400	05	.9287427	03	.1399984	01-	.2311163	03
.3205400	05	.8032656	03	.1192299	01-	.2347103	03
.3305400	05	.6963062	03	.1017950	01-	.2383044	03
.3405400	05	.6049082	03	.8711931	02-	.2418985	03
.3505400	05	.5266217	03	.7473405	02-	.2454926	03
.3605400	05	.4594110	03	.6425532	02-	.2490866	03
.3705400	05	.4015797	03	.5536767	02-	.2526807	03
3805400	05	.3517105	03	.4761207	02-	.2562748	03
.3905400	05	.3086161	03	.4137352	02-	.2598688	03
.4005400	05	.2712995	03	.3587465	02-	.2634629	03
.4105400	05	.2389214	03	.3116802	02-	.2670570	03
.4205400	05	.2107735	03	.2713091	02-	.2706511	03
.4305400	05	.1862418	03	.2368948	02-	.2738923	03
.4405400	05	.1647560	03	.2082535	02-	.2756177	03
.4505400	05	.1458660	03	.1832292	02-	.2773432	03
.4605400	05	.1292443	03	.1613461	02-	.2790686	03
.4705400	05	.1146063	03	.1421932	02-	.2807940	03
.4805400	05	.1016957	03	1255938	02-	.2820931	03
.4905400	05	.9020333	02	.1127375	02-	.2787483	03
.5005400	05	.7989696	02	.1010692	02-	.2754035	03
.5105400	05	.7066601	02	.9049115	03-	.2720587	03
.5205400	05	.6240913	02	.8091260	03-	.2687139	03
.5305400	05	.5496276	02	.7364510	03-	.2600057	03
.5405400	05	.4817213	02	.6732833	03-	.2492620	03
.5505400	05	.4197763	02	.6131324	03-	.2385183	03
.5605400	05	.3634995	02	.5559767	03-	.2277747	03
.5705400	05	.3133113	02	.4876170	03-	.2237573	03
.5805400	05	.2698985	02	.4162423	03-	.2258981	03
.5905400	05	.2328390	02	.3557175	03-	.2280388	03
.6005400	05	.2011547	02	.3044540	03-	.2301795	03
.6105400	05	.1740252	02	.2609656	03-	.2323203	03
.6205400	05	.1508088	02	.2231111	03-	.2354852	03
.6305400	05	.1309805	02	.1904848	03-	.2395545	03
.6405400	05	.1140345	02	.1630703	03-	.2436239	03
.6505400	05	.9951336	01	.1399669	03-	.2476932	03
.6605400	05	.8703788	01	.1204413	03-	.2517625	03
.6705400	05	.7620777	01	.1055420	03-	.2515545	03
.6805400	05	.6652808	01	.9628628	04-	.2407126	03
.6905400	05	.5771790	01	.8747527	04-	.2298706	03
.7005400	05	.4973417	01	.7910650	04-	.2190286	03
.7105400	05	.4253410	01	.7117747	04-	.2081867	03
.7205400	05	.3629598	01	.5941173	04-	.2128334	03
.7305400	05	.3112613	01	.4891435	04-	.2216905	03
.7405400	05	.2685535	01	.4058154	04-	.2305475	03
.7505400	05	.2330087	01	.3390765	04-	.2394046	03
.7605400	05	.2030804	01	.2876902	04-	.2457533	03
.7705400	05	.1767939	01	.2617586	04-	.2353014	03
.7805400	05	.1529503	01	.2369826	04-	.2248496	03
.7905400	05	.1314191	01	.2135485	04-	.2143977	03
.8005400	05	.1120713	01	.1914422	04-	.2039458	03
.8105400	05	.9520050	00	.1624230	04-	.2041973	03
.8205400	05	.8101413	00	.1354577	04-	.2083605	03
.8305400	05	.6916559	00	.1133812	04-	.2125236	03
.8405400	05	.5923419	00	.9523534	05-	.2166868	03

FIGURE 23

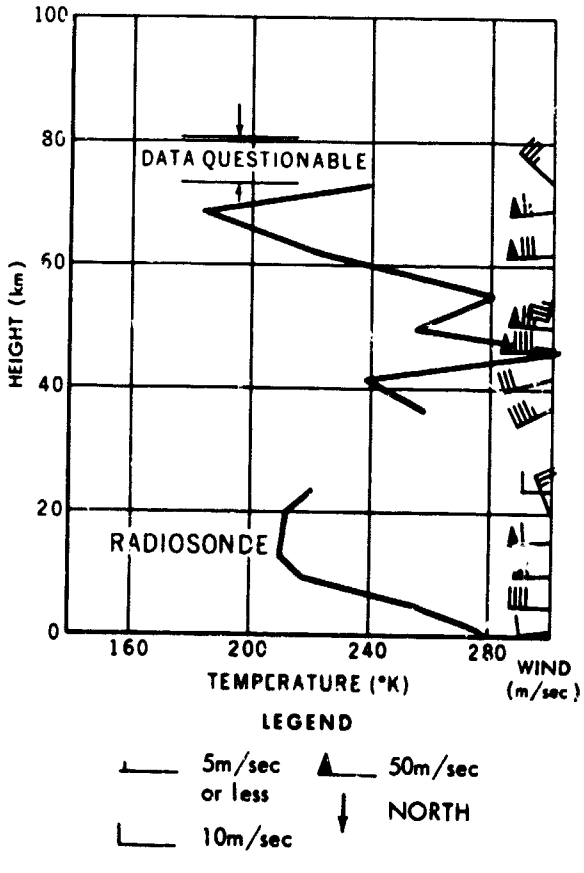
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
262.00698	43.808818	241.95939	42457.347
265.68618	54.826538	238.57479	48063.000
270.27807	148.54609	264.38278	57444.046
227.51558	100.64650	243.20858	67609.046
296.37478	103.47199	222.15139	73349.937
119.05209	51.014370	212.26008	78287.843
196.22400	98.051452	218.47559	83003.[illegible]
91.258529	172.84658	170.39279	87487.[illegible]



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meter		TEMPERATURE degrees Kelvin	
.3336000	05	.6200000	03	.9972211	02	.2166000	03
.3436000	05	.5308386	03	.8429634	02-	.2193875	03
.3536000	05	.4554130	03	.7141151	02-	.2221751	03
.3636000	05	.3914702	03	.6062425	02-	.2249626	03
.3736000	05	.3371486	03	.5157281	02-	.2277502	03
.3836000	05	.2909064	03	.4396118	02-	.2305378	03
.3936000	05	.2514635	03	.3754665	02-	.2333253	03
.4036000	05	.2177547	03	.3212965	02-	.2361129	03
.4136000	05	.1888917	03	.2754573	02-	.2389005	03
.4236000	05	.1641322	03	.2365903	02-	.2416880	03
.4336000	05	.1427284	03	.2059708	02-	.2444144	03
.4436000	05	.1240899	03	.1795226	02-	.2468106	03
.4536000	05	.1078519	03	.1564230	02-	.2492068	03
.4636000	05	.9370995	02	.1362547	02-	.2396030	03
.4736000	05	.8139702	02	.1186506	02-	.2389992	03
.4836000	05	.7070093	02	.1028902	02-	.2393918	03
.4936000	05	.6146928	02	.8843913	03-	.2421429	03
.5036000	05	.5352989	02	.7615113	03-	.2448940	03
.5136000	05	.4669002	02	.6568293	03-	.2476452	03
.5236000	05	.4076739	02	.5674876	03-	.2503962	03
.5336000	05	.3568511	02	.4911023	03-	.2531472	03
.5436000	05	.3126752	02	.4256810	03-	.2558983	03
.5536000	05	.2743666	02	.3695541	03-	.2586494	03
.5636000	05	.2410943	02	.3213207	03-	.2614005	03
.5736000	05	.2121523	02	.2798032	03-	.2641515	03
.5836000	05	.1867406	02	.2478616	03-	.2624748	03
.5936000	05	.1642286	02	.2197251	03-	.2603917	03
.6036000	05	.1442873	02	.1946020	03-	.2583087	03
.6136000	05	.1266306	02	.1721889	03-	.2562256	03
.6236000	05	.1110367	02	.1522114	03-	.2541426	03
.6336000	05	.9725489	01	.1344208	03-	.2520595	03
.6436000	05	.8509351	01	.1185920	03-	.2499765	03
.6536000	05	.7437282	01	.1045219	03-	.2478935	03
.6636000	05	.6493173	01	.9202692	04-	.2458104	03
.6736000	05	.5662602	01	.8094126	04-	.2437273	03
.6836000	05	.4931009	01	.7144335	04-	.2404541	03
.6936000	05	.4285489	01	.6305248	04-	.2367862	03
.7036000	05	.3716488	01	.5554111	04-	.2331183	03
.7136000	05	.3215905	01	.4882843	04-	.2294503	03
.7236000	05	.2776394	01	.4283998	04-	.2257824	03
.7336000	05	.2391312	01	.3750462	04-	.2221312	03
.7436000	05	.2055812	01	.3253614	04-	.2201281	03
.7536000	05	.1765026	01	.2819057	04-	.2181249	03
.7636000	05	.1513311	01	.2439427	04-	.2161218	03
.7736000	05	.1295699	01	.2108180	04-	.2141187	03
.7836000	05	.1107912	01	.1817609	04-	.2123552	03
.7936000	05	.9472337	00	.1544418	04-	.2136734	03
.8036000	05	.8106781	00	.1313666	04-	.2149916	03
.8136000	05	.6945027	00	.1118551	04-	.2163098	03
.8236000	05	.5955634	00	.9533916	05-	.2176281	03
.8336000	05	.5104408	00	.8284639	05-	.2146495	03
.8436000	05	.4352889	00	.7436367	05-	.2039271	03
.8536000	05	.3680404	00	.6636455	05-	.1932046	03
.8636000	05	.3082299	00	.5884540	05-	.1824822	03
.8736000	05	.2553959	00	.5180250	05-	.1717597	03
.8836000	05	.2100200	00	.4340887	05-	.1685545	03

FIGURE 24

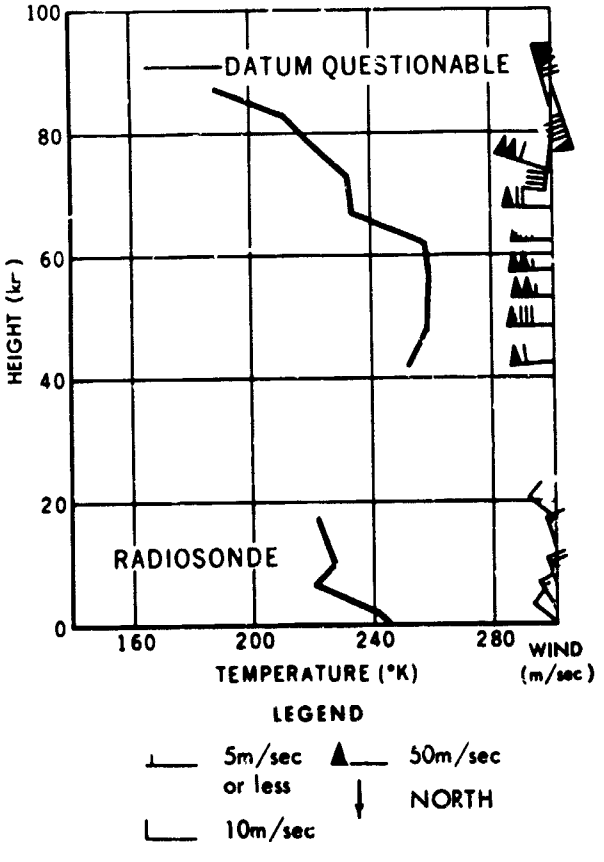
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
232.42788	45.897457	257.57629	37144.296
246.15689	30.189699	238.73149	42062.847
271.00738	89.221633	302.53408	46258.250
277.28488	75.514892	254.42720	50314.047
201.54100	36.999607	280.00747	55151.000
262.58689	77.385620	223.14878	62319.238
248.51348	65.737266	183.39270	69056.632
313.18399	34.554256	239.66099	73325.132
159.90309	280.56689	101.17469	76917.984
22.31157	446.54379	130.91799	80274.734



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2438370	05	.2600000	04	.4115021	01-	.2201200	03
.2538370	05	.2231400	04	.3485163	01-	.2230553	03
.2638370	05	.1918978	04	.2956270	01-	.2259906	03
.2738370	05	.1653592	04	.2516469	01-	.2289259	03
.2838370	05	.1427677	04	.2145162	01-	.2318612	03
.2938370	05	.1234964	04	.1832402	01-	.2347965	03
.3038370	05	.1070237	04	.1568378	01-	.2377318	03
.3138370	05	.9091546	03	.1345002	01-	.240667	03
.3238370	05	.8080890	03	.1155675	01-	.24360[illegible]	03
.3338370	05	.7040042	03	.9948325	02-	.2465378	03
.3438370	05	.6143538	0	.8579324	02-	.2494731	03
.3538370	05	.5369974	03	.7411849	02-	.2524084	03
.3638370	05	.4701317	03	.6414349	02-	.2553437	03
.3738370	05	.4100641	03	.5593278	02-	.2566590	03
.3838370	05	.3609474	03	.4973676	02-	.2528277	03
.3938370	05	.3155462	03	.4314975	02-	.2489963	03
.4038370	05	.2752930	03	.3911965	02-	.2451649	03
.4138370	05	.2396693	03	.3459816	02-	.2413335	03
.4238370	05	.2085724	03	.2982760	02-	.2436109	03
.4338370	05	.1823906	03	.2455077	02-	.2586186	03
.4438370	05	.1607282	03	.2043421	02-	.2740264	03
.4538370	05	.1426148	03	.1717803	02-	.2890341	03
.4638370	05	.1272435	03	.1472516	02-	.3040461	03
.4738370	05	.1135315	03	.1367726	02-	.2891848	03
.4838370	05	.1008184	03	.1266517	02-	.2773235	03
.4938370	05	.8906867	02	.1168908	02-	.2654623	03
.5038370	05	.7827057	02	.1070200	02-	.2547055	03
.5138370	05	.6869251	02	.9001398	03-	.2600840	03
.5238370	05	.6044756	02	.7935621	03-	.2653725	03
.5338370	05	.5332869	02	.6864254	03-	.2706610	03
.5438370	05	.4716423	02	.5954444	03-	.2759496	03
.5538370	05	.4178348	02	.5233178	03-	.2781617	03
.5638370	05	.3697095	02	.4766349	03-	.2702296	03
.5738370	05	.3259494	02	.4329264	03-	.2622976	03
.5838370	05	.2862708	02	.3920820	03-	.2543656	03
.5938370	05	.2504010	02	.3539028	03-	.2464335	03
.6038370	05	.2180777	02	.3185506	03-	.2385015	03
.6138370	05	.1890495	02	.2856484	03-	.2305695	03
.6238370	05	.1630821	02	.2550416	03-	.2227684	03
.6338370	05	.1400427	02	.2249697	03-	.2168676	03
.6438370	05	.1197598	02	.1977676	03-	.2109668	03
.6538370	05	.1019659	02	.1732286	03-	.2050660	03
.6638369	05	.8641358	01	.1511565	03-	.1991652	03
.6738369	05	.7287357	01	.1313640	03-	.19326[illegible]	03
.6838369	05	.6113454	01	.1136736	03-	.1873636	03
.6938369	05	.5115676	01	.9494836	04-	.1877041	03
.7038369	05	.4307313	01	.7469892	04-	.2008863	03
.7138369	05	.3666746	01	.5967414	04-	.2140685	03
.7238369	05	.3151777	01	.4831794	04-	.2272507	03
.7338369	05	.2729977	01	.4006173	04-	.2374037	03
.7438369	05	.2341850	01	.3402727	04-	.1988587	03
.7538369	05	.1943520	01	.4223537	04-	.1603138	03
.7638369	05	.1532051	01	.4383239	04-	.1017688	03
.7738369	05	.1141682	01	.3777198	04-	.1053012	03
.7838369	05	.8425408	00	.2571152	04-	.1141620	03
.7938369	05	.6361263	00	.1801425	04-	.1230227	03
.8038369	05	.4887637	00	.1312008	04-	.1297838	03

FIGURE 25

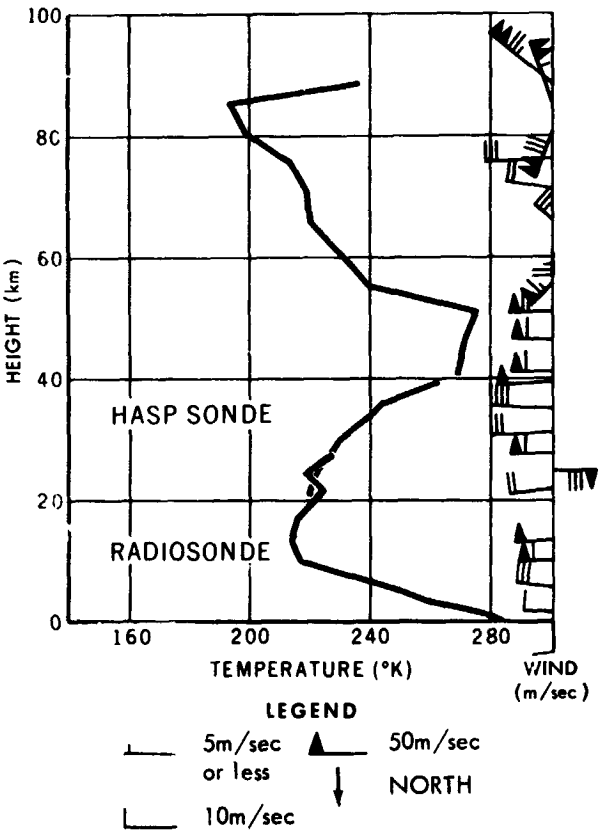
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
260.64889	61.721878	250.75329	42317.437
268.32989	79.007347	257.60409	47971.398
270.38507	104.85550	257.69129	52821.547
262.97647	104.81910	258.43267	57512.750
265.70349	68.107749	256.77227	62055.097
273.74368	70.241676	253.70109	67513.382
306.80477	111.20809	231.48858	73290.437
187.57319	40.743816	220.89549	78273.937
346.53317	171.06559	210.96269	83051.187
172.36470	36.968605	188.12139	87622.734
342.01950	306.11908	225.62030	91531.937



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.3336000	05	.6200000	03	.0972211	02-	.2146000	03
.3436000	05	.5310311	03	.8395465	02-	.2304128	03
.3536000	05	.4560605	03	.7085906	02-	.2262256	03
.3636000	05	.3926991	03	.5999428	02-	.2230385	03
.3736000	05	.3389957	03	.5092011	02-	.2318513	03
.3836000	05	.2933526	03	.4336652	02-	.2356642	03
.3936000	05	.2544562	03	.3701752	02-	.2394770	03
.4036000	05	.2212232	03	.3167852	02-	.2432899	03
.4136000	05	.1927580	03	.2717647	02-	.2471027	03
.4236000	05	.1683120	03	.2337962	02-	.2509040	03
.4336000	05	.1471677	03	.2034425	02-	.2540165	03
.4436000	05	.1287680	03	.1771553	02-	.2532282	03
.4535000	05	.1127452	03	.1543729	02-	.2544399	03
.4636000	05	.9878261	02	.1346140	02-	.2556516	03
.4736000	05	.8660679	02	.1174649	02-	.2568632	03
.4836000	05	.7597292	02	.1027431	02-	.2576110	03
.4936000	05	.6656474	02	.9014872	03-	.2576290	03
.5036000	05	.5849947	02	.7910156	03-	.2576470	03
.5136000	05	.5133638	02	.6941098	03-	.2576649	03
.5236000	05	.4505216	02	.6090994	03-	.2576829	03
.5336000	05	.3953874	02	.5343651	03-	.2577763	03
.5436000	05	.3470334	02	.4687274	03-	.2579343	03
.5536000	05	.3046302	02	.4112027	03-	.2580924	03
.5636000	05	.2674410	02	.3607822	03-	.2582504	03
.5736000	05	.2348207	02	.3165832	03-	.2584085	03
.5836000	05	.2061784	02	.2782753	03-	.2581229	03
.5936000	05	.1810079	02	.2446496	03-	.2577574	03
.6036000	05	.1588371	02	.2150563	03-	.2573919	03
.6136000	05	.1394494	02	.1890154	03-	.2570263	03
.6236000	05	.1223354	02	.1668198	03-	.2554835	03
.6336000	05	.1071660	02	.1485928	03-	.2512567	03
.6436000	05	.9367090	01	.1321032	03-	.2470298	03
.6536000	05	.8169868	01	.1172103	03-	.2428030	03
.6636000	05	.7107123	01	.1037827	03-	.2385762	03
.6736000	05	.6168278	01	.9169767	04-	.2343494	03
.6836000	05	.5345280	01	.7979412	04-	.2333708	03
.6936000	05	.4630366	01	.6923283	04-	.2329939	03
.7036000	05	.4010297	01	.6006269	04-	.2326169	03
.7136000	05	.3472536	01	.5205724	04-	.2322279	03
.7236000	05	.3006411	01	.4517013	04-	.2318449	03
.7336000	05	.2602208	01	.3918757	04-	.2313460	03
.7436000	05	.2250595	01	.3402673	04-	.2292151	03
.7536000	05	.1943949	01	.2982235	04-	.2270494	03
.7636000	05	.1676850	01	.2596807	04-	.2249648	03
.7736000	05	.1444488	01	.2258305	04-	.2228382	03
.7836000	05	.1242607	01	.1961360	04-	.2207165	03
.7936000	05	.1067454	01	.1700916	04-	.2186373	03
.8036000	05	.9157022	00	.1473119	04-	.2165582	03
.8136000	05	.7844003	00	.1274123	04-	.2144790	03
.8236000	05	.6709457	00	.1100504	04-	.2123998	03
.8336000	05	.5728540	00	.0529817	05-	.2094137	03
.8436000	05	.4876373	00	.8310452	05-	.2044233	03
.8536000	05	.4134672	00	.7222964	05-	.1994269	03
.8636000	05	.3491320	00	.6255807	05-	.1944305	03
.8736000	05	.2935277	00	.5398202	05-	.1894341	03
.8836000	05	.2468767	00	.4646287	05-	.1951936	03
.8936000	05	.2090307	00	.3556051	05-	.2047860	03
.9036000	05	.1784488	00	.2898326	05-	.2143785	03
.9136000	05	.1532389	00	.2383612	05-	.2239710	03

FIGURE 26

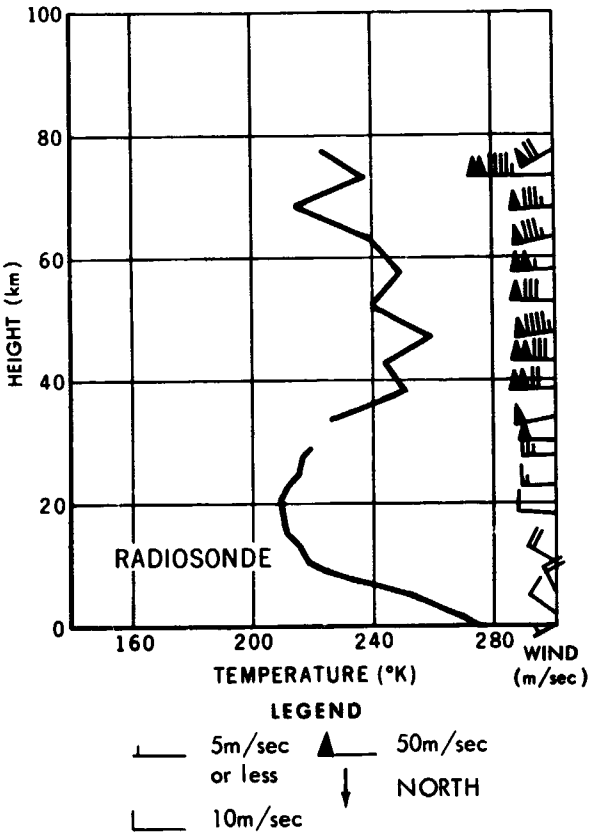
WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
272.15747	61.452530	268.70599	41274.347
274.21068	62.321479	271.26879	46610.097
261.49487	65.586547	275.41629	51168.648
219.41980	64.513748	240.00009	55569.847
180.40618	26.251798	231.82708	60863.097
314.42779	24.592708	220.14679	66444.593
290.04797	20.815719	219.13139	71234.046
265.47329	31.611558	213.91780	76220.437
206.19059	129.47969	199.31849	80953.187
338.90649	103.13029	194.28079	85013.687
302.24038	123.67859	236.12199	88816.648



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.2861200	05	.1400000	04	.2144656	01-	.2274200	03
.2961200	05	.1307758	04	.1824010	01-	.2306805	03
.3061200	05	.1044124	04	.1554905	01-	.2339410	03
.3161200	05	.9045213	03	.1328493	01-	.2372016	03
.3261200	05	.7851555	03	.1137541	01-	.2404621	03
.3361200	05	.6828720	03	.9761163	02-	.2437226	03
.3461200	05	.5950414	03	.8393401	02-	.2469832	03
.3561200	05	.5164665	03	.7231902	02-	.2502437	03
.3661200	05	.4543074	03	.6243422	02-	.2535042	03
.3761200	05	.3980190	03	.5400408	02-	.2567647	03
.3861200	05	.3493015	03	.4679969	02-	.2600253	03
.3961200	05	.3070584	03	.4063046	02-	.2632858	03
.4061200	05	.2703634	03	.3533730	02-	.2665464	03
.4161200	05	.2383794	03	.3088784	02-	.2688681	03
.4261200	05	.2103254	03	.2720417	02-	.2693484	03
.4361200	05	.1856217	03	.2396617	02-	.2698287	03
.4461200	05	.1638622	03	.2111914	02-	.2703091	03
.4561200	05	.1446910	03	.1861522	02-	.2707894	03
.4661200	05	.1277962	03	.1641245	02-	.2712765	03
.4761200	05	.1129141	03	.1445272	02-	.2721803	03
.4861200	05	.9981012	02	.1273288	02-	.2730502	03
.4961200	05	.8826655	02	.1122286	02-	.2740000	03
.5061200	05	.7809284	02	.9896443	03-	.2749008	03
.5161200	05	.6906091	02	.8850407	03-	.2758486	03
.5261200	05	.6092001	02	.8045269	03-	.2638017	03
.5361200	05	.5353250	02	.7292089	03-	.2557548	03
.5461200	05	.4684877	02	.6588956	03-	.2477078	03
.5561200	05	.4082426	02	.5927655	03-	.2399350	03
.5661200	05	.3548117	02	.5185210	03-	.2383909	03
.5761200	05	.3081064	02	.4532013	03-	.2368469	03
.5861200	05	.2673140	02	.3957789	03-	.2353029	03
.5961200	05	.2317159	02	.3453393	03-	.2337588	03
.6061200	05	.2006772	02	.3010693	03-	.2322148	03
.6161200	05	.1736150	02	.2626802	03-	.2302599	03
.6261200	05	.1500172	02	.2290583	03-	.2281672	03
.6361200	05	.1294582	02	.1984970	03-	.2260745	03
.6461200	05	.1115689	02	.1735357	03-	.2239818	03
.6561200	05	.9602200	01	.1507624	03-	.2218891	03
.6661200	05	.8253689	01	.1306363	03-	.2201113	03
.6761200	05	.7090005	01	.1123262	03-	.2198993	03
.6861200	05	.6089808	01	.9657327	04-	.2196873	03
.6961200	05	.5230171	01	.8302110	04-	.2194753	03
.7061200	05	.4491438	01	.7136376	04-	.2192632	03
.7161200	05	.3856251	01	.6141901	04-	.2187362	03
.7261200	05	.3309232	01	.5295972	04-	.2176906	03
.7361200	05	.2837854	01	.4563514	04-	.2166451	03
.7461200	05	.2431929	01	.3929718	04-	.2155995	03
.7561200	05	.2082603	01	.3381647	04-	.2145539	03
.7661200	05	.1781671	01	.2918084	04-	.2127099	03
.7761200	05	.1521521	01	.2528673	04-	.2096252	03
.7861200	05	.1296384	01	.2186688	04-	.2065404	03
.7961200	05	.1101958	01	.1886919	04-	.2034557	03
.8061200	05	.9344162	00	.1624665	04-	.2003709	03
.8161200	05	.7907683	00	.1387856	04-	.1985011	03
.8261200	05	.6683612	00	.1180401	04-	.1972605	03
.8361200	05	.5643328	00	.1002983	04-	.1960198	03
.8461200	05	.4760093	00	.8513952	05-	.1947791	03
.8561200	05	.4023631	00	.6978709	05-	.2008636	03
.8661200	05	.3424943	00	.5631844	05-	.2118658	03
.8761200	05	.2939355	00	.4594754	05-	.2228681	03
.8861200	05	.2541386	00	.3785765	05-	.2338704	03

FIGURE 27

WIND DIRECTION degrees	WIND SPEED m/sec	TEMPERATURE degrees Kelvin	HEIGHT meters
251.04699	49.156208	225.17968	34084.500
266.78720	120.79139	250.57888	38644.847
274.32379	130.62210	243.76129	43035.796
260.92868	95.027588	258.45147	47155.898
274.01498	79.796524	239.92878	52804.097
265.85369	104.42839	248.69580	58055.597
247.35418	84.380134	239.13909	63433.289
262.06008	85.245559	215.70478	68496.437
271.16949	147.49369	236.68159	73204.734
231.80848	67.915214	223.47479	77866.046



HEIGHT meters		PRESSURE nt/sq meters		DENSITY kg/cu meters		TEMPERATURE degrees Kelvin	
.3371160	05	.6000000	03	.8953968	02-	.2334500	03
.3471160	05	.5183956	03	.7897794	02-	.2286723	03
.3571160	05	.4480216	03	.6663348	02-	.2342419	03
.3671160	05	.3885483	03	.5644600	02-	.2398115	03
.3771160	05	.3380882	03	.4800066	02-	.2453811	03
.3871160	05	.2950756	03	.4104183	02-	.2504752	03
.3971160	05	.2577956	03	.3608024	02-	.2489226	03
.4071160	05	.2250448	03	.3169423	02-	.2473699	03
.4171160	05	.1962951	03	.2781987	02-	.2456173	03
.4271160	05	.1710773	03	.2440001	02-	.2442646	03
.4371160	05	.1491077	03	.2112820	02-	.2458645	03
.4471160	05	.1301350	03	.1820931	02-	.2489768	03
.4571160	05	.1137734	03	.1572334	02-	.2520891	03
.4671160	05	.9965715	02	.1360180	02-	.2552013	03
.4771160	05	.8740130	02	.1178769	02-	.2583136	03
.4871160	05	.7666611	02	.1047648	02-	.2549448	03
.4971160	05	.6712982	02	.9307289	03-	.2512757	03
.5071160	05	.5866743	02	.8254546	03-	.2476065	03
.5171160	05	.5117097	02	.7308085	03-	.2439373	03
.5271160	05	.4454194	02	.6458490	03-	.2402682	03
.5371160	05	.3874565	02	.5590685	03-	.2414438	03
.5471160	05	.3373267	02	.4833930	03-	.2431132	03
.5571160	05	.2939739	02	.4183948	03-	.2447826	03
.5671160	05	.2564433	02	.3625076	03-	.2464521	03
.5771160	05	.2239200	02	.3144030	03-	.2481215	03
.5871160	05	.1955870	02	.2752773	03-	.2475300	03
.5971160	05	.1707354	02	.2420378	03-	.2457529	03
.6071160	05	.1489009	02	.2126223	03-	.2439758	03
.6171160	05	.1297344	02	.1866129	03-	.2421987	03
.6271160	05	.1129253	02	.1636349	03-	.2404216	03
.6371160	05	.9817432	01	.1437974	03-	.2378510	03
.6471160	05	.8517133	01	.1272275	03-	.2332226	03
.6571159	05	.7368373	01	.1122961	03-	.2285942	03
.6671159	05	.6355982	01	.9886879	04-	.2239657	03
.6771159	05	.5466049	01	.8681987	04-	.2193373	03
.6871159	05	.4689100	01	.7539841	04-	.2166633	03
.6971159	05	.4025276	01	.6342033	04-	.2211186	03
.7071159	05	.3466127	01	.5353203	04-	.2255739	03
.7171159	05	.2993529	01	.4533761	04-	.2300292	03
.7271159	05	.2592765	01	.3852185	04-	.2344845	03
.7371159	05	.2249351	01	.3331150	04-	.2352455	03
.7471159	05	.1950275	01	.2923446	04-	.2324122	03
.7571159	05	.1688084	01	.2561652	04-	.2295739	03
.7671159	05	.1458590	01	.2241054	04-	.2267457	03
.7771159	05	.1258039	01	.1957375	04-	.2239124	03

FIGURE 28

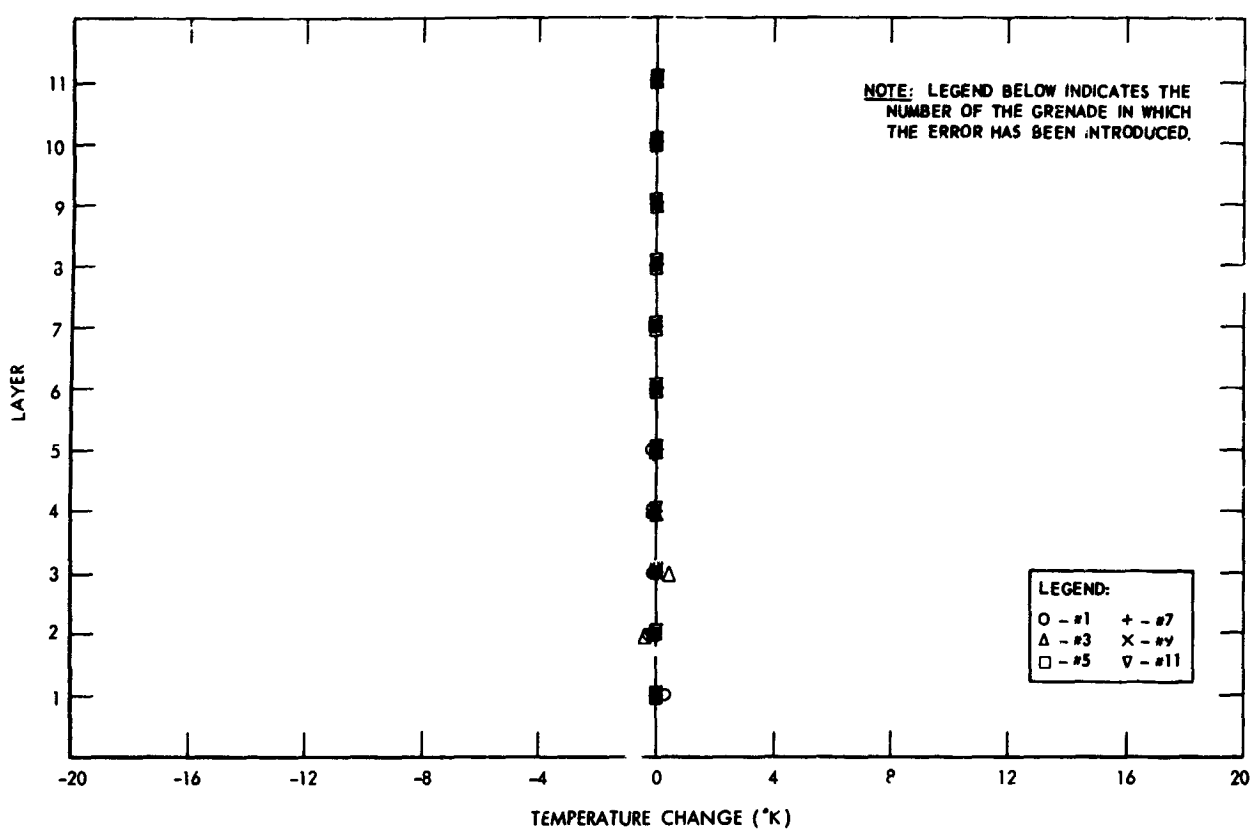


FIGURE 29

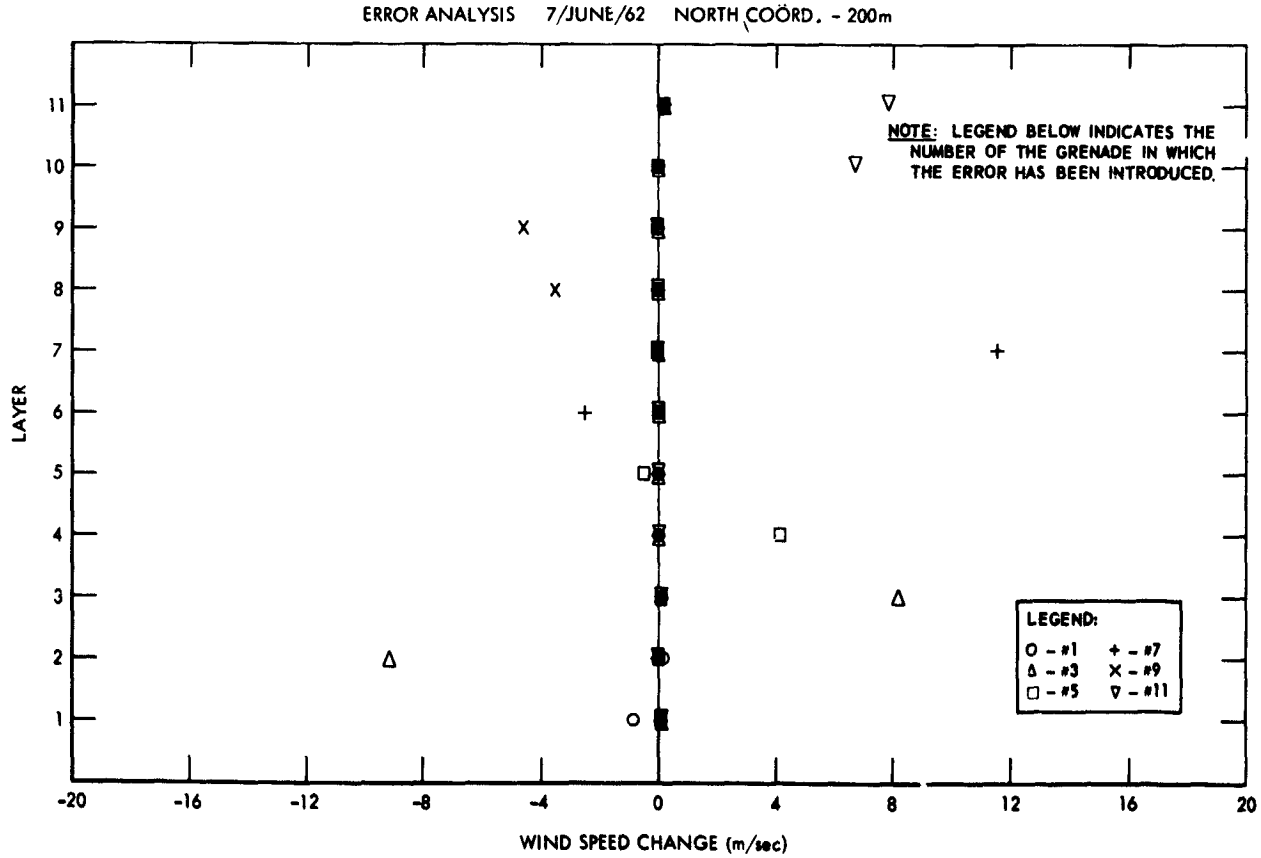


FIGURE 30

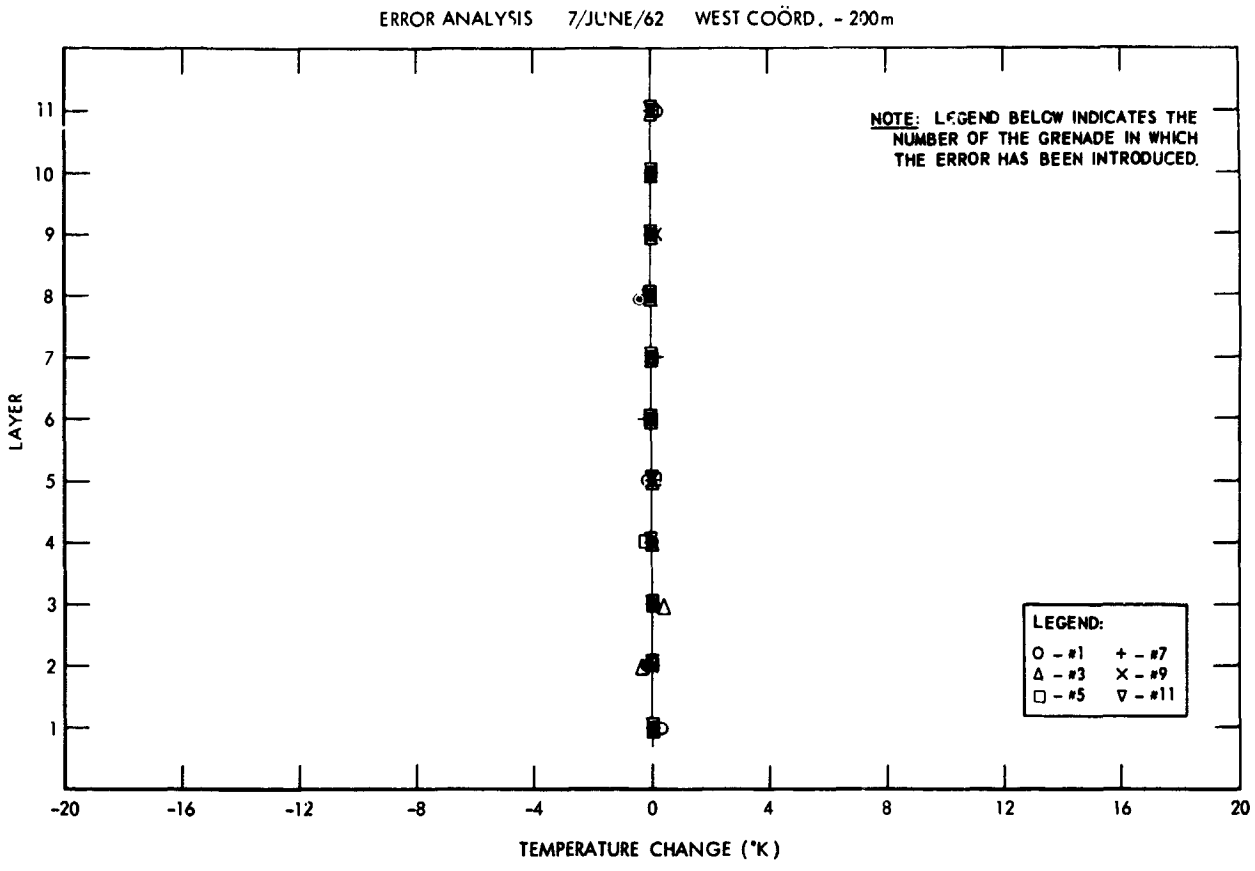


FIGURE 31

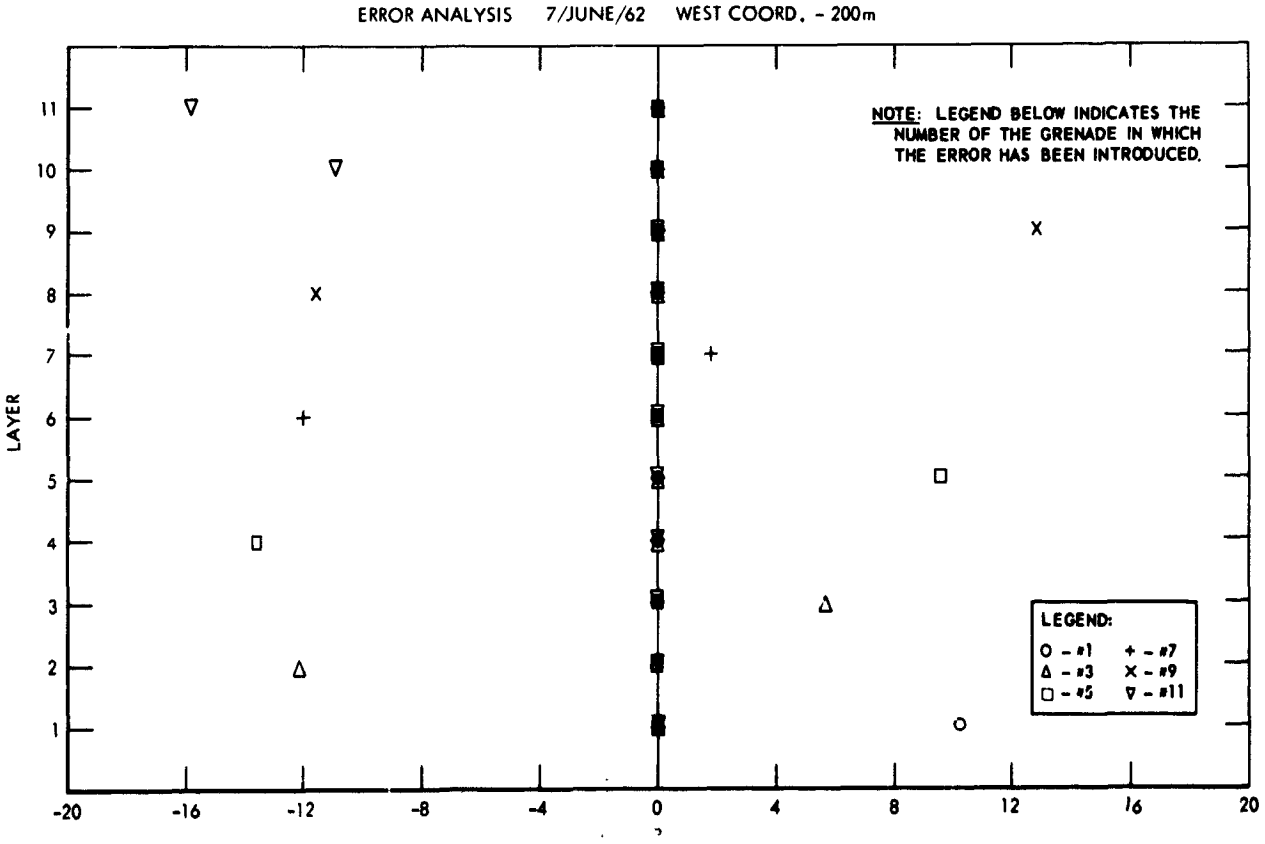


FIGURE 32

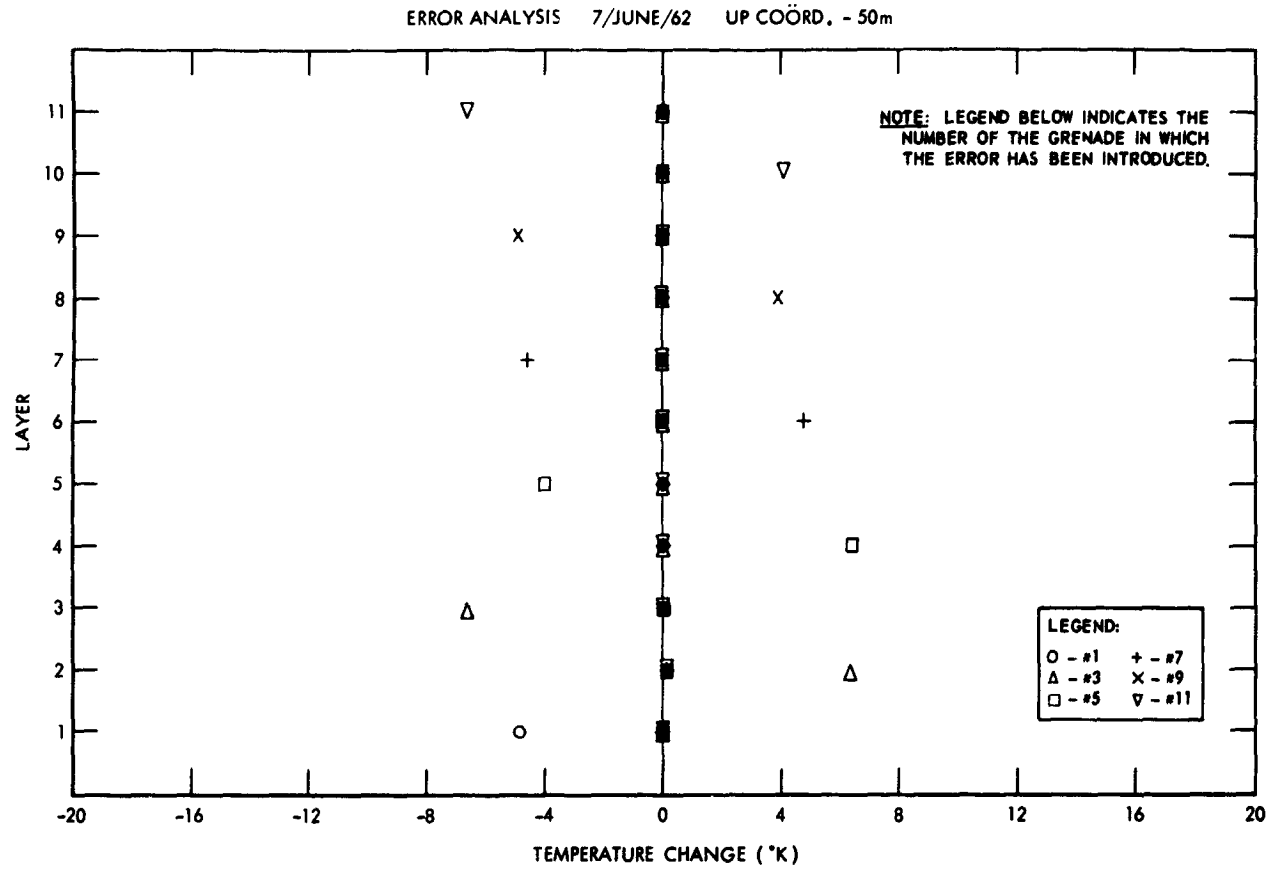


FIGURE 33

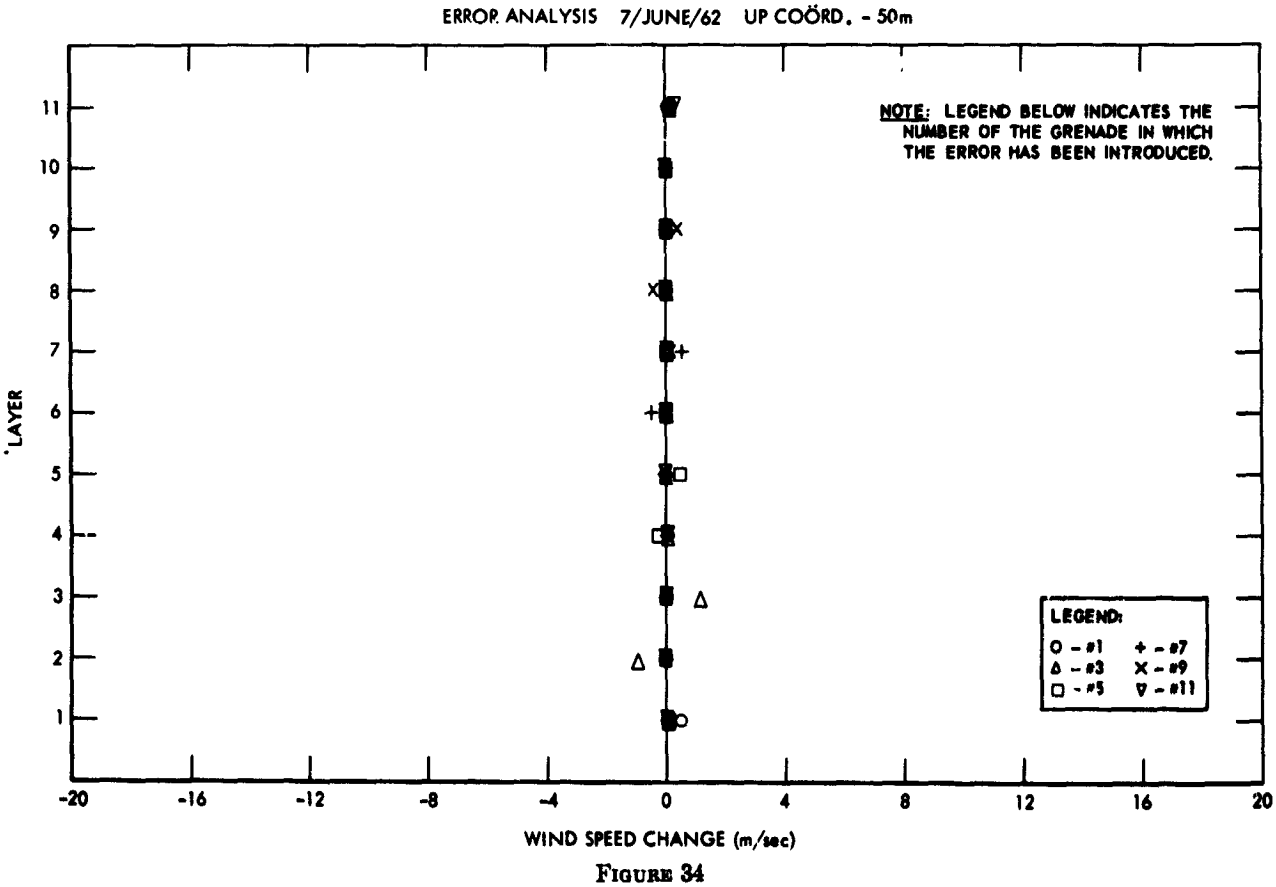


FIGURE 34

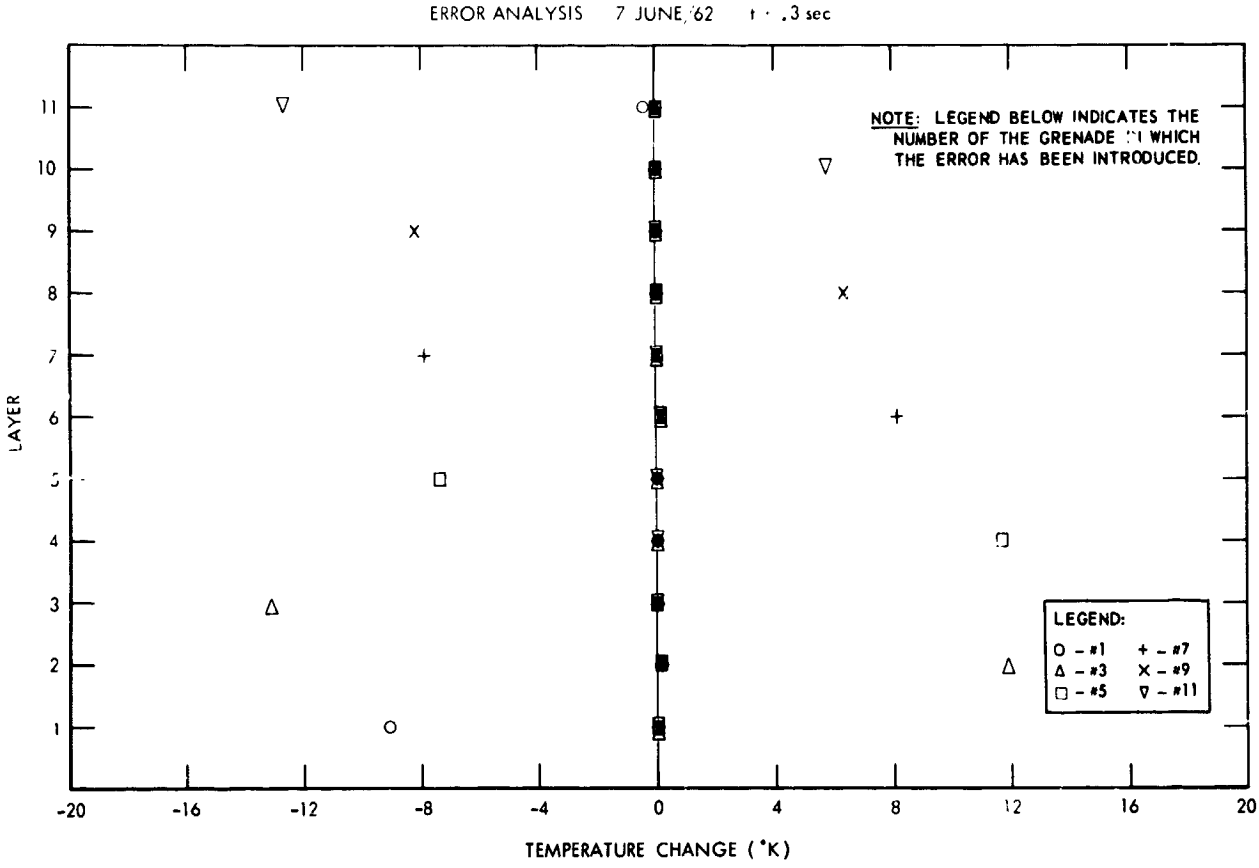


FIGURE 35

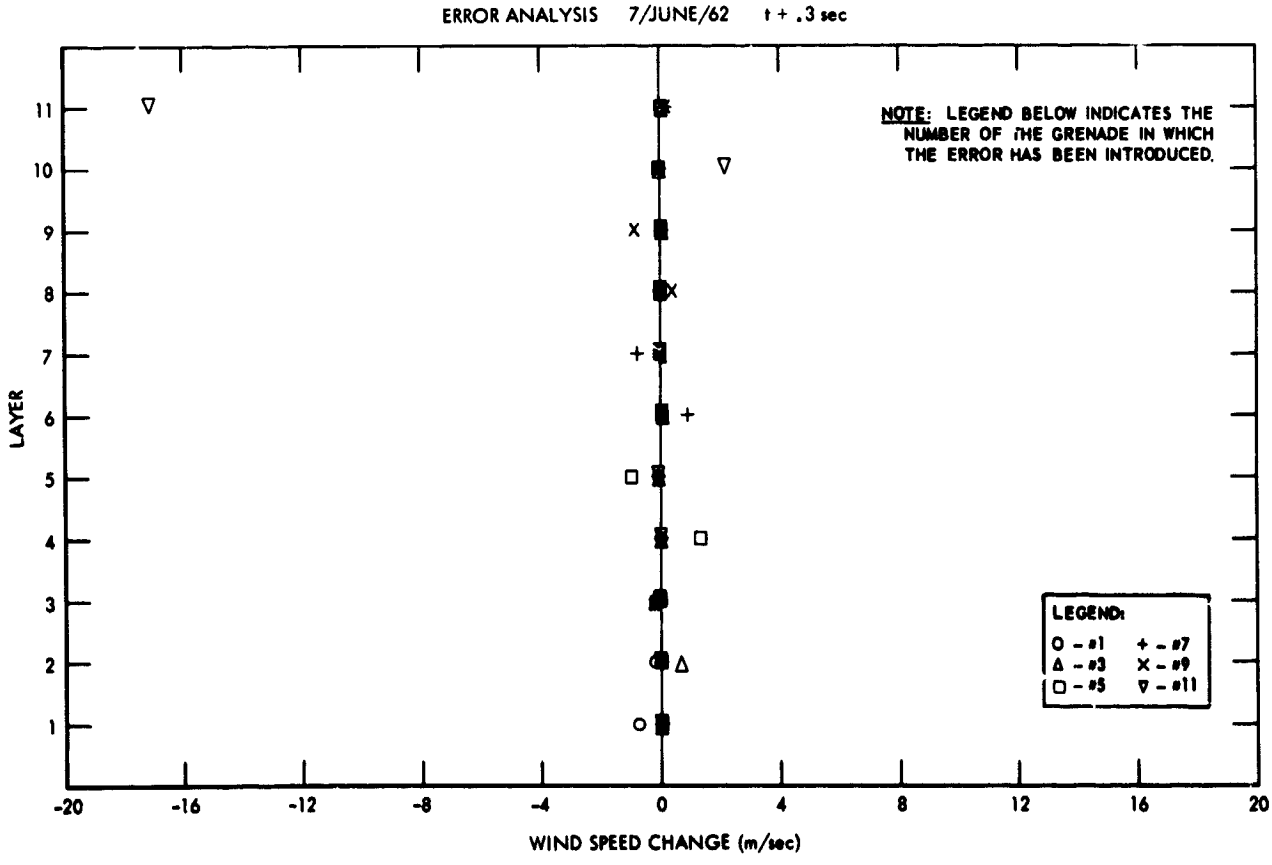


FIGURE 36

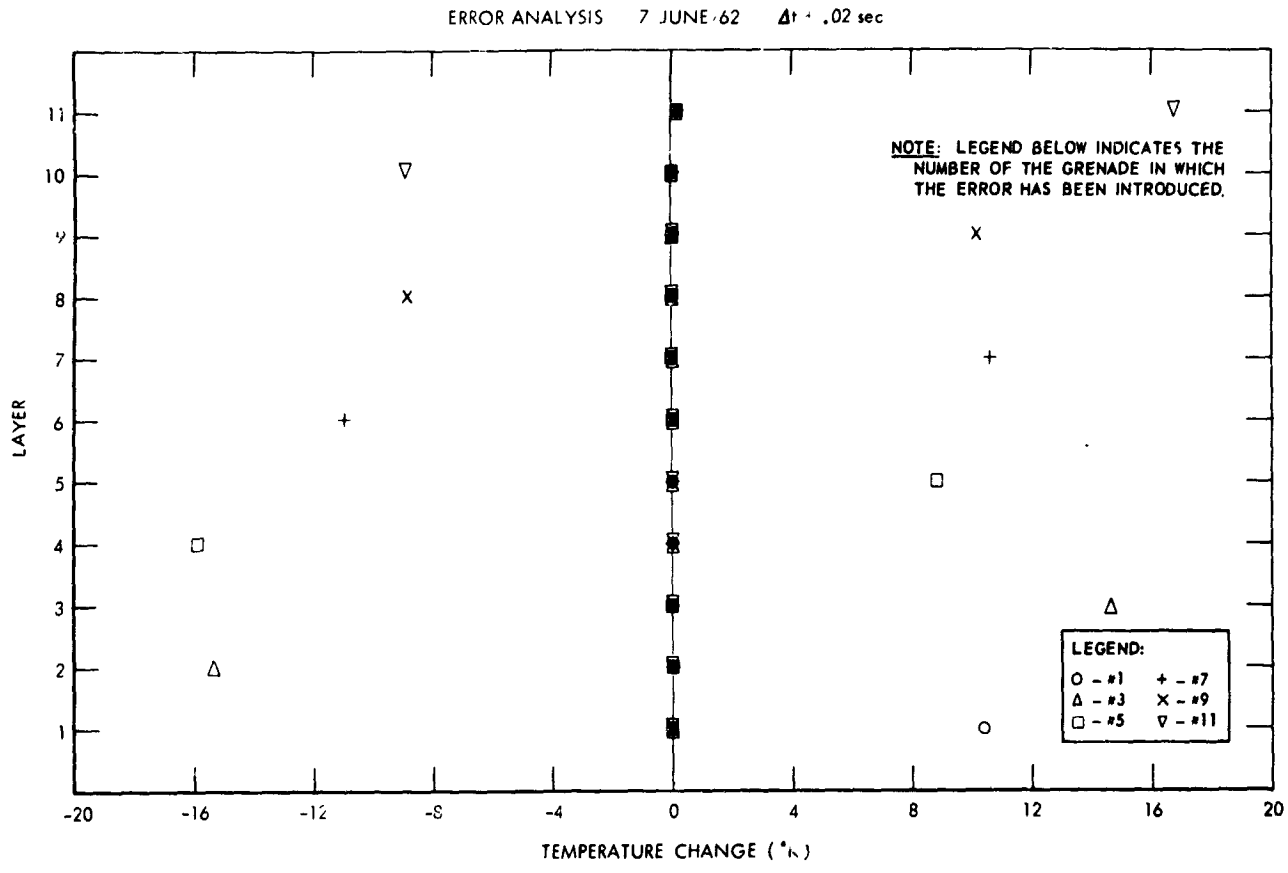


FIGURE 37

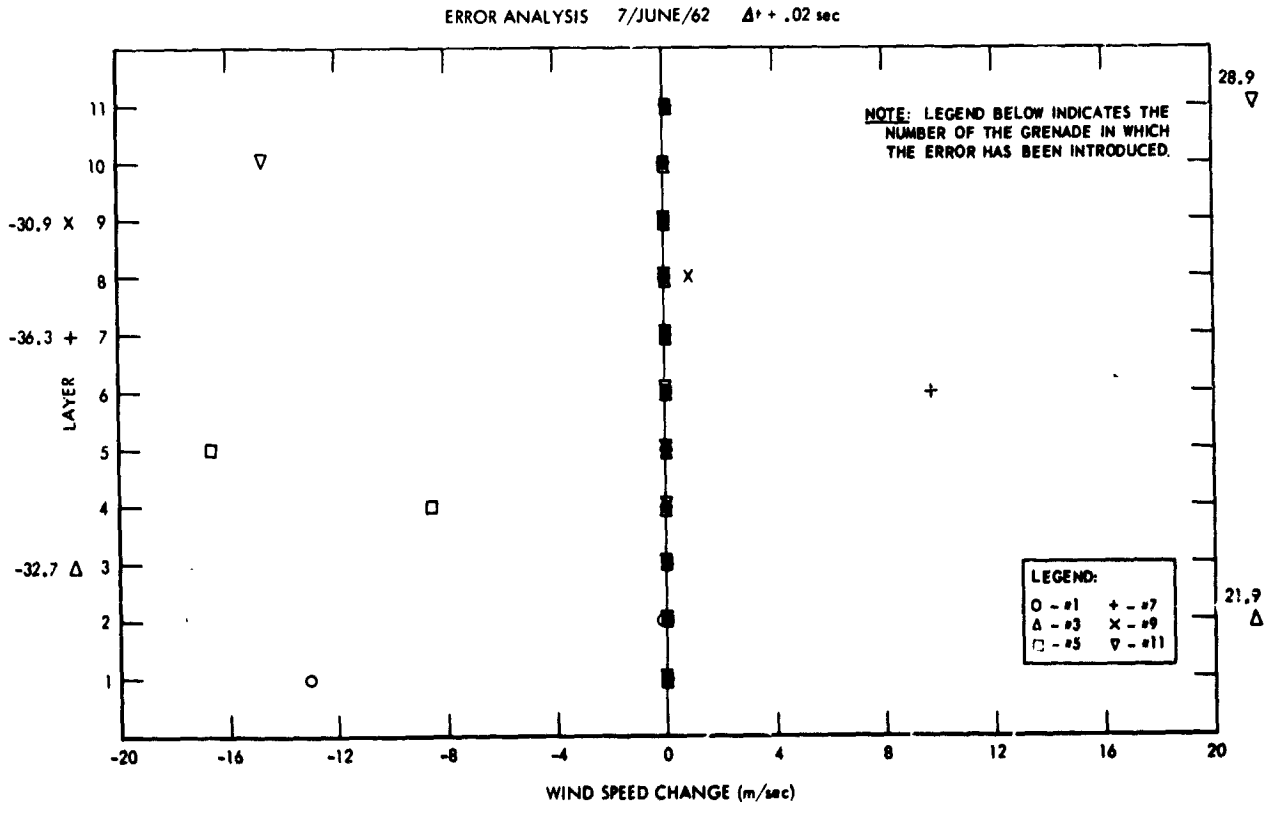


FIGURE 38

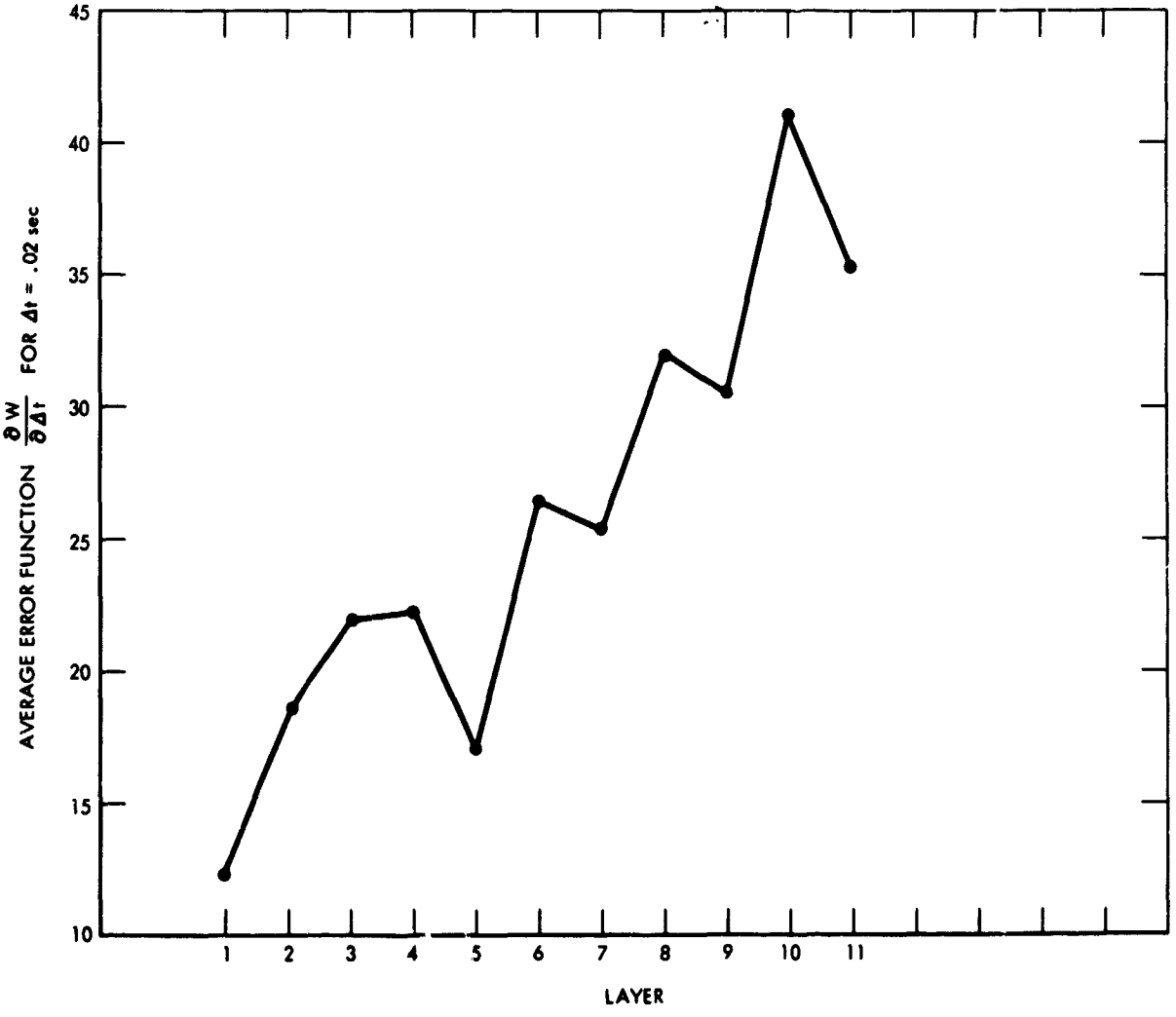


FIGURE 39.—Graph of $\frac{\partial T}{\partial \Delta t}$ vs. layer where $\Delta t = .02$ sec. (average of 16 firings).

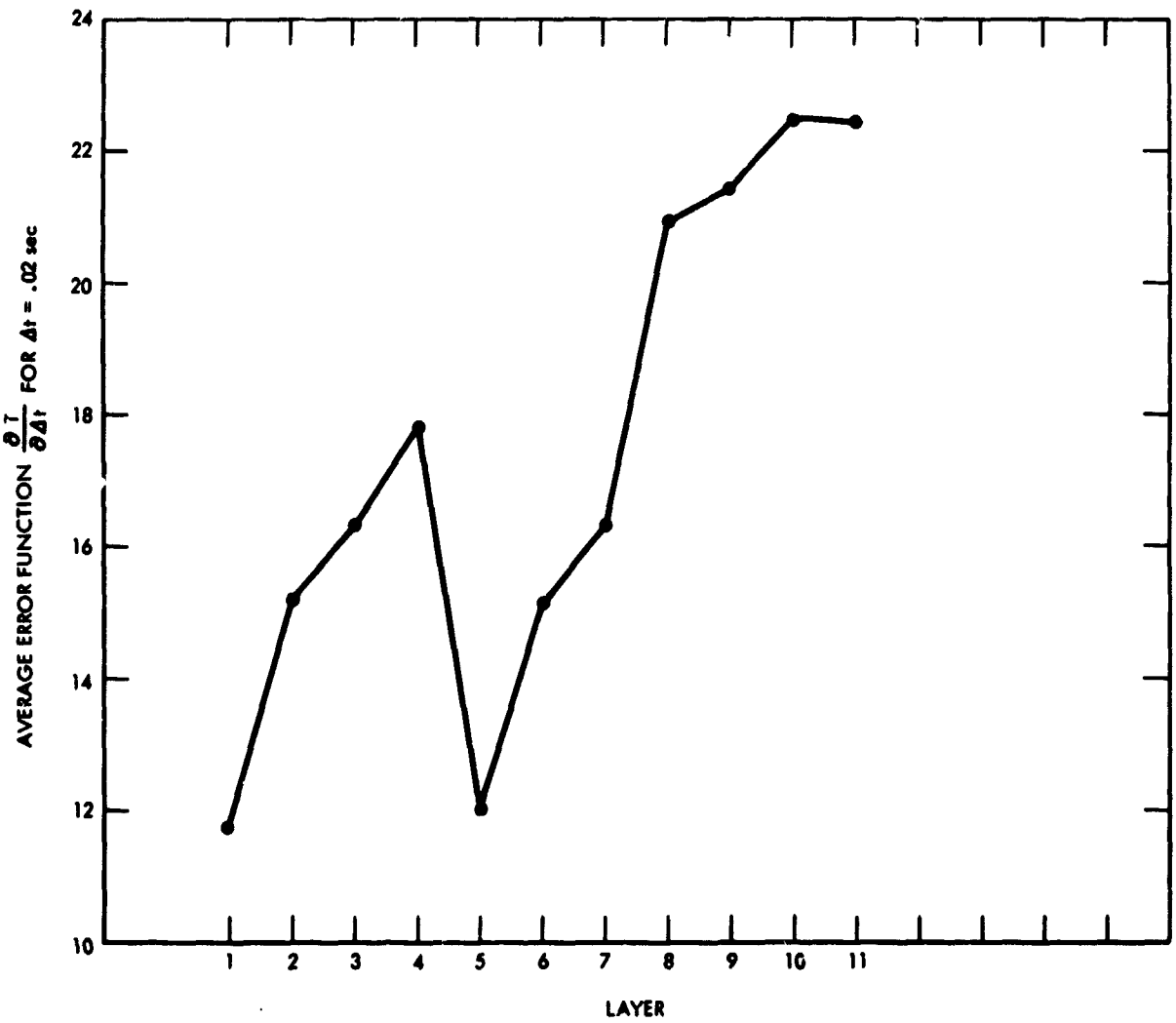


FIGURE 40.—Graph of $\frac{\partial W}{\partial \Delta t}$ vs. layer where $\Delta t = .02$ sec. (average of 16 firings).

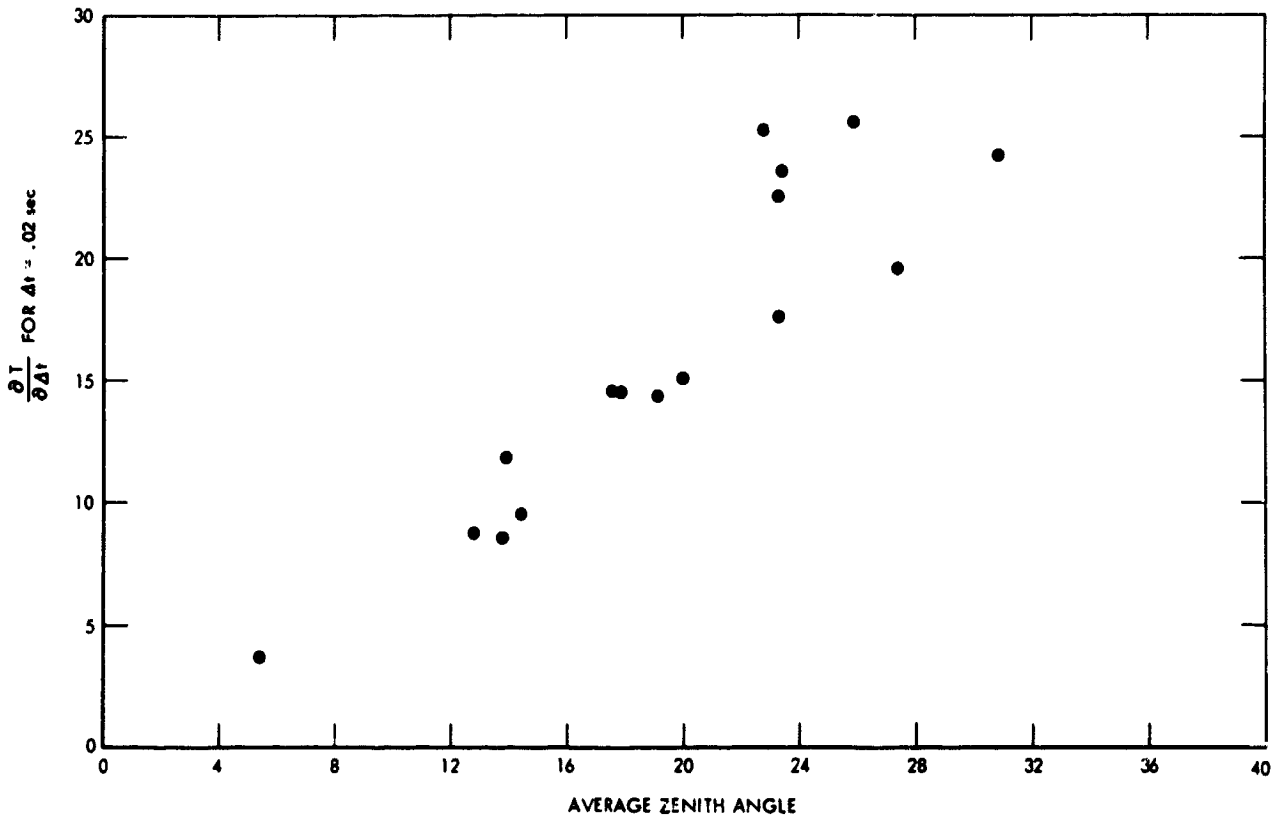


FIGURE 41.—Dependence of $\frac{\partial T}{\partial \Delta t}$ upon zenith angle, where $\Delta t = .02$ sec. for the 16 firings 1960–1963.

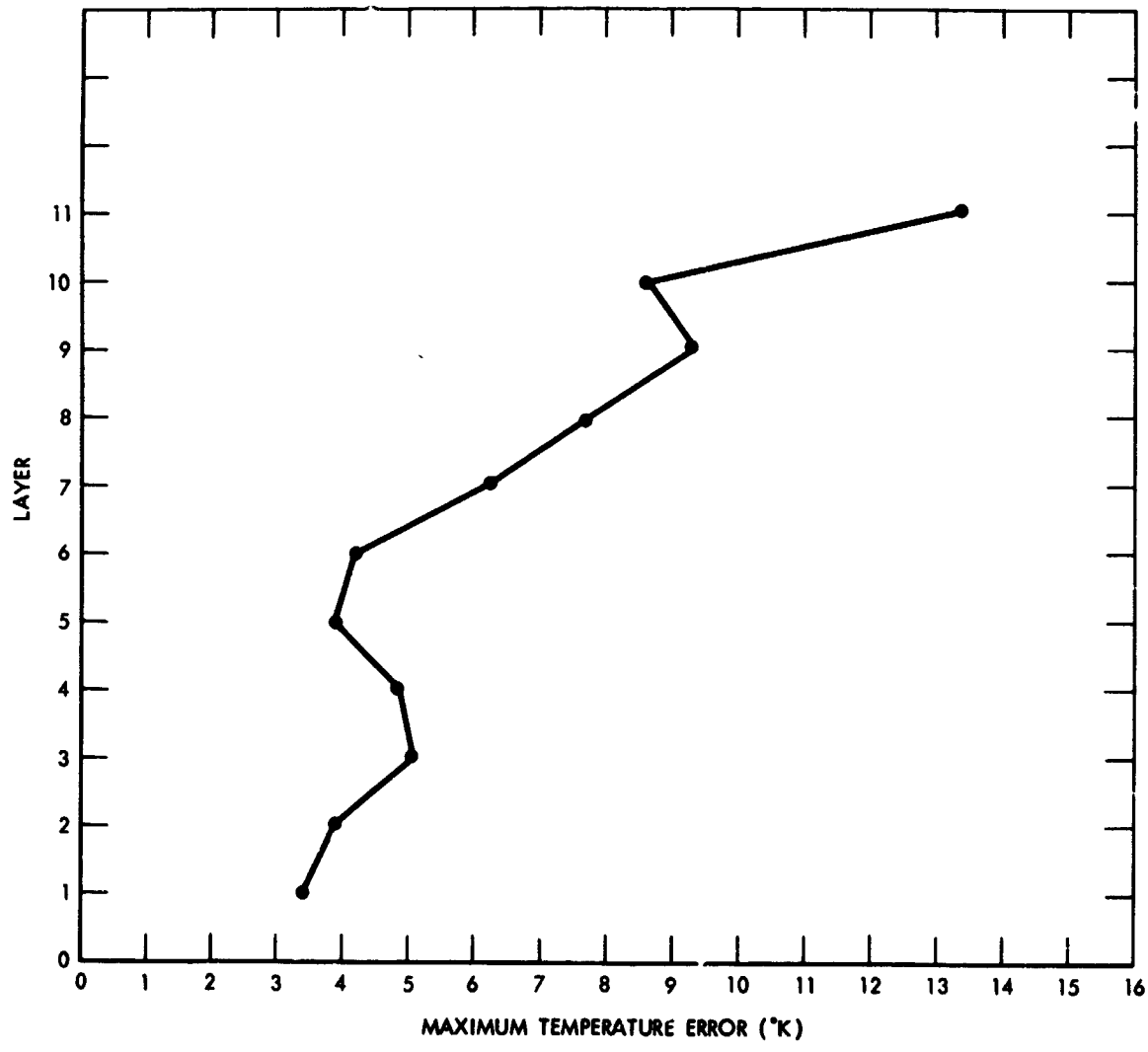


FIGURE 42.—Maximum temperature errors for the 16 firings 1960–1963.

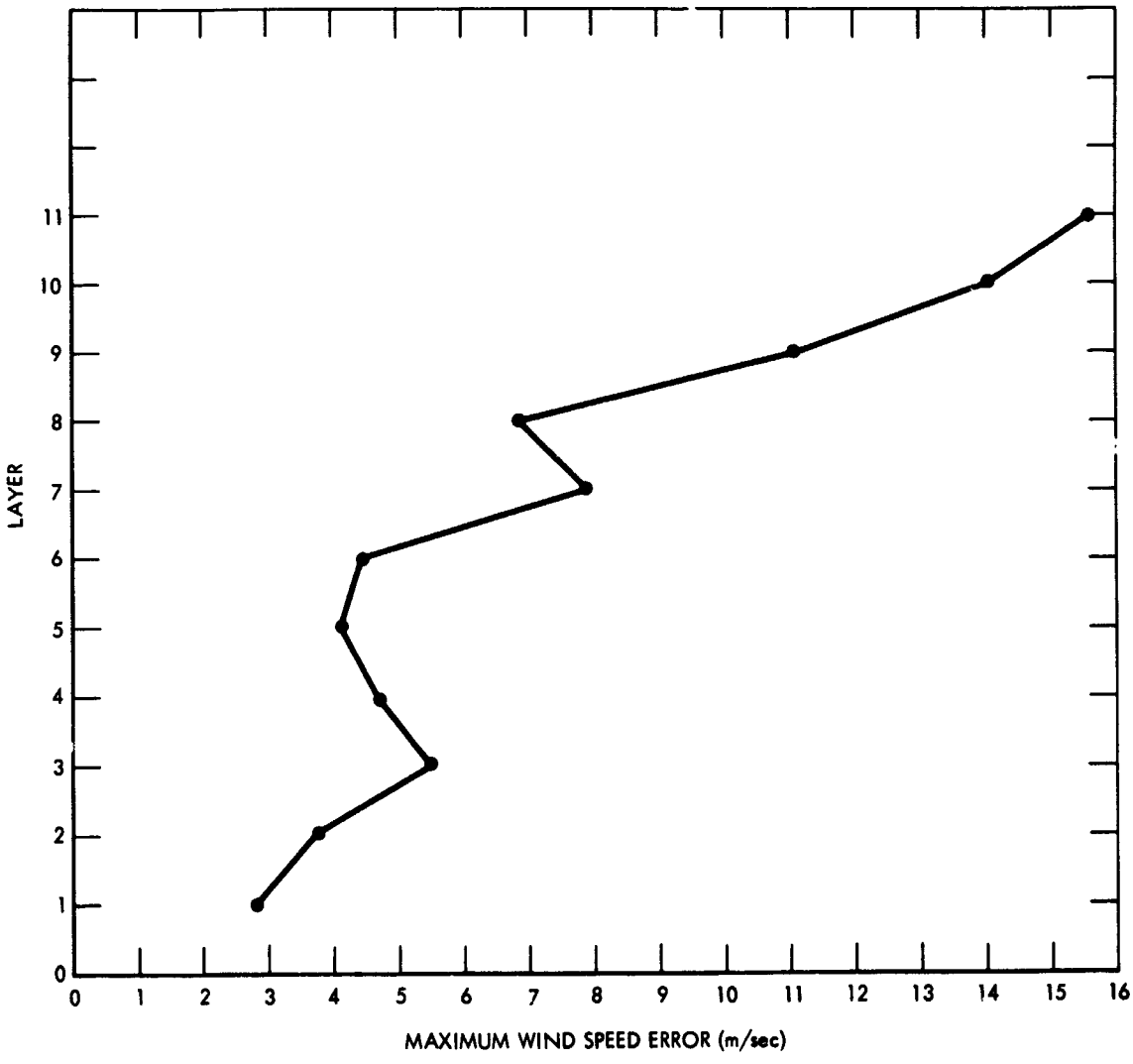


FIGURE 43.—Maximum wind errors for the 16 firings 1960–1963.

NEW KNOWLEDGE OF THE EARTH'S ATMOSPHERE FROM THE AERONOMY SATELLITE (EXPLORER XVII)*

N. W. SPENCER, G. P. NEWTON, C. A. REBER, L. H. BRACE AND R. HOROWITZ

The Explorer XVII satellite performed direct, very localized measurements of the total neutral particle density, the concentration of neutral particle masses 4, 14, 16, 28, and 32 and the temperature and concentration of thermal electrons, between the altitudes of 258 km and 920 km over those regions of the earth where the satellite was accessible to the Minitrack network, and in particular between $\pm 58°$ latitude. Pressure gages on the satellite showed that the total density at 280 km was about 50 percent lower than is given by the appropriate atmospheric models based on satellite drag measurements. Daily variations in total density are more strongly dependent on a_p (the magnetic index) than had been believed previously. Neutral mass spectrometers showed that *He* is the predominant neutral constituent above 600 km, 0 is predominant between 250 km and 600 km, and N_2 is predominant below 250 km. The scale heights of the various constituents agree in general with the corresponding model atmospheric scale heights. Langmuir probe results confirmed the global extent of thermal non-equilibrium ($T_e > Tg$) and provided high resolution of the diurnal variation of electron temperature and density at several stations. For example, the electron temperatures near the F_2 maximum over Blossom Point show a nighttime value of about 1100° K, followed by a mid-morning maximum of 2800° K and an afternoon plateau of 2200° K. A consistent and strong latitude effect, evident particularly at Blossom Point, caused a significant positive gradient in electron temperature (the order of 25° K/degree of latitude) and an inverse gradient in electron density in a manner approximately in accord with recent theories of Hanson and Dalgarno.

INTRODUCTION

The Explorer XVII satellite, Figure 1, was designed to provide direct measurements of aeronomic parameters as a basis for new studies of the physics of the Earth's upper atmosphere. Thus, instruments were selected for the satellite which would provide both total and relative concentration of the neutral particles, and high-resolution measurements of the electron temperature and density; all of considerable significance in studies of the physical processes controlling the upper atmosphere. These data would help also to (a) clarify and define the structural properties of the atmosphere, previously established primarily through inferences from satellite drag measurements, and (b) investigate the variability and dependence of the atmosphere on solar conditions.

The technological advance of measurement techniques was also an objective of the project, as part of a continuing effort to improve experimental capability. The application of laboratory-developed techniques required engineering as well as measurement technique refinement and adaptation to new environments. The atmospheric data to be obtained, if it were to be of maximum benefit consistent with its timeliness, required computer usage for processing the large quantity of data (the order of 2×10^9 bits of information).

EXPERIMENTS

The choice of experiments to obtain the desired data was based mainly upon laboratory vacuum technique experience; thus both thermionic and cold-cathode pressure gages, derived from proven laboratory sensors, were selected for the measurements of the total neutral particle density. Double-focusing magnetic mass spectrometers were chosen for the measurements of neutral particle concentrations. Two Langmuir probe experiments were employed for measurements of the electron temperature and ion density.

*Published as *Goddard Space Flight Center Document X-661-64-114*, May 1964.

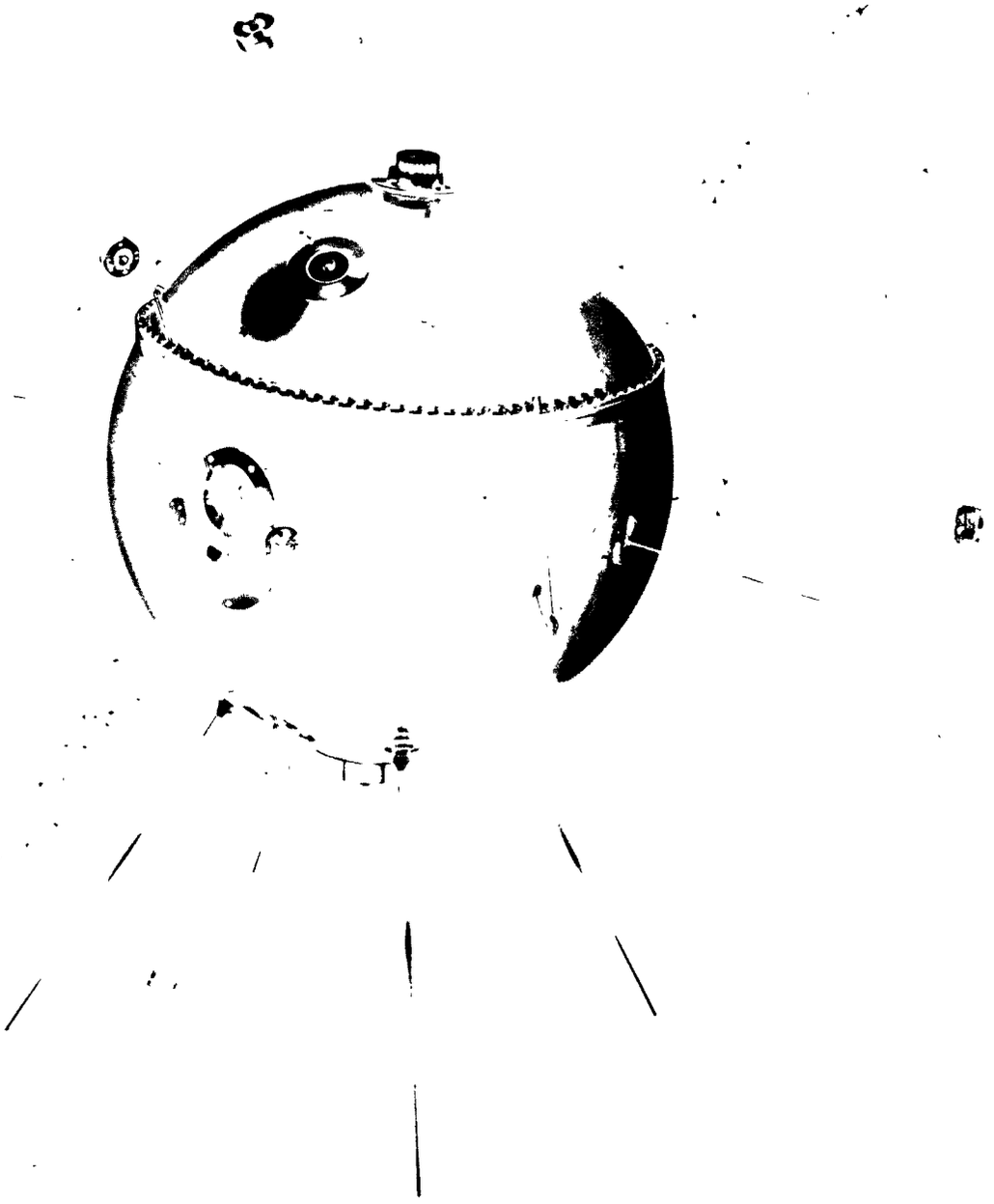


FIGURE 1.—Artist's conception of Explorer XVII in orbit.

It is of special significance from the point of view of determining local values of the ambient concentrations of constituents that proven laboratory vacuum system techniques were employed in construction of the satellite. For example, to minimize potential sources of contaminating gases, nearly all external surfaces were constructed of stainless steel, and all satellite joints were either welded or utilized copper shear gaskets for vacuum sealing. A significant effort was thus devoted to making the satellite-sensor system a true inside-out ultra-high vacuum system so that the experimenter could be confident that the atmospheric samples measured were not contaminated by materials carried aloft by the satellite.

Pressure Gage Experiments

The density measurements made by the Explorer XVII ionization gages extended to satellite usage the technique employed in rocket manometer experiments (Schultz, Spencer, Reifman, 1948) (Horowitz, Kleitman, 1953) to measure the total atmospheric density. The description of the instrumentation, the calibration technique, and the flight response of the Explorer XVII sensors have been reported (Newton, Pelz, Miller, Horowitz, 1963).

Four independent pressure gage experiments were employed: two Bayard-Alpert type (thermionic cathode) ionization gages and two Redhead type (cold-cathode, magnetic) gages. Each sensor was equipped with a special vacuum-sealed orifice that could be opened after the satellite was in orbit. Thus it was possible for the sensor to be properly cleaned, calibrated, sealed under vacuum, and opened on command to the space environment. This procedure, whose validity was previously established through rocket experience, assured the necessary high degree of vacuum cleanliness for the sensors.

The use of both cold and thermionic cathode gages was considered essential because of uncertainties in (a) the response of a hot-cathode gage in a sometimes predominately atomic oxygen environment (not subject to adequate laboratory calibration) and, (b) the general applicability of ionization gages to the high-velocity satellite environment. At the same time, a desirable redundancy was accomplished and valuable studies of the usefulness of the two fundamentally different sensors were made possible. Each pressure gage was provided with an appropriate electrometer amplifier and other electronic support devices which enabled conversion of the sensor output current to a voltage suitable for telemetry. The electronic systems also included provision for in-orbit current calibration of the amplifiers once during each operation of the gages.

Neutral Particle Mass Spectrometers

Two identical double focusing magnetic mass spectrometers were employed for the determination of the local concentrations of atmospheric helium (mass 4), atomic nitrogen (mass 14), atomic oxygen (mass 16), molecular nitrogen (mass 28) and molecular oxygen (mass 32). Although detailed descriptions of the spectrometers are provided elsewhere (Meadows, 1960) (Hall, Howden, Iwasaki, 1960), (Spencer, Reber, 1962), (Reber, Hall, in prep.), the significant features are summarized here. The external ion source, designed to reduce the interactions of the sampled particles with the sensor, was followed by an electrostatic ion lens which focused the relatively high energy ions on the entrance slit of the analyzer. In this way, it was possible for the instrument to accept particles from a 2π steradian solid angle, and up to 12 *ev* kinetic energy. In the magnetic analyzer the beam of ions was separated according to mass, a given mass falling on the appropriate collector electrode in spectrograph fashion.

Each of the spectrometers was provided with a sensitive electrometer and logarithmic amplifier for conversion of the collected ion current to the proper telemetry voltages. Electronic logic circuitry accomplished the required changes in sensitivity and mass-number selection. As with the pressure gages, a vacuum-seal arrangement was adopted which permitted exposure of the spectrometer ion source to the atmosphere after orbit was attained.

The primary data analysis task was to determine the relationship of the measured ion currents to the ambient neutral particle densities. Particles could enter the region of ionization in one of three ways: (1) directly with no collisions; (s) after suffering one or more collisions with surfaces in the source region; (3) by entering the spectrometer analyzer, becoming thermalized, and

subsequently being re-emitted. The relationships between the measured ion currents and the ambient atmosphere were computed on the basis of these three mechanisms and the laboratory gas calibrations. The validity of these calculations is demonstrated by the fact that the total mass-density measured by the spectrometer is in satisfactory agreement with that obtained independently by the companion pressure gage experiments described above.

Langmuir Probes

Two independent Langmuir probe systems, based on established techniques and previous rocket usage (Spencer, Brace, Carignan, 1962), (Brace, Spencer, Carignan, 1963), (Nagy, Brace, Carignan, Kanal, 1963) were employed to provide measurements of the ion concentration (N_i), and the electron temperature (T_e) of the ionosphere. Each probe system used a cylindrical electrode (projecting into the plasma) whose potential was varied with respect to the satellite shell. The resulting current to the probe was converted to a voltage suitable for telemetry.

Using the following equation, the temperature was derived from the electron current to the probe as it was swept from the satellite potential to the plasma potential:

$$\frac{d \log_e i_e}{dV} = \frac{e}{kT_e} \tag{1}$$

To localize the T_e measurement, the electron temperature probe was swept at a rate of 10 sweeps per second; and to maximize the resolution, the voltage was swept in two ranges, 0 to ¾ V and 0 to 1½ V, respectively. As a result, each temperature measurement was completed in less than 400 meters of the satellite path, and to that extent represents a point measurement. The telemetry sampling rate was sufficiently high to permit determination of temperature values as low as 400°K (although the lowest temperatures actually recorded were about 900°K).

The ion density probe was swept from −3 to +2 volts in a 2-second period, which was long compared with the satellite spin period of 0.7 seconds. N_i was derived from the ion current maxima which occurred each time the probe axis was perpendicular to the velocity vector.

Sensor Location

As noted above, the capability to provide measurements of the constituents of the space environment required close adherence to established laboratory vacuum techniques. Thus, in addition to providing a sealed housing which would not itself contaminate the local atmosphere, the eight sensors were located so as to provide maximum separation from each other.

Consideration of sensor orientation with respect to the direction of motion required that the mass spectrometers be located at the two ends of the spin axis to minimize changes in orientation during each mass-sampling sequence. The four pressure gages were distributed uniformly about the spherical satellite to insure that at least one gage would always experience pressure variations due to satellite spin, regardless of the spin axis orientation. One of each type of pressure gage was located on the satellite equator to enable a comparison of the gage responses under identical conditions. The two Langmuir probes were also located on the equator at points nearly diametrically opposite each other.

SUPPORTING SYSTEM

Interpretation of the data from the various sensors required detailed knowledge of the instantaneous angle between any sensor and the direction of motion of the satellite. This information was provided through the use of a multiple optical sensor arrangement, which enabled sensing the direction of the sun and/or moon and the instants of passage, during spin, of the Earth's horizons.

Direct measurement sensors like those employed in Explorer XVII provide time rates-of-change of data requiring high telemetry sampling rates. For example, the spectrometers and pressure gages required 60 samples per second and the "high speed" Langmuir probe required 180 samples per second. To meet these needs a pulse code modulation (PCM) telemetry system capable of 1000 samples per second was selected. This system had the additional advantage of providing a digital format which facilitated computer data processing.

The satellite was powered exclusively by silver-zinc cells, since solar cells presented the possibility of local contamination of the atmosphere. The

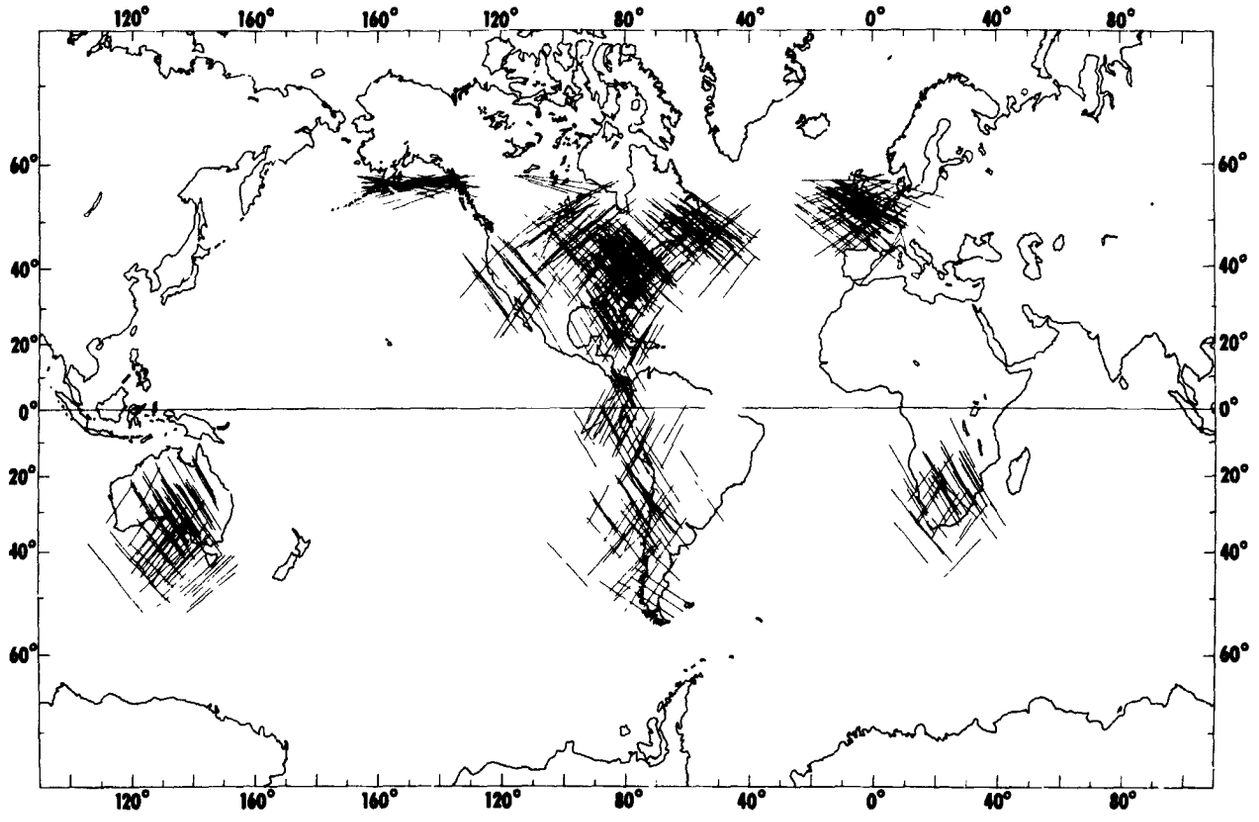


FIGURE 2.—Explorer XVII geographic coverage.

150 pounds of cells which were employed provided adequate energy to operate the entire satellite system for a total of 75 hours. A command-control system permitted the experiments to be turned on for four-minute periods, each of which was terminated by an internal programmer. Because a tape recorder was not employed, responses were confined to geographic regions of approximately 4000 kilometers diameter about each minitrack command station. Figure 2 illustrates the geographic coverage attained by showing the path of the satellite during each data-producing response. Table I summarizes some of the pertinent statistics of the satellite, and other information.

TABLE I.—*Explorer XVII Statistics*

Launch Date	April 3, 1963
Inclination	58°
Perigee	258 km
Apogee Range	920–870 km
Useful lifetime	100 days
Perigee motion	+39° to +58° to −18°
Data responses	650 on command
Telemetry	PCM—8640 bits/sec
Spin rate	90 RPM
Power supply	Chemical
Size and shape	1 meter sphere
Weight	410 pounds

EXPERIMENTAL RESULTS

Pressure Gage Results

As noted above, four independent gage systems were employed on the satellite, and all gages operated in orbit. Three of the gages gave useful and meaningful data during the active lifetime of the satellite. The fourth gage, however, experienced an apparent decrease in sensitivity after opening in orbit and no atmospheric structure data is available at present from this gage. Some of the results from the other three gages for five northern mid-latitude stations, representing approximately 25 percent of the total available data are presented here.

Figure 3 shows the measured atmospheric density over Grand Forks, Minnesota, during a pass at 2000 hours local time and demonstrates typical resolution of the density gage data. It may be

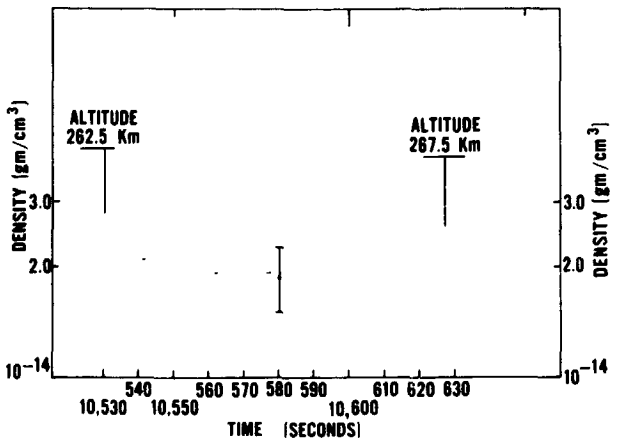


FIGURE 3.—Density data derived from a single pass over Grand Forks, Minn., demonstrating the resolution of the pressure gauge data.

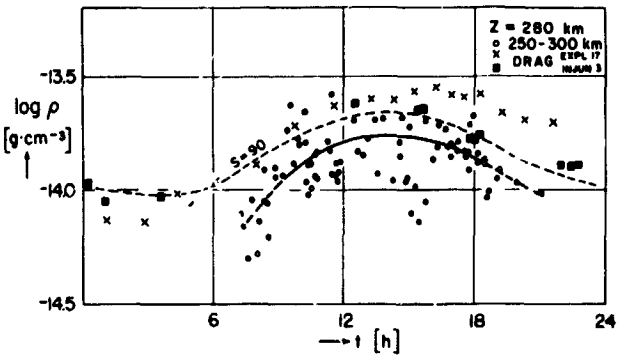


FIGURE 4.—Comparison of directly-measured and drag-derived atmospheric densities.

seen that the total altitude change was only 5 km for this pass, and at the average altitude of 265 km, the atmospheric density was 2×10^{-14} gm/cm^3. Each point shown is the average of the three independent density measurements. The resolution of the density data from each gage is such that the density was measured every 700 milliseconds or once every six kilometers along the satellite orbit. The observed small scatter of the density data makes quantitative density scale-height determination possible for passes possessing significant altitude changes. The magnitude of the error in the absolute value of the density is ±25 percent for this pass and can be attributed primarily to uncertainty in the absolute calibration of the sensors in the laboratory. At an altitude of 600 km, the error in the absolute value of the measured density is ±55 percent for some passes. Generally, the precision of the density measurements is better than ±20 percent.

Figure 4 (Newton, Horowitz, Priester, 1964) provides a comparison of the atmospheric densities directly measured by the gages in the altitude range 258 to 300 kilometers, to the density obtained from drag observations of Explorer XVII (Bryant, 1964) and Injun 3 (Jacchia, Slowey, 1964). All data are normalized to a height of 280 kilometers for comparison purposes, by using density gradients obtained from the Harris and Priester model with $S=90$.

The Injun 3 data, indicated by squares, were selected for quiet geomagnetic conditions ($A_p=2$), for the time interval of 18 February through 30 June 1963. During this period, the latitude of the satellite perigee varied between $-40°$ and $+60°$. The drag data from Explorer XVII, indicated by crosses, correspond to the time interval from 3 April to 6 July 1963, when the latitude of the satellite perigee was between $+58°$ and $-20°$. All Explorer XVII data in the figure have been reduced to quiet conditions ($A_p=0$) by using the preliminary linear reduction relation:

$$\log \rho = 0.006\, A_p$$

Further analysis of the gate data indicates, however, that the correlation between density and geomagnetic activity should be a steeper, nonlinear relation with the steepest portion applicable for low A_p (Newton, Horowitz, Priester, 1964). Application of the modified relation is expected to (a) remove some of the scatter from the gage data (which reflects real atmospheric variations), and (b) lower the average value of the directly measured densities by a small amount.

It is observed that the densities determined from drag are systematically 40 percent to 50 percent greater than the normalized densities measured by the gages, and that this separation is just outside the combined, stated uncertainties of the two sets of data. This difference is significant but at this time is not considered serious, since it could be accounted for by modest changes in the altitude to which the drag data are assigned, the drag coefficient, or the gage calibration constants.

Figure 5 shows measured atmospheric density-versus-altitude for the altitude range 258 to 600 km. These data result from approximately 60 passes for an A_p between 0 and 10, $F_{10.7}$ between

70 and 100, and most local times. It is seen that considerable variation in the atmospheric density occurs, resulting primarily from the differences in local time, a factor of 5 diurnal variation at 360 km being observed. The Harris and Priester model densities ($S=90$) for 0400 and 1400 hours are shown for comparison purposes.

Continued analysis of the Explorer XVII data is currently underway to further define:

(1) The quiet atmosphere and its variation with local time.

(2) The variations from the quiet atmosphere resulting from solar and geomagnetic disturbances.

(3) Other effects not now apparent.

Mass Spectrometer Experiment Data

Concentrations as a function of altitude are shown in Figure 6 (Reber and Nicolet, 1964) where Nicolet's model (private communication) is lincluded for reference. The data points shown are measured values, converted to ambient number densities. Measurements of each mass taken on the same satellite interrogation are made approximately one minute apart and are shown in the figure joined by straight lines. The numbers refer to orbit numbers. Table II lists pertinent information for the passes shown in the figure.

Of particular interest a.e the helium concentration vs. altitude, and the altitude regions where helium, atomic oxygen and molecular nitrogen are predominant. It is seen that helium is the major constituent above 600 km, molecular nitrogen is predominant below 250 km, and atomic

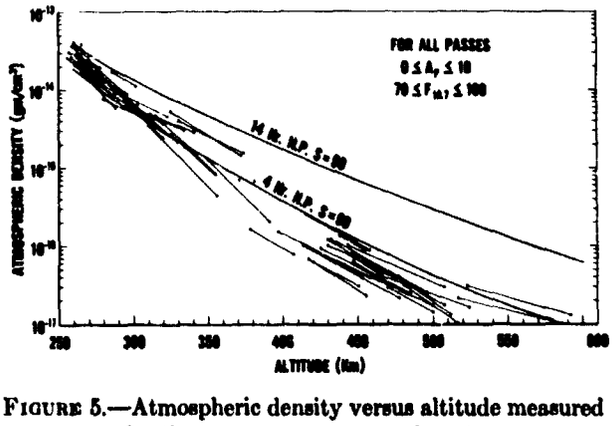


FIGURE 5.—Atmospheric density versus altitude measured by the pressure gauge experiment.

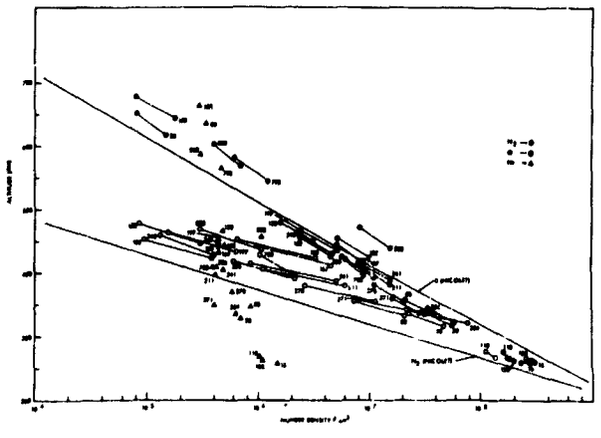


FIGURE 6.—Average daytime and nighttime concentrations of He, O, and N_2 from Explorer XVII mass spectrometer experiment.

oxygen is predominant between these levels. It should be noted that the scale heights of the constituents at higher altitudes correspond to temperatures of about 700° at night.

A possible deviation of the nighttime N^2 distribution from a diffusive equilibrium condition at altitudes less than 400 km is also suggested by this figure. One possible explanation is that the diffusion time in this altitude region is the same order of magnitude as the diurnal variation period. It should also be kept in mind, however, that these data represent a variety of times and geographic locations and thus do not accurately present an instantaneous vertical profile.

Figure 7 shows the variation of mean mass with altitude. The mean mass was computed using the major constituents, N_2, O and He. Hydrogen, which the instrument was not designed to measure, could significantly reduce the value of the mean mass at higher altitudes.

The variation of concentration ratios of helium-to-oxygen and atomic oxygen-to-molecular nitrogen with altitude is shown in Figure 8. The solid lines drawn through the points are averages; as in all the data presented, the points are measured values and are not averaged, nor do they reflect any smoothing. It can be seen here again that helium is the dominant component above 600 km and that molecular nitrogen is predominant below about 250 km.

Langmuir Probe Experiment Results

Figure 9 shows the detailed variation of T_e and

TABLE II.—*Tabulated Mass Spectrometer Data*

Pass & Station	*Date*	*Local Time*	*α*	*Geo. Lat.*	*Geo. Long.*
# 15 BP	4/4/63	21.15 hrs.	6°	38.5°	−75.0°
# 50 COL	4/6/63	0.65	16°	57.0°	−149.0°
# 80 COL	4/8/63	0.99	9°	55.0°	−147.0°
# 80 FTM	4/8/63	4.89	63°	18.0°	− 92.0°
#118 BP	4/10/63	18.81	70°	37.0°	− 72.0°
#120 GF	4/11/63	20.32	51°	51.0°	− 98.5°
#138 BP	4/12/63	2.51	12°	37.0°	− 84.0°
#152 BP	4/13/63	2.01	14°	39.5°	− 68.5°
#167 BP	4/14/63	1.65	20°	39.5°	− 75.0°
#182 BP	4/15/63	1.54	25°	37.0°	− 78.0°
#183 QUI	4/15/63	3.26	23°	4.5°	− 79.0°
#197 BP	4/16/63	1.43	27°	34.0°	− 81.5°
#211 BP	4/17/63	0.53	45°	41.5°	− 71.5°
#226 BP	4/18/63	0.48	53°	38.5°	− 74.0°
#241 BP	4/19/63	24.19	62°	38.0°	− 79.5°
#242 MOJ	4/19/63	0.64	54°	31.0°	−121.5°
#254 NFL	4/20/63	22.75	82°	49.0°	− 53.0°
#270 BP	4/21/63	23.30	80°	41.5°	− 71.5°
#271 GF	4/21/63	22.88	85°	45.0°	−101.0°
#708 NFL	5/20/63	7.18	39°	49.5°	− 49.5°
#795 OOM	5/26/63	15.81	63°	−34.0°	137.5°
#800 JOB	5/26/63	15.90	65°	−37.5°	19.0°
#888 JOB	6/1/63	13.24	33°	−27.0°	25.0°

The local sun time, angle of attack (α), geographic latitude and longitude are averaged over the four minute pass. The stations involved are: BP—Blossom Point, Md.; COL—College, Alaska; FTM—Fort Myers, Fla.; GF—Grand Forks, Minn.; QUI—Quito, Equador; MO.—Mojave, Calif.; NFL—Newfoundland; OOM—Woomera, Australia; JOB—Johannesburg, South Africa.

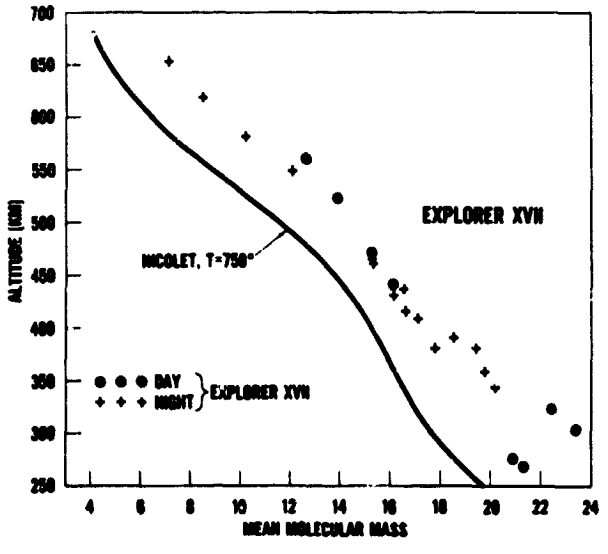


FIGURE 7.—Mean molecular mass versus altitude from mass spectrometer.

N_t, measured during a series of near perigee passes near the F_2 maximum over Blossom Point, Maryland. These data were obtained over the three month satellite lifetime during which the rotation of the orbit plane caused a complete diurnal variation to occur. The data are shown as points or pairs of points, the latter corresponding to the measured values at the beginning and at the end of individual satellite passes.

At first glance one is struck by the lack of correlation between the gross diurnal variations (smoothed curves) and the changes during the individual passes, especially in the afternoon. Since perigee passes such as these can exhibit very little altitude change, this in-pass variation must be largely latitude dependent (10° change in latitude within average pass). The average in-pass change in T_t implies a latitude gradient near Blossom Point of approximately 25° K per degree of latitude, coresponding to about a 10 percent change in T_t within a pass. With few exceptions the changes in T_t within a pass are accompanied by an inverse change in N_t which is even greater than 10 percent.

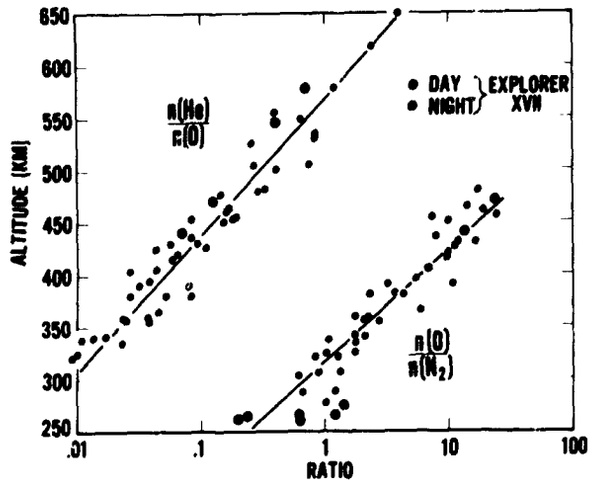


FIGURE 8.—Ratios of n(He)/n(O)n(N_2) versus altitude from mass spectrometer experiment.

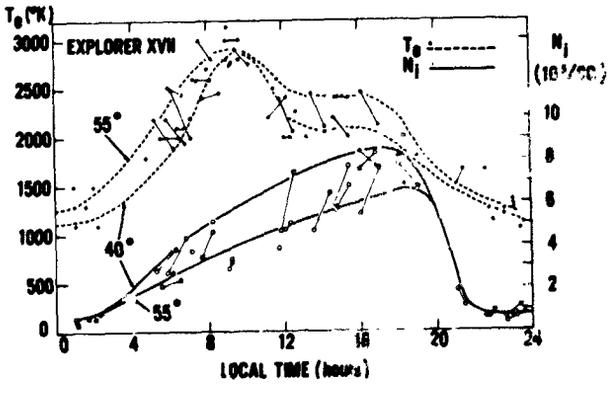


FIGURE 9.—Diurnal variation of T_e and N_i for magnetic latitudes 40–50°N and altitudes 258 to 350 km, from the Langmuir probes.

Plots similar to Figure 9 have been prepared from T_e and N_i measurements at two other latitudes (10° N at Quito, 60° N at College), and the resulting gross diurnal variation curves at all three latitudes are shown in Figure 10. These data also correspond to the region of the F_2 maximum (below 400 km).

The T_e variation at the F_2 maximum at all latitudes shown is characterized by a steep morning rise, a midmorning maximum, and afternoon plateau, and a gradual decrease near sunset. The nighttime values of T_e are somewhat variable but are always significantly above the neutral particle temperature (Harris and Priester, 1962), particularly at College, Alaska where the summer night at F_2 altitudes is short or non-existent.

The values of N_i rose gradually throughout the

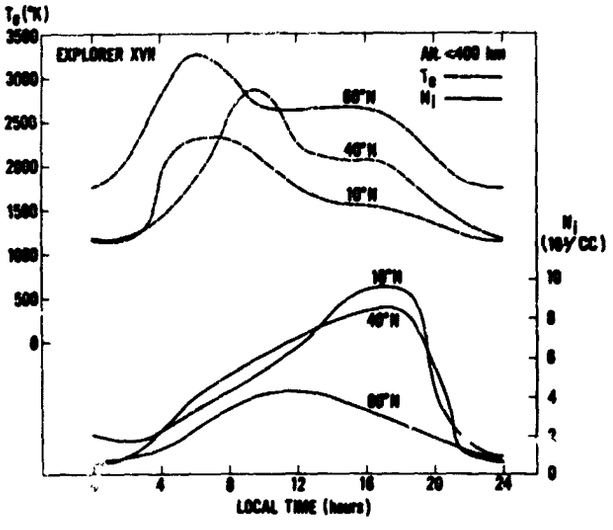


FIGURE 10.—Averaged T_e and N_i showing the diurnal variation above three selected stations; Quito (10°N), Blossom Point (40°N) and College (60°N).

day, reaching a maximum density in the late afternoon, except at College where the maximum occurred in the early afternoon.

It should be noted that the curves in Figure 10 represent direct "in situ" measurements above specific geographic locations during the late spring and summer of 1963, and therefore should not be considered models of the diurnal variation at other altitudes, longitudes, and seasons. However, the analysis of higher altitude data from these sites, as well as data from other sites is now in progress, and it is hoped that these data will help provide a better understanding of the global structure of the ionosphere.

In conclusion, it is clear from the simultaneous measurements that T_e and N_i are related in a generally inverse manner which agrees reasonably well with the inverse square relationship predicted by Hanson (1962) and Dalgarno, et al (1963). The elevated nighttime values of T_e show that there is a heat source for the electrons at night, although its energy content is only a few percent of that required to cause the electron temperatures observed in the daytime (Brace, Spencer, Dalgarno, 1964).

ACKNOWLEDGMENTS

The success of a satellite laboratory such as discussed above results, clearly, from the efforts of many individuals. The authors gratefully and

sincerely wish to acknowledge these efforts and specifically mention the major contributing group leaders: James S. Albus—aspect system, Joseph P. Corrigan—tracking and data systems, James L. Cooley—coordinator, Paul C. Donnelly—batteries, Donald F. Fitzpatrick—mechanical design, William D. Hoggard—environmental testing, Robert E. Kidwell—thermal systems, John N. Libby—electronic control, Chris C. Stephanides—spacecraft system and integration engineer, Virginia Zanner—computer data analysis.

REFERENCES

BRACE, L. H., N. W. SPENCER, G. R. CARIGNAN, Ionosphere electron temperature measurements and their implications, J. Geophys. Res., **68**, 5397–5412, 1963.

BRACE, L. H., N. W. SPENCER, A. DALGARNO, Explorer XVII electrostatic probe data and interpretation (Explorer XVII) American Geophysical Union, 45th Annual meeting, Washington, D.C., April 1964).

BRYANT, R., Densities obtained from drag on the Explorer XVII satellite, J. Geophys. Res., **69**, 1423–1425, 1964.

DALGARNO, A., M. B. MCELROY and R. J. MOFFETT, Electron temperatures in the ionosphere, Planet. Space Sci., **11**, 463–484.

HALL, L. G., P. F. HOWDEN and T. F. IWASAKI, Design of a mass spectrometer for use in a sateilite, presented at the 8th Annual Meeting of the ASTM Committee E–14 on Mass Spectroscopy.

HANSON, W. B., Electron temperatures in the upper atmosphere, Space Research, Proc. Intern. Space Sci. Symp. 3rd, Washington 1962, edited by W. Priester, pp. 282–302, North Holland Publishing Co., Amsterdam.

HARRIS, I. and W. PRIESTER, Theoretical models for the solar cycle variation of the upper atmosphere, J. Geophys. Res. **67**, 45–4591, 1962.

HOROWITZ, R. and D. KLEITMAN, Upper atmosphere research report No. XVIII, N.R.L. Report 4246, Naval Research Lab., Washington, D.C., Oct. 1953.

JACCHIA, L. G. and J. SLOWEY, Atmospheric heating of the auroral zones: A preliminary analysis of atmospheric drag of the Injun 3 satellite. SAO Special Report 136, Sept. 1963, J. Geophys. Res. **69**, 905–910, 1964.

MEADOWS, E. B., Design requirements of a mass spectrometer for satellite use, presented at 8th Annual Meeting of ASTM Committee E–14 on Mass Spectroscopy Atlantic City, N. J., June 1960.

NAGY, A. F., L. H. BRACE, G. R. CARIGNAN, M. KANAL, Direct Measurements bearing on the extent of thermal non-equilibrium in the ionosphere, J. Geophys. Res. **68**, 6401-6412, 1963.

NEWTON, G. P., D. J. PELZ, G. E. MILLER, and R. HOROWITZ, Response of modified Redhead magnetron and Rayard-Alpert vacuum gages aboard Explorer XVII, Transactions of the Tenth National Vacuum Symposium, edited by George H. Bancroft, pp. 208-212, The Macmillan Company, New York, 1963.

NEWTON, G. P., R. HOROWITZ, W. PRIESTER, Pressure gage experimental data and interpretation; AGU Presentation of Aeronomy Results from the Explorer XVII satellite, 45th Annual Meeting, Washington, D.C., April 23, 1964.

REBER, C. A., Preliminary results regarding neutral atmospheric composition from the Explorer XVII satellite, presented at the Third Western National Meeting, American Geophysical Union, Boulder, Colorado, Dec. 1963.

REBER, C. A. and M. NICOLET, Explorer XVII mass spectrometer data and interpretation, AGU Presentation of Aeronomy Results from the Explorer XVII satellite, 45th Annual Meeting, Washington, D.C., April 23, 1964.

REBER, C. A., and L. G. HALL, The Explorer XVII mass spectrometer, (in preparation).

SCHULTZ, F. V., N. W. SPENCER, and A. Reigman, Upper Air Research Program, Progress Report No. 2, Contract W-33-038 ac-14050. Engineering Research Institute, University of Michigan, pp. 129-137, July 1, 1948.

SPENCER, N. W., and C. A. REBER, A mass spectrometer for an Aeronomy satellite, Proceedings of the Third International Space Science Symposium, Washington, D.C., May 1962.

SPENCER, N. W., L. H. BRACE, G. R. CARIGNAN, Electron temperature evidence for non-thermal equilibrium in the ionosphere, J. Geophys. Res. **67**, 157-175, 1962.

N66 37968

THE REFLECTIVITY OF JUPITER IN THE ULTRAVIOLET*

THEODORE P. STECHER

Four spectral scans of Jupiter were obtained by means of objective grating rocket spectrometers. The geometrical reflectivity is interpreted in terms of Rayleigh scattering by obtaining a numerical relationship between reflectivity and optical depth. This gives an upper limit to the amount of molecular hydrogen above the cloud layer of 10.2 ± 1 km. atm. The effect of Raman scattering by molecular hydrogen increases this to a 10.6 ± 1 km. atm. An absorption feature is found to be consistent with its identification being photodissociation of H_2.

I. THE OBSERVATIONS

The spectrum of Jupiter in the ultraviolet was obtained from an Aerobee rocket on July 23, 1963 at 06^h02^mU.T. Three additional spectra were obtained in a similar manner on November 14, 1964 at 10^h28^mU.T. The instruments used were objective grating photoelectric stellar spectrometers similar to those described by Stecher and Milligan (1962). Spectral scanning is the consequence of the spin of the rocket. The observations were telemetered to the ground. The exit slit widths were between 50Å and 90Å. The total spectral range covered was from 1050Å to 4000Å. The instruments were calibrated in both a relative and an absolute sense prior to flight in the laboratory. The accuracy of the absolute calibration is believed to be considerably better than the limit set by noise in signal for these observations. A complete description of the instrumental techniques will be included in a forthcoming paper on stellar energy distributions.

The discussion of planetary photometry by Harris (1961) is now followed. In Figure 1 the geometrical reflectivity, p, as a function of wavelength is given between λ2000 and λ3000. The mean error for each value of p was determined in the standard manner from the observations, no measurable signal was detected below λ2000. The solar flux values used in obtaining p are those given by Tousey (1963) and were obtained by photographic spectrophotometry. The necessary geometry was obtained using the appropriate Ephemeris values. The phase angle of Jupiter was 11.72 degrees on July 23, 1963 and almost zero on November 14, 1964.

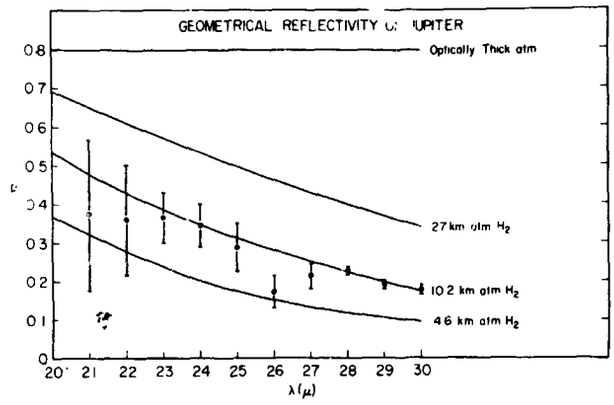


FIGURE 1.—Circles: Observed average geometrical reflectivity of Jupiter as a function of wavelength with mean errors as determined from four spectra. Solid curves: Computed geometrical reflectivities for various amounts of molecular hydrogen.

II. INTERPRETATION

The literature on Jupiter, in particular recent polarization measurements and molecular hydrogen quadrupole line measurements, indicates the possible presence of a Rayleigh atmosphere at ultraviolet wavelengths above the cloud layer. Upper limits for the number of atoms or molecules above the cloud layer may be derived immediately for each species if the reflectivity is known in terms of optical depth. Tables have been tabulated for the total intensity emerging from the top of a Rayleigh atmosphere by Coulson, Dave and

*Published as *Goddard Space Flight Center Document X-613-64-391*, December 1964.

Sekera (1960). This tabulation is for optical depths, τ, up to unity as a function of zenith distance of the sun, μ_0; zenith distance of the emergent rays, μ, and the azimuthal angle, φ. From these values p may be obtained as a function of τ.

For our purposes, Jupiter is assumed to be at opposition even through one observation was made at quadrature, since the phase integral is almost constant for small phase angles. This approximation makes less than three percent difference for that observation (Talley and Horak, 1956). The plane parallel atmosphere is made tangent to a unit sphere at each point on the illuminated hemisphere by the restraint $\mu_0=\mu$. The value of the emergent intensity is then projected onto a unit disk. Integration over the surface then gives the geometric reflectivity for a given value of optical depth. The aximuthal angle is 180°. Curves in the $p-\tau$ plane were constructed in this manner in the range $\tau=0$ to $\tau-1$ for ground reflectivities of 0, 0.25, and 0.8. The values of p for the optical thick case are given for Rayleigh scattering and for isotropic scattering by Harris (1961). The extrapolation from $\tau=1$ to $\tau=\infty$ was made with the aid of geometric reflectivities calculated in a similar manner from the available X-Y functions for isotropic scattering given by Mayers (1962) and by Sobouti (1963). Over the range of τ to be used in this paper the extrapolation error is probably negligible.

The total number of atoms or molecules in a column of unit area perpendicular to and above the cloud layer is now obtained from $\tau=n\sigma$ under the assumption of only one constituent. Here n is the number of atoms or molecules and σ is the Rayleigh scattering cross section per atom or molecule. The Rayleigh scattering cross section for molecular hydrogen is given by Dalgarno and Williams (1962)[1]. The Rayleigh scattering cross section for helium is given by Dalgarno (1962). The molecular scattering cross section for these gases does not follow the Rayleigh law in the ultraviolet. This is physically due to the increased probability of inducing a dipole moment as a resonance is approached. In the range of interest here the deviation of molecular hydrogen is greater than that of helium. This in principle would allow an abundance ratio determination but would require more accurate observational material at the shorter wavelengths.

In Figure 1 the observed geometric reflectivities as a function of wavelength (circles) are compared with ones calculated for a given atmosphere. The best fit to the data under the assumption of zero reflectivity in the clouds is a 10.2 km. atm. of H_2 or a 200 km. atm. of He or any combination of the two resulting in the same total reflectivity. If the clouds are assumed to form a Lambert surface with a reflectivity of 0.25 the total number of scattering atoms or molecules is reduced by a factor of ten and the form of the reflectivity curve becomes incorrect for Rayleigh scattering. The possibility that the clouds have a small reflectivity can not be excluded. The derived atmosphere is therefore an upper limit.

The above analysis has been carried out under the assumption that the scattering is coherent. In the case of most molecular gases including molecular hydrogen this is known not to be the case. Raman scattering from H_2 in the lowest vibrational state is five percent of the Rayleigh scattering at λ1216 (Dalgarno and Williams, 1962). Isotropic scattering may be used to approximate the effect by setting the particle albedo equal to 0.95 and comparing it with the result for a particle albedo of unity. In the case of an optically thick atmosphere one concludes that 43 percent of the incident photons emerge in the Raman lines. Brant (1963) found 34 percent of the incident Lyman-α photons in the Raman lines by means of an approximation to the Rayleigh phase function.

There is a half electron volt difference between the incident photon and the Raman scattered photon. This energy difference means that the Raman photon has a wavelength several hundred angstroms longer than the incident photon at the wavelengths under consideration here. This would appear as a second order effect except on the longer wavelength side of an absorption feature in the planet or when there is a strong wavelength variation in the source function. The absorption bands of ammonia begin at 2100A and result in an almost continuous absorption at shorter wavelengths reaching an optical depth of more than 10^5 at λ1216 (Kuiper 1952) (Watanabe, et al. 1953). This is apparently indicated by the data although the solar continuum becomes so

[1] The ordinate in Fig. 1, Ap. J. 136, 691, 1962 should read $\lambda^4\sigma(\lambda)$ $\times 10^{44}$ cm^4.

weak here that the determination is unreliable. The effect here is that only the Rayleigh component and not the Rayleigh plus Raman appears up to λ2200. Similarly, the absorption feature between λ2400 and λ2700 reduces the Raman scattered component between λ2700 and λ3000. Using isotropic scattering as an approximation and assuming that the ratio of the Raman scattering cross section to the Rayleigh cross section is a constant independent of wavelength the total scattering atmosphere would be considerably more.

Dalgarno and Williams (1964) have kindly communicated to me in advance of publication the Raman scattering cross section for H_2 between λ2000 and λ3000. They give $Q=2.75\times10^{-46}/\lambda'^4$ where λ' is the wavelength of the scattered radiation in centimeters. This is a smaller cross section than was anticipated and leads to only 0.4 km. atm. increase of H_2. Thus, we have an upper limit of a 10.6 km. atm. of molecular hydrogen. The total mean error is estimated to be ±one km. atm.

There are several possible causes of the absorption feature between λ2700 and λ2400 which appears to be real. If for the sake of conjecture, this absorption is assumed to be the forbidden photodissociation of molecular hydrogen, an empirical absorption coefficient may be derived. This may be done by following the path of the photon by using isotropic scattering for which $X-Y$ functions are available (Sobouti, 1963). A particle albedo of unity is equated numerically to the Rayleigh scattering cross section and the particle albedo giving the observed mean absorption depth is determined. If the Rayleigh scattering is considered to be due entirely to H_2 a cross section of 1×10^{-27} cm^2 is obtained for photodissociation of H_2. Dalgarno (1964) considers 10^{-27} to be a good estimate. Herzberg (1964) points out that H_2S would also absorb in this wavelength region so the identification is by no means certain.

SUMMARY AND DISCUSSION

Zabriskie (1962) determined at 4.6 km. atm. of H_2 from their quadrupole bands. Spinrad and Trafton (1963) empirically derived a 27 km. atm. from these bands. Foltz and Rank (1963) concluded from laboratory measurements and theoretical calculations on the quadrupole bands that both of these determinations are much too small and it should be on the order of a 270 km. atm. The latter value would give an optical depth of nine at λ3000 and correspond to a conservative atmosphere throughout the ultraviolet. This is not observed. Casual inspection of photographs of Jupiter taken on blue plates also indicate this not to be the case since the surface markings would be obscured near the edge of the disk. This large optical depth would also require U and B color intensities several times what is observed (Harris, 1961). Öpik (1962) derives an optical depth by inspection of published photographs which is in remarkably good agreement with that obtained in this paper. Öpik then proceeds to derive an atmosphere that is consistent with the molecular weight he obtained from a rediscussion of the occultation measurement of Baum and Code (1953). His resulting atmosphere consists mostly of helium giving a pressure of 11 atmospheres at the top of the cloud layer. Spinrad and Trafton (1963) find from the widths of methane lines a pressure equal to or less than 2.8 atmospheres at the top of the clouds. If we adopt 2.8 atmospheres at the top of the clouds. If we adopt 2.8 atmospheres as the upper limit then we are limited to less than 50 km. atm. of helium. The cross section for Rayleigh scattering in helium is five percent that of molecular hydrogen. A 50 km. atm. of helium is insufficient to give the observed reflectivity. The amount of molecular hydrogen then becomes at least 8 km. atm. for zero reflectivity clouds.

The conclusion of this paper is that Jupiter has less than 10.6 ± 1 km. atm. of molecular hydrogen above the clouds. If the reflectivity of the clouds is zero in this part of the ultraviolet and the upper limit to the pressure is 2.8 atmospheres at the cloud tops then Jupiter has at least a 8 km. atm. of molecular hydrogen. Inherent throughout this paper is the assumption that the atmosphere has the same physical depth at all wavelengths longer than 2000 Angstroms.

REFERENCES

Baum, W. A. and Code, A. D., 1953, A. J., *58*, 108.

Brant, J. C., 1963, *Planet and Space Sci.*, *11*, 725.

Coulson, K. L., Dave, J. V. and Sekera, Z., 1960. *Tables Related to Radiation Emerging from a Planetary Atmosphere with Rayleigh Scattering*, (Berkeley and Los Angeles, University of California Press).

DALGARNO, A., 1962, *G. C. A. Tech. Rep. No. 62-28-A*, Geophysics Corporation of America, Bedford, Mass.
DALGARNO, A., 1964, private communication.
DALGARNO, A. and WILLIAMS, D. A., 1962a, Ap. J. *136*, 690.
DALGARNO, A. and WILLIAMS, D. A., 1962b, M. N. *124*, 313.
DALGARNO, A. and WILLIAMS, D. A., 1964, private communication.
FOLTZ, J. V. and RANK, D. H. 1963, Ap. J., *138*, 1319.
HARRIS, D. L., 1961, Planets and Satellites; The Solar System, ed. G. Kuiper and B. Middleburst (Chicago: University of Chicago Press), *3*, 272.
HERZBERG, G., 1964, private communication.
KUIPER, G. P., 1952, *The Atmospheres of the Earth and Planets* (University of Chicago Press, Chicago), 2nd ed.
MAYERS, D. F., 1962, M. N., *123*, 471.
ÖPIK, E. J., 1962, Icarus, *1*, 200.
SOBOUTI, Y., 1963, Ap. J. Suppl. *7*, No. 72.
SPINRAD, H. and FRAFTON, L. M., 1963, Icarus, *2*, 19.
STECHER, T. P. and MILLIGAN, J. E., 1962, Ap. J., *136*, 1.
TALLEY, R. L. and HORAK, H. G., 1956, Ap. J., *123*, 176.
TOUSEY, R., 1963, Space Sci. Rev., *2*, 3.
WATANABE, K., SELIKOFF, M. and INN, E. C. Y., 1953, AFCRC Tech, Rep. No. 53–23, Geophys. Res. Paper No. 21.
ZABRISKIE, F. R., 1962, A. J. *67*, 168.

N66 37969

THE UPPER ATMOSPHERE*

JAMES C. G. WALKER

The past four years of rocket and satellite measurements have significantly increased our knowledge of the upper atmosphere. For example, the variation of upper atmosphere density, which can be deduced from the rate of decay of satellite orbits, has been mapped out in considerable detail. It is plain now that the presently available reference atmospheres are seriously inadequate in their description of the atmosphere above 60 miles, especially in describing the variation of atmospheric properties with time. In fact, the international Committee on Space Research (Cospar) is working on a revision of its 1961 reference atmosphere that will reflect this recent information.

So fast is our knowledge of the upper atmosphere changing, however, that the new Cospar model could already be out of date, in some respects, by the time it is published. All of which merely underlines a fact that is often overlooked: while a standard model of the upper atmosphere may provide useful guidance in the design of vehicles and experiments to work in that environment, it is not likely that the model at any time will represent the upper atmosphere as we think it is, let alone represent it as it really is. We do not yet know the upper atmosphere as it really is, although we are much closer than we were four years ago.

FROM 60 TO 40,000 MILES

In looking at the upper atmosphere we are, by definition, concerned with what lies between the lower atmosphere (troposphere, stratosphere, and mesosphere) and outer space. This vast region, which includes virtually all of the ionosphere, extends upward from the top of the mesosphere, at an altitude of about 60 miles, to the magnetopause at 40,000 miles. (The magnetopause separates the region of space dominated by the earth's magnetic field from the interplanetary region dominated by fields and particles of solar origin.)

However, because atmospheric pressure and density decrease exponentially with increasing altitude, our knowledge of atmospheric properties is almost entirely limited to altitudes below about 600 miles. At higher altitudes, the densities are so low that, with few exceptions, our present-day experimental techniques are not able to detect the presence of the atmosphere at all, and we must base our models of this region on theoretical extrapolations from measurements made at lower altitudes.

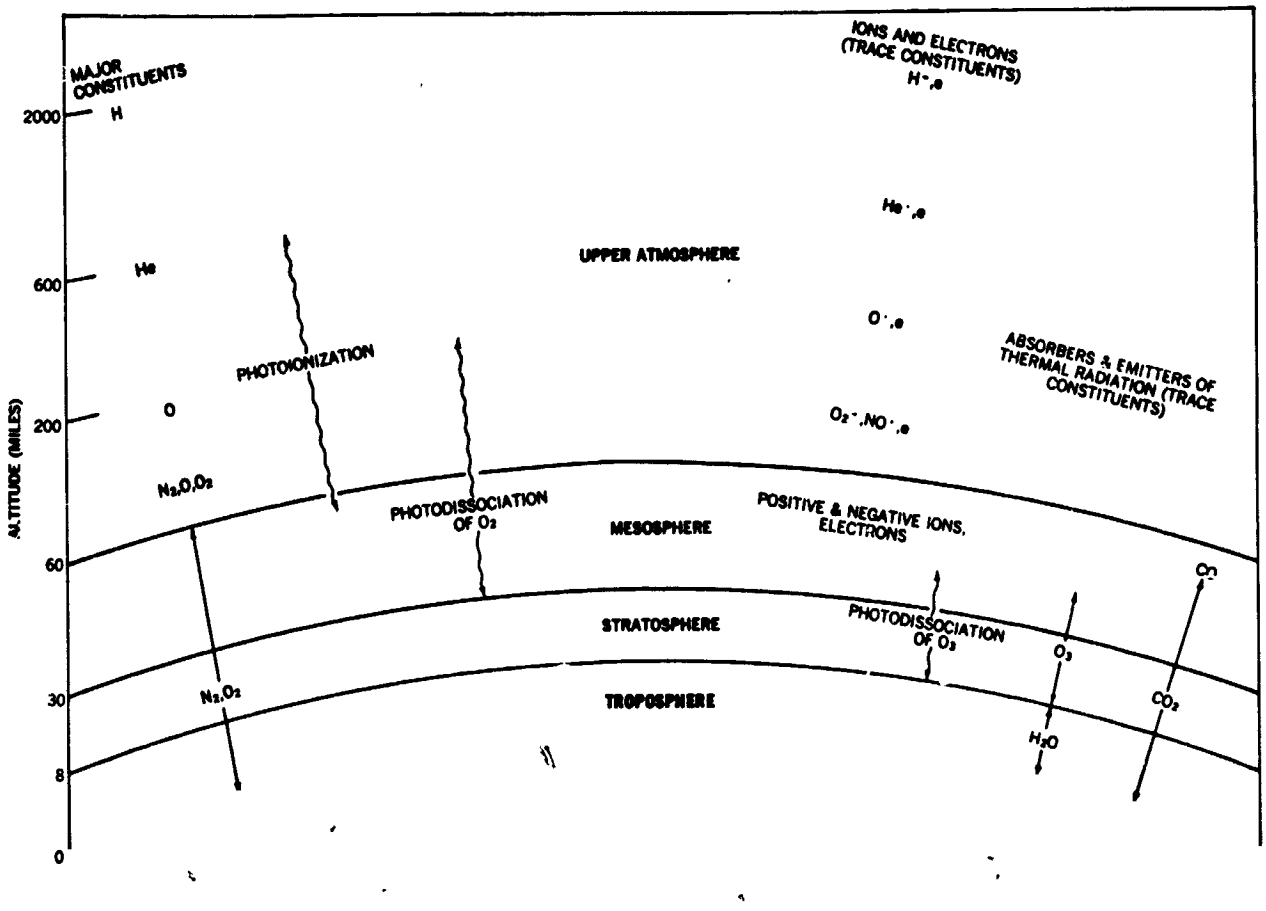
Thus, most of our discussion of the upper atmosphere of necessity refers to the region between 60 and 600 miles. Even within this range of altitudes the data coverage is not uniform. Very few satellites penetrate to altitudes below 120 miles, because at lower altitudes the atmospheric density and the resulting drag on the satellite are so high that the satellite can remain in orbit only a short while before falling back into the lower atmosphere. Accordingly, we must rely on sounding rockets to provide data on atmospheric properties between 60 and 120 miles, and the coverage in time and space which sounding rockets can provide is limited by the expense of rockets powerful enough to reach up into the upper atmosphere.

ONLY SMALL SAMPLES TO WORK WITH

In summary then, the data on upper atmosphere properties tend to be limited to isolated vertical profiles measured with sounding rockets and extending up to about 150 miles, and a considerable number of horizontal profiles measured with satellites and concentrated largely in the region from 180 to 400 miles.

Both the major atmospheric constituents and the trace constituents change with increasing altitude. How they change, why they change, and how their changes can be related to changes in temperature, pressure, and density, is what

*Published in *Space Aeronautics*, 42(4):56–63, October 1964.



SKETCH 1.—Upper atmosphere extends from mesopause out to magnetopause, as shown to scale in left section. A ten times expansion of scale in the right section shows how small is the layer below the mesopause that we call lower atmosphere.

recent rocket and satellite measurements have helped us to determine.

At the higher altitudes, light atoms (oxygen, hydrogen, and helium) are the major consituents. These constituents are atomic, rather than molecular, because their concentrations are so small at great altitudes that atoms seldom collide to form a molecule. The predominance of helium above about 600 miles is one of the new facts uncovered by recent satellite measurements; before this discovery there was thought to be a gradual transition from atomic oxygen to atomic hydrogen with increasing altitude. The important trace elements of the upper atmosphere are ions and electrons, resulting from ionization of the major constituents by solar ultraviolet radiation.

The altitude at which the upper atmosphere begins is defined by a temperature minimum. The temperature of the atmosphere reaches a maximum at 30 miles, the base of the mesosphere, then drops off sharply to the 60 mile base of the upper atmosphere. In the upper atmosphere, temperature rises steadily to around 1200 K (some 900 K above the mesospheric maximum). Because we have not yet found a way to measure upper atmosphere temperatures effectively, the temperature profiles have generally been deduced from measured composition and density profiles.

ENERGY PROCESSES KEY TO UNDERSTANDING

To understand the upper atmosphere—the complex behavior of temperature and composition as a function of altitude—we must consider its energetics. The energy transport processes, and the sources and sinks of energy, are what control the structure of the atmosphere. For the balance between energy gain and loss at any altitude determines the corresponding gas temperature. This temperature in turn controls the rate of change of pressure and density, through the law of

hydrostatic pressure balance. At the same time, energy transfer leads to chemical changes in the atmospheric constituents and thus a variation of chemical composition with altitude.

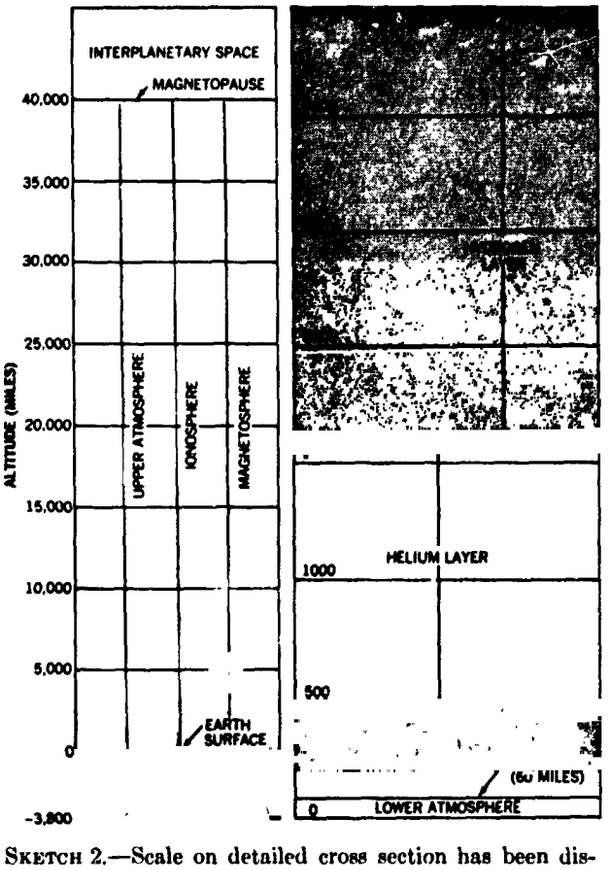
For all practical purposes, we can consider the sun to be the source of all the energy which influences the structure of our atmosphere. The flux of solar energy reaching the earth's surface is about a million times greater than the flux of energy conducted to the surface from radioactive heat sources in the deep interior. The energy fluxes carried by bombarding meteors and high energy galactic cosmic rays are also negligible.

SOLAR WIND MAY PLAY BIG ROLE

The atmosphere receives solar energy both from the sun's electromagnetic radiation and from the streams of protons and electrons flowing radially outward from the sun. Although the solar wind, as the latter streams are called, carries only one ten-millionth as much energy as the solar electromagnetic radiation, yet it may have an important influence on atmospheric properties at very high altitudes—particularly above 120 miles. This is an important question mark, as we shall see later when we consider how temperature varies with time.

Solar electromagnetic radiation varies in wavelength from infrared through visible, and into the ultraviolet. Although the energy carried by the solar ultraviolet radiation is only a small fraction of that carried by the visible in infrared, this energy is absorbed at fairly high altitudes where the atmospheric density is low and the energy transport processes are relatively inefficient. So there is a substantial temperature increase produced by the absorption of ultraviolet radiation.

There are two basic processes by which ultraviolet radiation is absorbed in the atmosphere. The most energetic photons, with wavelengths less than 1000 Angstroms, are able to knock an electron out of an atom or molecule when they are absorbed. The resulting formation of electrically charged particles, or ions, is known as photoionization. Photons with wavelengths greater than 1000 A do not have enough energy to ionize any of the atmosphere's major constituents. But they do have enough energy to break the weaker molecular bonds, specifically those which unite oxygen atoms into oxygen (O_2) or ozone (O_3) molecules.



SKETCH 2.—Scale on detailed cross section has been distorted for illustration purposes. Ultraviolet radiation from the gun is the principal source of energy in upper atmosphere: it reacts with major constituents by photoionization to produce ions and electrons, shown as Trace constituents at right above. Photoionization processes use up most of the sun's shorter wavelength ultraviolet radiation in the upper atmosphere. What remains to reach the lower atmosphere is less energetic and acts to dissociate molecular oxygen into atomic oxygen—around the mesopause, at 60 miles altitude—and to dissociate ozone at the base of the mesosphere. Very little ultraviolet radiation penetrates below the stratosphere. The solar radiation which penetrates into the lower atmosphere is in the visible and near-infrared portion of the spectrum. Neither of these are absorbed to any extent by the constituents of the atmosphere. The energy for the lower atmosphere comes from infrared radiation that is absorbed by the earth and reradiated in the far-infrared portion of the spectrum with maximum intensity at about 150,000 Angstroms. This radiation is strongly absorbed by water vapor and carbon dioxide in the troposphere and stratosphere. The absorption process is the energy source for the general circulation and weather phenomena that characterize the lower atmosphere. Our most recent knowledge of the upper atmosphere extends up to about 600 miles, but is limited at the lower end of the scale to around 120 miles by the requirement that the satellite last some months at least. Rocket soundings are used between 60 and 120 miles.

The resulting breakdown of the molecule is called photodissociation.

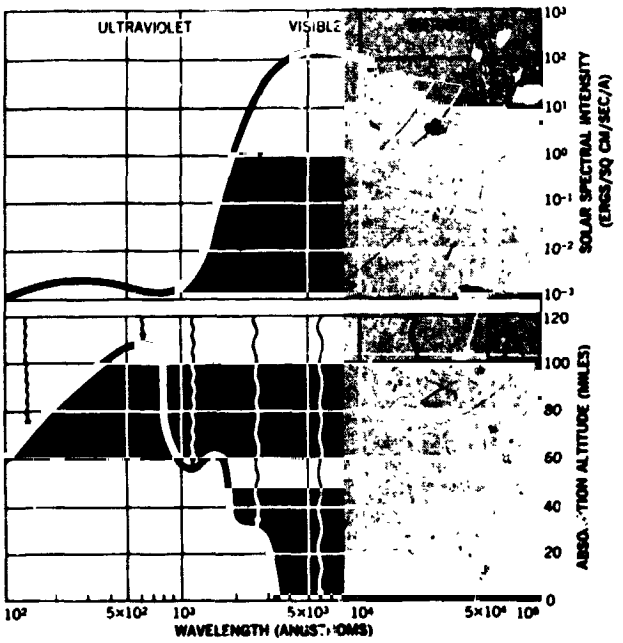
OZONE BREAKDOWN GIVES HEAT SOURCE

Photodissociation of oxygen is produced by the more energetic photons, with wavelengths between 1000 and 2000 A, and is most pronounced at about 60 miles altitude. The less energetic photons, with wavelengths of 2000–3000 A, penetrate even closer to the earth, into the stratosphere; there, they dissociate ozone, which is less tightly bound than oxygen. The temperature maximum at 30 miles attests to the importance of ozone photodissociation as a heat source. There is no corresponding maximum in the temperature profile around 60 miles where oxygen is photodissociated. But while oxygen absorption is not particularly important as a heat source, it does have a profound effect on the chemical composition of both the upper and lower atmosphere.

One result of photodissociation of oxygen is that much of the atmospheric oxygen at high altitudes is in the atomic state. The actual concentration of atomic oxygen is determined by the balance between the rate at which it is produced by photodissociation and the rate at which it recombines to form molecular oxygen. Theoretical treatment has so far failed to provide quantitative information on the relative concentrations of atomic and molecular oxygen in the upper atmosphere. Experimental measurements of atmospheric composition are difficult at high altitudes and at present are of doubtful reliability. Consequently, there is still some uncertainty about the composition of the upper atmosphere.

Below 60 miles, in the lower atmosphere, the atomic oxygen produced by photodissociation combines with molecular oxygen to form ozone. Present only in trace amounts, ozone is only one hundred-thousandth (10^{-5}) of the ambient density—even at 15 miles, where ozone concentration is a maximum.

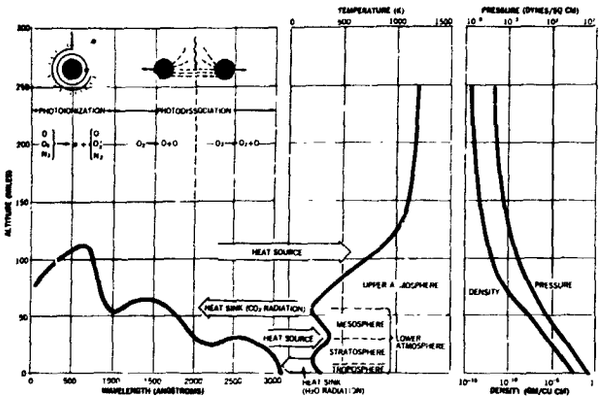
Although ozone is only a trace constituent it plays a very important role indeed. By absorbing strongly in the near ultraviolet (between 2000 and 3000 A) it is not only shields the troposphere and the ground from this biologically harmful solar radiation, but also provides the energy source responsible for the temperature maximum where the stratosphere and mesosphere meet at the stratopause, 30 miles high. The rate of radiation absorption at any altitude is proportional to the product of the concentration of absorbing constituents and the radiation flux. Very little absorption occurs at high altitudes, where the concentrations are small. But by the time ultraviolet reaches low altitudes where concentrations are high, the radiation is attenuated. Here, the rate of absorption is low because the radiation flux is low. The maximum rate of absorption thus occurs at an intermediate altitude. The altitude at which the rate of absorption of any given wavelength reaches a maximum is called the absorption altitude for that wavelength.



SKETCH 3.—Solar radiation controlling the structure of the atmosphere covers the range from extreme ultraviolet to infrared. Most of this radiant energy is in the visible and near-infrared, as the upper curve shows, and most of the energy in this region penetrates to the ground, rather than being absorbed by the atmosphere. Ultraviolet is absorbed at both high and medium altitudes. Comparison of upper and lower curves shows that greatest amount of absorbed ultraviolet is in the 2000–3000 A wavelength range; it is absorbed at altitudes of 25 to 30 miles.

ION-ELECTRON MAXIMUM 180 MILES UP

For extreme ultraviolet radiation of wavelengths shorter than 1000 A, the absorbing constituents are molecular nitrogen and oxygen, and atomic oxygen, and the absorption altitude is around 110 miles. The absorption occurs by way



SKETCH 4.—Variation of temperature with altitude is not undirectional as are pressure and density variations. Reversals in temperature curve are due to endothermic and exothermic reactions, the latter drawing on the energy created by the absorption of ultraviolet radiation. At the highest altitudes (100 miles or so) photoionization occurs when the most energetic ultraviolet photons (wavelengths below 1000 A) knock an electron out of the oxygen or nitrogen atom. The rapid temperature rise above 60 miles reflects this heat source. The temperature minimum at 60 miles is a consequence of the heat sink provided by the emission of thermal radiation by CO_2. The temperature rise of about 30 miles reflects the photodissociation of ozone.

of photoionization, producing free electrons along with the ions of the absorbing constituents. Because the rate of absorption is a function of altitude, it is to be expected that the concentration of ions and electrons will also be a function of altitude with a maximum at some altitude. This is, in fact, the case. Ions and electrons are present in detectable amounts at all altitudes above 40 miles, with the maximum electron and ion concentration occurring at about 180 miles. This layer of ionization is what is called the ionosphere. The region below the 180-mile ionization maximum is called the bottomside ionosphere; the region above is called the topside.

The greatest electron number density of the ionosphere is about 10^6 per cu cm, which amounts to only a thousandth of the neutral particle number density at the same altitude. Thus the ionosphere comprises only a small fraction of the total number of particles in the upper atmosphere.

Because it is central to an understanding of upper atmosphere energy sources, we would like to have a quantitative theory of the ionosphere which would allow us to calculate its properties and the way they vary in time without having to refer the measurements of these properties. We don't yet have such a theory. But, in the past three years, rocket measurements of the solar flux in the extreme ultraviolet region have made it possible, for the first time, to attempt to formulate one.

The theory is based on "airglow" measurements. Absorption of solar ultraviolet radiation by photoionization not only plays a central role in the formation of the ionosphere, but also leads to the excitation of atoms and molecules when ions and electrons from photoionization recombine. These excited species can radiate in their characteristic spectral lines and bands, which lie mostly in the visible region. This is called the airglow.

Because the airglow is excited predominantly in the chemical reactions and particle collision processes which follow the absorption of solar ultraviolet energy, airglow measurements constitute a valuable source of information on these reactions and collision processes and their occurrence in the upper atmosphere. Ground-based measurements of the airglow are fairly straightforward at night, and the first measurements of this kind were made 80 years ago. In fact, such measurements provided the earliest information about conditions in the upper atmosphere. During the day, however, sunlight scattered in the troposphere is many times brighter than the airglow, and it has not been possible until very recently to obtain any daytime airglow data. In the last two years, sophisticated ground-based and rocket techniques have yielded the first daytime airglow measurements. From these measurements we are gaining new insight into the reactions and energy conversion processes which take place at great altitudes when the atmosphere is sunlit.

AIRGLOW IS ENERGY SINK

It has been assumed that the theory of the daytime airglow would involve a simple extension of the theory of the night airglow, but the data now available have made it plain that completely unexpected energy transformation phenomena occur during the day.

We are particulary interested in the airglow because it represents an energy sink for the upper atmosphere. Only a part of the absorbed solar ultraviolet energy is released to the atmosphere

in the form of heat; the rest of the energy escapes down to the ground or out into space in the form of airglow radiation.

We have been discussing ultraviolet radiation as if it were the only source of energy in the upper atmosphere. That this may not be exactly correct will soon be evident as we discuss the energy transport mechanisms and the implications of recent atmospheric density measurements.

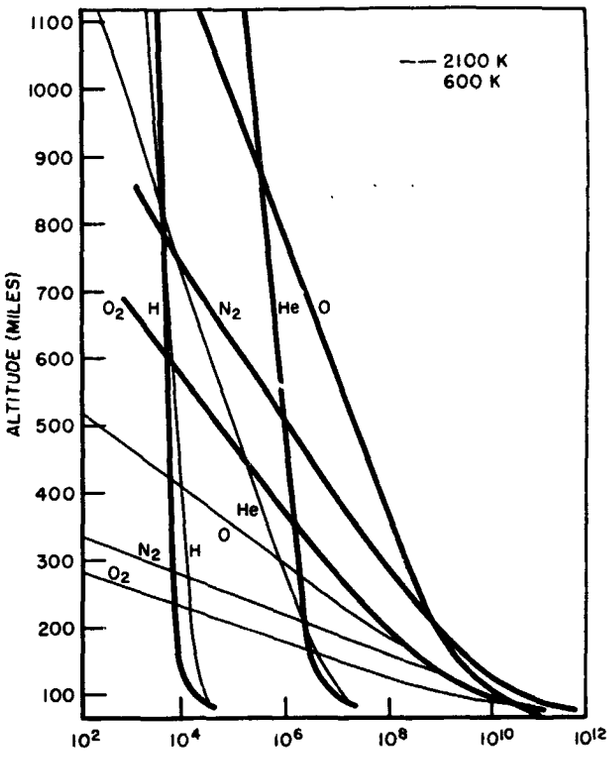
Cooling of the upper atmosphere by the emission of thermal radiation is negligible. The atmospheric molecules which have spectra in the infrared, and can therefore emit thermal radiation at atmospheric temperatures, are water and carbon dioxide. But these molecules are very rare in the upper atmosphere where they are subject to photodissociation by solar ultraviolet radiation.

NO MIXING MECHANISM AVAILABLE

Convection also turns out to be unimportant above 60 miles because temperature increases with altitude, making the atmosphere stable. This inhibition of convection has a profound effect on the composition at great heights. In the absence of convection, and the turbulence which results from convection, there is nothing to keep the atmosphere well-mixed and reasonably homogeneous. Instead, a regime of diffusive equilibrium is maintained. The heavier atmospheric constituents settle out at fairly low altitudes under the action of gravity, while the lighter atmospheric constituents, which are less tightly bound by the earth's gravitational field, rise above them and become increasingly important as altitude increases.

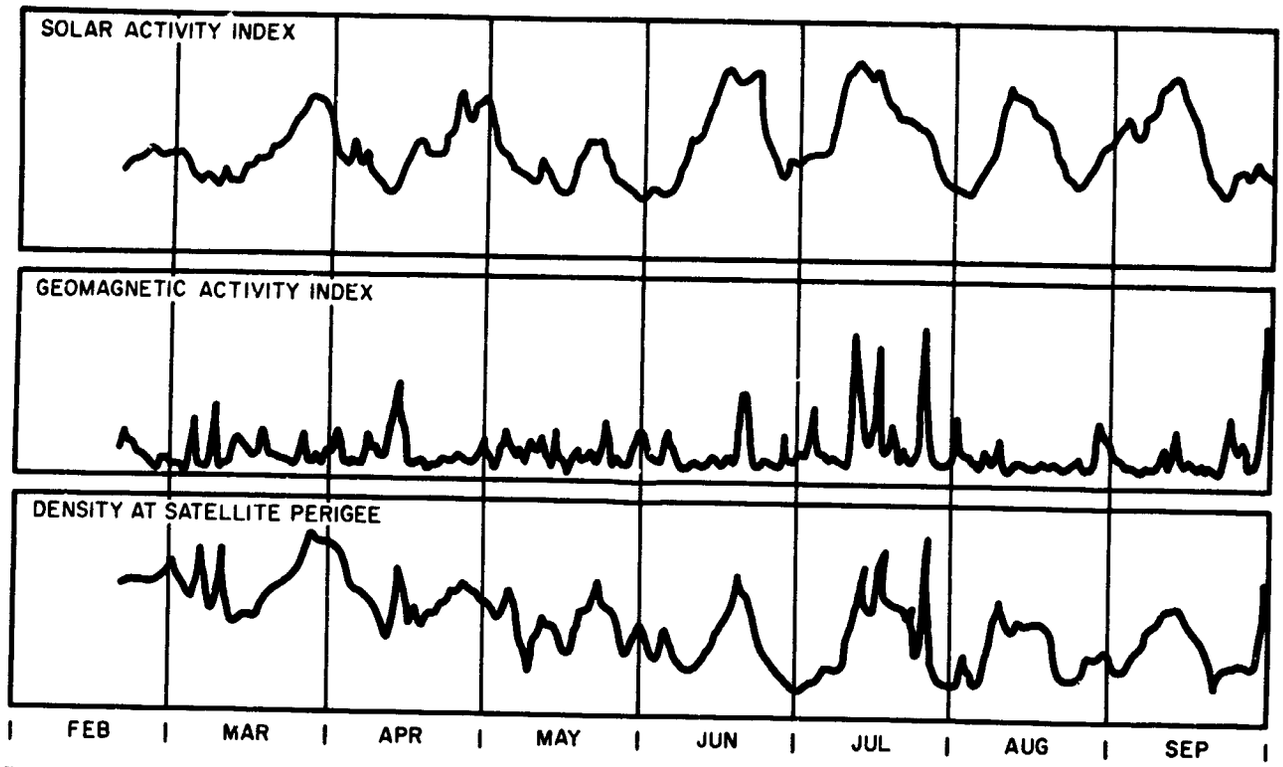
The number densities of these constituents, both light and heavy, decrease with altitude at a rate that depends on the temperature. However, these densities are so low that the temperature at these altitudes cannot be measured directly. Instead, the temperature must be decuced indirectly from measurements of other atmospheric properties. Above 120 miles, atmospheric densities can be deduced from the orbital decay of satellites as the result of atmospheric drag.

Drag studies have been our most prolific source of information on the upper atmosphere, and they have revealed enormous variations of density with time. Because the density at any altitude is sensitive to the temperature, the density variations are believed to reflect variations in temperature that are caused by changing heat sources. Most of the variations can be explained, at least qualitatively, as the result of changes in the ultraviolet heat source caused by changes in the solar ultraviolet flux.



SKETCH 5.—Altitude variation of number density of major constituents in upper atmosphere, plotted for two different temperatures. These curves show how lightest constituents persist to highest altitudes and how rate of fall-off in density is a function of temperature. The relationships, determined theoretically using hydrostatic balance and the ideal gas law, have been confirmed qualitatively by satellite measurements. This confirmation supports the validity of our present indirect method of deducing upper atmosphere temperatures from density measurements.

One phenomenon, however, cannot be explained on this basis and is therefore of particular interest. The satellite drag densities show sharp and sudden increases at times of geomagnetic storms, which show up as spikes in the geomagnetic activity index. Since magnetic storms (unusually large variations in earth's magnetic field, measured at its surface) occur about a day after intense solar flares, it has long been believed that these storms are caused by the impact on the earth's magnetic field of streams of charged particles emitted by the sun at the time of the flare. (These particles



SKETCH 6.—Densities deduced from drag measurements on 12-ft-diameter balloon satellite, Explorer 9 (bottom curve) during February through September 9, 1961, show close correlation with ground-based-measurements of solar, or sunspot, activity (top curve) and disturbance of the earth's magnetic field (center curve). Solar radio flux (10.7 cm), which is used as index of solar activity here, appears to be roughly proportional to flux of extreme ultraviolet radiation which is the heat source (by photoionization) for the upper atmosphere. Curve of temperature at satellite perigree would be essentially identical in shape with density curve.

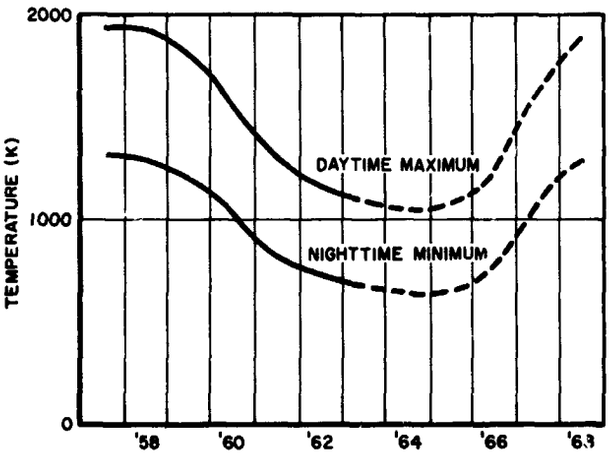
take just about a day to travel from the sun to the earth.) Because of the coincidence between the density spike and the spike in magnetic storm activity, it is presumed that the heat sources which are responsible for the sudden increase in upper atmosphere temperature and density derive their energy from these streams of charged solar particles.

ULTRAVIOLET NOT SOLE ENERGY SOURCE

The geomagnetic activity effect, as this phenomenon is called, provides the strongest evidence we have for an energy source in the upper atmosphere other than that provided by the absorption of solar ultraviolet radiation. Because magnetic storms are comparatively infrequent, this does not constitute evidence for more than an intermittent additional heat source, and it is not yet clear whether this additional heat source (called the corpuscular heat source because it derives its energy from the solar wind) is important in determining the average conditions in the upper atmosphere.

We have mentioned how the inadequacy of our knowledge of upper atmosphere composition may introduce errors in the temperatures deduced from satellite drag data. In an attempt to obtain direct measurements of upper atmosphere temperatures, attention was turned early in the space age to rocket and satellite experiments which could determine the temperature of the electrons in the ionosphere above 60 miles. Measuring the temperature of the electrons in the upper atmosphere is considerably easier than measuring the temperature of the neutral particles (which have no charge), and techniques borrowed from plasma physics have been applied with some considerable success.

It was not clear, when the first measurements were made, whether or not the electron temperature would be equal to the neutral gas temperature. Because the mass of the electron and the



SKETCH 7.—Temperature in isothermal region above 120 miles varies during the solar sunspot cycle. Solid line shows temperature deduced from densities measured by satellite drag. Broken line shows predicted values for remainder of solar cycle (after Harris and Priester). Maximum temperature of 2200K occurs in afternoon during solar cycle maximum. Minimum temperature of 600K at dawn occurs at solar cycle minimum (solar maximum and minimum occur about 11 years apart). Recent discovery of this time variation of temperature is one reason Cospar standard atmosphere model is now undergoing revision.

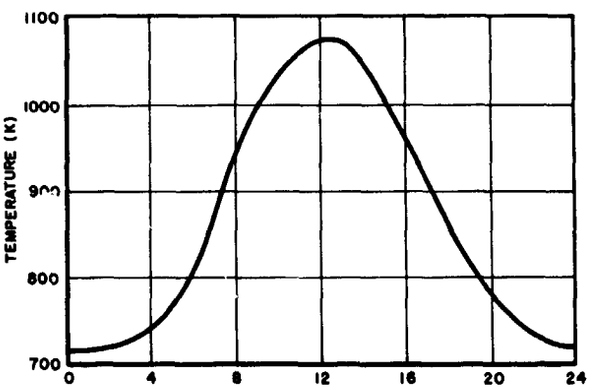
mass of the neutral particles are very different, the electron gas in the upper atmosphere is in poor thermal contact with the neutral particle gas. The early measurements indicated that the electron and neutral gas temperatures were equal above 250 miles but that during the day the electron temperature was considerably higher than the neutral gas temperature at lower altitudes. This finding was in accordance with theoretical studies which predicted that photoionization—concentrated at about 110 miles—heats the electron gas more than the neutral gas.

More recent electron temperature measurements, using the Ariel 1 and Explorer 17 satellites and the ground-based radar backscatter technique, have indicated that the early measurements were misinterpreted. The electron temperature is, in fact, greater than the neutral gas temperature at all altitudes and at all times of day.

This misinterpretation of the early measurements illustrates the danger of trusting the reference atmospheres. The early electron temperature measurements were made in 1960 and 1961 and the results were compared with the gas temperatures contained in the 1962 U.S. Standard Atmosphere—which more or less reflects conditions as they were in 1958. The comparison showed that electron and gas temperatures were approximately the same. However, by 1961 the neutral gas temperature had declined with the decrease in solar activity to values considerably below those given by the standard atmosphere.

So if the early measurements are reinterpreted in the light of our more recent knowledge of neutral gas temperatures, they agree with the latest electron temperature measurements in indicating that electron temperatures are invariably higher than the neutral gas temperatures. It is therefore not possible to determine neutral gas temperatures in the upper atmosphere by measuring the electron temperatures. Just this year, however, it has been realized that the electron temperature data present a more exciting possibility. The electron temperature, it turns out, is so sensitive to variations in the upper atmosphere heat source that it is possible to deduce the absolute magnitude of the heat source from measurements of the electron temperature.

The first work along these lines has been carried out using the Ariel 1 data, and a similar analysis is now being applied to the Explorer 17 results. These analysis provide our first direct information on upper atmosphere heat sources. It is still too



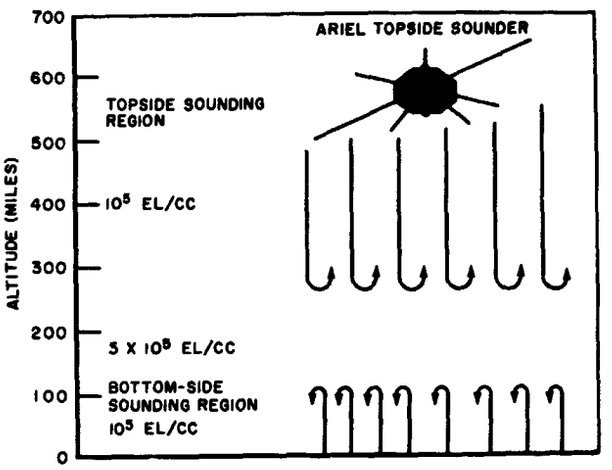
SKETCH 8.—Daily variation of upper atmosphere temperature based on satellite drag measurements of density (after Harris and Priester) in 1962. A recent theoretical treatment of this variation showed it could not be explained on the basis of an ultraviolet heat source; an additional heat source, of approximately equal mangitude, had to be postulated. Presumably this may be a corpuscular source, derived from the solar wind. Very little is yet known about this corpuscular heat source and its importance is controversial. Electron temperature measurements made by Ariel 1 and Explorer XVII may help.

early to tell for sure, but it appears that the daytime heat sources deduced from these data are consistent with the heating that would be expected to result from the absorption of solar extreme ultraviolet radiation. There is no clear evidence in the daytime results of a non-ultraviolet or corpuscular heat source. On the other hand, the electron temperature measurements show clearly that there must be a nighttime source of energy which, while only one percent of the daytime heat source, is still enough to raise electron temperatures a few hundred degrees above the neutral gas temperatures. Since this heat source cannot be attributed to the action of solar ultraviolet radiation, it provides strong evidence for the existence of a non-ultraviolet heat source at night. This nighttime energy source may well be provided by a flux of low energy electrons bombarding the atmosphere from above.

ONLY CONDUCTION IS IMPORTANT

Whether the upper atmosphere heat source is provided by ultraviolet radiation or bombarding electrons or both, the neutral gas temperature is ultimately determined by the rate at which heat transport processes can remove the heat which the sources release. Since both convection and radiation cooling are unimportant, thermal conduction is all that is left to remove heat from the upper atmosphere. In a gas, conduction is a notably inefficient process of heat transport. So it doesn't require a large heat source to maintain the very high temperatures which are found at great altitudes. In fact, the solar extreme ultraviolet radiation, which is believed to provide the major source of heat at high altitudes, carries an energy flux of only about 3 ergs/sq cm/sec. For comparison, the energy flux in the visible region of the solar spectrum—which is responsible for heating the ground and the lower atmosphere—is a million times as large.

The heat which is deposited by the solar ultraviolet radiation at about 110 miles is conducted from this high temperature region down to the very low temperature (200 K) mesopause. At the mesopause a heat sink is provided by infrared-radiating molecules of water and carbon dioxide—which are relatively abundant at low altitudes. The balance between the rate at which heat is conducted to the mesopause and the rate at which heat is put into the upper atmosphere is what determines the temperature of this higher region. The theoretical problem is in principle quite straightforward, although sophisticated computational techniques are made necessary by the fact that the heat sources vary with time and with the density and composition of the upper atmosphere, and the fact that these latter quantities in turn depend on the temperature.



SKETCH 9.—Electron density in ionopshere falls of above and below electron density maximum at 200 miles altitude. Numbers on chart indicate density in electrons per cu cm. Density measurements, made from the ground by radio wave reflections, have long been confined to bottom-side sounding region because no radio waves are reflected to earth from above the density maximum. With advent of satellites capable of making radio soundings from above, electron density measurements of the topside have become possible.

A recent theoretical treatment of the diurnal variation of upper atmosphere temperature showed that the observed temperatures could not be explained on the basis of an ultraviolet heat source alone. It was necessary to postulate an additional heat source, presumably corpuscular, equal in magnitude to the ultraviolet. Although it is too early to say for sure, this conclusion may conflict with the deductions from electron temperature measurements. In any event, we need more positive evidence—not simply a disagreement between experimental and theoretical values of temperature—to prove the importance of the corpuscular heat source.

The problem with postulating a corpuscular source is that the charged particles of the solar wind cannot penetrate beyond the boundary of

the earth's magnetic field at about 40,000 miles altitude and thus cannot interact directly with the upper atmosphere. Although the Mariner 2 measurements have told us what the energy flux in the solar wind is, there have been no similar measurements which could tell us how much of the solar wind energy is able to cross the boundary and travel down through the earth's magnetosphere, to show up as a heat source in the upper atmosphere at altitudes of 100–200 miles.

TURBULENT INTERACTION NOT YET PROVEN

It is presumed that there must be a turbulent interaction between the solar wind and the earth's magnetic field leading to the production of hydromagnetic waves, or fluxes of energetic charged particles, which travel down through the magnetosphere and deposit their energy in the upper atmosphere. But there are no quantitative theoretical treatments of any of these processes which would let us estimate the magnitude of the corpuscular heat source.

Several attempts have been made to locate flaws in the theoretical treatment of the heat conduction equation in the upper atmosphere in the hope that these flaws might account for the disagreement between the measured temperatures and the temperatures calculated with an ultraviolet heat source only. To date none of these attempts have been successful, and the discrepancy stands as one of the most interesting recent results in upper atmosphere physics. What it amounts to is this: we do not yet understand the physics of the upper atmosphere well enough to be able to calculate, from first principles, temperatures which agree with the measurements.

Calculation of the temperature is the fundamental goal of atmospheric structure theory because the temperature profile controls the pressure and density profiles through the requirement of hydrostatic balance. So a calculation of the temperature profile leads at once to the determination of overall atmospheric structure. In addition, because the energy sources responsible for temperature changes also lead to changes in chemical composition, an adequate understanding of the temperature and the energy sources which it reflects necessarily implies a thorough understanding of the chemical composition of the atmosphere. This goal is closely related to the immediate practical problem of predicting the conditions which are likely to be encountered by our space probes in the atmosphere of the other planets.

Selected Bibliography

Upper atmosphere physics has developed so rapidly that textbooks naturally lag behind current developments. One of the few books available is Space Physics, *edited by D. P. LeGalley and A. Rosen; Wiley, '64. / Another book available: J. A. Ratcliffe (ed.),* Physics of the Upper Atmosphere; *Academic Press, '60. / Professional journals are the most useful sources of current information. Jacchia has reviewed satellite drag measurements of upper atmosphere density in* Reviews of Modern Physics; *Vol. 35, pp. 973-991, '63. / The results of inospheric experiments using rockets and satellites have been described by Bourdeau in* Space Scienec Reviews; *Vol. 1, pp. 683-728, '63. / Readers seeking more details on current research will find the* Journal of Geophysical Research, Planetary & Space Science, *and* Space Research *journals most useful.*

N64-19010

INFERENCES OF STRATOSPHERIC AND MESOSPHERIC CIRCULATION SYSTEMS FROM ROCKET EXPERIMENTS*

G. WARNECKE† AND W. NORDBERG

Recently performed rocket grenade experiments at Wallops Island and Churchill and MRN data are interpreted to show the existence and movement of conventional circulation systems throughout the upper stratosphere and mesosphere. Evidence of moving pressure systems exists up to an altitude of 70 km where the nature of the circulation changes abruptly. Also, the pattern of the pressure systems in the stratosphere (below 50km) seems to be different from the pattern inferred at higher altitudes (50—70 km.)

INTRODUCTION

Since the early rocket grenade experiments more than ten years ago and the successful firings during the IGY period, the experiments have been continued by NASA at Wallops Island. Thus a large number of soundings is available giving already a climatological survey on the gross features of the winds and the thermal structure in the upper stratosphere and in the mesosphere (1) (2). Experiments have also been performed in Australia, Japan, Sweden, Italy, and France and the United States program will be extended to include simultaneous launchings at Churchill, Wallops Island, and Ascension Island. (3) The number of rocket stations reporting data above the normal Meteorological Rocket Network (MRN) levels is gradually increasing promising an improving global data coverage for the levels between 50 and 90 kilometers during the forthcoming International Geophysical Year (IQSY). In this report, interesting features are derived from the first simultaneous rocket grenade experiments performed at Churchill, Manitoba, and Wallops Island, Virginia, and from simultaneous Meteorological rockets launched at the various MRN sites over North America. While the earlier sporadic experiments enabled a study of climatological gross features of the mesospheric circulation, these simultaneous observations have encouraged us to investigate the dynamic structure of the mesosphere on a smaller scale by means of synoptic presentation of flow patterns for selected days over the North American continent. At the higher altitudes, the acoustic temperature and wind measurements were supplemented by other sounding techniques such as wind measurements by means of sodium vapor release.

THE VERTICAL CHANGE IN STRATOSPHERIC AND MESOSPHERIC WIND STRUCTURE

Considering the seasonal variations of stratospheric and mesospheric winds as measured at Wallops Island over a period of 3 years, one finds a remarkable change in the behavior of mesospheric winds around the height of 70 km. Below this level, the wind follows a consistent and predictable seasonal pattern while above 70 km the wind structure becomes very irregular and does not show the well-known regular features observed at levels up to 70 km (2).

This fact is apparent in figure 1, where all rocket grenade winds for the years 1960 through 1963 are presented in polar diagrams for the lower, middle, and upper mesosphere. The winds in the 45–55 km layer show the same general features as in the stratosphere: strong and relatively steady westerly winds with maximum deviations of $\pm 25°$ about an average direction of 265° and light but equally steady easterly winds in summer. During the transitions between winter and summer sea-

*Published as *Goddard Space Flight Center Document X-651-64-22*, January 1964; also published in Proceedings of 5th International Space Science Symposium, Florence, Italy, 1964, North Holland Publishing Company, Amsterdam.

†Goddard Space Flight Center; on leave from Institut für Meteorologie and Geophysik der Freien Universität Berlin, Berlin, Germany, as National Academy of Sciences-National Research Council Research Associate with National Aeronautics and Space Administration.

sons, there is a large variability in wind strength and direction; however, this variability is still part of a consistent seasonal pattern which has been the subject of several previous analyses (4). In the middle mesosphere (60–70 km) one finds qualitatively the same behavior, although varia-

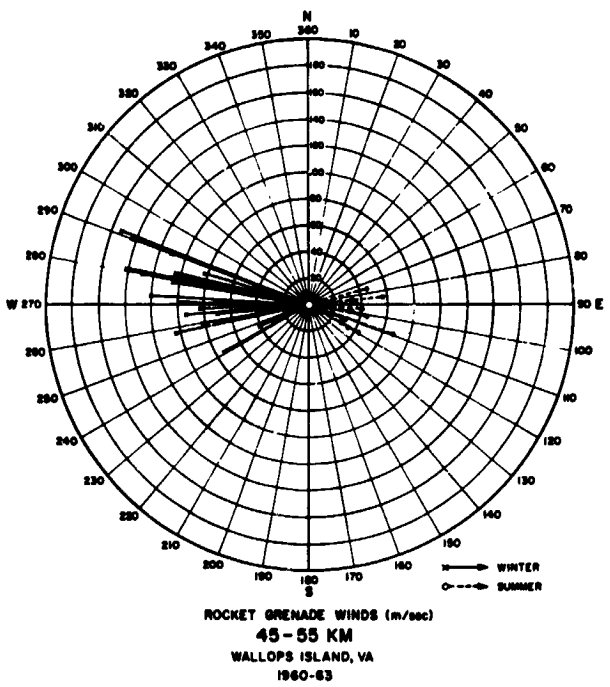


FIGURE 1A.

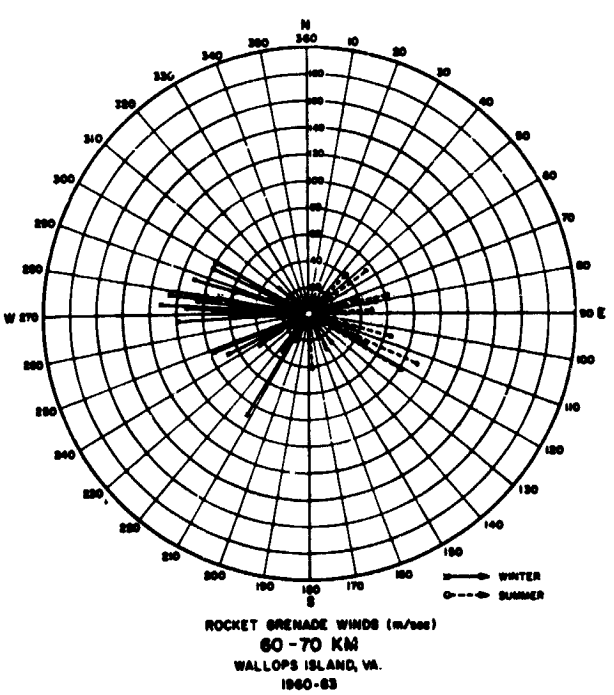


FIGURE 1B.

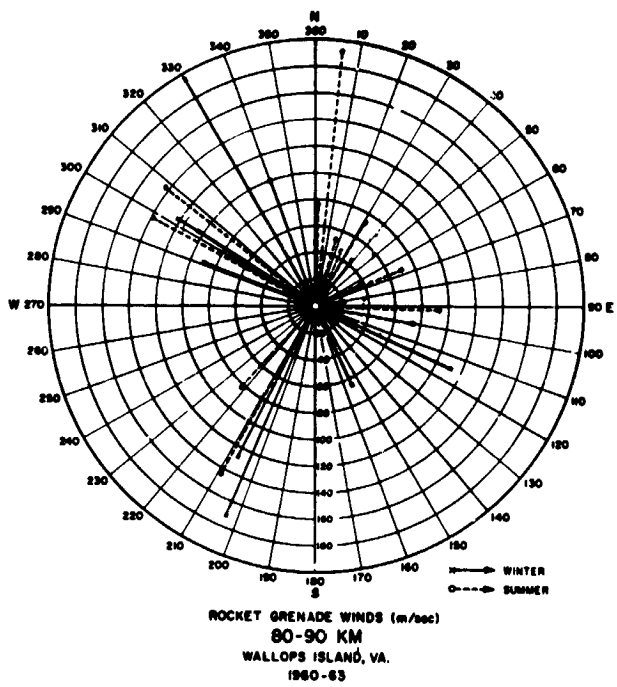


FIGURE 1C.

tions in the wind directions are somewhat larger. In the upper mesosphere (80–90 km), however, the picture changes completely. The winds both in winter and summer are blowing from all directions and in highly variable strengths. The same characteristics are shown by the results of recent sodium vapor release experiments at Wallops Island (5). In figure 2 wind direction profiles from five experiments reaching below 70 km are reproduced, and the drastic change in wind structure between 70 and 80 km is apparent.

From these rocket observations with both the grenade and sodium release methods, one may conclude that above a transition layer between 70 and 80 km the dynamic behavior of the mesosphere becomes so complex that synoptic representations of the flow patterns on a Time scale of days will be meaningless. On that scale, the motions above this layer appear to be highly "disorganized" and it seems that for synoptic presentations Time scales of the order of an hour or so will have to be employed at altitudes higher than 80 km. Therefore, such an abrupt change in the wind structure may very well be due to the increasing importance of tidal motions and possibly in the fact that the atmospheric structure as a whole undergoes basic and major changes at the 80 km level.

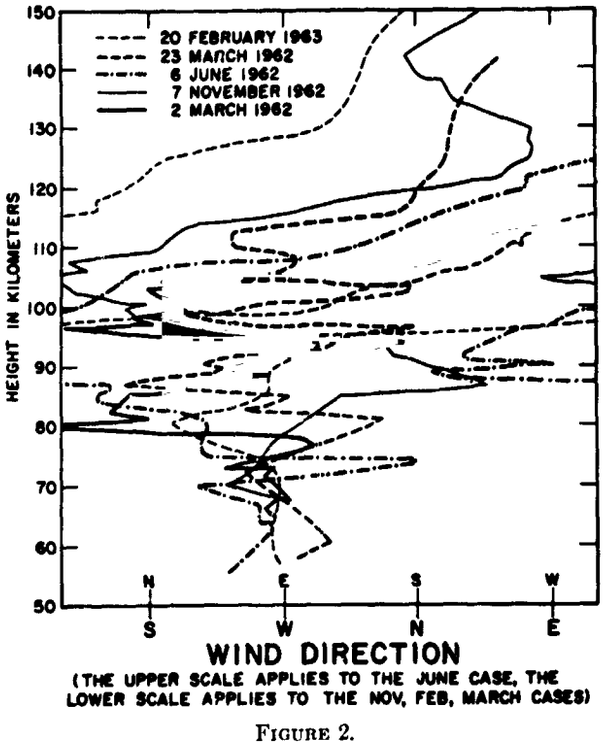


FIGURE 2.

Nevertheless, up to this level the mesospheric motions are still organized in the usual meteorological sense, and synoptic presentations are quite justified and well representative of the strea patterns over periods of several days.

INFERRED MESOSPHERIC SYNOPTIC CIRCULATION SYSTEMS UP TO 70 KILOMETERS

Several investigators have already extended presentation of synoptic weather maps up to the 0.5 or 0.4 mb level (about 55 km) by using Meteorological Rocket Network data (6), (7), (8), (9). In this paper, an attempt is made to construct synoptic maps up to the 0.05 mb level (68 km) for a few special days by making use of recent grenade and MRN data.

Data from the first simultaneous rocket grenade soundings at Churchill and Wallops Island, in early December, 1962, (10) were analyzed in addition to Meteorological Rocket Network data (11) also available for that time period. Variation of flow and pressure patterns for various upper stratospheric and mesospheric levels over the American continent were investigated by constructing synoptic maps for 4 December and 6 December, 1962. In these maps, the pressure fields at constant heights are presented. Using the geostrophic approximation, the isobars were spaced such that each pressure field is in accordance with the observed winds. The pressure values of the isobars were determined by the observed pressures over the discrete points of observation.

The circulation in the lower stratosphere (100 mb or 16 km) around 5 December, 1962, was characterized by the displacement of the cold polar vortex to Siberia, the entire North American continent being under a rather uniform westerly stratospheric flow with a slowly developing trough moving gradually eastward over the United States (figures 3) (12). The middle stratosphere (30 mb or 23 km) (figure 4) shows a quite different stream pattern over the entire western hemisphere (13) with a weak but, nevertheless, unusual quasi-stationary anticyclone over Canada and a zonally oriented trough over the United States along 40 deg. latitude. In the upper stratosphere (10 mb or 30 km) the same contour pattern appears with greater intensity: the easterly wind at the southern flank of the Canadian anticyclone exceeds 25 m/sec (50 knots), whereas a stratospheric jet stream with winds up to 50 m/sec (100 knots) is located over the Gulf coast (figure 5). There are no significant changes of the circulation pattern at 30 km within the period of December 4–6, 1962, so that figure 5 applies to this interval. The North American high pressure systems represent an eastwardly displaced Aleutian anticyclone that governed the Northern hemispheric circulation throughout November (14) and is still apparent at the 5 mb surface, which was constructed for 5 December, 1962, (figure 6). The 40 km pressure maps (figure 7) show the pronounced trough over the Great Lakes area moving slowly southward, whereas the existence of the Canadian anticyclone is not necessarily indicated at this level. Its center may well have shifted to the Pacific region, Alaska and Canada being in the regime of the polar vortex. The anticyclone can still be observed ten kilometers higher (figure 8), and here there is a definite indication that its center has shifted northward, causing a nearly northerly wind over Churchill on 4 December and a southwesterly wind over Alaska. Although the stream pattern remains essentially the same at

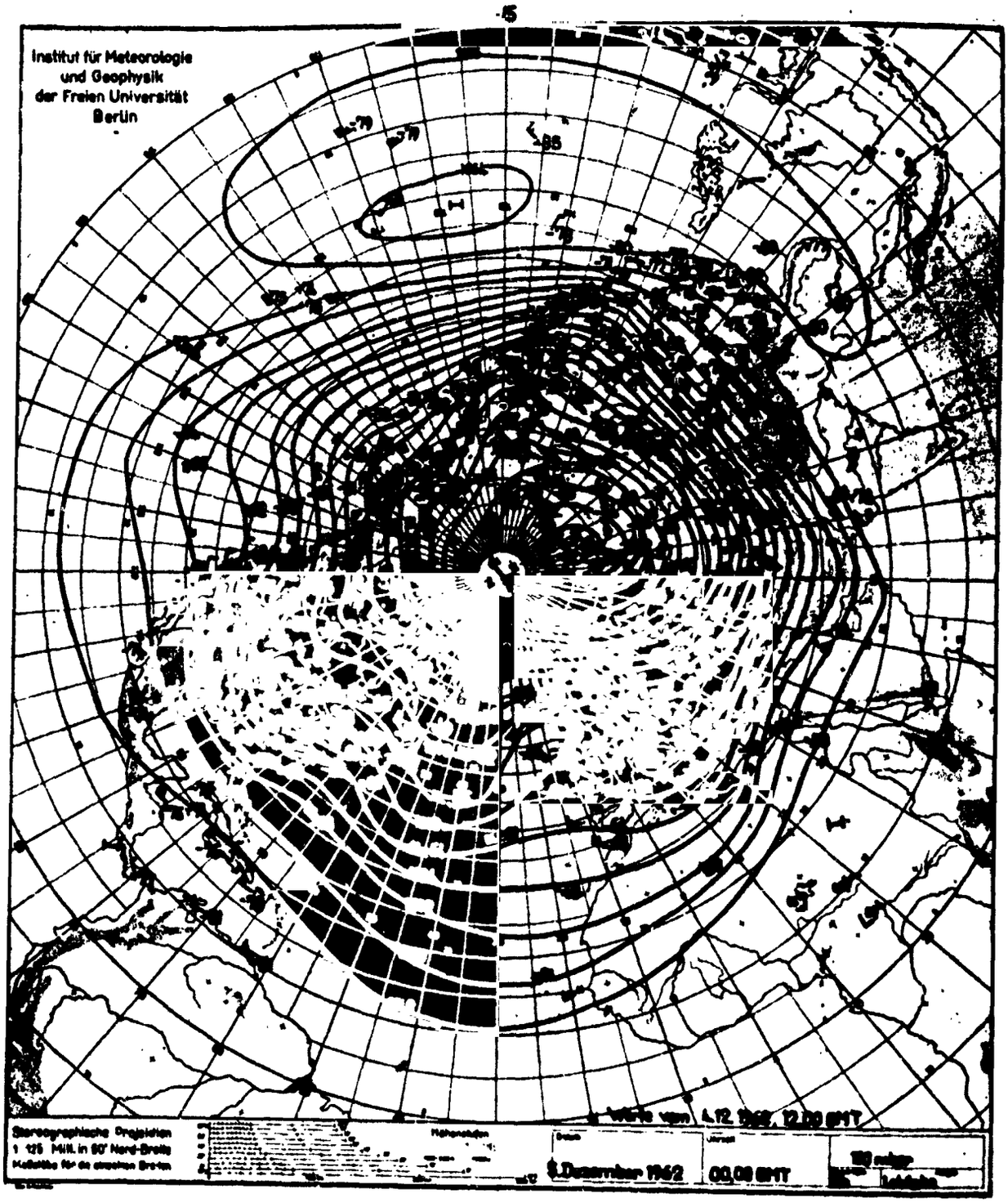


FIGURE 3

60 km (figure 9), there is a distinct further shift in the position of the center of the anticyclone. It is apparently displaced to the Arctic, as indicated by the southeasterly wind over Alaska. The pronounced trough, which has persisted at all altitude levels from 23 km up, is still located over southern Canada; and there is some indication that its vertical axis is slightly inclined to the

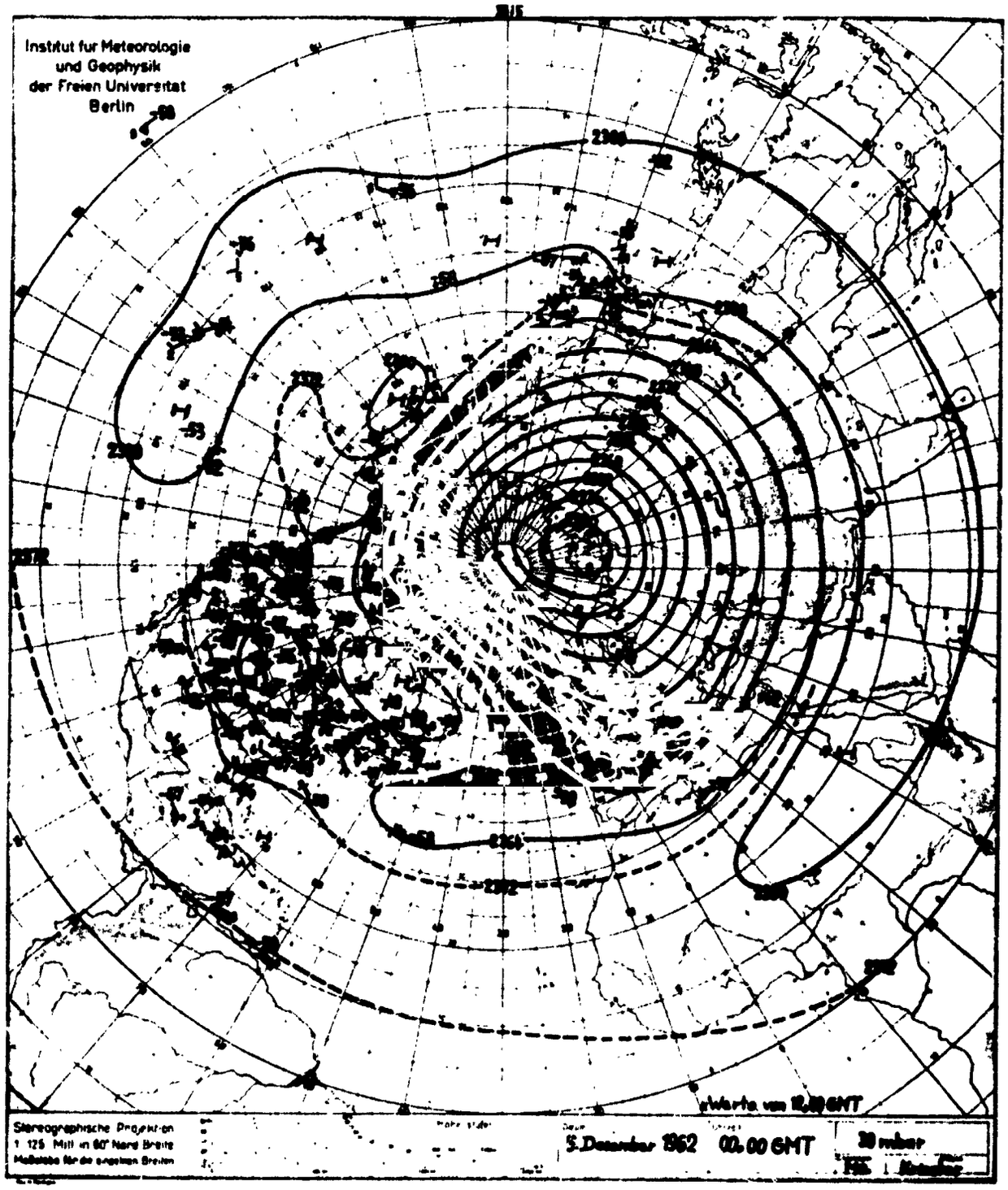


FIGURE 4

north, such that at 60 km the trough has taken the position of the ridge in the stratosphere (20 to 35 kilometers).

At 60 km (figure 10) data are very sparse and are mainly based on grenade soundings at Churchill and Wallops Island. There is evidence that the zonally oriented trough over North America still exists, as indicated by strong winds from the

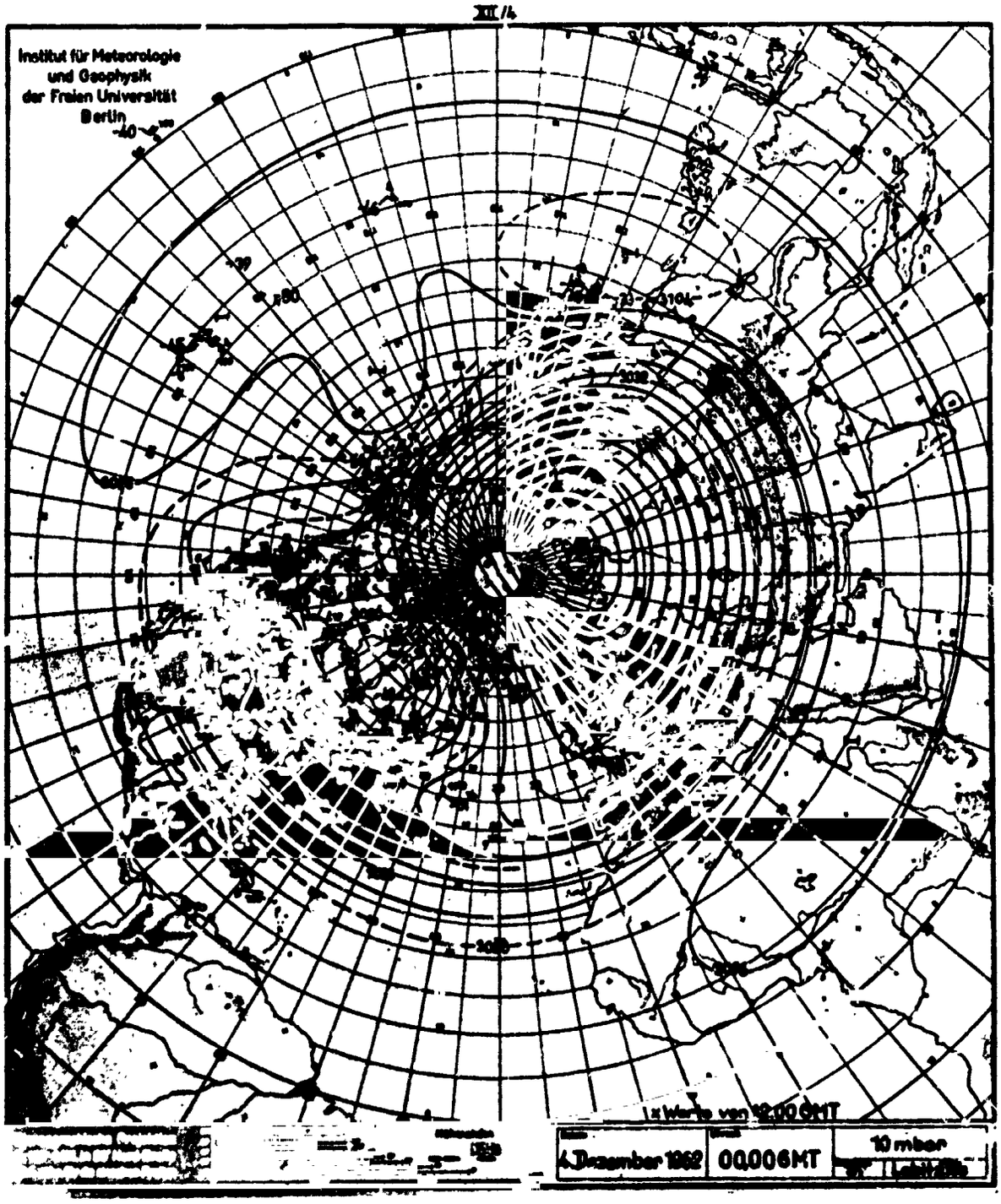


FIGURE 5

west south-west over the United States and very light northwesterly wind inside the trough over Churchill on 4 December, 1962. In general, it is interesting to note the alternation between low and high pressure systems on a vertical scale between the 30 and 60 km levels as shown in figures 5–10. This seems to confirm the theoretical expectation of such alternations previously expressed by Paetzold (15).

Differences in the flow and pressure patterns between 4 December and 6 December were found to be insignificant at 40, 50, and 60 km. This is

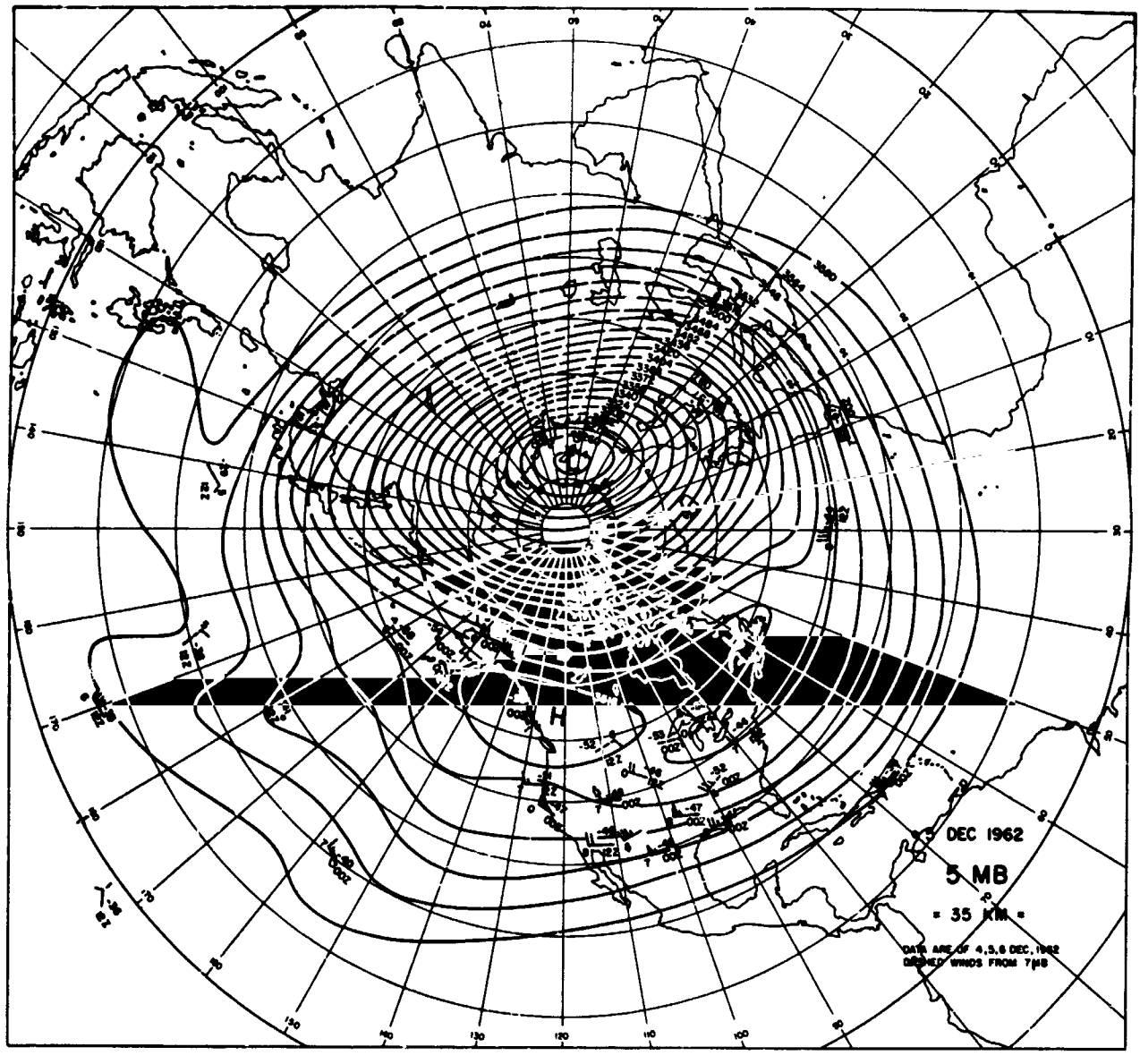


FIGURE 6

not surprising because the patterns in the stratosphere also remained nearly unchanged. At 68 km, however, a significant change seems to have taken place between 4 December and 6 December over Churchill and 1 December and 6 December over Wallops Island. There is a 90 degree rotation of the wind vectors at both stations, but at Churchill this rotation is clockwise, while at Wallops Island it is counter clockwise. This may be interpreted as a southwestward motion of the low pressure system (figures 10A, 10B).

Of course, the sparsity of data permits only a very crude analysis even in this case, which could be considered well documented by rocket sounding standards. Nevertheless, this case shows that significant nonuniformities exist in the slopes of the constant pressure surfaces which are generally inclined downward toward the winter pole. If in the past such uniformity was assumed, it was only because the available data were not sufficient to observe any detail. It is to be expected that systematic rocket observations of the mesosphere extended over the hemisphere and spaced over distances in the order of 1000 km will reveal up to about 70 km circulation systems of similar variability and variety as observed at lower altitudes.

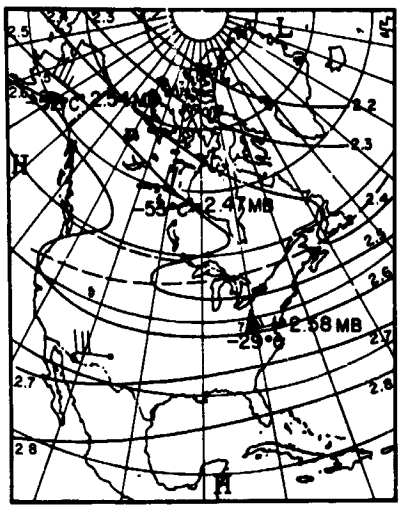


FIGURE 7A.

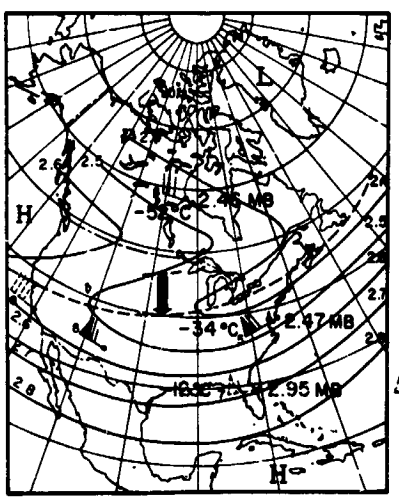


FIGURE 7B.

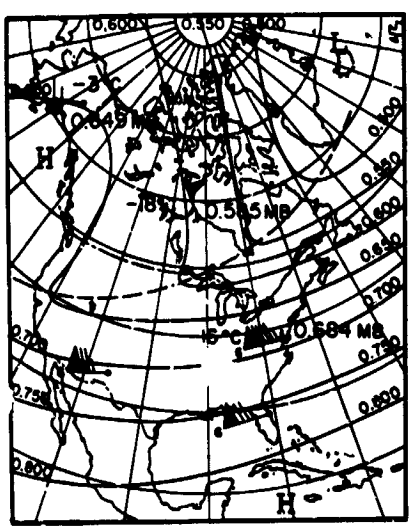


FIGURE 8A.

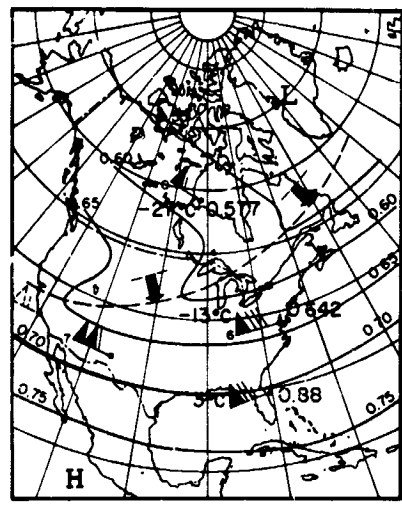


FIGURE 8B.

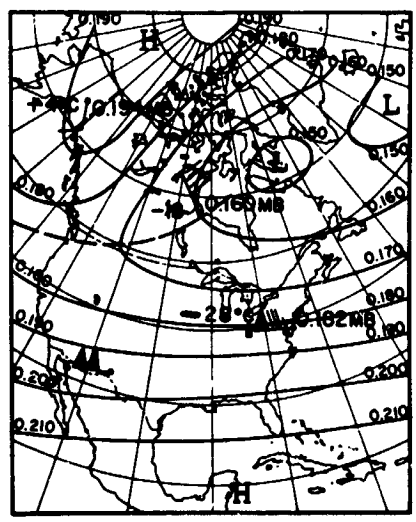


FIGURE 9A.

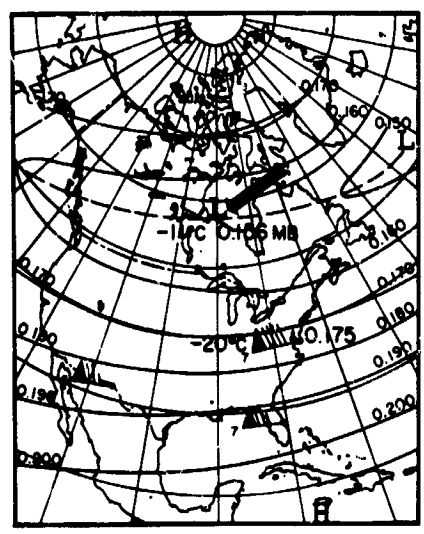


FIGURE 9B.

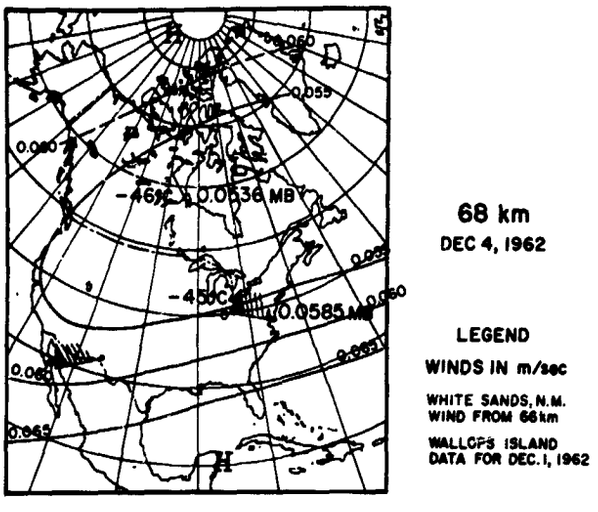


FIGURE 10A.

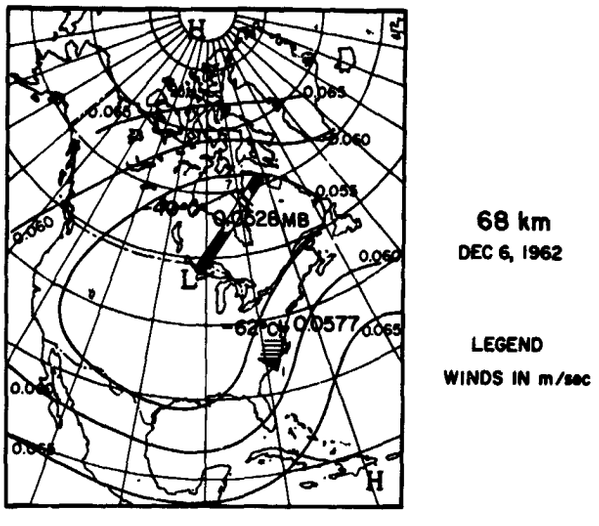


FIGURE 10B.

REFERENCES

1. NORDBERG, W., and STROUD, W., "Seasonal, Latitudinal and Diurnal Variations in the Upper Atmosphere," NASA, Technical Note TN D–703, Washington, D. C., April 1961.
2. NORDBERG, W., and SMITH, W., "Rocket Measurements of the Structure of the Upper Stratosphere and Mesosphere," Institut fur Meteorologie und Geophysik der Freien Universitaet Berlin, Meteorologische Abhandlungen, Band XXXVI (1963) p. 391–400.
3. NORDBERG, W., "Rocket Soundings in the Mesosphere," Astronautics and Space Engineering, Vol. 2, May 1964, American Inst. for Aeronautics and Astronautics, Publisher.
4. MIERS, B. T., "Zonal Wind Reversal between 30 and 80 km over Southwestern United States," Journal Atm. Sciences, Vol. 20, pp. 87–93, (1963).
5. MANRING, et al., "An Experimentally Determined Model for the Periodic Character of Winds from 85 to 35 km," GCA Technical Rep. No. 63–27–N Contract No. NAS W–712, Geophysics Corporation of America, November 1963.
6. SCHERHAG, R., "Eine erste synoptische Hoehenwetterkarte der 0.5 mbar-Flaeche," Berliner Wetterkarte, Beilage 33, Mai 1961 (1961).
7. WARNECKE, G., "Eine synoptische Hoehenwetterkarte fuer das 0.5 mbar-Niveau vom Sommer 1960," Berliner Wetterkarte, Beilage 37, Mai 1961 (1961).
8. KEEGAN, T. J., "Synoptic Patterns at 100,000 to 200,000 Feet," Air Force Surveys in Geophysics No. 140 AFCRL–62–273 (II), 195–210 (1962).
9. FINGER, F. G., and TEWELES, S., "Synoptic Charts Based on Rocketsonde Data," Institut fur Meteorologie und Geophysik der Freien Universitaet Berlin, Meteorologische Abhandlungen, Band XXXVI (1963), pp. 379–390.
10. SMITH, W., KATCHEN, L., SACHER, P., SWARTS, P., THEON, J., "Temperature, Pressure, Density and Wind Measurements with the Rocket Grenade Experiment, 1960–1963, NASA, Goddard Space Flight Center, May 1964.
11. "Data Report of the Meteorological Rocket Network," December 1962. Prepared by U. S. Army Signal Missile Support Agency. White Sands Missile Range, New Mexico.
12. SCHERHAG, R. et Coll., "Taegliche Hoehenkarten der loo mb Flaeche sowie monatliche Mittelkarten fuer das Jahr 1962," Institut fuer Meteorologie und Geophysik der Freien Universitaet Berlin, Meteorologische Abhandlungen, Band XXIII, Teil IV.
13. SCHERHAG, R. et Coll., "Preliminary Daily Northern Hemisphere 30 mb Synoptic Weather Maps of the Year 1962," Institut fuer Meteorologie und Geophysik der Freien Universitaet Berlin, Meteorologische Abhandlungen, Band XXV, Part IV.
14. SCHERHAG, R. et Coll., "Preliminary Daily Northern Hemisphere 10 mb Synoptic Weather Maps of the Year 1962," Institut fuer Meteorologie und Geophysik der Freien Universitat Berlin, Meteorologische Abhandlungen, Band XXVI, Part IV.
15. PAETSOLD, H. K., "On the Problem of Solar Activity Effects on the Lower Terrestrial Atmosphere," Institut fur Meteorologischie und Geophysik der Freien Universitat Berlin, Meteorologische Abhandlungen, Band XXXVI (1963).

VIII. GENERAL

N65 37970

"EXAMPLES OF CERTAIN DATA REDUCTION AND MAPPING PROCEDURES UTILIZING TIROS III FIVE-CHANNEL RADIOMETER DATA"*

LEWIS J. ALLISON AND GUENTER WARNECKE†

Selected examples of current data reduction and mapping procedures are discussed for the purpose of demonstrating their utility as well as their limitations.

INTRODUCTION

In order to analyze and accurately interpret the TIROS radiation data, certain data reduction and mapping processes are described which should be of interest to the researcher and meteorologist alike.

A brief description of the 5-channel scanning radiometer is included for the benefit of the reader who may be unfamiliar with the TIROS radiation experiment.

1. THE FIVE CHANNEL SCANNING RADIOMETER

The TIROS III scanning radiometer consists of a five-channel instrument. The five channels are sensitive in the following spectral regions:[1]

Channel	Spectral Region	
1	6.0 to 6.5μ	(radiation within the ν_2 fundamental water vapor band)
2	8 to 12μ	(radiation emitted in the atmospheric "window")
3	0.2 to 6μ	(total reflected solar radiation (albedo))
4	8 to 30μ	(total infrared radiation of the earth-atmosphere system)
5	0.55 to 0.75μ	(reflected solar radiation within the visible part of the spectrum)

The 5-channel radiometer employs a chopper that causes each sensor to view alternately and at a rapid rate, in two directions 180 degrees apart. At least one direction always views outer space which serves as a zero radiation reference. The response from each channel is proportional to the difference in the irradiation of the sensor bolometer from the two directions. When both optics view outer space, the resultant signal is designated as the "space viewed level". The bi-directional axes of the channels are parallel to one another and inclined to the spin axis by 45 degrees (Figure 1).[2]

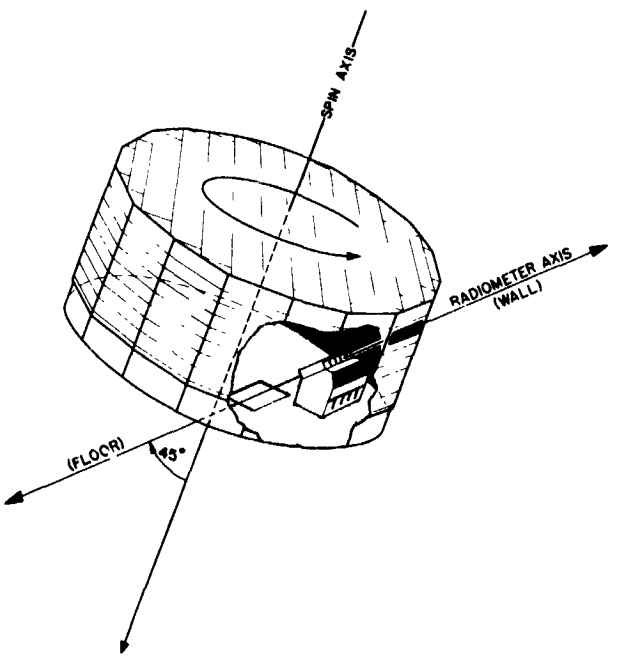


FIGURE 1.—Position of the Tiros Radiometer within the satellite.

*Published as *Goddard Space Flight Center Document X-651-64-132*, April 1964.

†Goddard Space Flight Center; on leave from the Institut für Meteorologie und Geophysik der Freien Universitat Berlin, Berlin, Germany as National Academy of Sciences – National Research Council Research Associate with the National Aeronautics and Space Administration.

2. REMARKS ON THE DATA REDUCTION AND MAPPING PROCEDURES

2.1 Equivalent Blackbody Temperature

The calibration of the infrared channels is in terms of equivalent temperatures, T_{BB}, of a blackbody target filling the 5° field of view, whereas the calibration of the solar channels is in terms of that portion of radiant emittance from a target filling the field of view to which the channel responds through its filters and other optical elements.[2]

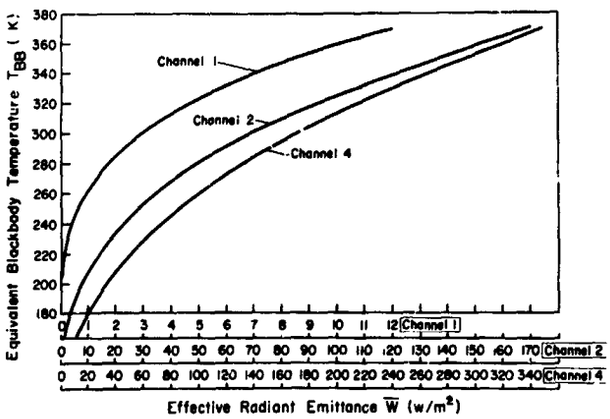


FIGURE 2.—The effective radiant emittance of the thermal channels of the Tiros III radiometer versus equivalent blackbody temperature.

The family of curves of equivalent blackbody temperatures vs effective radiant emittance for TIROS III (Figure 2), were obtained by integrating the Planck function over the spectral response curves for Channels 1, 2 and 4.[1,3] The estimated relative accuracy of equivalent blackbody temperature for Channels 1, 2 and 4 is ±2°K, while the estimated absolute accuracy is ±4° to ±8°K, up to orbits 118, 875 and 130 respectively, after which degradation of the instrumental response becomes appreciable.[1,4]

A degradation correction nomogram for Channel 2 measurements (Figure 3), based upon degradation models inferred from the relevant data obtained by the respective radiometer has been included.[4] The nomogram was developed from a computer program which determined the time rate of change of the measured total emitted radiant power between latitudes 55°N–55°S. The results showed that Channel 2 degraded continuously following the TIROS III launch. The degradation for Channel 2 was determined to be symmetrical, i.e., independent of the side (wall or floor aperture) of the satellite. The nomogram (Figure 3) may be used with some degree of confidence up through orbit 875. Beyond this subjectively determined limiting orbit, the correction nomogram becomes correspondingly uncertain. A more detailed discussion of the factors effecting the response of the scanning radiometer may be found in Reference 5.

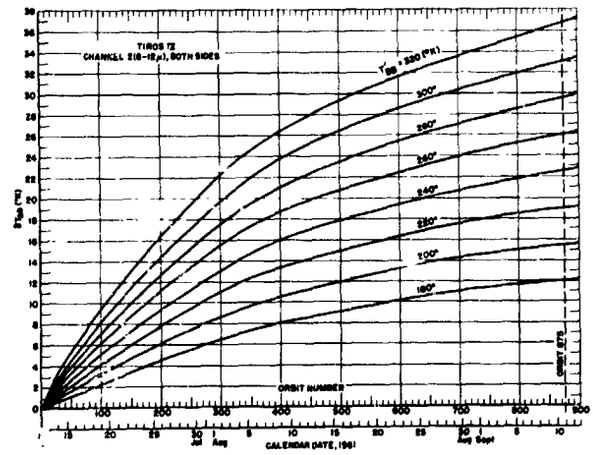


FIGURE 3.—The empirically-developed correction nomogram for instrumental degradation of Channel 2, Tiros III.

2.2 The Radiation Data Processing Procedure

The medium resolution radiometer data that are recorded at the read-out station are demultiplexed, demodulated and fed to an analog-to-digital converter at the Goddard Space Flight Center.[2] This converter produces a magnetic "Radiation Data Tape" made up of 36 bit words suitable for an IBM 7094 computer. Analog records may be produced on a recording oscillograph for special hand analysis. The computer produces a Final Meteorological Radiation Tape (FMRT) when combined with the inputs from the Radiation Data Tape, Orbital Tape and Calibration. The most convenient method of presentation of these FMRT data is to make grid point maps from an IBM 7094 computer. Several problems arising in the composition and averaging of the original radiation data are discussed below.

2.3 Radiation Data Resolution

The data resolution of computer produced radiation maps, printed on a mercator map base is directly affected by the following:

2.3.1 Nadir Angle Variation

Radiation data contained on the FMR Tapes are sampled approximately 8 times per second. Thus, at the nominal 10 rpm spin rate, approximately 48 samples are taken during each satellite rotation. Data sampled during the earth-viewed portion of the scan having the smallest nadir angle will be the most densely located, with progressively greater separation occurring at the higher nadir angles. Additionally, the size of the scan spot varies considerably with the nadir angle and with the height of the satellite. For example, at an 800 km satellite height (Figure 4) the elongated scan spot resolved by the radiometer at a 56.6° scan nadir angle, would enclose approximately 10 times the scan spot area at a scan-nadir angle of zero.[6] Thus the higher the radiometer nadir angle, the lower the data resolution.

Figure 5 shows that the nominal 5 degree field of view of the radiometer pertains to the 50 percent of peak-response level, and that there is some residual response to about ±4 degrees from the optical axis.[7] The percent power scan spots of Figure 4 are directly derived from this response curve.

2.3.2 Grid Print Map Scale

The mesh size of the multi-resolution mercator mapping program is 0.5 inches. Thus, the number of degrees of longitude per mesh interval is in effect the selection of the map scale. A resolution of 0.625°, 1.25° and 2.50° of longitude per mesh interval corresponds to a map scale of approximately 1:5 million, 1:10 million and 1:20 million, respectively, near the equator. On a mercator projection the map scale, the higher the

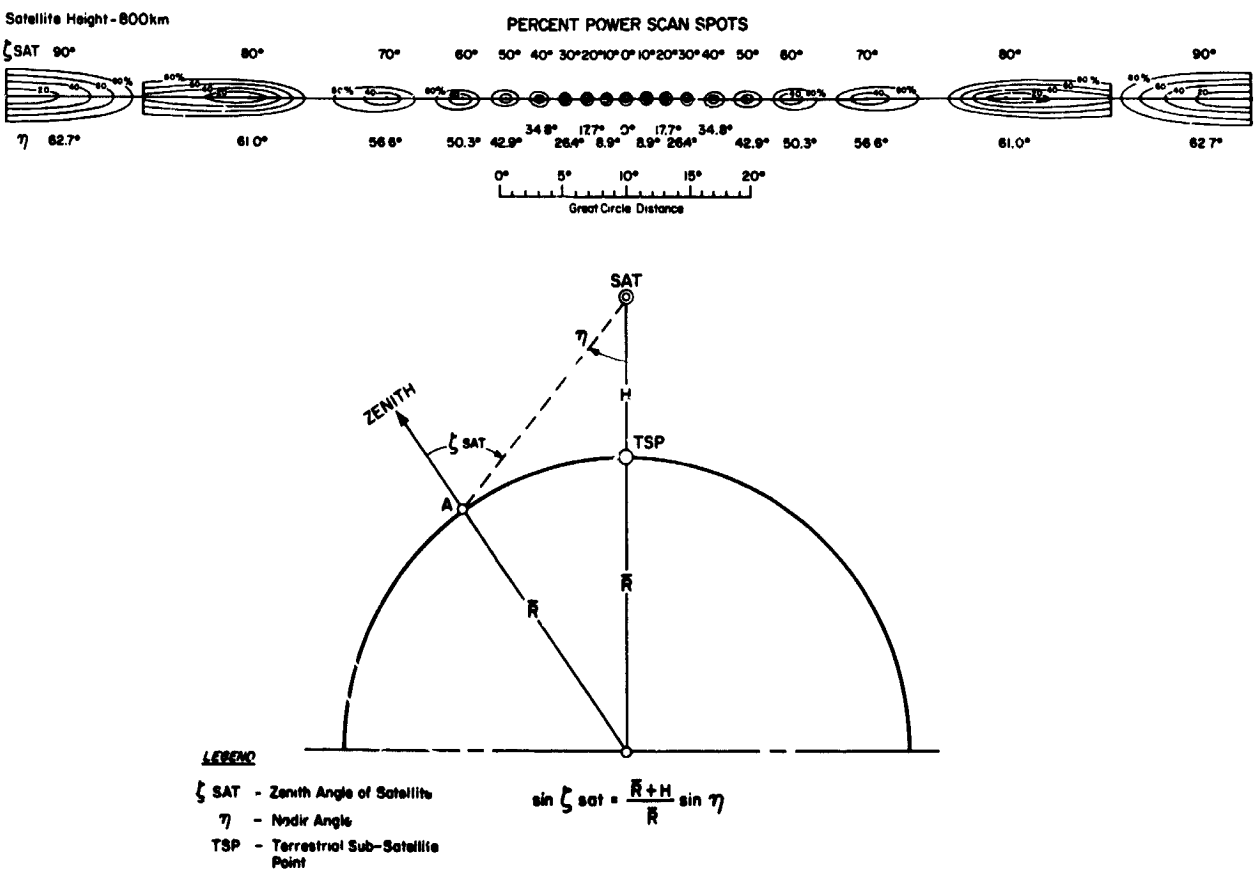


FIGURE 4.—Satellite zenith and nadir angle relationship. a.—Scan spot size as a function of zenith and nadir angle of the sensor and the power distribution within the scan spot. b.—Zenith angle and nadir angle relationship.

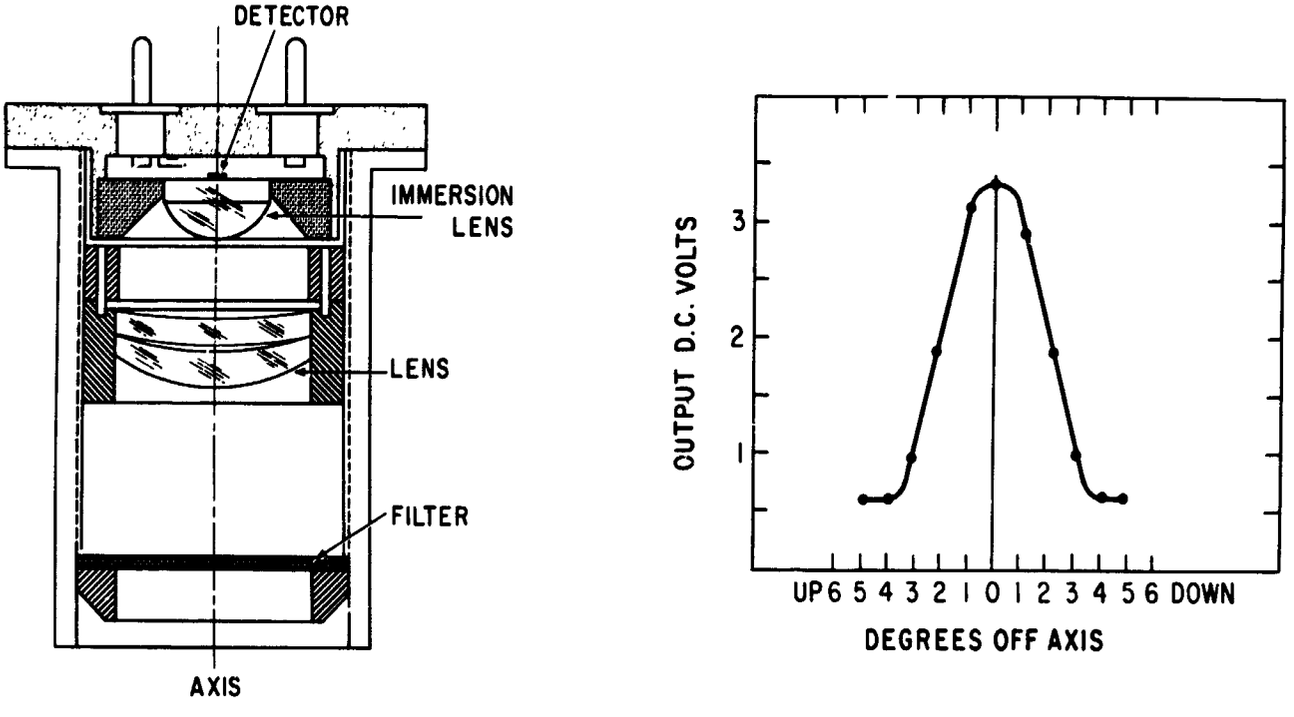


FIGURE 5.—Cross-section through a detector unit of the Tiros III radiometer and response curve within the field of view

resolution, the lower the number of measurements averaged with the ½ inch area around each grid point.

In order to illustrate the effect of map scale on data resolution, several TIROS III television pictures and radiation charts are shown in the following. The well-defined cloud system at 10°N, 40°W on July 16, 1961, which was identified by S. Fritz[8] as being the cloud mass from which Hurricane Anna, later developed in July, 1961, was selected as the example.

Figure 6a shows the "window" radiation analysis based upon hand-plotted data from a computer listing, which contains the geographical location, the nadir and azimuth angles, and the simultaneous readings of all five radiation channels for every single scan spot. The location of each scan spot is indicated by a cross in Figure 6a. This is the most detailed information available from the FMR Tape.

Figure 6b is a composite of the "window" radiation field shown in the previous figure and the gridde TIROS photograph of the same area. The latitude-longitude grid of this picture was obtained from a computer program.

The agreement between the picture information and the radiation field in the "window" region is surprisingly good, even at the fringes of the

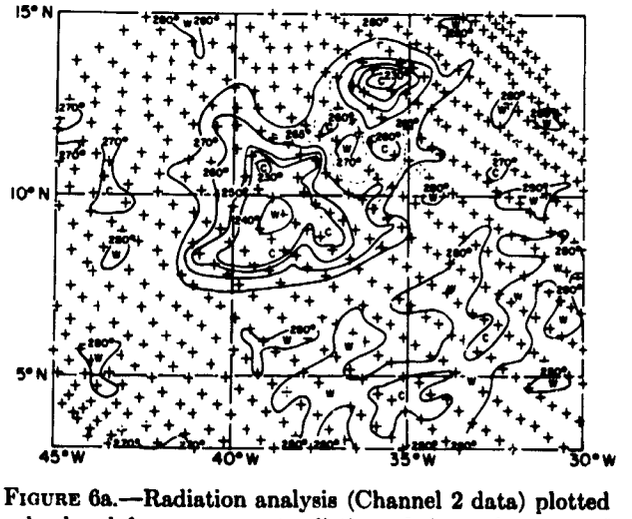


FIGURE 6a.—Radiation analysis (Channel 2 data) plotted by hand from a computer listing of the early stage of Hurricane Anna, which was used for the composite in Figure 6b.

photograph. Considering the range of accuracy of satellite attitude determination and the computer gridding programs, it is interesting to note that the radiation data fit the television pictures within one degree of latitude and longitude. An equivalent blackbody temperature difference of more than 50°K was indicated between the cloudless area and the highest cloud tops.

A similar radiation analysis (Figure 6c) was later performed by L. F. Whitney, Jr. of the

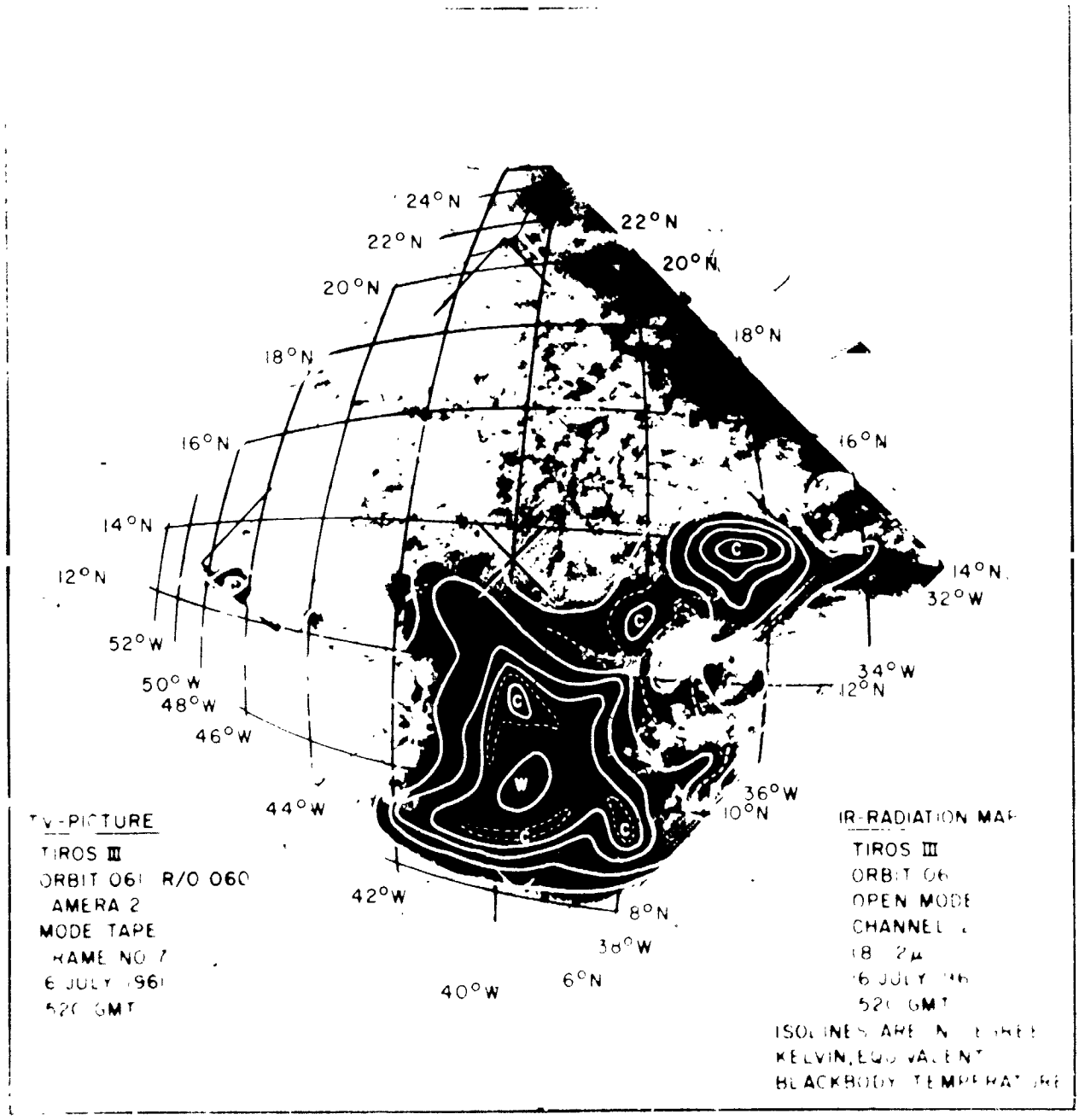


FIGURE 6b.—Composite analysis of Channel 2 radiation data and a gridded television picture of the early stage of Hurricane Anna.

National Weather Satellite Center, U.S. Weather Bureau, using the highly accurate Fujita method.[6] This method employs analog records produced on a recording oscillograph for special hand analysis rather than the automatically processed FMR Tapes. It should be considered, however, that this picture represents another frame of the same series, one minute later in time. From Figures 6b and 6c one can judge, that, at least in this case, the evaluation of the FMRT data does not result in any significant loss of details when compared with the results of the Fujita technique.

Figure 7 shows the radiation analysis of the same storm area based on three different types of computer produced grid maps at three different grid spacings, viz., 0.625°, 1.25°, and 2.50° longitude per mesh interval, corresponding to approximate map scale of 1:5 million, 1:10 million and

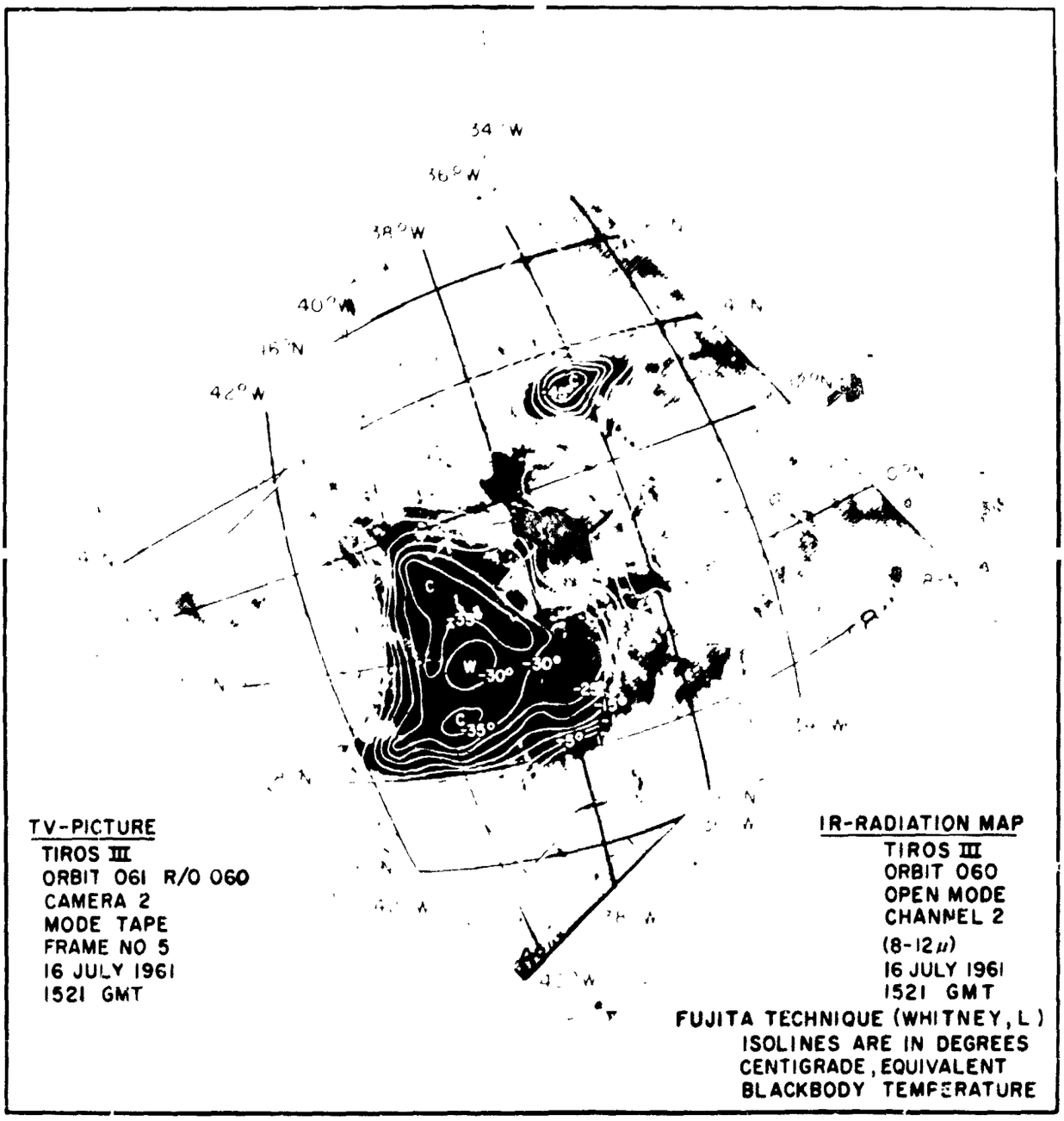


FIGURE 6c.—Composite analysis of Channel 2 radiation data and a gridded television picture of the early stage of Hurricane Anna, using the Fujita Technique by L. F. Whitney, Jr., U.S. Weather Bureau.

1:20 million respectively. The data points are again indicated in each figure. From the left to the right it becomes clear that the appropriate selection of the map scale for each radiation study is of great importance, as detailed features in the isoline configuration and extreme values of equivalent blackbody temperatures become "washed out" by the averaging processes involved at smaller map scales. Thus, for the analysis of more detailed synoptic features, a grid spacing of 0.625° of longitude per mesh interval is the most useful one, but the accuracy in data location is considerably less than by use of hand-plotted maps of the FMRT computer listings and of maps produced by the more precise Fujita method.

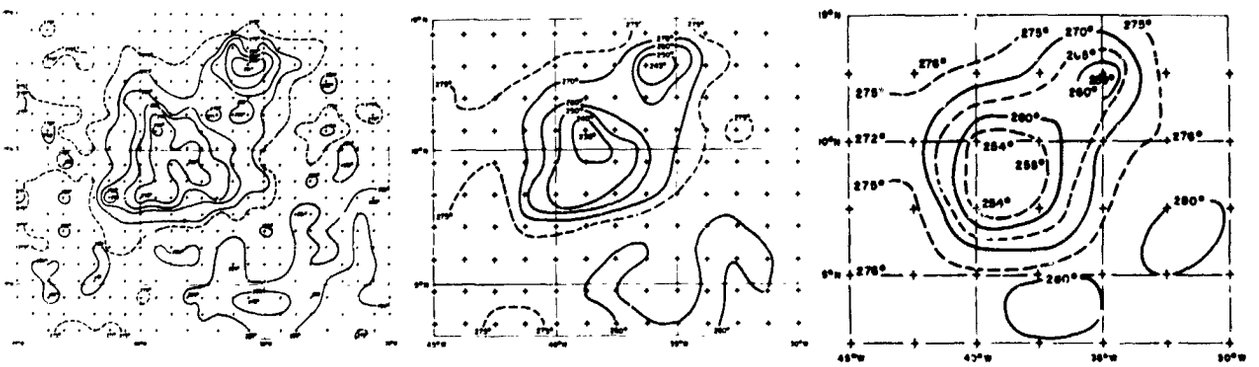


FIGURE 7.—Analysis of Channel 2 radiation data of the early stage of Hurricane Anna, using different resolutions in computer-produced grid print maps (Open mode data only).
a.—0.625° longitude/mesh interval
b.—1.250° longitude/mesh interval
c.—2.50° longitude/mesh interval

2.4 Radiometer Scanning Modes

Figure 1 showed the basic relationships between the satellite's spin axis and the radiometer axis.

As the satellite rotates on its axis, the radiometer scan pattern is defined by the intersection of a 45° half-angle cone and the earth. This pattern ranges from nearly a circle (closed mode) when the satellite nadir angle is less than17°, to two hyperbola-like branches in the "wall" and "floor" directions, when the spin vector is approximately perpendicular to the orbital radius vector.[1]

Therefore the three patterns of earth scanning modes, (Figure 8) can be distinguished as follows.[9]

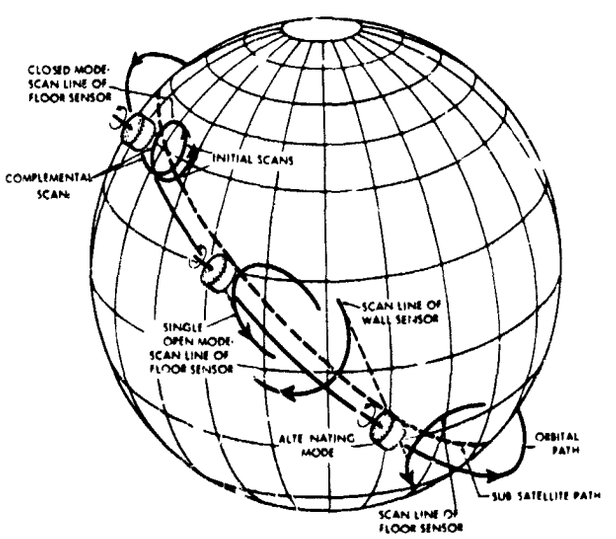


FIGURE 8.—Different types of scan modes of the Tiros radiometer generally occurring in one orbit.

1. Single Open Mode: Some scan spots of a spin cycle are space viewed and the remainder are earth viewed through the wall (side) sensor only or through the floor (baseplate) sensor only.

2. Closed Mode: All the scan spots throughout a number of spin cycles of the satellite are earth-viewed, either through the wall or the floor sensor.

3. Alternating Open Mode: The scan spots of a spin cycle are alternately viewed through the wall sensor and the floor sensor.

The IBM 7094 computer can easily be programmed to distinguish between different modes and to use or omit certain scanning modes. For example, in Figure 9 for orbits 56 and 58, those parts with single open mode data are grey shaded, and it is obvious that large and important parts of the data coverage are lost if closed mode data are omitted. Some examples of actual scan lines and scan spots are indicated on orbit 56. Additionally in this figure, the computer grid print map of orbit 57 is introduced, showing the full data coverage along the orbital path. However, there is a significant difference in accuracy of the geographical location between open and closed mode data, which reduces the applicability of closed mode data unless a very detailed investigation of possible short-term aberrations of the pseudo-spin rate value used in the computer reduction of the closed mode data is performed for every individual case. In the following

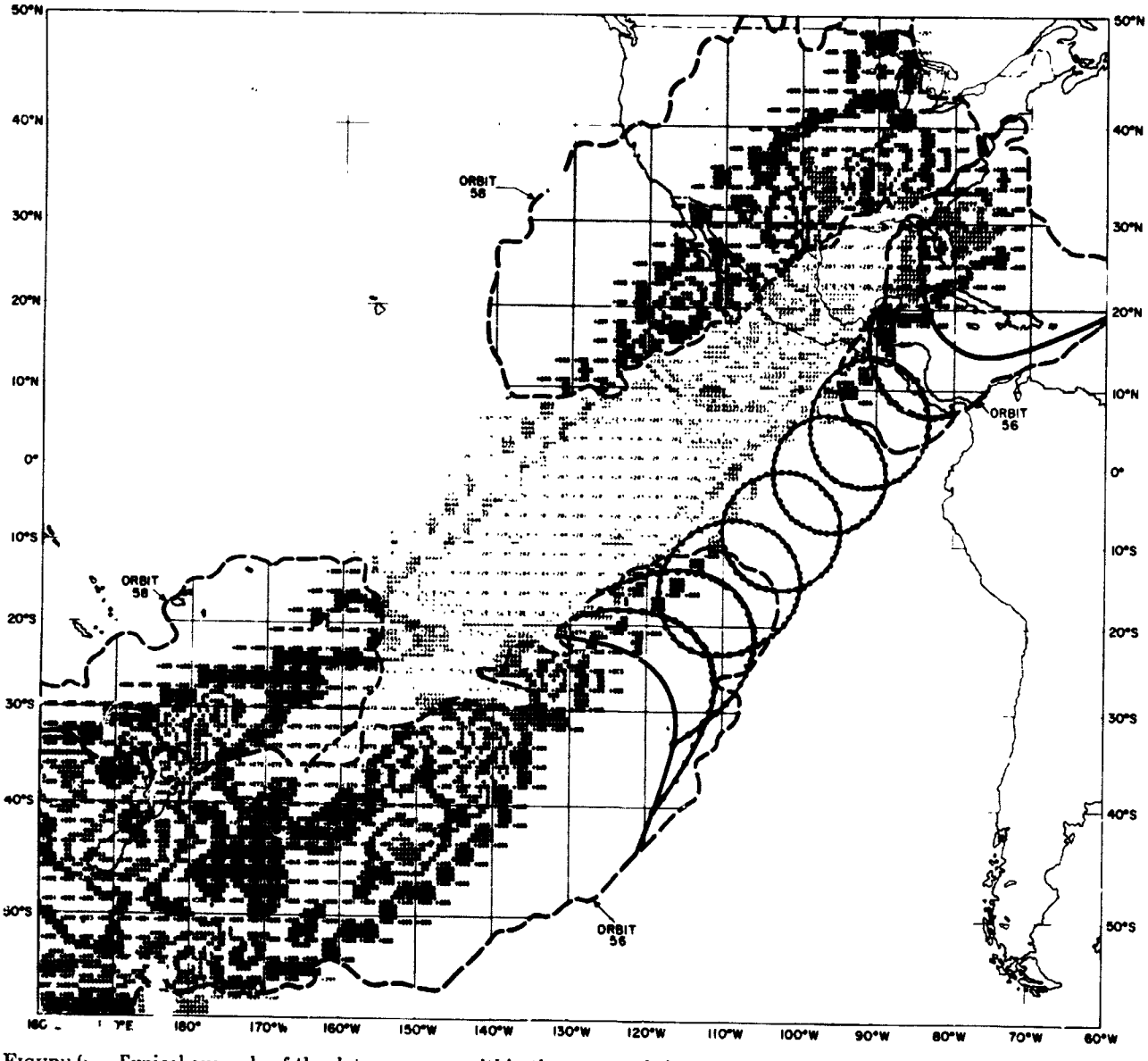


FIGURE 9.--Typical example of the data coverage within the open and closed scanning modes for three orbits of Tiros III. Some typical scan lines and the distribution of computer-selected scan spots along these lines are added for demonstration of some characteristic features of scan line geometry.

sections these practical difficulties will be discussed more in detail.

An example of typical closed mode data mislocation is given in Figure 10. This figure includes 10 minutes of closed mode data and shows the radiation "cloud" pattern of the early stage of Hurricane Anna at the same location within minutes of the same time as shown in Figure 7, which included open mode data only. This longer time period permitted the complemental scans of the closed mode to follow by several minutes the initial scans through the storm area (see Figure 8), resulting in an apparent shift of data leading to the figure. The floor aperture of the satellite was viewing the storm at this time, and a 6 scan spot shift in a counterclockwise direction was confirmed by diagnostic methods. The computer program averaged the displaced scan spot data with the preceding accurately positioned open mode scan spots and presented and erroneously elongated cloud structure.

A separate example of analog presentations of open and closed mode patterns is shown in Figures 11 and 12. Line A identifies the measurement

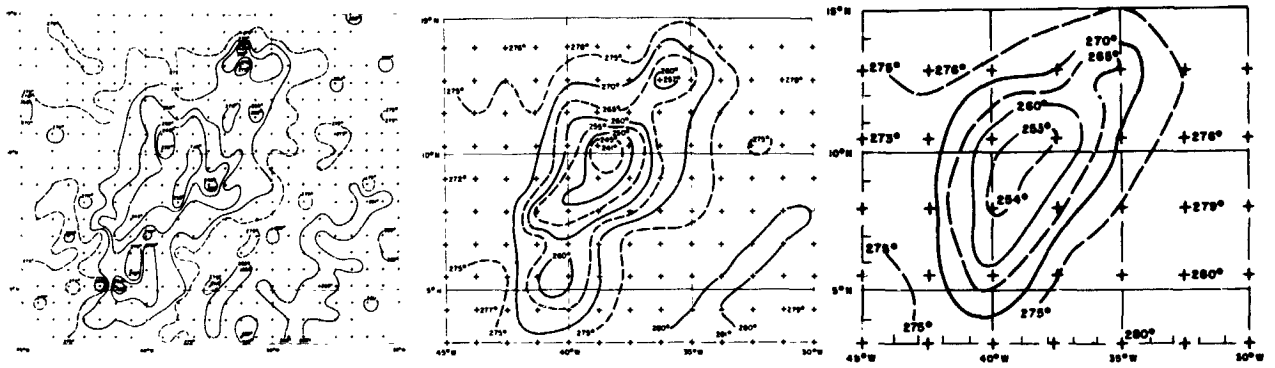


FIGURE 10.—Analysis of Channel 2 radiation data of the early stage of Hurricane Anna, using different resolutions in computer-produced grid print maps. (Open and closed mode data).
a.—0.625° longitude/mesh interval
b.—1.250° longitude/mesh interval
c.—2.50° longitude/mesh interval

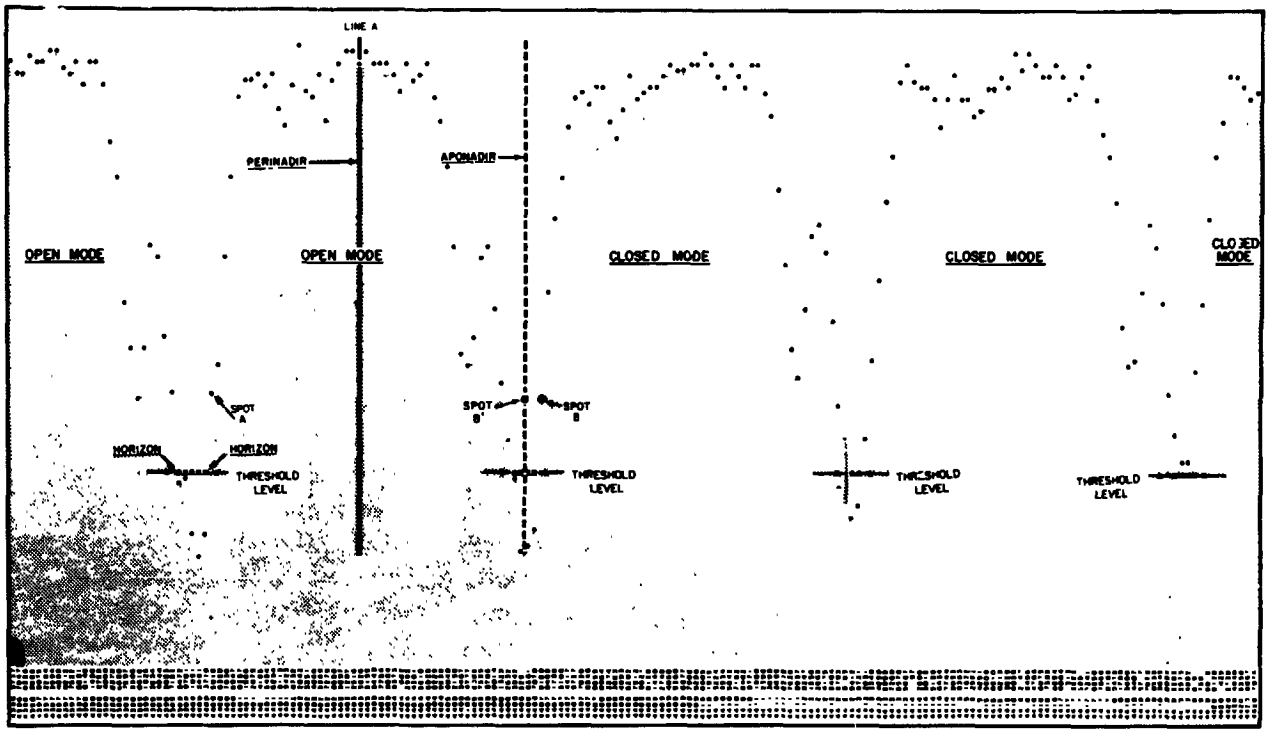


FIGURE 11.—Digitized form of the analog signal from Channel 2 during the transition from the single open to closed scanning modes. The numbers along the bottom indicate the minutes and seconds from the beginning of the data. The ordinate is the radiometer output on a relative scale.

obtained when viewing the perinadir in an open mode swath. This line is centered between the horizons of the swath and all the scan spots are positioned symmetrically about this line by the computer program, in accordance with the spin rate. Normally there can be little or no error in positioning these scan spots since each swath is processed individually and the horizons for each are determined quite accurately. For the closed mode swaths, mislocation errors do occur simply because there are no horizons detected by the satellite and the location of the data must be determined solely by the spin rate.

The spin rate used in the data reduction program is measured by means of counting cycles of a standard clock at the read-out station between

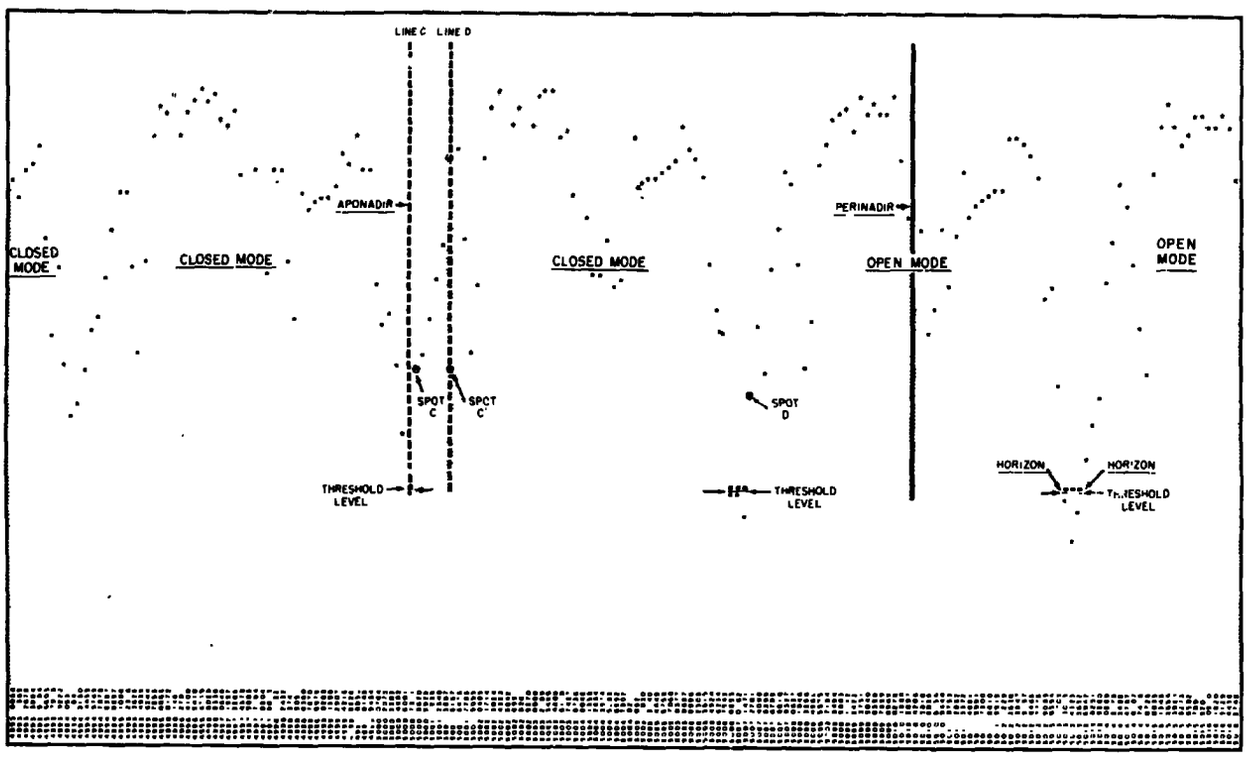


FIGURE 12.—Digitized form of the analog signal from Channel 2 during the transition from closed to single open scanning modes. The numbers along the bottom indicate the minutes and seconds from the beginning of the data. The ordinate is the radiometer output on a relative scale.

sun pulses and is assumed to be constant throughout the closed mode. For a spin rate determination error of 0.01°/sec (or 1 part in 6000), the spin angle error at the end of 10 minutes would be 6°, or about equal to one scan spot.[9]

The computer program is such that the first scan spot (Spot B) after the last open mode swath is positioned on the aponadir (Spot B'), a shift backward of 2½ scan points or about 15° of the spin angle.

All succeeding scan spots in the swath are shifted proportionally which results in a shift of the radiation data by 2½ scan spots (approx. 200 km). Figure 13 illustrates this mislocation in relation to geographic position, giving, however, examples for points close to the centers of the swaths. The measurement at point H of the first closed mode swath, which should be contiguous to that of point G within the last open mode swath, will be mislocated by the computer and appears in the wrong position H'. Comparing the last closed with the first open mode swath, point I should be contiguous to the measurement at point J; in this instance however, it was mislocated by the computer and shifted into the wrong location I'.

The 1½ to 2½ scan spot shift that occurs at the beginning of the closed mode is always counter to the rotation of the swath. If the spin rate used in the computer program is less than the actual rate, the mislocation of successive scan spots will be counter to the scan rotation. If the assumed spin rate is greater than the actual, the mislocation will be in the direction of the scan rotation.

The spin rate used for the 7 orbits in this study were all less than the actual rate, so that by the end of the closed mode, the mislocation averaged six data points or about 36° of the spin angle counter to the direction of scan rotation. Therefore, all scan spots were shifted backwards so that by the end, they were 6 spots too far in back of the true position. Thus, for example, spot C′ was erroneously located on the aponadir, (Figure 12) instead of on its correct position which should have been 6 spots ahead of the aponadir.

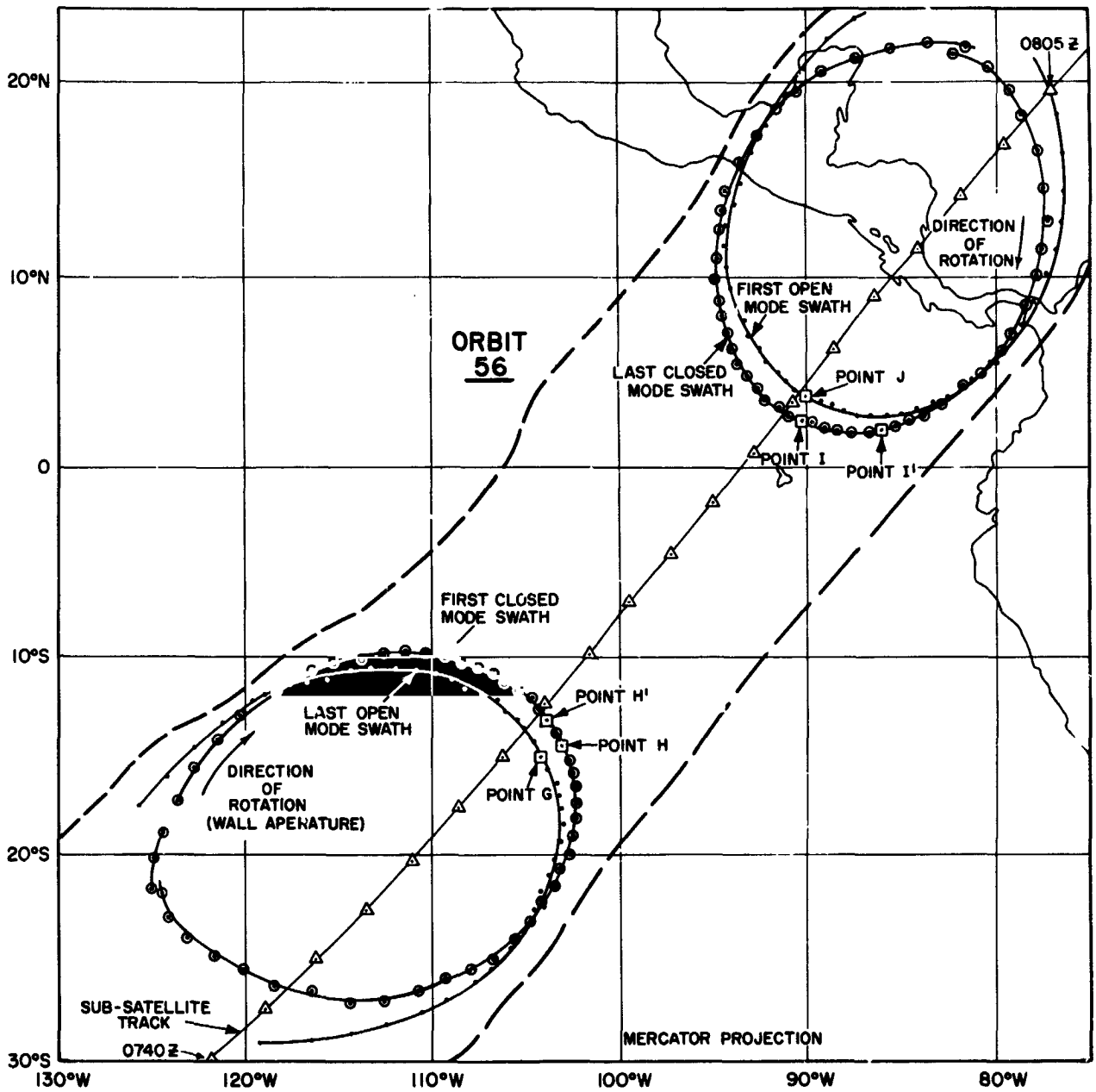


FIGURE 13.—An example of the scan spot shift at the entrance and exit from the closed scanning mode due to computer program techniques and inaccuracy in spin rate determination.

In order to improve the accuracy of location of the closed mode data, steps have been initiated to calculate the psuedo-spin rate values which should be individually applied to each closed mode sequence in the automatic reduction of the data.

2.5 Measurements Falling Below the Space-earth Discriminant

For the reasonable interpretation of the recorded radiometer signals it is necessary to distinguish very carefully between earth-scanned data and space-viewed or space-contaminated data. Therefore, the computer logic was set to make the distinction by means of an arbitrarily defined threshold value below which the data should be defined as "space viewed". Theoretically, this space-earth discriminant should be identical with the space viewed level and be located far from all atmospheric data. Practically, however, it is necessary to define the threshold value above the noise level of the radiometer system. Normally, there is no interference with

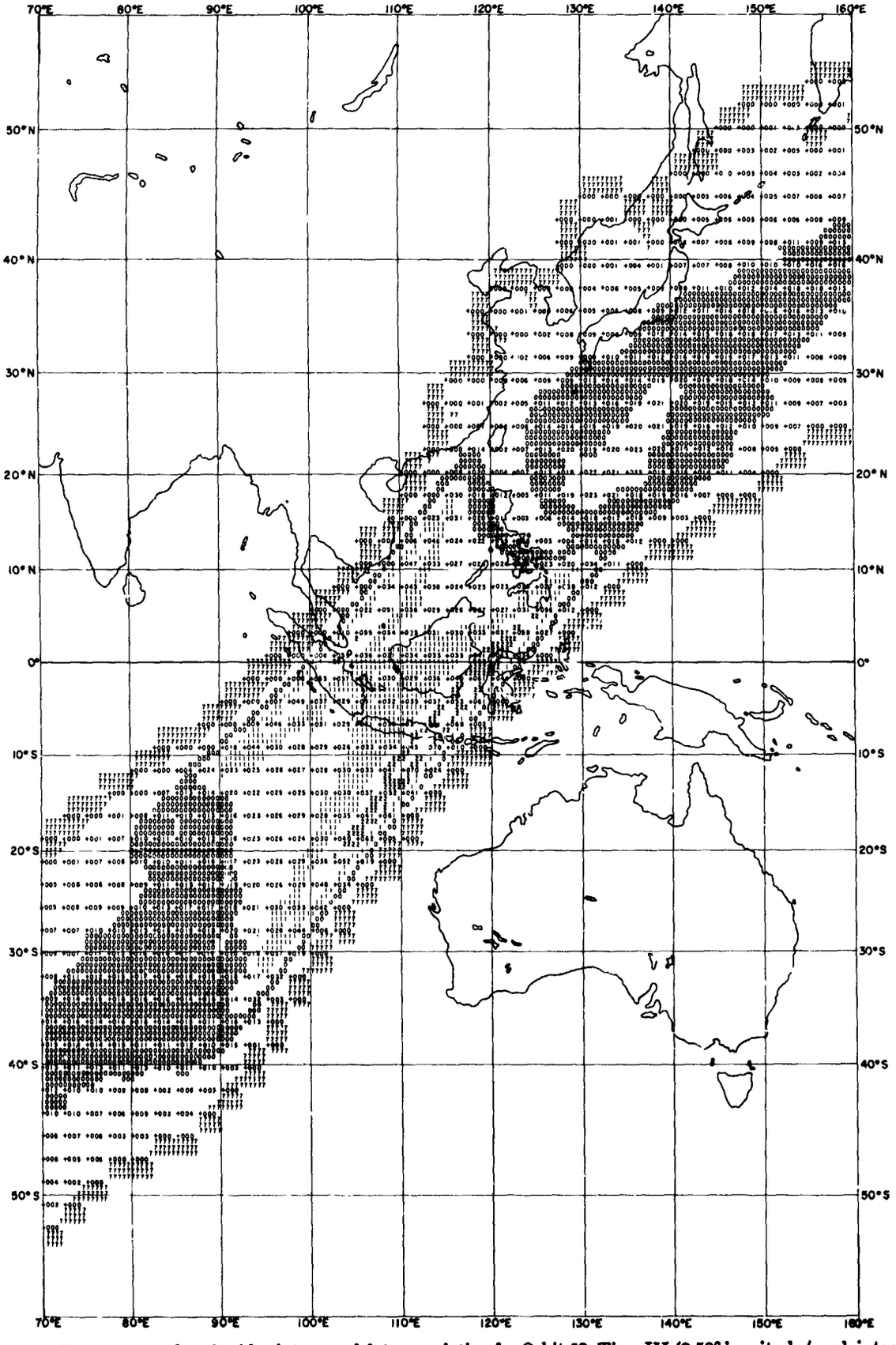


FIGURE 14.—Computer-produced grid print map of data population for Orbit 62, Tiros III (2.50° longitude/mesh interval).

real atmospheric measurements. But in the case of very low radiation intensities the actual equivalent blackbody temperatures may be lower than the threshold value. This occurs mainly over areas of very high opaque cloud systems such as tropical storms.

If a measurement appears below the space-earth discriminant, according to the TIROS III Radiation Data Users' Manual[1], it will be flagged by a minus sign on the FMR Tape and in the computer listings. Three or more consecutive values below the threshold will be assumed to define the end of a swath by the computer logic. Also, all data making up the terminal swath of the closed mode are flagged with minus signs for diagnostic purposes. This means that over large high-cloud areas or when an unusually high noise level occurs just in the transition between closed and open scanning modes, the computer logic may not be able to distinguish properly between real earth-scanned and space-viewed data, resulting in a large number of measurements being flagged with minus signs.

Generally, the computer will be ordered to reject all data flagged by a minus sign in order to avoid the use of questionable data. The com-

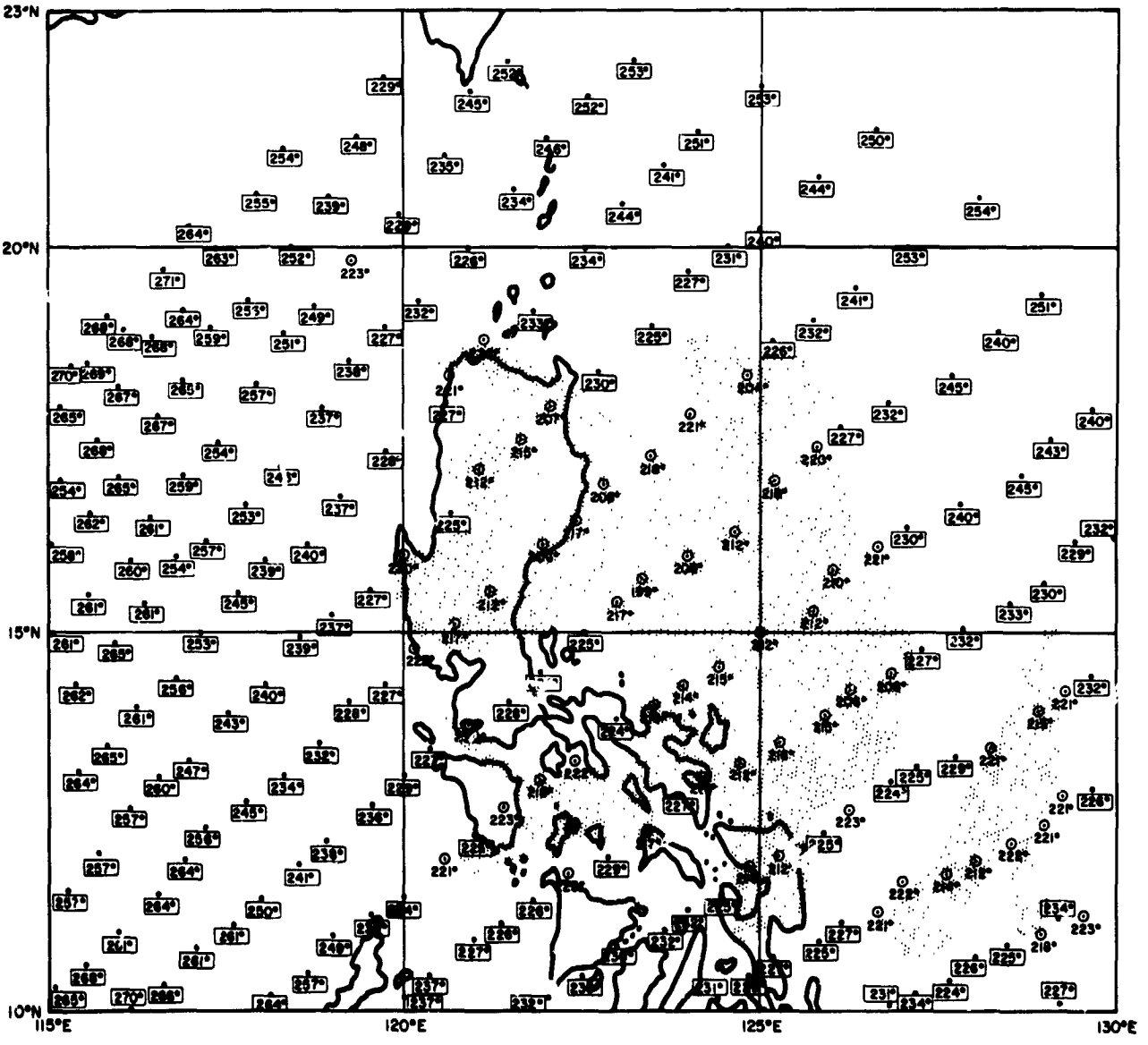


FIGURE 15.—Location of scan spots over a tropical disturbance close to the Philippines, based upon computer listings. (Encircled data in the shaded area were rejected by computer logic in producing the grid print maps).

puter has, however, an option to accept these data for grid print maps. Therefore, a careful check of the FMR Tape data is necessary under such circumstances to permit the analyst in advance to select the areas where the flagged data are meaningful for mapping.

An example of such an occurrence is given in Figure 14. By inspecting the population count, which shows how many data points are contributing to each grid point in the computer produced grid print map, for orbit 62 an irregularity is noted over the Philippine Islands. Approximately 20 data points were rejected by the computer in every grid square in this area. Figure 15 shows all the single data points taken from the computer listings. All data accepted by the computer mapping program are marked by a rectangle, while the rejected "flagged" data points are encircled. The final (Figure 16) analysis of this radiation field shows that just the most interesting low radiation temperature data (very high reaching dense cloud system) over Tropical Storm Flossie had been rejected. Thus, the effect of the artificially defined space-earth discriminate on the composition of radiation data over certain meteorologically significant areas

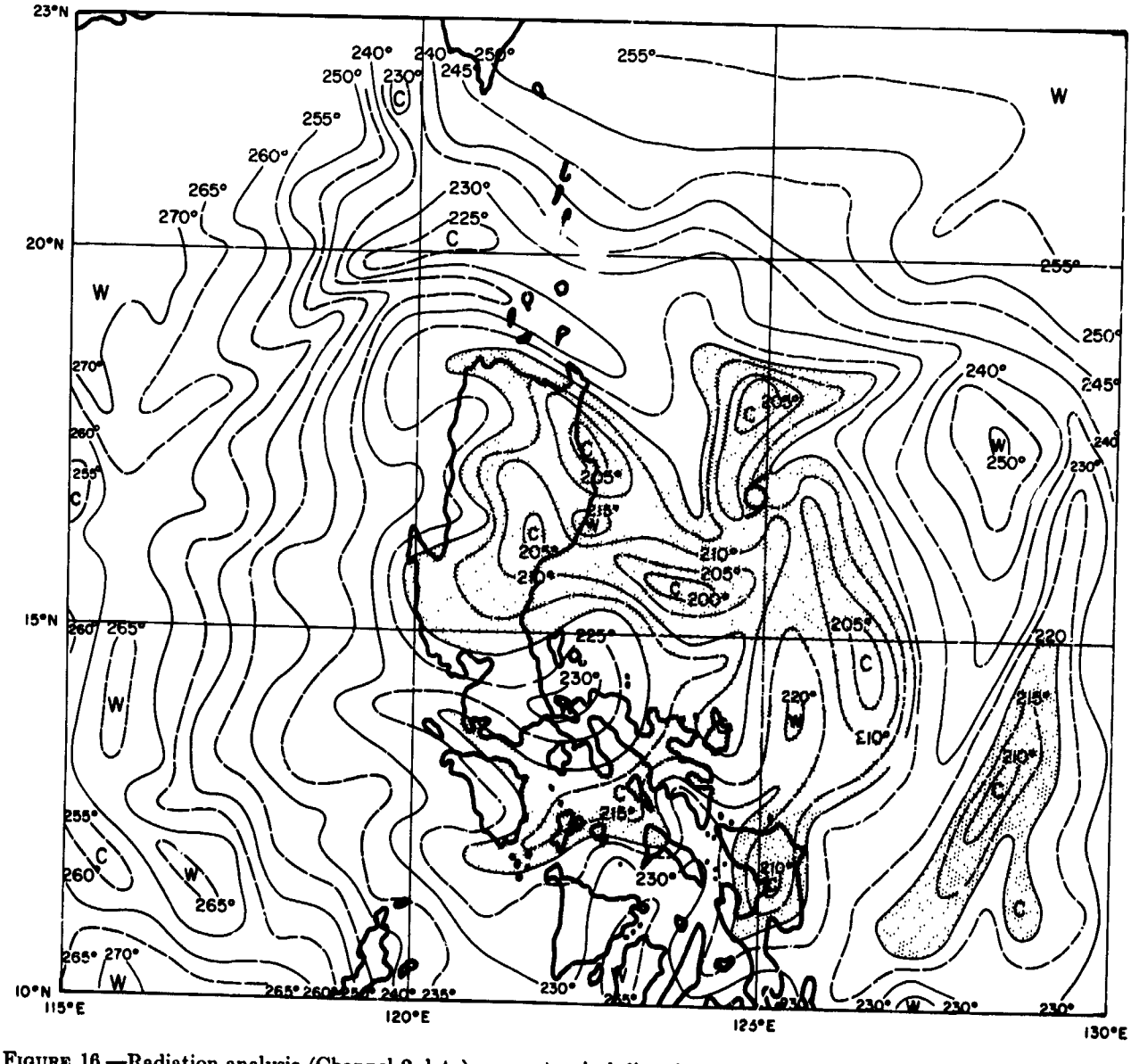


FIGURE 16.—Radiation analysis (Channel 2 data) over a tropical disturbance close to the Philippines, based upon computer listings. (Data within the grey shaded area were rejected by computer logic in producing the grid print maps.)

should always be accounted for in detailed studies. In certain cases, perhaps, the space-earth discriminant could be set somewhat lower, but this depends on such conditions, as signal-to-noise ratio, radiometer housing temperatures, etc. which are discussed in Reference 10.

3. CONCLUSIONS

This paper has attempted to review briefly the satellite radiation experiment and to discuss the current TIROS radiation data reduction processes. With improvement and modification of the computer processing procedures, it will become possible to reduce the location (closed mode) uncertainties of the data and also to accept low but real measurements so that these data could be used in operational weather analyses.

REFERENCES

1. TIROS III Radiation Data Users' Manual, Goddard Space Flight Center, NASA, August 1962.
2. BANDEEN, W. R., "Data Processing From Meteorological Satellites," NASA SP-16, Nat'l. Aero. and Space Admin., Dec. 1962.
3. NORDBERG, W., W. R. BANDEEN, B. J. CONRATH, V. KUNDE, and I. PERSANO, "Preliminary Results of Radiation Measurements from the TIROS III Meteorological Satellite," Jrl. Atmosp. Sci., 19, 20–30, 1962.
4. BANDEEN, W. R., R. E. SAMUELSON, I. P. STRANGE, "TIROS III Radiation Data Users' Manual Supplement, Correction Models for Instrumental Response Degradation," NASA, Goddard Space Flight Center, Dec. 1, 1963.
5. BANDEEN, W. R., V. KUNDE, W. NORDBERG, and H. P. THOMPSON, "TIROS III Meteorological Satellite Radiation Observations of a Tropical Hurricane," Tellus—to be published (1964).
6. FUJITA, T., "Outline of a Theory and Examples for Precise Analysis of Satellite Radiation Data," Research Paper 15, University of Chicago, Feb. 1963.
7. ASTHEIMER, R. W., R. DE WAARD, and E. A., JACKSON, "Infrared Radiometric Instruments on TIROS II", Jrl. of Optical Society, 51, 1386–93, Dec. 1961.
8. FRITZ, S., "Satellite Pictures and the Origin of Hurricane Anna," Monthly Weather Review, Vol. 90, pp. 507–513, Dec. 1962.
9. BANDEEN, W. R., "TIROS II Radiation Data Users' Manual Supplement," NASA Goddard Space Flight Center, Greenbelt, Md., May 15, 1962.
10. TIROS IV Radiation Data Catalog and Users' Manual, NASA/Goddard Space Flight Center, Greenbelt, Md., 15 December 1963.

ACKNOWLEDGEMENT

The authors are indebted to Mr. W. R. Bandeen, Mr. George W. Nicholas, Mr. Harold P. Thompson, of Goddard Space Flight Center and Mr. L. F. Whitney, Jr., U. S. Weather Bureau for their contributions to the presented analysis work.

NGS-35-71

SYMMETRIC EULER ANGLE DECOMPOSITION OF THE TWO-ELECTRON FIXED-NUCLEUS PROBLEM*

A. K. BHATIA† AND A. TEMKIN

1. INTRODUCTION

The decomposition of the Laplacian operator,

$$\nabla^2=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)+\frac{1}{r^2}\left(\frac{1}{\sin\vartheta}\frac{\partial}{\partial\vartheta}\sin\vartheta\frac{\partial}{\partial\vartheta}+\frac{1}{\sin^2\vartheta}\frac{\partial^2}{\partial\varphi^2}\right),$$

where the coefficient of the second term is proportional to the square of the angular momentum operator, is the basic relation between kinetic energy and angular momentum in the quantum mechanics of the one body problem (or the relative motion of two particles). When acting on a wave function which is an eigenfunction of total angular momentum ℓ, the Laplacian simplifies to

$$\nabla^2=\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial}{\partial r}\right)-\frac{\ell(\ell+1)}{r^2},$$

in which form it is clear that the effect of this decomposition is to reduce the Schödinger equation with a spherically symmetric potential from a three dimensional partial differential equation to a one dimensional (ordinary) differential equation. As such this relation is of fundamental mathematical importance and is familiar to everyone who utilizes quantum mechanics at all.

The analogous procedure when more than one particle is involved, in particular two identical particles in an external force field, although known, is not as well known, nor is it as well developed. When the external field is that of a fixed nucleus the wave function is expanded in eigenfunctions of the total angular momentum of the two particles multiplied by functions of the three remaining independent variables. The total angular momentum eigenfunctions are functions of the three Euler angles only. These angles are not unique, but in some way they must describe the orientation of the instantaneous plane formed by the two particles and the center of coordinates (nucleus) in space. The remaining three coordinates then describe the positions of the particles in this plane, and the functions of these variables are the generalized radial functions. Hylleraas' original papers[1] in effect contained the reduced or radial equations for total S states in terms of the residual coordinates r_1, r_2, r_{12}. In this case, the total orbital angular momentum is a constant function, and hence the reduction of a six-dimensional to a three-dimensional partial differential equation is independent of how one defines the Euler angles.

The standard treatment of the general problem is due to Breit.[2,3] He used the Euler angles that Hylleraas [1] originally introduced: namely the two spherical angles of one of the particles in the space fixed coordinate system and a second azimuthal angle between the r_1–z plane and the r_2–r_1 plane. Breit's remaining coordinates were chosen as r_1, r_2, and θ_{12}, the latter being the angle between $\mathbf{r}_1$ and $\mathbf{r}_2$. If, however, one wants to describe two electron atoms or ions in the approximation that the nuclei are fixed, then one has an additional requirement of which there is no analogue in the one-body problem. And that is the Pauli principal: the requirement that the spatial function be either symmetric or antisymmetric under the exchange of the particle coordinates. It is clear that the Hylleraas-Breit choice of Euler angles (which we hereinafter refer to as the Hylleraas-Breit angles), being quite unsymmetrical with respect to the two particles, is not optimum in this respect.[4] In fact the construction the linear combinations of angular momentum functions with the appropriate exchange properties is a very difficult task which

*Published as *Goddard Space Flight Center Document X-640-64-68*, March 1964.

†National Academy of Sciences—National Research Council Resident Research Associate.

depends not only on the Euler angles but on θ_{12} as well. It is not surprising, therefore, that Breit's original work[2] was limited to P-states, and work thereafter has always been limited to specific angular momentum states.[5]

However, a treatment by Holmberg[6] using a symmetrical choice of Euler angles (which we shall call Holmberg's angles) has in the interim been carried out. With these angles the description of exhange as well as parity (which latter property is also simply describably with th *II-B* angles) is simple (although these properties are only alluded to in Holmberg's paper). One of the purposes of the present paper is to examine these properties and relate them more clearly to the construction of the total wave function and further clarify other aspects of Holmberg's important paper. Consequently, we shall be enabled to derive the general radial equations for arbitrary angular momentum (ℓ) for the case of two identical particles in the field of a fixed nucleus. Holmberg's treatment applies to three particles of the same mass.

It is clear that in treating three particles of equal mass Holmberg had in mind the three nucleon problem whereas we are interested in two-electron atoms, ions, and diatomic molecules. The application of this formalism to two electron atoms and ions is clear, and the decomposition amounts to a rigorous reduction of the Schrödinger equation. It should only be remarked here that the scattering of electrons from one electron atoms and ions is also a special class of these problems. We have therefore worked out the connection between the boundary conditions for electron-atom scattering and Holmberg's angles (Section VIII).

Inasmuch as Holmberg's paper refers to the three nucleon problem, reference should also be made to the papers of Derrick and Blatt.[7] These papers deal much more realistically with the three nucleon problem in that full account is taken of an internucleon potential which is considerably more complicated than a central potential. As regards the actual choice of coordinates, Derrick and Blatt define axes along the moments of inertia of the three-body system. As such th[e] will depend on the lengths of the interparticle distances and therefore are quite different fr[o]m Holmberg's angles (Section II).

II. HOLMBERG'S ANGLES

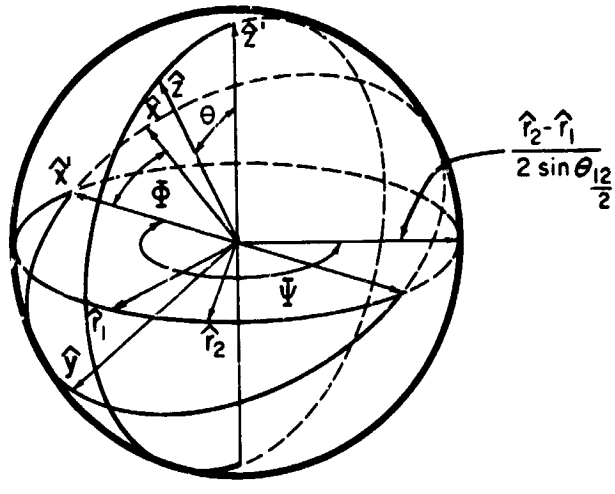


FIGURE 1.—Perspective drawing of (Holmberg's) Euler angles and the unit vectors of the problem.

Figure 1 contains a perspective drawing of Homberg's angles which define the particle plane with respect to the space fixed x, y, and z axes. The rotated axes x', y', z' are then defined by

$$\hat{\mathbf{z}}' = \frac{\hat{\mathbf{r}}_1 \times \hat{\mathbf{r}}_2}{|\mathbf{r}_1 \times \mathbf{r}_2|} \tag{1}$$

$$\hat{\mathbf{x}}' = \frac{\hat{\mathbf{z}}' \times \hat{\mathbf{z}}'}{|\mathbf{z} \times \mathbf{z}'|} \tag{2}$$

$$\hat{\mathbf{y}}' = \hat{\mathbf{z}}' \times \hat{\mathbf{x}} \tag{3}$$

The Holmberg Euler angles are then

$$\theta \equiv \text{angle between } \hat{\mathbf{z}} \text{ and } \hat{\mathbf{z}}' \tag{4}$$

$$\Phi \equiv \text{angle between } \hat{\mathbf{x}}' \text{ and } \hat{\mathbf{x}} \tag{5}$$

$$\Psi \equiv \text{angle between } \hat{\mathbf{x}}' \text{ and } (\hat{\mathbf{r}}_2 - \hat{\mathbf{r}}_1) \tag{6}$$

The ranges and planes of these angles are:

$0 \leq \theta \leq \pi$	in z–z' plane
$0 \leq \Phi \leq 2\pi$	in x–y plane
$0 \leq \Psi \leq 2\pi$	in x'–y' plane

As is usual a cap on a vector is used to represent a unit vector in the given direction. In particular $\hat{\mathbf{i}}$, $\hat{\mathbf{j}}$, $\hat{\mathbf{k}}$ are the three unit vectors along the (space fixed), x, y, and z axes respectively, and thus are synonymous with $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$. Similarly $\hat{\mathbf{i}}'$, $\hat{\mathbf{j}}'$, $\hat{\mathbf{k}}'$ and $\hat{\mathbf{x}}'$, $\hat{\mathbf{y}}'$, $\hat{\mathbf{z}}'$ are identical.

It is clear from the figure that $\hat{x}$, being in the x–y plane has components:

$$\hat{x}' = \hat{i}\cos\Phi + \hat{j}\sin\Phi \tag{7}$$

Since $\hat{x}'$ is perpendicular to the z–z' plane, it is perpendicular to every line in that plane going through the origin. This includes specifically the line of intersection of the z–z' plane with the x–y plane. However, the azimuthal angle of that intersecting line is the azimuth of $\hat{z}'$ itself, and since $\hat{x}'$ has azimuth Φ, $\hat{z}'$ has azimuth $\frac{3\pi}{2}+\Phi$ (cf. Figure 1). The polar angle of $\hat{z}'$ is clearly θ, therefore we have the important relation:

$$\hat{z}' = \hat{i}\sin\theta\sin\Phi - \hat{j}\sin\theta\cos\Phi + \hat{k}\cos\theta \tag{8}$$

The relations between Holmberg's angles and the spherical angles of the individual particles are obtained by substituting Eqs. (7) and (8) into the *lhs* of Eqs. (1) and (2) and using the ordinary decomposition of $\mathbf{r}_1$ and $\mathbf{r}_2$ in the *rhs*:

$$\hat{r}_1 = \hat{i}\sin\vartheta_1\cos\varphi_1 + \hat{j}\sin\vartheta_1\sin\varphi_1 + \hat{k}\cos\vartheta_1$$

$$\hat{r}_2 = \hat{i}\sin\vartheta_2\cos\varphi_2 + \hat{j}\sin\vartheta_2\sin\varphi_2 + \hat{k}\cos\vartheta_2$$

One obtains

$$\sin\theta_{12}\cos\theta = \sin\theta_1\sin\vartheta_2\sin(\varphi_1-\varphi_1) \tag{9}$$

$$\sin\theta_{12}\sin\theta\sin\Phi = \sin\vartheta_1\sin\varphi_1\cos\theta_2 - \cos\vartheta_1\sin\vartheta_2\sin\varphi_2 \tag{10}$$

$$\sin\theta_{12}\sin\theta\cos\Phi = \sin\vartheta_1\cos\varphi_1\cos\vartheta_2 - \cos\vartheta_1\sin\vartheta_2\cos\varphi_2 \tag{11}$$

$$2\sin(\tfrac{1}{2}\theta_{12})\cos\psi = \sin\varphi_2\cos(\varphi_2-\Phi) - \sin\vartheta_1\cos(\varphi_1-\Phi) \tag{12}$$

$$2\sin(\tfrac{1}{2}\theta_{12})\sin\psi = (\cos\vartheta_2-\cos\vartheta_1)\sin\theta - \sin\theta_1\sin(\varphi_1-\Phi) + \sin\vartheta_2\sin(\varphi_2-\Phi) \tag{13}$$

$$\cos\theta_{12} = \cos\vartheta_1\cos\vartheta_2 + \sin\vartheta_1\sin\vartheta_2\cos(\varphi_1-\varphi_2) \tag{14}$$

The latter relation is, of course, the well known expansion for the angle between two vectors.

It is also of interest to give the vectors $\hat{r}_1$ and $\hat{r}_2$ in the particle plane: (primed co-ordinate system):

$$\hat{r}_1 = \hat{i}'\sin(\psi-\tfrac{1}{2}\theta_{12}) - \hat{j}'\cos(\psi-\tfrac{1}{2}\theta_{12}) \tag{15}$$

$$\hat{r}_2 = \hat{i}'\sin(\psi+\tfrac{1}{2}\theta_{12}) - \hat{j}'\cos(\psi+\tfrac{1}{2}\theta_{12}) \tag{16}$$

$$\hat{z} = \hat{j}'\sin\theta + \hat{k}'\cos\theta \tag{17}$$

$$\hat{y} = \hat{i}'\sin\Phi + \hat{j}'\cos\theta\cos\Phi - \hat{k}'\sin\theta\cos\Phi \tag{18}$$

$$\hat{x} = \hat{i}'\cos\Phi - \hat{j}'\cos\theta\sin\Phi + \hat{k}'\sin\theta\cos\Phi \tag{19}$$

The following relations which are also very useful can now simply be derived by computing $(\hat{r}_1\cdot\hat{z})$, $(\hat{r}_2\cdot\hat{z})$ etc. in the primed system.

$$\cos\vartheta_1 = -\sin\theta\cos(\psi-\tfrac{1}{2}\theta_{12}) \tag{20a}$$

$$\cos\vartheta_2 = -\sin\theta\cos(\psi+\tfrac{1}{2}\theta_{12}) \tag{20b}$$

$$\sin\vartheta_1\cos\varphi_1 = \cos\Phi\sin(\psi-\tfrac{1}{2}\theta_{12}) + \cos\theta\sin\Phi\cos(\psi-\tfrac{1}{2}\theta_{12}) \tag{21a}$$

$$\sin\vartheta_2\cos\varphi_2 = \cos\Phi\sin(\psi+\tfrac{1}{2}\theta_{12}) + \cos\theta\sin\Phi\cos(\psi+\tfrac{1}{2}\theta_{12}) \tag{21b}$$

$$\sin\theta_1\sin\varphi_1 = \sin\Phi\sin(\psi-\tfrac{1}{2}\theta_{12}) - \cos\theta\cos\Phi\cos(\psi-\tfrac{1}{2}\theta_{12}) \tag{22a}$$

$$\sin\vartheta_2\sin\varphi_2 = \sin\Phi\sin(\psi+\tfrac{1}{2}\theta_{12}) - \cos\theta\cos\Phi\cos(\psi+\tfrac{1}{2}\theta_{12}) \tag{22b}$$

III. PROPERTIES UNDER PARITY AND EXCHANGE

The operation of parity corresponds to the simultaneous inversion of both particles' coordinates: $\mathbf{r}_1 \rightarrow -\mathbf{r}_1$, $\mathbf{r}_2 \rightarrow -\mathbf{r}_2$. It can be seen from figure 1 that this places $\mathbf{r}_1$ and $\mathbf{r}_2$ facing the opposite direction, but the cross product and hence $\hat{z}'$ will not change as a result of this operation. Thus the z–z' plane will not change and $\hat{x}'$ will not change. On the other hand $(\hat{r}_2-\hat{r}_1)$ goes into the negative of itself, so that ψ gets increased by π. In other words under *parity*

$$\begin{aligned}\theta &\rightarrow \theta\\ \Phi &\rightarrow \Phi\\ \psi &\rightarrow \psi+\pi\end{aligned} \tag{23}$$

Exchange corresponds to the transformation $\mathbf{r}_1 \rightleftarrows \mathbf{r}_2$. From the analytical definitions $\hat{z}'$ and $\hat{x}'$, Eqs. (1) and (2), the new primed axes will go into the negative of themselves. Also $(\hat{r}_2-\hat{r}_1)$ goes into negative of itself. Clearly the inversion of the z'

axis corresponds to the transformation $\theta\to\pi-\theta$. Noting that Φ is the angle in the x–y plane and measured as positive with respect to the z axis, which is fixed, we see that $\Phi\to\pi+\Phi$. The simultaneous inversion of $\hat{\mathbf{x}}'$ and $(\mathbf{r}_2-\mathbf{r}_1)$ means that the modulus of the angle ψ remains the same. However, since ψ is an angle in the x'–y' plane, it is measured as positive with respect to the z' axis. Since the latter goes into the negative of itself, it becomes clear that $\psi\to2\pi+\psi$. Thus we have under *exchange*

$$\begin{aligned}\theta&\to\pi-\theta\\ \Phi&\to\Phi+\pi\\ \psi&\to2\pi+\psi\end{aligned}\tag{24}$$

The significance of these transformations relates to the transformation properties of the vector spherical harmonics under the same operation. These functions, which are the eigenfunctions of the angular momentum (next section), are the basic functions in terms of which the complete wave function is expanded. They can be written

$$D_l^{m,k}(\theta,\Phi,\psi)=\frac{\sqrt{2(2l+1)}}{4\pi}e^{i(m\Phi+k\psi)}d_l^{m,k}(\theta)\tag{25}$$

where the normalization has been so chosen that the function is identical with what is given in Section IV and the $d_l^{mk}(\theta)$ agree with those given by Wigner.[8] Only the dependence on θ is nontrivial:

$$d_l^{m,k}(\theta)=(-1)^{\frac{1}{2}(|k-m|+k-m)}\sqrt{\frac{4\pi}{2(2l+1)}}N_{lmk}\sin^{|k-m|}\frac{\theta}{2}\times$$
$$\times\cos^{|k+m|}\frac{\theta}{2}F\left(-l-\frac{\beta}{2}-1,\ l+\frac{\beta}{2},\ 1+|k-m|,\ \sin^2\frac{\theta}{2}\right),\tag{26}$$

where β and N_{lmk} are defined in Eqs. (45) and (46). $F(a,b;c;z)$ is the hypergeometric function in the notation of Magnus and Oberhettinger.[9] The important property of $d_l^{mk}(\theta)$, proved in Wigner's book,[8] is:

$$d_l^{m,k}(\pi-\theta)=(-1)^{l-m}d_l^{m,-k}(\theta)\tag{27}$$

The phase in (26) is such that this relation holds for all values of m and k.

Letting P and E_{12} represent parity and exchange, we have from (23) and (24):

$$PD_l^{m,k}(\theta,\Phi,\psi)=D_l^{m,k}(\theta,\Phi,\pi+\psi)$$

$$E_{12}D_l^{m,k}(\theta,\Phi,\psi)=D_l^{m,k}(\pi+\vartheta,\pi+\Phi,2\pi-\psi)$$

which using (27) reduce to

$$PD_l^{m,k}(\theta,\Phi,\psi)=(-1)^kD_l^{m,k}(\theta,\Phi,\psi)\tag{28}$$

$$E_{12}D_l^{m,k}(\theta,\Phi,\psi)=(-1)^lD_l^{m,-k}(\theta,\Phi,\psi)\tag{29}$$

The simplicity of Eq. (29) is the essential feature which recommends Holmberg's angles to the description of the two electron problem.

IV. ANGULAR MOMENTUM

The components of the total angular momentum are readily expressed in terms of the particles spherical angles. Thus, for example

$$\frac{-i}{\hbar}M_z=\sin\varphi_1\frac{\partial}{\partial\theta_1}+\cot\vartheta_1\cos\varphi_1\frac{\partial}{\partial\varphi_1}+\sin\varphi_2\frac{\partial}{\partial\vartheta_2}$$
$$+\cot\vartheta_2\cos\varphi_2\frac{\partial}{\partial\varphi_2}\tag{30}$$

The particles' angles ϑ_1, φ_1, ϑ_2, φ_2 via (9)–(14) are implicit functions of the four angles θ, Φ, ψ, θ_{12}. Thus the problem of finding M_z in these angles is a straightforward problem of partial differentiation. We can write

$$\frac{-i}{\hbar}M_z=A_\theta\frac{\partial}{\partial\theta}+A_\Phi\frac{\partial}{\partial\Phi}+A_\psi\frac{\partial}{\partial\psi}+A_{\theta12}\frac{\partial}{\partial\theta_{12}}$$

where

$$A_\varkappa=\sin\varphi_1\frac{\partial x}{\partial\vartheta_1}+\sin\varphi_2\frac{\partial x}{\partial\vartheta_2}+\cot\vartheta_1\cos\varphi_1\frac{\partial x}{\partial\varphi_1}$$
$$+\cot\vartheta_2\cos\varphi_2\frac{\partial x}{\partial\varphi_2}\tag{31}$$

and $\varkappa$ can be any one of the angles θ, Φ, ψ, or θ_{12}.

Using then Eqs. (9)–(14), one finds that the following relations fall out quite easily:

$$A_{\theta 12}=0 \tag{32}$$

$$A_\theta=-\cos\Phi \tag{33}$$

$$A_\Phi=\sin\Phi\cot\theta \tag{34}$$

$$A_\psi=-\frac{\sin\Phi}{\sin\theta} \tag{35}$$

Thus

$$M_x=\hbar/i\left[\cos\Phi\frac{\partial}{\partial\theta}-\sin\Phi\cot\theta\frac{\partial}{\partial\Phi}+\frac{\sin\Phi}{\sin\theta}\frac{\partial}{\partial\psi}\right] \tag{36}$$

One can, of course, proceed in a completely analogous way to get the remaining components of the angular momentum, however, let us note from Eqs. (20) that

$$\frac{\partial\vartheta_1}{\partial\Phi}=\frac{\partial\vartheta_2}{\partial\Phi}=0 \tag{37}$$

and from (10) and (11)

$$\frac{\partial\varphi_1}{\partial\Phi}=\frac{\partial\varphi_2}{\partial\Phi}=1 \tag{38}$$

Since

$$\frac{\partial}{\partial\Phi}=\sum_{j=1}^{2}\left(\frac{\partial\varphi_j}{\partial\Phi}\cdot\frac{\partial}{\partial\varphi_j}+\frac{\partial\vartheta_j}{\partial\Phi}\cdot\frac{\partial}{\partial\vartheta_j}\right)$$

substitution of (37) and (38) yields

$$\frac{\partial}{\partial\Phi}=\frac{\partial}{\partial\varphi_1}+\frac{\partial}{\partial\varphi_2}$$

However, since

$$M_z=\frac{\hbar}{i}\left(\frac{\partial}{\partial\varphi_1}+\frac{\partial}{\partial\varphi_2}\right)$$

we therefore have the z-component of $\mathbf{M}$:

$$M_z=\frac{\hbar}{i}\frac{\partial}{\partial\Phi} \tag{39}$$

The remaining component of the angular momentum may be derived from the commutation relation $[M_z, M_x]=i\hbar M_y$
Straightforward substitution yields:

$$M_y=\frac{\hbar}{i}\left[\sin\Phi\frac{\partial}{\partial\theta}+\cos\Phi\cot\theta\frac{\partial}{\partial\Phi}-\frac{\cos\Phi}{\sin\theta}\cdot\frac{\partial}{\partial\psi}\right] \tag{40}$$

These relations are independent of θ_{12} corresponding to the statement that the angular momentum only depends on the (three) Euler angles θ, Φ, ψ. The forms of the three operators is the same as one gets with the Hylleraas-Breit angles.[3] The square of the angular momentum is likewise the same. One finds directly from the sum of the squares that

$$\mathbf{M}^2=-\hbar^2\left[\frac{\partial^2}{\partial\theta^2}+\frac{1}{\sin^2\theta}\left(\frac{\partial^2}{\partial\Phi^2}+\frac{\partial^2}{\partial\psi^2}\right)+\cot\theta\frac{\partial}{\partial\theta}-\frac{2\cot\theta}{\sin\theta}\frac{\partial^2}{\partial\Phi\partial\psi}\right] \tag{41}$$

The vector spherical harmonics, which have been given for the restricted range $m\geq k\geq 0$ in Eq. (25), are the simultaneous eigenfunctions of $\mathbf{M}^2$ with eigenvalue $\hbar^2\ell(\ell+1)$ and $\hbar m$

$$\mathbf{M}^2 D_\ell^{m,k}(\theta,\Phi,\psi)=\hbar^2\,\ell(\ell+1)\,D_\ell^{m,k}(\theta,\Phi,\psi)$$

$$\mathbf{M}_z D_\ell^{m,k}(\theta,\Phi,\psi)=\hbar m\,D_\ell^{m,k}(\theta,\Phi,\psi) \tag{43}$$

They are given in a completely general, normalized form in Pauling and Wilson.[10] With our phase [8] they are explicity

$$D_\ell^{m,k}(\theta,\Phi,\psi)=(-1)^{\frac{1}{2}[|k-m|+k-m]}N_{\ell mk}\sin^{|k-m|}\frac{\theta}{2}\cos^{|k+m|}\frac{\theta}{2}\times e^{i(m\Phi-k\psi)}F\left(-\ell+\frac{1}{2}\beta-1,\ \ell+\beta/2,\ 1+|k-m|,\ \sin^2\frac{\theta}{2}\right) \tag{44}$$

where

$$\beta=|k+m|+|k-m|+2 \tag{45}$$

and the normalization constant is

$$N_{\ell mk}=\left[\frac{(2\ell+1)(\ell+\frac{1}{2}|k+m|+\frac{1}{2}|k-m|)!(\ell-\frac{1}{2}|k+m|+\frac{1}{2}|k-m|)!}{8\pi_2(\ell-\frac{1}{2}|k+m|-\frac{1}{2}|k-m|)!(\ell+\frac{1}{2}|k+m|-\frac{1}{2}|k-m|)!}\right]^{\frac{1}{2}}\times\frac{1}{|k-m|!} \tag{46}$$

In addition to the usual magnetic quantum number m, the vector spherical harmonics depend on the quantum number k, an integer whose range of values is the same as m: $-\ell\leq m,\ k\leq\ell$.

The physical significance of k derives from the fact that the $D_l^{m,k}$ are the eigenfunctions of the spherical top (for which (42) is the Schrödinger equation), and k is the angular momentum quantum number about the body-fixed axis of rotation. With regard to the applications that we contemplate here, k can be considered a degeneracy label which must be adjusted such that other requirements are fulfilled.

V. CONSTRUCTION OF THE TOTAL WAVE FUNCTION

We shall confine ourselves here strictly to the atomic problem which implies that the potential energy as well as the kinetic energy commutes with the total angular momentum. In this case the total wave function for a given l must be a linear superposition of the degenerate $D_l^{m,k}$. In addition, m will be fixed for a given magnetic substate and the "radial" equations will be independent of m (cf. Appendix II).

Considering, for the moment, the residual coordinates as r_1, r_2, θ_{12}, we can therefore expand the total wave function in the form:

$$\psi lm(\mathbf{r}_1, \mathbf{r}_2) = \sum_{k=-l}^{l} g^k(r_1, r_2, \theta_{12}) D_l^{m,k}(\theta, \Phi, \psi) \tag{47}$$

The parity operations, Eq. (23), only effects the Euler angles, and from Eq. (28) it only multiplies the $D_l^{m,k}$ function by $(-1)^k$. Therefore by restricting the sum to even and odd values of k, we guarantee that the superpositions have even and odd parity respectively:

$$\Psi_{lm}^{\text{even}}(\mathbf{r}_1, \mathbf{r}_2) = \sum_{k \text{ even}}'' g_l^k(r_1, r_2, \theta_{12}) D_l^{m,k}(\theta, \Phi, \psi) \tag{48a}$$

$$\Psi_{lm}^{\text{odd}}(\mathbf{r}_1, \mathbf{r}_2) = \sum_{k \text{ odd}}'' g_l^k(r_1, r_2, \theta_{12}) D_l^{m,k}(\theta, \Phi, \psi) \tag{48b}$$

where the double prime on the summation emphasizes that the sum goes over every other value of k.

In deriving the radial equations (next section) we shall exploit the invariance of the radial equations with respect to m, by choosing $m=0$. When the Hamiltonian is written in terms of the Euler angles and the remaining variables, there will occur terms involving $\frac{\partial}{\partial\Phi}$ and $\frac{\partial}{\partial\psi}$. By virtue of $m=0$, the former terms vanish, but the latter terms would bring down the imaginery coefficient ik. In order to avoid complex equations, it is therefore convenient to construct real angular momentum functions. Let

$$D_l^{\varkappa+} = \frac{1}{\sqrt{2}} \frac{(D_l^{0,\varkappa} + D_l^{0,-\varkappa})}{(1+\delta_{0\varkappa}(\sqrt{2}-1))} = \frac{\sqrt{2l+1}}{2\pi} \frac{\cos \varkappa\psi}{(1+\delta_{0\varkappa}(\sqrt{2}-1))} d_l^{0,\varkappa}(\theta) \tag{49a}$$

$$D_l^{\varkappa-} = \frac{1}{\sqrt{2}i}(D_l^{0,\varkappa} - D_l^{0,-\varkappa}) = \frac{\sqrt{2l+1}}{2\pi} \cdot \sin \varkappa\psi d_l^{0,\varkappa}(\theta) \tag{49b}$$

for $\varkappa > 0$, where

$$\varkappa = |k|, \tag{50}$$

The important property of these linear combinations is that they are eigenfunctions of exchange:

$$E_{12} D_l^{\varkappa\pm} = \pm(-1)^l D_l^{\varkappa\pm} \tag{49c}$$

This then constitutes a set of real, orthonormal vector spherical harmonics. These real vector spherical harmonics are still eigenfunctions of parity with eigenvalue $(-1)^\varkappa$.

The property of exchange is a mite more complicated than parity in the sense that it affects not only the Euler angles, but the residual coordinates as well. The beauty and importance of Holmberg's angles, however, is that there is no mixing, and independent of whether we consider the residual variables r_1, r_2, θ_{12} or r_1, r_2, r_{12}, the effect of exchange on the residual coordinates is simply $r_1 \rightleftarrows r_2$.

Finally then, if we construct

$$\Psi_{l0}(\mathbf{r}_1, \mathbf{r}_2) = \sum_{\varkappa}'' [f_l^{\varkappa+}(r_1, r_2, \theta_{12}) \cdot D_l^{\varkappa+}(\theta, \Phi, \psi) + f_l^{\varkappa-}(r_1, r_2, \theta_{12}) D_l^{\varkappa-}(\theta, \Phi, \psi)], \tag{51}$$

the operation of exchange on this sum then gives, with the use of (49c),

$$E_{12}\psi_{l0}(\mathbf{r}_1, \mathbf{r}_2) = \psi_{l0}(\mathbf{r}_2, \mathbf{r}_1) = \sum_{\varkappa}'' [f_l^{\varkappa+}(\mathbf{r}_2, \mathbf{r}_1, \theta_{12})(-1)^l D_l^{\varkappa+}(\theta, \Phi, \psi) + f_l^{\varkappa-}(\mathbf{r}_2, \mathbf{r}_1, \theta_{12})(-1)^{l+1} D_l^{\varkappa-}(\theta, \Phi, \psi)]$$

Thus if

$$f_l^{\varkappa+}(r_2, r_1, \theta_{12}) = \pm(-1)^l f_l^{\varkappa+}(r_1, r_2, \theta_{12}) \quad (52a)$$

$$f_l^{\varkappa-}(r_2, r_1, \theta_{12}) = \pm(-1)^{l+1} f_l^{\varkappa-}(r_1, r_2, \theta_{12}), \quad (52b)$$

the function $\Psi_{\ell 0}$ of Eq. (51) is a real, space symmetric (upper sign) or space antisymmetric (lower sign), eigenfunction of $\mathbf{M}^2$ and M_z corresponding to the quantum numbers ℓ and m with $m=0$. The space symmetric and antisymmetric solutions correspond to singlet and triplet spin states respectively. Furthermore the restriction to $m=0$ is sufficient for deriving the radial equations.

We have shown that the $m=0$ function can be written in manifestly real form, Eq. (51). However, in that form, it is not obvious what the generalization is to arbitrary m states. The generalization is nevertheless simply obtained.

Let

$$g_l^{\varkappa} = \frac{1}{\sqrt{2}}(f_l^{\varkappa+} - i f_l^{\varkappa-}) \quad \varkappa > 0 \quad (53)$$

and

$$g_l^{-\varkappa} = \frac{1}{\sqrt{2}}(f_l^{\varkappa+} + i f_l^{\varkappa-})$$

then the form (51) reduces to that of Eq. (47) for $m=0$. For arbitrary m one then need only replace the $D_l^{0,k}$ functions by the appropriate $D_l^{m,k}$ functions, the radial g_l^k functions remaining the same.

Alternatively one can define generalizations of the $D_l^{\varkappa\pm}$, Eq. (49), for arbitrary m.

$$D_l^{(m,\varkappa)+} \equiv [\sqrt{2} + \delta_{0\varkappa}(2-\sqrt{2})]^{-1}(D_l^{m,\varkappa} + D_l^{m,-\varkappa}) \quad m \neq 0$$

$$D_l^{(m,\varkappa)-} \equiv [\sqrt{2}i]^{-1}(D_l^{m,\varkappa} - D_l^{m,-\varkappa}) \quad (54)$$

with $D^{(0,\varkappa)\pm} \equiv D_l^{\varkappa\pm}$ for $m=0$. Note for $m \neq 0$ that the modified spherical harmonics, $D_l^{(m\varkappa)\pm}$ are no longer real.

The complete function for arbitrary m can then be written

$$\psi_{\ell m}(\mathbf{r}_1, \mathbf{r}_2) = \sum' [f_l^{\varkappa+} D_l^{(m,\varkappa)+} + f_l^{\varkappa-} D_l^{(m,\varkappa)-}] \quad (55)$$

Note that by virtue of the action of the raising and lowering operators (Appendix II), the phase of the individual terms except possibly for an overall minus sign, is correct as it stands.

VI. THE KINETIC ENERGY

Just as in the case of the angular momentum, the kinetic energy can be obtained by a straightforward process of partial differentiation. In this case, however, since second partial derivatives are involved, the differentiation is a much longer job, and, as we shall see, the partial derivatives involving θ_{12} no longer cancel out.

We start then with the kinetic energy in spherical coordinates

$$\nabla_1^2 + \nabla_2^2 = \frac{1}{r_1}\frac{\partial^2}{\partial r_1^2} r_1 + \frac{1}{r_2}\frac{\partial^2}{\partial r_2^2} r_2 + \frac{1}{r_1^2}\left[\frac{1}{\sin\vartheta_1}\frac{\partial}{\partial\theta_1}\sin\vartheta_1\frac{\partial}{\partial\vartheta_1} + \frac{1}{\sin^2\vartheta_1}\frac{\partial^2}{\partial\varphi_1^2}\right] + \frac{1}{r_2^2}\left[\frac{1}{\sin\vartheta_2}\frac{\partial}{\partial\vartheta_2}\sin\vartheta_2\frac{\partial}{\partial\vartheta_2} + \frac{1}{\sin^2\vartheta_2}\frac{\partial^2}{\partial\varphi_2^2}\right] \quad (56)$$

The first two terms are, of course, unaffected by the transformation. The angular differentiations then involve the transformation from the variables ϑ, φ_1, ϑ_2, φ_2 to θ, Φ, ψ and θ_{12}.

Consider the coefficient of the r_1^{-2} term. After some regrouping, we can write

$$\frac{1}{\sin\vartheta_1}\frac{\partial}{\partial\vartheta_1}\sin\vartheta_1\frac{\partial}{\partial\vartheta_1} + \frac{1}{\sin^2\vartheta_1}\frac{\partial^2}{\partial\varphi^2} = \sum_{\alpha=1}^{\psi_1}\left[\left(\frac{\partial\varkappa_\alpha}{\partial\vartheta_1}\right)^2 + \frac{1}{\sin^2\vartheta_1}\left(\frac{\partial\varkappa_\alpha}{\partial\varphi_1}\right)^2\right]\frac{\partial^2}{\partial\varkappa_\alpha^2} + \sum_\alpha\left[\frac{\partial^2\varkappa_\alpha}{\partial\vartheta_1^2} + \frac{1}{\sin^2\vartheta_1}\frac{\partial^2\varkappa_\alpha}{\partial\varphi_1^2} + \cot\vartheta_1\frac{\partial\varkappa_\alpha}{\partial\vartheta_1}\right]\frac{\partial}{\partial\varkappa_\alpha} + 2\sum_{\alpha>}\sum_{\beta=1}\left[\frac{\partial\varkappa_\alpha}{\partial\vartheta_1}\frac{\partial\varkappa_\beta}{\partial\vartheta_1} + \frac{1}{\sin^2\vartheta_1}\frac{\partial\varkappa_\alpha}{\partial\varphi_1}\frac{\partial\varkappa_\beta}{\partial\varphi_1}\right]\frac{\partial^2}{\partial\varkappa_\alpha\partial\varkappa_\beta} \quad (57)$$

where for $\alpha=1, 2, 3, 4$, $\varkappa_\alpha$ refers to θ, Φ, ψ, θ_{12}. The problem thus reduces to finding each of the square brackets separately in terms of the Euler angles and θ_{12}. The results are given in Table 1.

The kinetic energy thus becomes

$$\nabla_1^2 + \nabla_2^2 = \left\{\frac{1}{r_1}\frac{\partial^2}{\partial r_1^2} r_1 + \frac{1}{r_2}\frac{\partial^2}{\partial r_2^2} r_2 + \left(\frac{1}{r_1^2} + \frac{1}{r_2^2}\right)\frac{1}{\sin\theta_{12}}\frac{\partial}{\partial\theta_{12}}\sin\theta_{12}\frac{\partial}{\partial\theta_{12}}\right\} + \frac{F_1}{r_1^2} + \frac{F_2}{r_2^2} \quad (58)$$

where

$$F_1=\frac{1}{\sin^2\theta_{12}}\Big[\sin^2\Big(\psi+\frac{1}{2}\theta_{12}\Big)\frac{\partial^2}{\partial\theta^2}+\cos^2\Big(\psi+\frac{1}{2}\theta_{12}\Big)\cot\theta\frac{\partial}{\partial\theta}+\cos^2\Big(\psi+\frac{1}{2}\theta_{12}\Big)\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\Phi^2}$$
$$+\sin(2\psi+\theta_{12})\frac{\cot\theta}{\sin\theta}\frac{\partial}{\partial\Phi}-\sin(2\psi+\theta_{12})\frac{1}{\sin\theta}\frac{\partial^2}{\partial\theta\partial\Phi}+\sin(2\psi+\theta_{12})\cot\theta\frac{\partial^2}{\partial\psi\partial\theta}$$
$$-2\cos^2\Big(\psi+\frac{1}{2}\theta_{12}\Big)\frac{\cot\theta}{\sin\theta}\frac{\partial^2}{\partial\psi\partial\Phi}\Big]-\frac{\partial^2}{\partial\psi\partial\theta_{12}}+A_1\frac{\partial^2}{\partial\psi^2}+B_1\frac{\partial}{\partial\psi} \quad (59)$$

and

$$A_1=\frac{1}{\psi}+\frac{\cot^2\theta}{\sin^2\theta_{12}}\cos^2\Big(\psi+\frac{1}{2}\theta_{12}\Big) \quad (60)$$

$$B_1=-\frac{\cos\psi\sin(\psi+\theta_{12})}{\sin^2\theta_{12}}-\frac{\cot^2\theta}{\sin^2\theta_{12}}\sin(2\psi+\theta_{12})$$
$$+\frac{\cos\theta_{12}}{2\sin\Big(\frac{\theta_{12}}{2}\Big)}\Bigg(1-\frac{1}{\cos\Big(\frac{1}{2}\theta_{12}\Big)}\Bigg) \quad (61)$$

The expressions for F_2, A_2 and B_2 can be obtained by replacing θ_{12} by $-\theta_{12}$ in the above formulae (including the appropriate partial derivatives). This is equivalent to operating with E_{12}.

$$F_2(\theta,\Phi,\psi,\theta_{12})=F_1(\theta,\Phi,\psi,-\theta_{12}) \quad (62)$$

It is clear, since all the coefficients are independent of Φ, that M_z commutes with the kinetic energy. We have also explicitly verified that $[M_z,\triangle 1^2+\triangle 2^2]=0$.

Note that the partial derivative involving θ_{12} and no other angles has been placed in the curly brackets with the radial derivatives. This is because this term, as the radial derivatives themselves, do not affect the orbital angular momentum, and are the terms which act on total S states[1,2,3]

In fact; in the action of the remaining terms on the angular momentum eigenfunction rests the bulk of the reduction of the Schrödinger equation to its 3-dimensional "radial" form. With this reduction in mind (cf. next section), it is convenient to write F_1 in terms of operators whose effect on the angular momentum eigenfunctions is particularly simple. One can show

$$F_1=\frac{1}{2}\sin^2\theta_{12}\Big[-\frac{1}{\hbar^2}\mathbf{M}^2+\cos\theta_{12}(\sin 2\psi\Lambda_2$$
$$-\cos 2\psi\Lambda_1)+\sin\theta_{12}(\sin 2\psi\Lambda_1+\cos 2\psi\Lambda_2)\Big]$$
$$-\frac{\partial^2}{\partial\theta_{12}\partial\psi}+\Big(\frac{1}{\psi}-\frac{1}{2\sin^2\theta_{12}}\Big)\frac{\partial^2}{\partial\psi^2}$$
$$+\Bigg(\frac{\cos\theta_{12}}{2\sin\Big(\frac{1}{2}\theta_{12}\Big)}-\cot\theta_{12}-\frac{1}{2\sin\theta_{12}}\Bigg)\frac{\partial}{\partial\psi} \quad (63)$$

where

$$\Lambda_1=2\frac{\partial^2}{\partial\theta_2}+\frac{\partial^2}{\partial\psi^2}+\frac{1}{\hbar^2}\mathbf{M}^2$$
$$\Lambda_2=2\frac{\cot\theta}{\sin\theta}\frac{\partial}{\partial\Phi}-(1+2\cot^2\theta)\frac{\partial}{\partial\psi}$$
$$-\frac{2}{\sin\theta}\frac{\partial^2}{\partial\theta\partial\Phi}+2\cot\theta\frac{\partial^2}{\partial\theta\partial\psi} \quad (64)$$

and $\mathbf{M}^2$ is the total angular momentum squared operator given in Eq. (41). F_2 is again derivable from F_1 by replacing θ_{12} by $-\theta_{12}$.

VII. THE REDUCED OR RADIAL EQUATIONS, ATOMIC CASE

The essential properties of the combinations of the operators appearing in F_1 and F_2, Eq. (62) are the following (cf. Appendix I):

$$(\sin 2\psi\Lambda_2-\cos 2\psi\Lambda_1)D_l^{\varkappa+}=-2\varkappa(\varkappa-1)B_{l\varkappa}(1-\delta_{0\varkappa})D_l^{(\varkappa-2)+}+\delta_{1\varkappa}(1-2A_\varkappa^l)D_l^{\varkappa+}$$
$$-\frac{(1+3\delta_{0\varkappa})}{(1+\delta_{0\varkappa}(\sqrt{2}-1))}\frac{A_\varkappa^l A_{\varkappa+1}^l}{2B_{l,\varkappa+2}}D_l^{(\varkappa+2)+} \quad (65a)$$

$$(\sin 2\psi\Lambda_1+\cos 2\psi\Lambda_2)D_l^{\varkappa+}=-2\varkappa(\varkappa-1)B_{l\varkappa}(1-\delta_{2\varkappa})(1-\delta_{0\varkappa})D_l^{(\varkappa-2)-}-\delta_{1\varkappa}(1-2A_\varkappa^l)D_l^{\varkappa-}$$
$$+\frac{1+3\delta_{0\varkappa}}{(1+\delta_{0\varkappa}(\sqrt{2}-1))}\frac{A_\varkappa^l A_{\varkappa+1}^l}{2B_{l,\varkappa+2}}D_l^{(\varkappa+2)-} \quad (65b)$$

$$(\sin 2\psi\Lambda_2-\cos 2\psi\Lambda_1)D_\ell^{\varkappa-}=-2\varkappa(\varkappa-1)B_{\ell\varkappa}(1-\delta_{2\varkappa})(1-\delta_{0\varkappa})D_\ell^{(\varkappa-2)-}-\delta_{1\varkappa}(1-2A_\varkappa^\ell)D_\ell^{\varkappa-}$$
$$-\frac{(1-\delta_{0\varkappa})}{(1+\delta_{0\chi}(\sqrt{2}-1))}\frac{A_\varkappa^\ell A_{\varkappa+1}^\ell}{2B_{\ell,\varkappa+2}}D_\ell^{(\varkappa+2)-} \quad (66a)$$

$$(\sin 2\psi\Lambda_1+\cos 2\psi\Lambda_2)D_\ell^{\varkappa-}=2\varkappa(\varkappa-1)B_{\ell\varkappa}(1-\delta_{0\varkappa})D_\ell^{(\varkappa-2)+}-\delta_{1\varkappa}(1-2A_\varkappa^\ell)D_\ell^{\varkappa+}$$
$$-\frac{(1-\delta_{0\varkappa})}{(1+\delta_{0\varkappa}(\sqrt{2}-1))}\frac{A_\varkappa^\ell A_{\varkappa+1}^\ell}{2B_{\ell\,\varkappa+2}}D_\ell^{(\varkappa+2)+} \quad (66b)$$

where

$$A_\varkappa^\ell=\frac{(\varkappa-\ell)(\varkappa+\ell+1)}{2(\varkappa+1)} \quad (67)$$

and

$$B_{\ell\varkappa}=\frac{[(\ell-\varkappa+1)(\ell-\varkappa+2)(\ell+\varkappa)(\ell+\varkappa-1)]^{\frac{1}{2}}}{4\varkappa(\varkappa-1)} \quad (68)$$

Recall that $\varkappa$ is the absolute value of k, Eq. (50).

We have proved in Appendix II that these relationships are not altered if one replaces the $D_\ell^{\varkappa\pm}$ functions by the $D_\ell^{(m,\varkappa)\pm}$ functions of Eq. (54). As such it becomes quite simple to derive the reduced equations from the original Schrödinger equation

$$H\psi_{\ell m}=E\psi_{\ell m} \quad (69)$$

for any m, and to see that the radial equations are independent of m. One obtains

$$\left[L_{\theta 12}+\frac{2m}{\hbar^2}(E-V)\right]f_\ell^{\varkappa+}-\left(\frac{1}{r_1^2}+\frac{1}{r_2^2}\right)\left[\left\{\frac{\ell(\ell+1)-\varkappa^2}{2\sin^2\theta_{12}}+\frac{\varkappa^2}{4}-\frac{\cot\theta_{12}}{2\sin\theta_{12}}\delta_{1\varkappa}(1-2A_\varkappa^\ell)\right\}f_\ell^{\varkappa+}\right.$$
$$+\frac{\cot\theta_{12}}{\sin\theta_{12}}(\varkappa+1)(\varkappa+2)B_{\ell,\varkappa+2}f_\ell^{(\varkappa+2)+}+\frac{\cot\theta_{12}}{4\sin\theta_{12}}\frac{(1+3\delta_{0,\varkappa-2})}{(1+\delta_{0,\varkappa-2}(\sqrt{2}-1))}\frac{A_{\varkappa-1}^\ell A_{\varkappa-2}^\ell}{B_{\ell\varkappa}}f_\ell^{(\varkappa-2)+}\Big]$$
$$+\left(\frac{1}{r_1^2}-\frac{1}{r_2^2}\right)\left[\varkappa\left(\frac{\cos\theta_{12}}{2\sin\theta_{12}}-\cot\theta_{12}-\frac{1}{2\sin\theta_{12}}\right)f_\ell^{\varkappa-}-\varkappa\frac{\partial f_\ell^{\varkappa-}}{\partial\theta_{12}}-\frac{\delta_{1\varkappa}(1-2A_\varkappa^\ell)}{2\sin\theta_{12}}f_\ell^{\varkappa-}\right.$$
$$\left.+\frac{(\varkappa+1)(\varkappa+2)}{\sin\theta_{12}}B_{\ell,\varkappa+2}f_\ell^{(\varkappa+2)-}-\frac{1}{4\sin\theta_{12}}\frac{(1-\delta_{0,\varkappa-2})}{(1+\delta_{0,\varkappa-2}(\sqrt{2}-1))}\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell}{B_{\ell\varkappa}}f_\ell^{(\varkappa-2)-}\right]=0 \quad (70a)$$

$$\left[L_{\theta 12}+\frac{2m}{\hbar^2}(E-V)\right]f_\ell^{\varkappa-}-\left(\frac{1}{r_1^2}+\frac{1}{r_2^2}\right)\left[\left\{\frac{\ell(\ell+1)-\varkappa^2}{2\sin^2\theta_{12}}+\frac{\varkappa^2}{4}+\frac{\cot\theta_{12}}{2\sin\theta_{12}}\delta_{1\varkappa}(1-2A_\varkappa^\ell)\right\}f_\ell^{\varkappa-}\right.$$
$$+\frac{\cot\theta_{12}}{\sin\theta_{12}}(1-\delta_{2,\varkappa+2})(\varkappa+1)(\varkappa+2)B_{\ell,\varkappa+2}f_\ell^{(\varkappa+2)-}+\frac{\cot\theta_{12}}{4\sin\theta_{12}}\frac{(1-\delta_{0,\varkappa-2})}{(1+\delta_{0,\varkappa-2}(\sqrt{2}-1))}\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell}{B_{\ell\varkappa}}f_\ell^{(\varkappa-2)-}\Big]$$
$$+\left(\frac{1}{r_1^2}-\frac{1}{r_2^2}\right)\left[-\varkappa\left(\frac{\cos\theta_{12}}{2\sin\left(\frac{\theta_{12}}{2}\right)}-\cot\theta_{12}-\frac{1}{2\sin\theta_{12}}\right)f_\ell^{\varkappa+}+\varkappa\frac{\partial f_\ell^{\varkappa+}}{\partial\theta_{12}}-\frac{\delta_{1\varkappa}(1-2A_\varkappa^\ell)}{2\sin\theta_{12}}f_\ell^{\varkappa+}\right.$$
$$\left.-\frac{(1-\delta_{2,\varkappa+2})}{\sin\theta_{12}}(\varkappa+1)(\varkappa+2)B_{\ell,\varkappa+2}f_\ell^{(\varkappa+2)+}+\frac{1}{4\sin\theta_{12}}\frac{(1+3\delta_{0,\varkappa-2})}{(1+\delta_{0,\varkappa-2}(\sqrt{2}-1))}\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell}{B_{\ell\varkappa}}f_\ell^{(\varkappa-2)+}\right]=0 \quad (70b)$$

$L_{\theta 12}$ is the S wave part of the kinetic energy, and only the term containing it survives in the description of S states:[1,2,3]

$$L_{\theta 12}=\frac{1}{r_1}\frac{\partial^2}{\partial r_1^2}r_1+\frac{1}{r_2}\frac{\partial^2}{\partial r_2^2}r_2$$
$$+\left(\frac{1}{r_1^2}+\frac{1}{r_2^2}\right)\frac{1}{\sin\theta_{12}}\frac{\partial}{\partial\theta_{12}}\left(\sin\theta_{12}\frac{\partial}{\partial\theta_{12}}\right) \quad (71)$$

Eqs. (70) are the "radial" equations, which it has been our purpose to derive. They pertain to both types of parity and exchange states. Parity is determined by the evenness or oddness of $\varkappa$. If, for example, ℓ is even, and we want to describe a state of even parity, Eqs. (70) couple the functions $f_\ell^{\varkappa+}$ and $f_\ell^{\varkappa-}$ for $\varkappa=0,2,4\ldots\ell$. This involves $\ell/2$ pairs plus one function (for $\varkappa=0$, D_ℓ^{0-} is zero hence f_ℓ^{0-} can be taken to be zero)

or $\ell+1$ functions. The odd parity equations for the same ℓ correspond to the coupling of the function with $\varkappa=1,3,\ldots,\ell-1$. This relates $\ell/2$ pairs or ℓ functions to each other. Both even and odd parity together therefore involve $(2\ell+1)$ functions corresponding to the $(2\ell+1)$ degeneracy of the vector spherical harmonics for a given m. For ℓ odd, there are ℓ functions involved in the even parity equations and $\ell+1$ functions in the odd parity equations.

For a given parity and ℓ, both singlet and triplet (space symmetric and antisymmetric) states are described by the same set of equations. The differences in the solutions devolve from the different boundary conditions which must be applied, Eqs. (52). One of the key virtues of the functions $f_\ell^{\varkappa\pm}(r_1, r_2, \theta_{12})$ is that they are either symmetric or antisymmetric; thus they may be confined to the region, say, $r_1\geq r_2$. If, for example, the exchange character of $f_\ell^{\varkappa+}$ is symmetric (which, according to (52), implies that $f_\ell^{\varkappa-}$ is antisymmetric), then these properties may be embodied in the boundary conditions [11]:

$$\left[\frac{\partial}{\partial n} f_\ell^{\varkappa+}(r_1, r_2, \theta_{12})\right]_{r_1=r_2}=0 \tag{72}$$

where $\frac{\partial}{\partial n}$ represents the normal derivative, and

$$[f_\ell^{\varkappa-}(r_1, r_2, \theta_{12})]_{r_1=r_2}=0 \tag{73}$$

and the solution from there on involves only the region $r_1>r_2\geq 0$. Such equations have distinct advantages from the point of view of numerical solutions.[12]

One can define, however, an asymmetric function in terms of which the radial equations can be more simply written. Letting

$$F_\ell^{\varkappa}(r_1,r_2,\theta_{12})\equiv f_\ell^{\varkappa+}(r_1,r_2,\theta_{12})+f_\ell^{\varkappa-}(r_1,r_2,\theta_{12}) \tag{74}$$

and

$$\tilde{F}_\ell^{\varkappa}(r_1, r_2, \theta_{12})\equiv f_\ell^{\varkappa+}(r_1, r_2, \theta_{12})-f_\ell^{\varkappa-}(r_1, r_2, \theta_{12}) \tag{75}$$

so that, from eq. (74),

$$\tilde{F}_\ell^{\varkappa}(r_2, r_1, \theta_{12})=\pm(-1)^\ell F_\ell^{\varkappa}(r_1, r_2, \theta_{12}) \tag{76}$$

for singlet (upper sign) and triplet (lower sign) cases.

one obtains by adding (70a) and (70b):

$$\begin{aligned}
&\left[L_{\theta 12}+\frac{2m}{\hbar^2}(E-V)\right]F_\ell^{\varkappa}-\left(\frac{1}{r_1^2}+\frac{1}{r_2^2}\right)\left[\left(\frac{\ell(\ell+1)-\varkappa^2}{2\sin^2\theta_{12}}+\frac{\varkappa^2}{4}\right)F_\ell^{\varkappa}-\frac{\cot\theta_{12}}{2\sin\theta_{12}}\delta_{1\varkappa}(1-2A_\varkappa^\ell)\tilde{F}_\ell^{\varkappa}\right.\\
&+\frac{\cot\theta_{12}}{\sin\theta_{12}}(\varkappa+1)(\varkappa+2)B_{\ell,\varkappa+2}\left\{F_\ell^{\varkappa+2}-\frac{\delta_{2,\varkappa+2}}{2}(F_\ell^{\varkappa+2}-\tilde{F}_\ell^{\varkappa+2})\right\}+\frac{\cot\theta_{12}}{4\sin\theta_{12}}\frac{1}{(1+\delta_{0,\varkappa-2}(\sqrt{2}-1)}\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell}{B_{\ell\varkappa}}\{F_\ell^{\varkappa-2}\\
&\left.+\delta_{0,\varkappa-2}(2\tilde{F}_\ell^{\varkappa-2}+F_\ell^{\varkappa-2})\}\right]+\left(\frac{1}{r_1^2}-\frac{1}{r_2^2}\right)\left[-\varkappa\left(\frac{\cos\theta_{12}}{2\sin\left(\frac{\theta_{12}}{2}\right)}-\cot\theta_{12}-\frac{1}{2\sin\theta_{12}}\right)\tilde{F}_\ell^{\varkappa}+\varkappa\frac{\partial\tilde{F}_\ell^{\varkappa}}{\partial\theta_{12}}\right.\\
&-\frac{\delta_{1\varkappa}(1-2A_\varkappa^\ell)}{2\sin\theta_{12}}F_\ell^{\varkappa}+\frac{(\varkappa+1)(\varkappa+2)}{\sin\theta_{12}}B_{\ell,\varkappa+2}\left\{-\tilde{F}_\ell^{\varkappa+2}+\frac{1}{2}\delta_{2,\varkappa+2}(F_\ell^{\varkappa+2}+\tilde{F}_\ell^{\varkappa+2})\right\}\\
&\left.\left.+\frac{1}{4\sin\theta_{12}}\frac{1}{(1+\delta_{0,\varkappa-2}(\sqrt{2}-1))}\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell}{B_{\ell\varkappa}}\left\{\tilde{F}_\ell^{\varkappa-2}+\delta_{0,\varkappa-2}(2F_\ell^{\varkappa-2}+\tilde{F}_\ell^{\varkappa-2})\right\}\right]=0
\end{aligned} \tag{77}$$

These equations, depending as they do on F_ℓ and $\tilde{F}_\ell$, are more analogous to the form of the P-wave equation of Breit.[2] The question may arise in connection with these as well as Breit's equations, of whether they are well-defined, since they involve two functions $F_\ell^{\varkappa}$ and $F_\ell^{\varkappa}$ and yet there is only one equation (for a given $\varkappa$). This question, in fact would appear to be particularly relevant as the previous form of our equations, (70) do constitute a coupled set for a given $\varkappa$. To see that both situations are meaningful and in particular that (77) is well-defined, consider a numerical solution of (77). In that case the space of the independent variables is divided into a grid of points, and $F_\ell^{\varkappa}$ is the collection of numbers associated with these grid points. $F_\ell^{\varkappa}$ can

therefore be considered a vector with as many components as there are grid points. The differential equation is replaced by a matrix which operates on the vector $F_l^\varkappa$. Now everytime an $\tilde{F}_l^\varkappa$ occurs in the equation, it is completely clear what has to be done: namely one must let the matrix counterpart of its coefficient in the differential equation operate on that component of $F_l^\varkappa$ which is its reflection point defined by (76). This is a completely unambiguous prescription which is tantamount to saying that the set (77) is well-defined by itself. The reason that (70) is composed of two equations for each $\varkappa$ whereas (77) is not is due to the fact that the functions $F_l^\varkappa$ are asymmetric and therefore must be solved for in the whole r_1, r_2, θ_{12} space. On the other hand the $f_l^{\varkappa\pm}$ functions are either symmetric or antisymmetric, and therefore they are restricted to the $r_1 \geq r_2$, θ_{12} (or equivalently to the $r_1 \leq r_2$, θ_{12}) space. Since this is only half the independent variable space, it is necessary that there be double the number of functions to recover the same information. This is again to say that (70) and (77) are completely equivalent. (Nevertheless a redundant equation with $F_l^\varkappa$ and $\tilde{F}_l^\varkappa$ interchanged may readily be derived.)

We have stated that (70) has certain advantages from the point of view of numerical integration. However, it should also be stated that the form (77) will probably be more advantageous for ordinary variational calculations. This is because if one adopts a specific analytic form of $F_l^\varkappa$, one need only interchange r_1 and r_2 in the expression to obtain $\tilde{F}_l^\varkappa$.

The differences in the $f_l^{\varkappa\pm}$ description from that of $F_l^\varkappa$ gives rise to characteristic differences in the formulation of boundary conditions for scattering problems (cf. the next section).

The restriction of these equations to the atomic case (two identical particles in a fixed central field) has implicitly been made by assuming that the potential is a function of the residual coordinates,

$$V = V(r_1, r_2, r_{12}), \tag{78}$$

so that V commutes with the angular momentum and therefore appears as an additional diagonal term in the radial equations. The inter-particle distance r_{12} is related to the independent radial coordinates that we have thus far considered, r_1, r_2, θ_{12} via the law of cosines:

$$r_{12}^2 = r_1^2 + r_2^2 - 2r_1 r_2 \cos \theta_{12}.$$

Alternatively, however, one can consider r_1, r_2, and r_{12} as the independent coordinates and derive radial equations involving them. Those coordinates, in fact, have certain advantages since the three singularities in the potential occur at their null points. As such they can describe the wave function in the region of close interaction very well. These variables, therefore, are particularly suited to calculation of lowlying bound states of two-electron atoms (where on the whole the electrons are quite close to each other and to the nucleus) and such successful calculations have been done ever since the early work of Hylleraas.[1]

When one considers the equation in the form we have previously given them, involving θ_{12}, one is naturally led to expand the "radial" wave function in terms of Legendre polynomials of cos θ_{12}[11]. The expansion is then truncated after some P_n (cos θ_{12}) and convergence is sought as a function of n. In these classes of two electron problems, this constitutes the idea of configuration interaction in its most general form. Recently this idea has come under some criticism [13,14,15] principally because such a relative partial wave expansion necessarily converges slowly where the electron-electron interaction is large (r_{12} small). The argument has some validity for the above-mentioned low-lying bound states. However, the argument can easily get distorted and exaggerated, for instance when applied to the low-energy scattering of electrons from hydrogen.[14] The point there is that the long-range correlation coming from the induced potential in the atom is at least as important as the short range correlations [16] and yet is only poorly approximated by the conventional Hylleraas type of expansion. This situation has been discussed elsewhere.[12]

These reservations notwithstanding, however, it is nevertheless true that the most accurate three body calculations have been made using the r_1, r_2, r_{12} coordinates, or linear combinations of them,[17] on the low-lying states of helium [18] and its isoelectronic ions.[17] We therefore give below the radial equations in terms of r_1, r_2, r_{12}. The equations are in their asymmetric form corres-

ponding to Eq. (77), since it is assumed that they will be utilized in connection with variational calculations with analytic expansions of the radial wave functions.

$$\left[L_{r12}+\frac{2m}{\hbar^2}(E-V)\right]F_\ell^\varkappa-\left(\frac{1}{r_1^2}+\frac{1}{r_2^2}\right)\Bigg[\left\{\frac{(\ell(\ell+1)-\varkappa^2)}{\zeta^2}2r_1^2r_2^2+\frac{\varkappa^2}{4}\right\}F_\ell^\varkappa-\delta_{1\varkappa}(1-2A_\varkappa^\ell)(r_1^2+r_2^2-r_{12}^2)\frac{r_1r_2}{\zeta^2}\tilde{F}_\ell^\varkappa$$

$$+(\varkappa+1)(\varkappa+2)B_{\ell,\varkappa+2}(r_1^2+r_2^2-r_{12}^2)\frac{2r_1r_2}{\zeta^2}\left\{F_\ell^{\varkappa+2}-\frac{\delta_{2,\varkappa+2}}{2}(F_\ell^{\varkappa+2}-\tilde{F}_\ell^{\varkappa+2})\right\}$$

$$+\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell(r_1^2+r_2^2-r_{12}^2)}{2B_{\ell\varkappa}[1+\delta_{0,\varkappa-2}(\sqrt{2}-1)]}\frac{r_1r_2}{\zeta^2}\{F_\ell^{\varkappa-2}+\delta_{0,\varkappa-2}(2\tilde{F}_\ell^{\varkappa-2}+F_\ell^{\varkappa-2})\}\Bigg]+\left(\frac{1}{r_1^2}-\frac{1}{r_2^2}\right)\Bigg[-\varkappa\Bigg\{\frac{1}{2}\frac{r_1^2+r_2^2-r_{12}^2}{\sqrt{r_1r_2[r_{12}^2-(r_1-r_2)^2]}}$$

$$-\frac{r_1^2+r_2^2-r_{12}^2+r_1r_2}{\zeta}\Bigg\}\tilde{F}_\ell^\varkappa+\frac{\varkappa\zeta}{2r_{12}}\frac{\partial\tilde{F}_\ell^\varkappa}{\partial r_{12}}+(\varkappa+1)(\varkappa+2)B_{\ell,\varkappa+2}\frac{2r_1r_2}{\zeta}\left\{-\tilde{F}_\ell^{\varkappa+2}+\frac{1}{2}\delta_{2,\varkappa+2}(F_\ell^{\varkappa+2}+\tilde{F}_\ell^{\varkappa+2})\right\}$$

$$+\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell}{2B_{\ell\varkappa}}\frac{r_1r_2}{(1+\delta_{0,\varkappa-2}(\sqrt{2}-1))}\frac{1}{\zeta}\{\tilde{F}_\ell^{\varkappa-2}+\delta_{0,\varkappa-2}(2F_\ell^{\varkappa-2}+\tilde{F}_\ell^{\varkappa-2})\}-\delta_{1\varkappa}(1-2A_\varkappa^\ell)\frac{r_1r_2}{\zeta}\tilde{F}_\ell^\varkappa\Bigg]=0 \quad (79)$$

Here

$$\zeta=+[r_{12}^4-(r_1^2-r_2^2)^2+2r_{12}^2(r_1^2+r_2^2)]^{\frac{1}{2}} \quad (80)$$

The quantity whose square root ζ is can easily be shown to be positive definite. In the equation (79) the $F_\ell^\varkappa$ is understood to be a function of r_1, r_2, r_{12}:

$$F_\ell^\varkappa=F_\ell^\varkappa(r_1, r_2, r_{12})$$

In addition L_{r12} is the kinetic energy counterpart of the S-wave $L_{\theta 12}$ in terms of r_1, r_2, r_{12}:

$$L_{r12}=\frac{1}{r_1}\frac{\partial^2}{\partial r_1^2}r_1+\frac{1}{r_2}\frac{\partial^2}{\partial r_2^2}r_2+\frac{2}{r_{12}}\frac{\partial^2}{\partial r_{12}^2}r_{12}$$

$$+(r_1^2+r_{12}^2-r_2^2)\frac{1}{r_1r_2}\frac{\partial^2}{\partial r_1\partial r_{12}}$$

$$+(r_2^2+r_{12}^2-r_1^2)\frac{1}{r_1r_2}\frac{\partial^2}{\partial r_2\partial r_{12}} \quad (81)$$

The equations (79) can readily be put in the form of coupled equations for a given $\varkappa$. In that form they would be closest to the form originally given by Holmberg,[6] (although as we have stated his equations apply to three equal mass particles). One salient difference between the two sets of equations, however, is that the present ones are manifestly real, whereas one term in Holmberg's equations is imaginary.[6] It is clear that the equations as well as the solutions must be reducible to completely real form for any given angular momentum state. The accomplishment of this in the present case comes from the explicit construction of real vector spherical harmonics, Eq. (49).

VIII. BOUNDARY CONDITIONS FOR SCATTERING

In this section we derive the asymptotic forms of the radial functions corresponding to the scattering of an electron from a one electron atom in its ground state. The Coulomb modifications when the target system is an ion instead of an atom can readily be made and will have no effect on the angular integrations with which we are here concerned.

As we have seen in the foregoing sections culminating in the last section, the selection of a symmetric choice of Euler angles (Holmberg's angles) has allowed for a completely general derivation of the radial equations. From the point of view of a scattering problem, however, a symmetric choice of angles is not the most advantageous since here we are concerned with an intrinsically asymmetric situation. Thus if we consider that region of configuration space where r_1 is large and r_2 small, corresponding to electron 1 being scattered from the atom to which electron 2 is bound, the wave function *in this region alone* will not be symmetric. However, in terms of the Hylleraas-Breit angles, the spherical angles of one of the particles being defined as two of the Euler angles. the wave function in this asymmetric region is easier to describe. Nevertheless this is a complication of detail only, since all the

angular integrations may readily be performed as we shall now show.

We start with the statement that the complete wave function must have the asymptotic form:

$$\lim_{r_1\to\infty}\psi_{\ell 0}(\mathbf{r}_1,\mathbf{r}_2)=\frac{1}{r_1}\sin\left(kr_1+\delta\ell-\frac{\ell\pi}{2}\right)Y_{\ell 0}(\Omega_1)\frac{R_{1s}}{r_2}(r_2)Y_{00}(\Omega_2) \quad (82)$$

where $(R_{1s}(r_2)/r_2)\,Y_{00}(\Omega_2)$ is the ground state of the one-electron atom (hydrogen). On the other hand, from Eq. (51).

$$\lim_{r_1\to\infty}\psi_{\ell 0}(\mathbf{r}_1,\mathbf{r}_2)=\frac{1}{r_1}\sin\left(kr_1+\delta\ell-\frac{\ell\pi}{2}\right)\frac{R_{1s}(r_2)}{r_2}\times \times\sum_{\varkappa}''[\alpha_\ell^{\varkappa+}(\theta_{12})D_\ell^{\varkappa+}(\theta,\Phi,\psi)+\alpha_\ell^{\varkappa-}(\theta_{12})D_\ell^{\varkappa-}(\theta,\Phi,\psi)] \quad (83)$$

where

$$\alpha_\ell^{\varkappa\pm}(\theta_{12})=Y_{00}\int_0^\pi\int_0^{2\pi}\int_0^{2\pi}D_\ell^{\varkappa\pm}(\theta,\Phi,\psi)Y_{\ell 0}(\vartheta_1)\sin\theta d\theta\,d\Phi\,d\psi \quad (84)$$

It should be noted that (82) refers to the state of parity $(-1)^\ell$ as long as we are considering elastic scattering from the ground (1s) state. This then defines the evenness of oddness of the values over which $\varkappa$ goes in the summation in Eq. (83).

The quadrature in (84) can readily be performed by recalling from Section II that ϑ_1 is the angle between $\hat{\mathbf{z}}$ and $\hat{\mathbf{r}}_1$ whose spherical angles in the primed coordinate systems are given in Eq. (20). One can then use these spherical angles to expand $P_\ell(\cos\vartheta_1)$ via the addition theorems for spherical harmonics.

In its real form [9] this gives in the present case

$$P_\ell(\vartheta_1)=P_\ell(\pi/2)P_\ell(0)+2\sum_{m=1}^{\ell}(-1)^m\frac{(\ell-m)!}{(\ell+m)!}P_\ell^m(\pi/2)P_\ell^m(\theta)\cos m\left(\psi-\frac{1}{2}\theta_{12}\right) \quad (85)$$

In (85) we have written both the Legendre and associated Legendre polynomials as functions of the angle but what we mean in all cases is that the angle is to be substituted into the transcendental form of the function. For example $P_1(\beta)=\cos\beta$ and *not* $P_1(\beta)=\beta$. The sign of the P_ℓ^m is that of Magnus and Oberhettinger [9] (which differs by $(-1)^m$ from that of Morse and Feshbach [3]). To complete the quadrature in (84) we note that

$$d_\ell^{0,\varkappa}(\theta)=\left[\frac{(\ell-\varkappa)!}{(\ell+\varkappa)!}\right]^{\frac{1}{2}}P_\ell^\varkappa(\theta) \quad (86)$$

Substitution into (84) now yields the desired result:

$$\alpha_\ell^{\varkappa+}(\theta_{12})=(-1)^\varkappa P_\ell^\varkappa\left(\frac{\pi}{2}\right)\left[\frac{(\ell-\varkappa)!}{(\ell+\varkappa)!}\right]^{\frac{1}{2}}\left\{\frac{\delta_{0\varkappa}}{\sqrt{2}}+(1-\delta_{0\varkappa})\cos\left(\varkappa\frac{\theta}{2}12\right)\right\} \quad (87a)$$

$$\alpha_\ell^{\varkappa-}(\theta_{12})=(-1)^\varkappa P_\ell^\varkappa\left(\frac{\pi}{2}\right)\left[\frac{(\ell-\varkappa)!}{(\ell+\varkappa)!}\right]^{\frac{1}{2}}\sin\left(\varkappa\frac{\theta}{2}12\right) \quad (87b)$$

The radial function themselves thus approach

$$\lim_{r_1\to\infty}f_\ell^{\varkappa\pm}(r_1,r_2,\theta_{12})=\frac{1}{r_1}\sin\left(kr_1+\delta_\ell-\ell\frac{\pi}{2}\right)\frac{R_{1s}(r_2)}{r_2}\alpha_\ell^{\varkappa\pm}(\theta_{12}) \quad (88)$$

in which form we see that r_1, r_2 dependence of all the $f_\ell^{\varkappa\pm}$ functions is independent of $\varkappa$, so that none of them vanishes in the asymptotic region. Since in all cases the θ_{12} dependence is trivial, it may be worthwhile to define new functions whose asymptotic behaviour is strictly the r_1, r_2 dependence in (88).

For bound state problems, it is clear that all the radial functions must vanish in all asymptotic regions.

IX. OTHER APPLICATIONS

In addition to two electron atomic or ionic systems the present equations apply to double *mu* or *pi* mesic atoms, although as the mass of the identical particles get heavier, the correction for the center of mass becomes more important. Also for the spinless bosons (*pi* mesons) only the space symmetric solutions will presumably be relevant.

The equations can also be applied to two different particles of the same mass (positron-hydrogen scattering, for example). In this case, the potential, V will no longer be symmetric hence the solutions will not be symmetric which implies that boundary conditions like (72) must be

changed to matching conditions of the asymmetric solutions along the line $r_1 = r_2$.[19] This has the effect of giving one solution where formerly there were two, in accord with the distinguishability of the particles.

The major further application of this approach is to two-electron diatomic molecules. In this case, the extension from one [20] to two electrons is non-trivial. However, the analysis has been completed and will be published elsewhere.[21]

TABLE 1.—*Coefficients of the Angular Derivatives in the Kinetic Energy.*[a] *Cf. Eq. (49)*

Coefficient	Derivative	Coefficient	Derivative
$\dfrac{\sin^2\left(\psi \mp \frac{\theta_{12}}{2}\right)}{\sin^2 \theta_{12}}$	$\dfrac{\partial^2}{\partial\theta^2}$	$\cot \theta_{12}$	$\dfrac{\partial}{\partial\theta_{12}}$
$\dfrac{\cos^2\left(\psi + \frac{\theta_{12}}{2}\right)}{\sin^2 \theta_{12} \sin^2\theta}$	$\dfrac{\partial^2}{\partial\Phi^2}$	$\dfrac{-\sin (2\psi + \theta_{12})}{\sin^2 \theta_{12} \sin \theta}$	$\dfrac{\partial^2}{\partial\theta\partial\Phi}$
A_1	$\dfrac{\partial^2}{\partial\psi^2}$	$\dfrac{\cot \theta \sin (2\psi + \theta_{12})}{\sin_2 \theta_{12}}$	$\dfrac{\partial^2}{\partial\theta\partial\psi}$
1	$\dfrac{\partial^2}{\partial\theta_{12}^2}$	0	$\dfrac{\partial^2}{\partial\theta\partial\theta_{12}}$
$\dfrac{\cot \theta \cos^2\left(\psi + \frac{\theta_{12}}{2}\right)}{\sin^2 \theta_{12}}$	$\dfrac{\partial}{\partial\theta}$	$\dfrac{-2 \cos \theta \cos^2\left(\psi + \frac{\theta_{12}}{2}\right)}{\sin^2 \theta_{12} \sin^2\theta}$	$\dfrac{\partial^2}{\partial\Phi\partial\psi}$
$\dfrac{\cos \theta \sin (2\psi + \theta_{12})}{\sin^2 \theta \sin^2 \theta^{12}}$	$\dfrac{\partial}{\partial\Phi}$	0	$\dfrac{\partial^2}{\partial\Phi\partial\theta_{12}}$
B_1	$\dfrac{\partial}{\partial\psi}$	-1	$\dfrac{\partial^2}{\partial\psi\partial\theta_{12}}$

[a] A_1 and B_1 are given in Eqs. (60) and (61).

APPENDIX I

In this appendix, we prove the Eqs. (65) and (66). For $m=0$, $\frac{\partial}{\partial\Phi}$ terms give zero. Therefore, we can write

$$\Lambda_1 \rightarrow \frac{\partial^2}{\partial\theta^2} - \cot\theta\frac{\partial}{\partial\theta} - \cot^2\theta\frac{\partial^2}{\partial\psi^2} \tag{I-1}$$

$$\Lambda_2 \rightarrow 2\cot\theta\frac{\partial^2}{\partial\psi\partial\theta} - \frac{\partial}{\partial\psi} - 2\cot^2\theta\frac{\partial}{\partial\psi} \tag{I-2}$$

and

$$D_\ell^{\varkappa+} = N_{\ell\varkappa}\cos\varkappa\psi\sin^\varkappa\theta F_\varkappa \tag{I-3}$$

$$D_\ell^{\varkappa-} = N_{\ell\varkappa}\sin\varkappa\psi\sin^\varkappa\theta F_\varkappa \tag{I-4}$$

where

$$N_{\ell\varkappa} = \sqrt{\frac{2\ell+1}{8\pi^2}\frac{(\ell+\varkappa)!}{(\ell-\varkappa)!}}\frac{1}{\varkappa!}\frac{1}{2^\varkappa} \tag{I-5}$$

and

$$F_\varkappa = F\left(\varkappa-\ell,\ \varkappa+\ell+1,\ 1+\varkappa,\ \sin^2\frac{\theta}{2}\right) \tag{I-6}$$

Now

$$\begin{aligned}\Lambda_1 D_\ell^{\varkappa+} = N_{\ell\varkappa}\cos\varkappa\psi[&2\varkappa(\varkappa-1)\sin^{\varkappa-2}\theta\cos^2\theta F_\varkappa \\ &-\varkappa\sin^\varkappa\theta F_\varkappa + 2\varkappa\sin^\varkappa\theta\cos\theta A_\varkappa^\ell F_{\varkappa+1} \\ &+\sin^{\varkappa+2}\theta A_\varkappa^\ell A_{\varkappa+1}^\ell F_{\varkappa+2}]\end{aligned} \tag{I-7}$$

where we have used the well known relations for the derivatives of hypergeometric functions:

$$\frac{dF_\varkappa}{d\theta} = A_\varkappa^\ell\sin\theta F_{\varkappa+1} \tag{I-8}$$

$$\frac{d^2}{d\theta^2}F_\varkappa = A_\varkappa^\ell\cos\theta F_{\varkappa+1} + A_\varkappa^\ell A_{\varkappa+1}^\ell\sin^2\theta F_{\varkappa+2} \tag{I-9}$$

A relation between $F_\varkappa$, $F_{\varkappa+2}$ can be obtained from the differential equation satisfied by the hypergeometric function

$$\cos\theta F_{\varkappa+1} = F_\varkappa - \sin^2\theta\frac{A_{\varkappa+1}^\ell}{2(\varkappa+1)}F_{\varkappa+2} \tag{I-10}$$

Using (I-10) in (I-7), we find

$$\begin{aligned}\Lambda_1 D_\ell^{\varkappa+} = N_{\ell\varkappa}\cos\varkappa\psi\Big[&2\varkappa(\varkappa-1)\sin^{\varkappa-2}\theta\cos^2\theta F_\varkappa \\ &-\varkappa\sin^\varkappa\theta F_\varkappa + 2\varkappa\sin^\varkappa\theta A_\varkappa^\ell F_\varkappa \\ &+\left(1-\frac{\varkappa}{\varkappa+1}\right)\sin^{\varkappa+2}\theta A_\varkappa^\ell A_{\varkappa+1}^\ell F_{\varkappa+2}\Big]\end{aligned} \tag{I-11}$$

Similarly,

$$\begin{aligned}\Lambda_2 D_\ell^{\varkappa+} = -N_{\ell\varkappa}\sin\varkappa\psi\Big[&2\varkappa(\varkappa-1)\sin^{\varkappa-2}\theta\cos^2\theta F_\varkappa \\ &-\varkappa\sin^\varkappa\theta F_\varkappa + 2\varkappa A_\varkappa^\ell\sin^\varkappa\theta F_\varkappa \\ &-\frac{\varkappa}{\varkappa+1}\sin^{\varkappa+2}\theta A_\varkappa^\ell A_{\varkappa+1}^\ell F_{\varkappa+2}\Big]\end{aligned} \tag{I-12}$$

Multiply (I-12) by $\sin 2\psi$ and (I-11) by $\cos 2\psi$ and subtract to get

$$\begin{aligned}(\sin 2\psi\Lambda_2 - \cos 2\psi\Lambda_1)D_\ell^{\varkappa+} = &-N_{\ell\varkappa}\cos(\varkappa-2)\psi G \\ &-\frac{N_{\ell\varkappa}}{N_{\ell\varkappa+2}}A_\varkappa^\ell A_{\varkappa+1}^\ell D_\ell^{(\varkappa+2)+}\end{aligned} \tag{I-13}$$

where

$$\begin{aligned}G = [&2\varkappa(\varkappa-1)\sin^{\varkappa-2}\theta\cos^2\theta - \varkappa\sin^\varkappa\theta \\ &+2\varkappa A_\varkappa^\ell\sin^\varkappa\theta]F_\varkappa \\ &+\frac{1-\varkappa}{2(1+)\varkappa}\sin^{\varkappa+2}\theta A_\varkappa^\ell A_{\varkappa+1}^\ell F_{\varkappa+2}\end{aligned} \tag{I-14}$$

Let $\varkappa\rightarrow\varkappa+1$ in (I-10), then

$$\cos\theta F_{\varkappa+2} = F_{\varkappa+1} - \frac{\sin^2\theta}{2(\varkappa+2)}A_{\varkappa+2}^\ell F_{\varkappa+3}$$

Substituting in the above for $F_{\varkappa+1}$ and $F_{\varkappa+3}$ by using (I-10) we get after some rearrangement

$$\begin{aligned}\Big[\cos^2\theta\frac{(\varkappa+2)}{(\varkappa+1)} + &\frac{\varkappa+2}{2(\varkappa+1)^2}\sin^2\theta A_{\varkappa+1}^\ell \\ &+\frac{\sin^2\theta A_{\varkappa+2}^\ell}{2(\varkappa+1)}\Big]F_{\varkappa+2} \\ &-\sin^4\theta\frac{A_{\varkappa+2}^\ell A_{\varkappa+3}^\ell}{4(\varkappa+1)(\varkappa+3)}F_{\varkappa+4} = \frac{\varkappa+2}{\varkappa+1}F_\varkappa\end{aligned}$$

Letting $\varkappa \to \varkappa - 2$ multiplying by $2(\varkappa-1)^2 \sin^2\theta$ and rearranging we have for $\varkappa \geq 2$

$$G = 2\varkappa(\varkappa-1)\sin^{\varkappa-2}\theta F_{\varkappa-2} \qquad \text{(I-15)}$$

Also we find directly from (I-11) and (I-12)

$$G = \tfrac{1}{2}\sin^2\theta A_0^{\ell} A_1^{\ell} F_2 \quad \varkappa = 0 \qquad \text{(I-16)}$$

$$G = -(1-2A_{\varkappa}^{\ell})\sin\theta F_1 \quad \varkappa = 1 \qquad \text{(I-17)}$$

Finally then with the substitution of the above in (I-13) we obtain for $\varkappa \geq 2$

$$(\sin 2\psi\Lambda_2 - \cos 2\psi\Lambda_1) D_{\ell}^{\varkappa+} = -2\varkappa(\varkappa-1) B_{\ell\varkappa} D_{\ell}^{(\varkappa-2)+} - \frac{A_{\varkappa}^{\ell} A_{\varkappa+1}^{\ell}}{2B_{\ell,\varkappa+2}} D_{\ell}^{(\varkappa+2)+} \qquad \text{(I-18)}$$

where $B_{\ell\varkappa}$ has already been defined in Eq. (68).

The special cases $\varkappa = 0$, 1 can be determined from (I-13), (I-16), (I-17). With proper normalization

$$(\sin 2\psi\Lambda_2 - \cos 2\psi\Lambda_1) D_{\ell}^{0+} = -\sqrt{2}\frac{A_0^{\ell} A_1^{\ell}}{B_{\ell 2}} D_{\ell}^{2+} \qquad \text{(I-19)}$$

and

$$(\sin 2\psi\Lambda_2 - \cos 2\psi\Lambda_1) D_{\ell}^{1+} = (1-2A_1^{\ell}) D_{\ell}^{1+} - \frac{A_1^{\ell} A_2^{\ell}}{2B_{\ell 3}} D_{\ell}^{3+} \qquad \text{(I-20)}$$

We can combine (I-18, I-19, I-20) to get Eq. (65)a. If $\ell - \varkappa$ is odd, one can show that the + and − superscripts become reversed as is implicit in (65a)

Similarly, we can prove (65b), (66a), (66b).

APPENDIX II

We can form the raising and lowering operators $M_{\pm}=M_x\pm iM_y$ by using Eqs. (36) and (37). It can be proved easily

$$\frac{i}{\hbar}M+D_l^{m,k}=-\sqrt{(l-m)(l+m+1)}D_l^{m+1,k} \quad \text{(II-1)}$$

$$\frac{i}{\hbar}M-D_l^{m,k}=\sqrt{(l-m)(l-m+1)}D_l^{m-1,k} \quad \text{(II-2)}$$

It is well known that $M_{\pm}$ commutes with H, and in particular with the kinetic energy. One can show explicitly that they commute with the relevant part of the kinetic energy given below:

$$[\sin 2\psi\Lambda_2-\cos 2\psi\Lambda_1,\ M_{\pm}]=0 \quad \text{(II-3)}$$

$$[\sin 2\psi\Lambda_1+\cos 2\psi\Lambda_2,\ M_{\pm}]=0 \quad \text{(II-4)}$$

It may be useful to give the following relations

$$[\Lambda_1,\ M_{\pm}]=\pm\frac{2\hbar}{\sin\theta}e^{\pm i\Phi}\Lambda_2 \quad \text{(II-5)}$$

$$[\Lambda_2,\ M_{\pm}]=\mp\frac{2\hbar}{\sin\theta}e^{\pm i\Phi}\Lambda_1 \quad \text{(II-6)}$$

Below we give the results of the raising and lowering operators on the exchange vector spherical harmonics. These results may be derived from Eqs. (II 1), (II 2) and the definition (54).

$$\frac{i}{\hbar}M_+D_l^{(m,\kappa)\pm}=-\sqrt{(l-m)(l+m+1)}D_l^{(m+1,\kappa)\pm} \quad \text{(II-7)}$$

$$\frac{i}{\hbar}M_-D_l^{(m,\kappa)\pm}=\sqrt{(l+m)(l-m+1)}D_l^{(m-1,\kappa)\pm} \quad \text{(II-8)}$$

Operating $M\pm$ on Eqs. (65a), we find

$$\begin{aligned}M_+&(\sin 2\psi\Lambda_2-\cos 2\psi\Lambda_1)D_l^{\kappa+}\\&=-2\kappa(\kappa-1)B_{l\kappa}(1-\delta_{0\kappa})M_+D_l^{(\kappa-2)+}\\&\quad+\delta_{1\kappa}(1-2A_\kappa^l)M_+D_l^{\kappa+}\\&\quad-\frac{-1+3\delta_{0\kappa}}{(1+\delta_{0\kappa}(\sqrt{2}-1))}\frac{A_\kappa^lA_{\kappa+1}^l}{2B_{l,\kappa+2}}\times M_+D_l^{(\kappa+2)+}\end{aligned} \quad \text{(II-9)}$$

Using Eqs. (II 3) and (II 7), we get

$$\begin{aligned}(\sin 2\psi\Lambda_2&-\cos 2\psi\Lambda_1)D_l^{(1,\kappa)+}\\&=-2\kappa(\kappa-1)B_{l\kappa}(1-\delta_{0\kappa})D_l^{(1,\kappa-2)+}\\&\quad+\delta_{1\kappa}(1-2A_\kappa^l)D_l^{(1,\kappa)+}\\&\quad+\frac{(1+3\delta_{0\kappa})}{(1+\delta_{0\kappa}(\sqrt{2}-1))}\frac{A_\kappa^lA_{\kappa+1}^l}{2B_{l,\kappa+2}}D_l^{(1,\kappa+2)+}\end{aligned} \quad \text{(II-10)}$$

With repeated operation of M_+, it follows by induction that the above relation is true for any m. Similarly we can obtain Eqs. (65b), (66a) and (66b) for any m. Thus the radial equations are independent of m.

It can also be shown that

$$\psi_{l-m}=(-1)^{\kappa-m}\psi_{lm}^* \quad \text{(II-11)}$$

In order to show this, it is only necessary to state the easily derivable relation

$$[D_l^{(m,\kappa)\pm}]^*=(-1)^{\kappa-m}D_l^{(-m,\kappa)\pm} \quad \text{(II-12)}$$

REFERENCES

1. E. A. Hylleraas, Z. Physik **48**, 469 (1928); Z. Physik **54**, 347 (1929).
2. G. Breit, Phys. Rev. **35**, 569 (1930).
3. A clear exposition of Breit's work is contained in P. M. Morse and H. Feshbach, *Method of Theoretical Physics* (McGraw-Hill, New York, 1953), p. 1719 et seq.
4. U. Fano, private communication. One of us (A. T.) acknowledges valuable discussions with Dr. Fano in 1959 at the outset of the formulation of these ideas.
5. The D-wave equations in B-H angles have been worked out by H. Feshbach, M.I.T. thesis (1942, unpublished and by one of us, A. Temkin (1959, unpublished). The wave functions for several states have been derived by C. Schwartz, Phys. Rev. 123, 1700 (1961).
6. B. Holmberg, Kungl. Fysiografiska Sallskapets I Lund Förhandlingar **26**, 135 (1956).

7. G. H. Derrick and J. M. Blatt, Nuclear Phys. **8**, 310 (1958). G. H. Derrick, Nuclear Phys. **16**, 405 (1960). The characteristic of the three (or more) body problem of finite masses is that six coordinates can be eliminated, the three additional constants of the motion coming from the center of mass. Two papers dealing with the elimination of the six coordinates of rotation and translation from a system of N particles are:
J. O. Hirschfelder and E. Wigner, Proc. Nat. Acad. Sci. **21**, 113 (1935); C. F. Curtiss, J. O. Hirschfelder, and F. T. Adler, J. Chem. Phys. **18**, 1638 (1950).
8. E. Wigner, *Group Theory*, (Academic Press, New York and London, 1959), p. 216. cf. also, A. R. Edmonds, *Angular Momentum in Quantum Mechanics*, (Princeton University Press, Princeton, 1957), Chapter 4.
9. W. Magnus and I. Oberhettinger, *Formulas and Theorems for the Functions of Mathematical Physics*. (Chelsea Publishing Co., New York, 1949).
10. L. Pauling and E. B. Wilson, Introduction to Quantum Mechanics, (McGraw-Hill Book Co., Inc., New York and London, 1935), p. 280.
11. A. Temkin, Phys. Rev. 126, 130 (1962), P. Luke, R. Meyerott, W. Clendenin, Phys. Rev. **86**, 401 (1952).
12. A. Temkin and E. Sullivan, Phys. Rev. 129, 1250 (1963).
13. A. W. Weiss, Phys. Rev. **122**, 1826 (1961).
14. C. Schwartz, Phys. Rev. **126**, 1015 (1962).
15. A. W. Weiss and J. B. Martin, Phys. Rev. **132**, 2119 (1963). For a somewhat different finding from those of reference 13 and 14 and the preceding paper of this reference, cf. W. Lakin, A.E.C. Report NYO-10, 430 (1963, unpublished).
16. A. Temkin, Phys. Rev. Letters, **6**, 354 (1961).
17. C. L. Perkeris, Phys. Rev. **112**, 1649 (1958).
18. C. L. Perkeris, Phys. Rev. **115**, 1216 (1959). C. Schwartz, Phys. Rev. **128**, 1146 (1962).
19. A. Temkin, Proc. Phys. Soc. (London) **80**, 1277 (1962).
20. A. Temkin, J. Chem. Phys. 39, 161 (1963).
21. A. Temkin and A. K. Bhatia (to be published).

Sec A64-25147

SPACE: HIGHLIGHTS OF RECENT RESEARCH*

ROBERT JASTROW AND A. G. W. CAMERON

Space science is the collection of scientific problems to which space vehicles can make some specific contributions not achievable by ground-based experiments. At present this field includes broad segments of the traditional disciplines of the earth sciences, physics, and astronomy. In future years the biological sciences will join this group in an important role, as our explorations of the moon and planets provide us with opportunities for studying the conditions under which physical life may have developed. This article reviews some highlights of recent space research in the physical sciences.

GEODESY

Important results have been achieved in determining the internal structure of our own planet with the aid of near earth satellites. A satellite's orbit is determined by the distribution of mass within the earth. If the earth were a perfect sphere, under the attraction of the mass point at the earth's center of gravity the satellite would move in an ellipse whose plane would have a constant orientation in space.

Actually, the plane of a satellite's orbit rotates slowly in space, due to the additional force of attraction exerted by the equatorial bulge. Studies of the orbital rotation rates of a number of satellites have yielded a very precise value for the height of the equatorial bulge. These indicate a discrepancy between the observed value for the flattening and the value expected on the basis of an assumption of hydrostatic equilibrium. It has been suggested by Munk and MacDonald that these results imply that the interior of the earth is not in hydrostatic equilibrium but has a mechanical strength sufficient to maintain the earth's shape in spite of the stresses at the base of the mantle.

There are other departures of the geoid from the shape expected in the case of hydrostatic equilibrium, in addition to the discrepancy in the flattening. These departures are of very great significance because they represent variations in gravity which depend on the entire distribution of mass within the planet, which are more significant for the gross structure of the planet than the simple topographical variations—for example, mountains—which represent the distribution of the mass at the surface alone.

Detailed analysis of these gravitational variations yields a figure of the earth having a positive anomaly, or a lump, in the region of the western Pacific, near Indonesia and the Philippines, a large depression or negative anomaly in the Indian Ocean, and a negative anomaly in the Antarctic (Fig. 1).

Although these depressions and elevations are relatively minute, they are exceedingly significant because they represent variations in the force of gravity, or in the amount of matter per square centimeter, in the regions in question. For example, the depression in the Indian Ocean is only 60 meters deep, but it signifies that the force of gravity there is, relatively, so weak that the waters of the sea are not drawn together to the depth that one would expect if the whole earth were subject to a uniform gravitational force.

These anomalies are correlated with the rate at which heat flows through the body of the earth to the surface. The correlation is such that, where the geoid is anomalously high, the heat flow is anomalously low. On the average, the flow of heat outward through the crust of the earth is 60 erg cm^{-2} sec^{-1}. In the depression of the geoid near India, the flow of heat is substantially higher, 80 erg cm^{-2} sec^{-1}. At the elevation of the geoid in the western Pacific, the flow of heat is substantially lower, about 40 erg cm^{-2} sec^{-1}.

We expect this kind of correlation if there is a

*Published in *Science*, 145 (3637): 1129–1139, September 11, 1964.

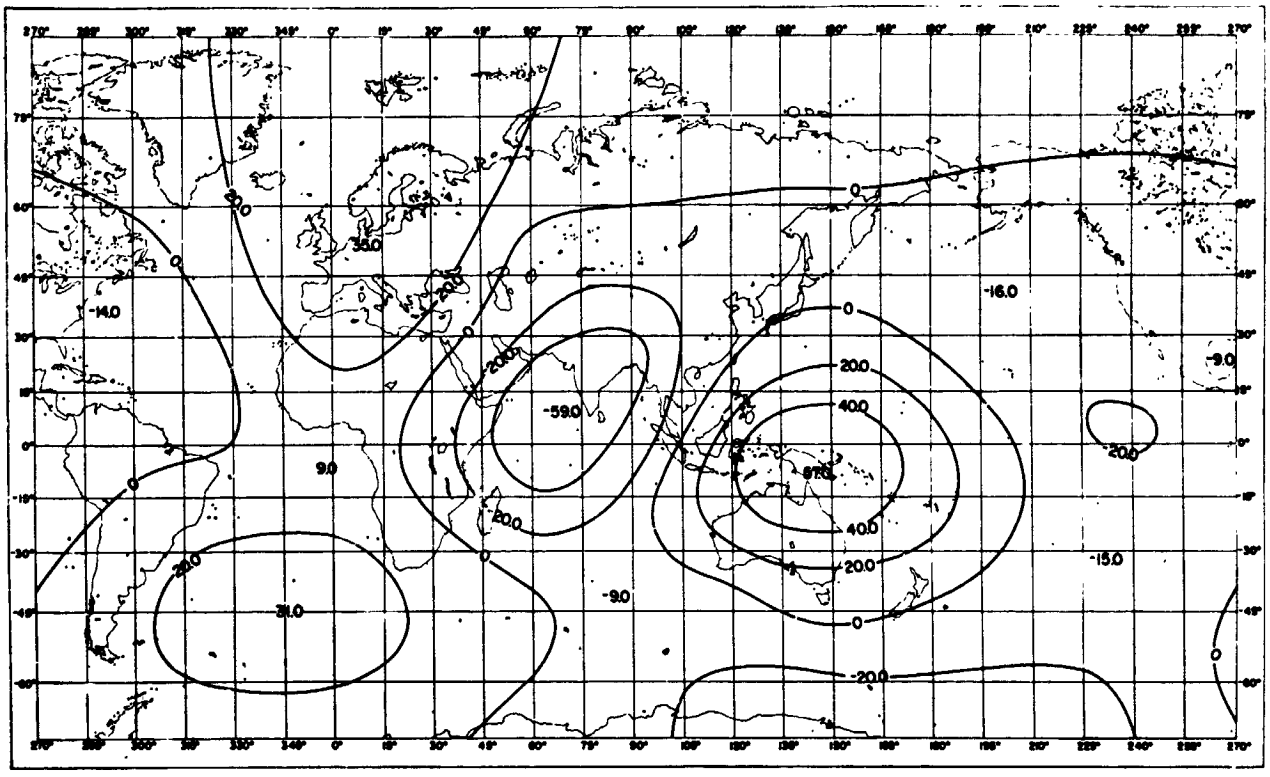


FIGURE 1.—Geoid heights (in meters) relative to an ellipsoid with a flattening of 1/298.3. The major features include a negative anomaly in the Indian Ocean and a positive anomaly centered near Indonesia and the Philippines. [From G. J. F. MacDonald, *Science* 143, 921 (1964)].

mass transport, or convection of matter, from the deep interior of the earth to the surface in these regions. If there were an upward motion through the interior of the earth which carried relatively warm material from below to the surface, this upward-moving column would have a lower density than its surroundings, and therefore the mass per square centimeter in the column, and the gravitational force on the surface of the earth about it, would be lower than the average. At the same time, the heat which the warm column transports upward would add to the normal release of radioactive heat throughout the mantle and crust, so that above the same upward-moving column there would be an exceptionally high rate of heat flow through the surface. The converse would hold for a descending column, which would carry a relatively dense, and therefore relatively cold, material from the surface layer to the interior of the earth. Above the cold and dense column the gravitational force would be relatively great, and a bump would appear in the sea level there. This may be the cause of the elevation in the western Pacific.

METEOROLOGY

In geocentric order, the next major area of investigation in space science concerns the atmosphere and the control exerted over it by the sun. This field of research includes questions related to the circulation of the winds in the lower atmosphere and to the vertical structure of the atmosphere at higher altitudes.

Eight Tiros satellites have been launched in the past 4 years, all carrying vidicon cameras for the global study of the cloud cover; Tiros II, III, IV, and VII carried, in addition, a set of infrared detectors for measuring the intensity of infrared radiation emitted from the earth-atmosphere system.

The cloud-cover photographs have already yielded results that are of great interest when correlated with ground observations, and they indicate that use of satellites may lead to a substantial improvement in weather forecasting by providing global and nearly continuous coverage of regions of weather activity. The matter of global coverage is critically important, because the success of weather forecasting has been found

to increase rapidly with the size of the region covered by the observations; yet at present large parts of the globe are very poorly covered. These are regions in which weather activity can develop and grow without detection before moving out into the inhabited areas. They include the polar regions, the major deserts, and the southern oceans. Satellite coverage will greatly strengthen the hand of the meteorologist by filling in these blank portions of the global weather map, and it may be expected to have important consequences for the economies of this country and the world.

The cloud photographs may also be important for the basic objectives of long-range forecasting and the understanding of the causes of weather activity, because clouds have a strong influence on both the amount of solar energy admitted to the earth-atmosphere system and the amount of energy returned to space from the surface.

The energy balance of the earth-atmosphere system is the difference between the incoming solar radiation, mostly in the visible region of the spectrum, and the outgoing terrestrial radiation, in the infrared. The latitudinal variation of the energy balance shows an excess of incoming solar radiation over outgoing radiation near the equator, and a deficiency at the poles. It is this variation of the energy balance with latitude that drives the atmospheric heat engine.

That part of the visible radiation from the sun which is not reflected by clouds or scattered in the atmosphere reaches the surface of the earth and is absorbed, heating the ground to a temperature in the neighborhood of 235°K. For a glowing body at a temperature of 235°K, most of the energy is radiated at wavelengths in the far infrared. This infrared radiation is strongly absorbed by several constituents of the atmosphere, including water, carbon dioxide, and ozone. The absorption of infrared radiation from the ground by these molecules heats the lower atmosphere, which reradiates the absorbed energy, partly upward to outer space and partly downward, providing additional heating of the surface.

The additional heating of the surface by the return of infrared radiation from the atmosphere is analogous to the action of the glass panes of a greenhouse, and is called the "greenhouse effect." It is sufficient to raise the temperature of the surface of the earth by about 55°K, so that the average temperature of the surface of our planet becomes 290°K.

Clouds are strong absorbers of infrared radiation and influence the infrared transmission and the local greenhouse effect. (This is in addition to the primary effect of clouds as reflectors of incident visible solar radiation.)

Thus, the cloud cover has an important effect on the deposition of energy in the atmosphere because it influences both the inflow and the outflow of energy through the atmosphere. Thus far the characteristics of clouds—amount, types, and approximate heights—have all been measured by ground-based observers. Satellite observation by television cameras introduces the possibility of obtaining extensive data on global cloud cover in relatively short periods.

A. Arking, of the Goddard Institute for Space Studies and New York University, has recently carried out the first statistical analysis of Tiros cloud-cover data, using a digitizing technique applied to Tiros III photographs. The Tiros III results are compared in Fig. 2 with a climatological mean cloudiness for the Northern Hemisphere that was compiled by K. Telegadas and J. London, then at New York University, from ground observations extending over a 50-year period. The Southern Hemisphere data of Fig. 2 were compiled by H. Landsberg of the U.S. Weather Bureau.

The results in Fig. 2 show that the cloud cover in middle latitudes is the same in the Northern and Southern hemispheres. However, in tropical latitudes there is an asymmetry, with a local maximum of the cloud cover in the tropics centered at 10° N. This is the average position of the "thermal equator" during the period 12 July to 30 September.

These results are preliminary, but the approach seems promising, and it is hoped that improved techniques will make it possible to detect fine-scale temporal and geographical variations in the distribution of the energy balance.

UPPER ATMOSPHERE

The physical processes which control the upper atmosphere are determined largely by the absorption of solar ultraviolet radiation by the atoms and molecules existing at great heights. Although the ultraviolet component of the solar radiation

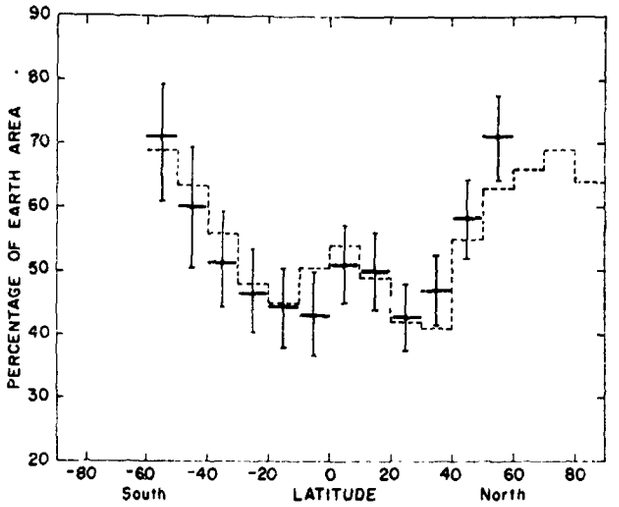


FIGURE 2.—The latitudinal distribution of cloud cover. (Solid horizontal bars.) Results derived from Tiros III photographs taken from 12 July to 30 September 1961; (vertical lines) estimated uncertainty in the Tiros derived data; (dashed histogram) climatological mean cloud cover derived from ground observations, as compiled by Telegadas and London for the Northern Hemisphere and by Landsberg for the Southern Hemisphere. The broad features of the latitudinal distribution are consistent with the known pattern of the general circulation. The air rising at the thermal equator produces a relative maximum in the cloud cover, while, on the average, there is downward motion of cool, dry air at latitudes 30° north and south of the thermal equator, which explains the realtive minimum of cloudiness. (From A. Arking, *Science* 143, 569 1964)

is only a small fraction of the total flux of solar energy, the absorption cross sections in the far ultraviolet are so large that radiation at these wavelengths has been effectively removed from the incident spectrum by the time the incident flux has penetrated to a height of 100 kilometers. The ultraviolet radiation is the principal source of heating of the thin upper air and the major determining factor in its structure.

At lower altitudes the air is composed of oxygen and nitrogen, and we can measure the proportions of these rather accurately. At the highest altitudes these gases have partially settled out of the air through diffusion. The lighter gases dominate the composition of the air at sufficiently high altitudes. Of these gases hydrogen is the lightest, and for this reason it was once believed to be the dominant constituent of the air above the oxygen-nitrogen layer. The emergence of the hydrogen atmosphere was thought to come at an altitude of about 1200 kilometers. However, in July 1961, Marcel Nicolet of Belgium suggested, on the basis of an initial examination of the density data of Echo I, that between the oxygen-nitrogen atmosphere and the hydrogen atmosphere there lies a layer of helium. The helium layer was discovered experimentally a short time later by R. Bourdeau, of the NASA Goddard Space Flight Center, and the finding has since been confirmed in other experiments (Fig. 3).

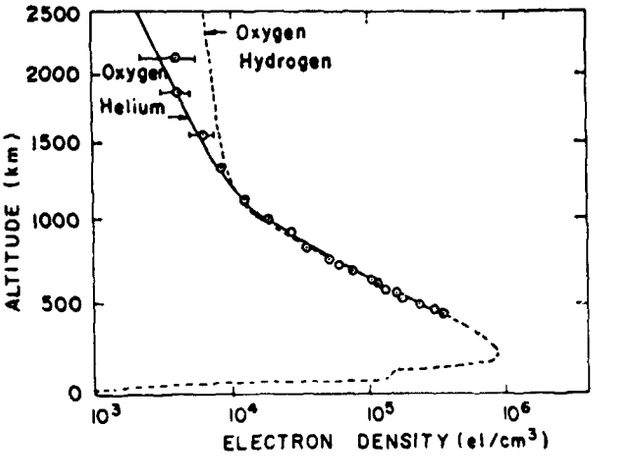


FIGURE 3.—Electron density profile measured by a Scout rocket, and providing evidence for a helium layer in the upper atmosphere. (Dotted line) Density distribution for a scale height derived from an oxygen-hydrogen mixture; (solid line) calculated density distribution for an oxygen-hydrogen mixture. [From S. J. Bauer and J. E. Jackson, J. Geophys. Res. 67, 1676 (1962)]

Our knowledge of properties of the atmosphere at altitudes above 200 kilometers is mainly derived from measurements of atmospheric drag on satellites. The period of revolution of a satellite decreases steadily at a rate proportional to the drag force exerted by the atmosphere, which is, in turn, proportional to air density; a measurement of the rate of change of period therefore gives the value of the air density suitably averaged around the orbit.

Data on satellite drag have been a very valuable source of information on atmospheric properties. L. G. Jacchia of the Smithsonian Astrophysical Observatory was the first to discover, by careful analysis of time variations in the drag, that the upper atmosphere is extremely responsive to solar control, deviating in density from the mean by as

much as a factor of 100 and deviating in temperature by hundreds of degrees at times of solar activity.

The significance of the correlation between solar activity and the properties of the earth's upper atmosphere can be described as follows. The surface of the sun is the scene of great activity, especially during the maximum of the sunspot cycle, when it is marked by sunspots and by hot, dense regions with temperatures of some millions of degrees, which are located in the solar corona above the sunspot areas. When such an active region faces toward the earth in the course of the sun's rotation, extreme ultraviolet radiation emitted from these active regions is absorbed in the upper atmosphere. The precise correlation between solar activity and density was discovered by Jacchia and by W. Priester of Bonn University Observatory and the Goddard Institute for Space Studies. Their results suggest that the amount of energy transferred to the earth is sufficient to heat the atmosphere appreciably, causing an upward expansion and a large increase in the density of the exceedingly thin air at high altitudes. This discovery provided the first direct evidence of the influence of solar surface activity on atmospheric properties.

The continuing analysis of the correlation has given us a rather full picture of solar control over the upper atmosphere. It indicates that the atmosphere is appreciably heated by the ultraviolet radiation emitted at times of general solar surface activity, and is further heated by interaction of the earth with the clouds of solar particles which are emitted from the sun following solar surface eruptions. The arrival of the clouds of particles at the earth is signified by the onset of geomagnetic disturbances or "magnetic storms." It is found that increases in the temperature of the atmosphere occur shortly after the commencement of the magnetic storms. Thus, it appears that both ultraviolet radiation and corpuscular streams are sources of energy for the upper atmosphere (Figs. 4 and 5). The question of the energy sources for the upper atmosphere is the most important single problem for upper-atmosphere physics at this time. The continuing investigation of this matter and, in particular, of the roles played by particle and radiation sources, respectively, will be one of the main areas of ex-

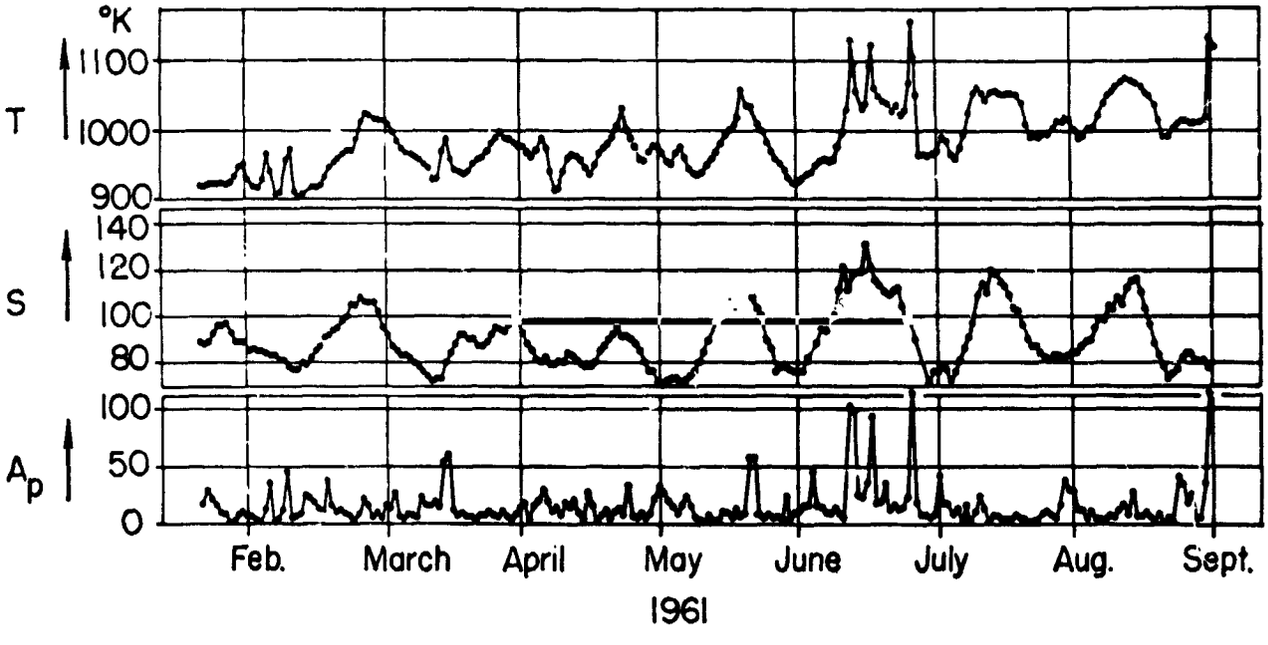


FIGURE 4.—Temperature variation correlated with solar radio emission and geomagnetic indices for the interval from February through September 1961. (Upper curve) Exospheric densities and temperatures, derived from Explorer IX drag measurements by L. G. Jacchia and J. Slowey (1962); (middle curve) flux of the solar radiation at 20-centimeter wavelengths, an indicator of solar activity measured at the Heinrick Hertz Institut in Berlin; (lower curve) geomagnetic activity index (A_p). [The Explorer IX data are adapted from L. G. Jacchia and J. Slowey. *Smithsonian Astronomical Observatory Spec. Rept. No.* 84 (1962); reproduced courtesy of W. Priester.]

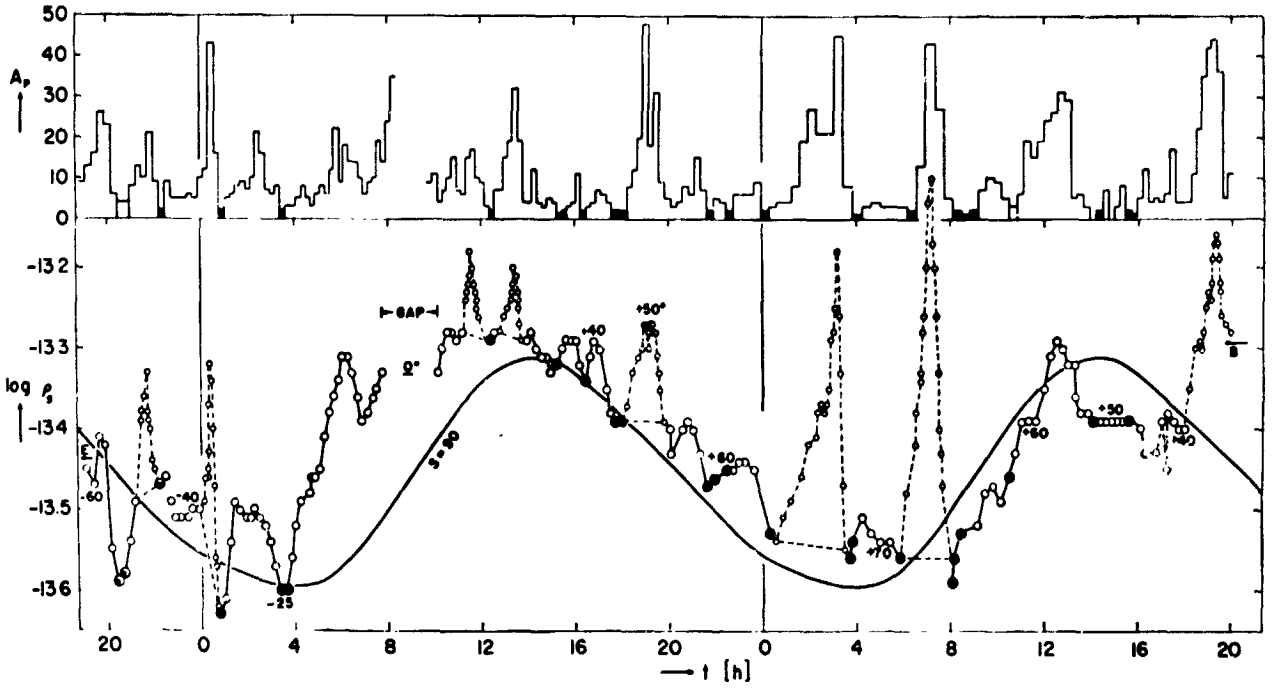


FIGURE 5.—Density variations at 250 kilometers above sea level as a function of local time t, determined by L. G. Jacchia and J. Slowey (1963) of the satellite Injun III from 15 December 1962 through 29 June 1963. During this time the geographic latitude of the perigee cover the range from $+70°$ to $-60°$, as indicated by the numbers on the density curve. (Solid curve) The Harris-Priester model for the proper level of solar activity; (histogram at top) daily geomagnetic indices A_p; (open circles) densities during magnetic storms; (solid circles) densities during magnetically quiet days ($A_p<2$). The response of the atmosphere to solar activity, as indicated by violent solar storms, is much greater within the auroral zone than outside it. [Adapted from L. G. Jacchia and J. Slowey, *J. Geophys. Res.* 69, 905 (1964)].

perimental and theoretical effort in the next several years.

MAGNETOSPHERE

The evidence cited suggests that corpuscular streams from the sun transfer appreciable amounts of energy to the atmosphere. How does the transfer of energy in the atmosphere occur?

The general answer seems to be connected with the properties of the outermost layer of the atmosphere. The density of the upper air merges into the density of the interplanetary gas at an altitude of about 100,000 kilometers, marking the boundary of the atmosphere. Early in 1958, however, J. A. Van Allen of the State University of Iowa discovered, by analyzing Geiger-counter data from Explorer I, that there was an additional layer of energetic charged particles in the upper atmosphere. These charged particles are trapped in the atmosphere by the earth's magnetic field, and the atmospheric layer which they constitute is therefore called the magnetosphere.

During the last few years three important developments have substantially changed our earlier impressions about the character of the magnetically trapped particles and their geophysical effects.

First, B. O'Brien, also of the State University of Iowa, using measurements from the Injun I satellite, discovered that the flux of charged particles coming down from the trapped-particle region was so large that, if this flux consisted of previously trapped particles which had just been dislodged by solar disturbances, it would drain the whole magnetosphere in about an hour. He also found that when a solar disturbance occurred, both the flux of untrapped descending particles and the number of trapped particles increased. Thus he concluded that the leakage of trapped particles from the Van Allen belts cannot be the principal source of the electrons which pass down through the atmosphere. He decided that, while a few charged particles are trapped during or after a solar disturbance, most pass directly into the atmosphere without spending an appreciable amount of time in the trapped-particle region.

Apparently, the charged particles which are observed in auroral displays and other atmospheric phenomena are those which come directly down the lines of force into the atmosphere.

Second, a large population of low-energy protons, having a range from 100,000 to several million electron volts, was discovered by A. H. Davis and J. M. Williamson of the Goddard Space Flight Center. The concentration of these protons reaches a maximum at a distance of 3.5 earth radii. At that distance their density is about one per cubic centimeter. This value for density of the trapped protons has interesting implications. As a result of the magnetic field gradient and curvature effects, the trapped protons drift westward in the magnetic field, with an associated electric current that produces magnetic effects. These effects have been calculated by S. Akasofu of the University of Alaska, S. Chapman of the universities of Colorado and Alaska, and (in unpublished work) R. A. Hoffman of the Goddard Space Flight Center. They find that the changes in the intensities of these trapped protons produce magnetic perturbations large enough to explain most magnetic storms observed on the earth, and also the very large perturbations of the geomagnetic field in space in the neighborhood of the proton belt. The relation between the trapped-proton drift current and the geomagnetic storms was suggested by S. F. Singer of the University of Maryland in 1956.

The third development was the discovery of a substantial flux of electrons with very high energies (in the neighborhood of 1 million volts) at a distance of 3 or 4 earth radii, presumably produced by beta decay of albedo neutrons resulting from cosmic ray interactions in the atmosphere. These electrons penetrate the Geiger counters with high efficiency, and when allowance is made for their presence, the estimate of the total flux of electrons is reduced from the earlier estimated value of 10^{10} cm^{-2} sec^{-1} srad^{-1} to the currently accepted value of 10^{8} cm^{-2} sec^{-1} srad^{-1}.

THE MAGNETOPAUSE

The connection between the magnetosphere and the transfer of corpuscular energy to the atmosphere is probably to be found in the properties of the atmosphere near the magnetopause, the boundary separating the interplanetary medium from the region around the earth, in which the geomagnetic field is dominant. The sharply defined surface of the magnetopause marks the termination of both the trapped-particle region and the geomagnetic field. Satellite measurements of the geomagnetic field by L. J. Cahill of the National Aeronautics and Space Administration and of the University of New Hampshire show that the magnetopause has a thickness on the order of 100 kilometers and occurs at a distance of 8 to 10 earth radii on the sunlit side of the earth.

The sharpness of the magnetospheric boundary is illustrated by Figs. 6 and 7. Figure 6 represents the magnetic-field measurements obtained by Cahill and P. G. Amazeen of the University of New Hampshire, using a three-component magnetometer flown on Explorer XII. At the magnetopause a sudden drop occurs and outside the magnetosphere the magnetic field is highly variable in magnitude and direction. Figure 7 shows the counting rates of charged particle detectors flown on Explorer XIV by Van Allen, L. A. Frank, and E. Macagno of the State University of Iowa.

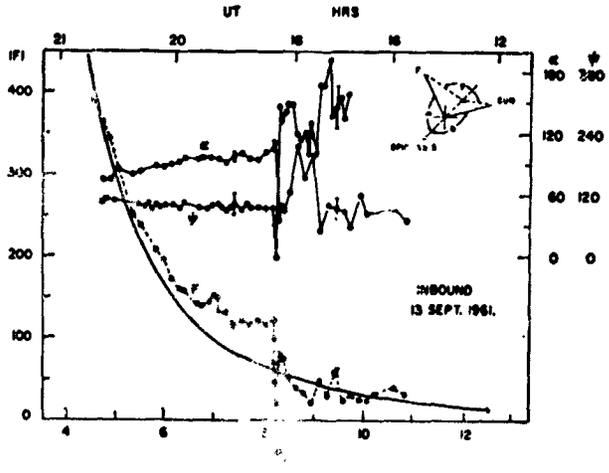


FIGURE 6.—Explorer XII measurements showing the abrupt termination of the geomagnetic field at the magnetopause. F is the magnitude of the magnetic field and α and β refer to its direction. Within the magnetosphere the field is closely described by the high-altitude extrapolation of the earth's approximately dipole field, shown as the solid curve. Outside the magnetopause, which occurred at 8.2 earth radii (R_E) during this flight, the field varies in magnitude and direction. [From L. J. Cahill and P. G. Amazeen, *J. Geophys. Res.* 68, 1841 (1963)].

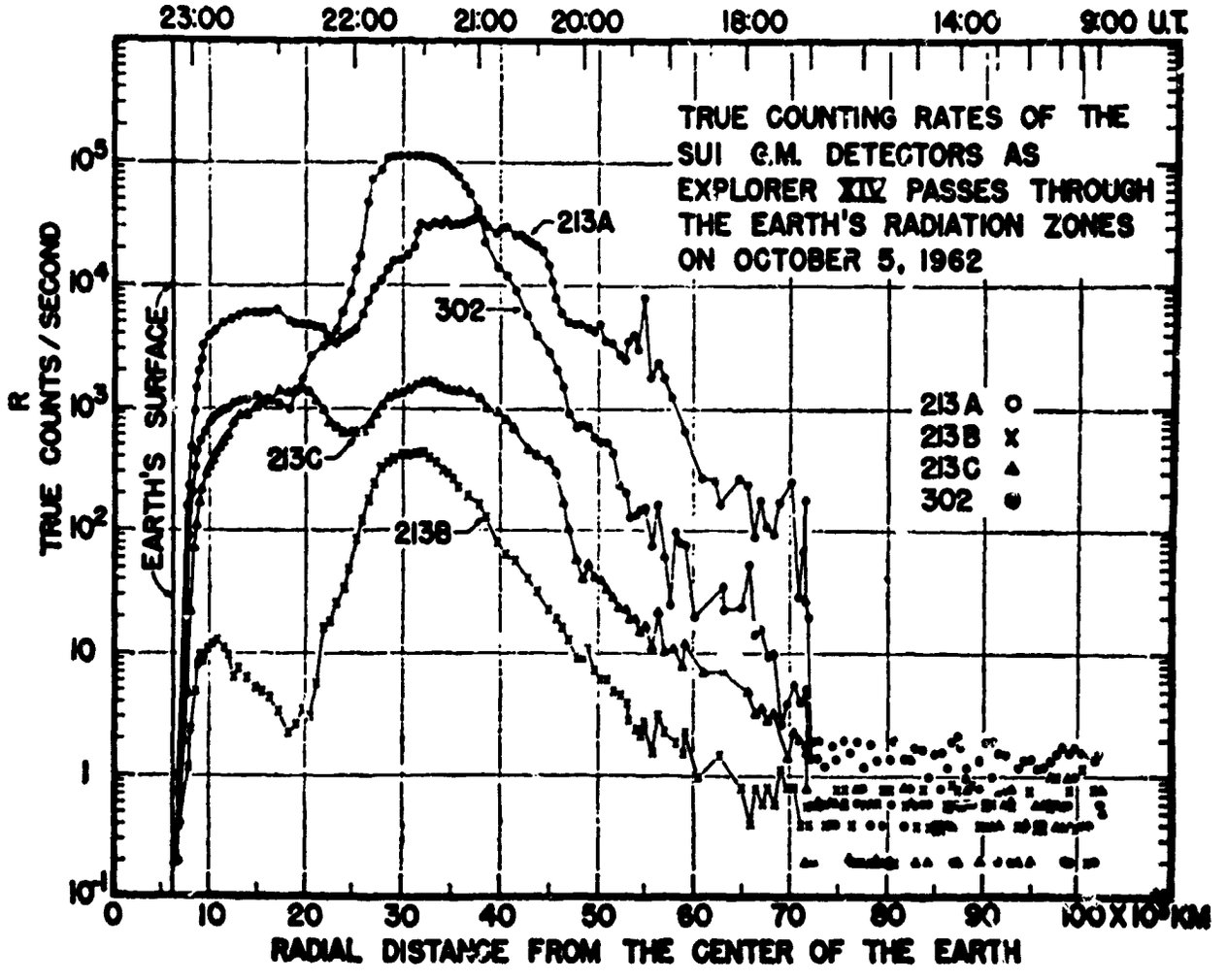


FIGURE 7.—Explorer XIV data showing the abrupt termination of the magnetically trapped particle belts at the magnetopause. The detector labeled 213A accepts the particles of lowest energy; it records principally 50-kev electrons. Its counting rate drops sharply at a radius of 72,000 kilometers, or 12.5 R_E, which corresponds to the location of the magnetopause during this flight. [From L. A. Frank, J. A. Van Allen, E. Macagno, *J. Geophys. Res.* 68, 3545 (1963).]

The detector which accepts the particles of lowest energy is labeled 213A, and its counting rate reflects principally the flux of 50-kev electrons. At the magnetopause the counting rate of this detector drops, by a factor of approximately 20, to a value which is approximately independent of altitude and is produced by the cosmic ray background in space.

Within the magnetopause there are no substantial fluxes of energetic particles other than those of the magnetically trapped particles illustrated in Fig. 8. Detectors flown on Mariner II indicate that the sun is the source of a particle stream which flows through interplanetary space continually, although with variable velocity and intensity. The Mariner II detectors, and also the plasma probe flown on Explorer X by B. Rossi, H. S. Bridge, A. J. Lazarus, A. Bonetti, and F. Scherb of the Massachusetts Institute of Technology, have shown that these particles move radially outward from the sun at velocities varying from 300 to 600 kilometers per second and at an average flux of 10^8 sec^{-1} outside the magnetopause. The interplanetary stream of solar particles, called the solar wind, cannot penetrate the magnetic field of the earth but divides and flows around it as the waters of a stream divide around a boulder (Fig. 8). The closest distance of approach of the solar wind to the earth is about 10 earth radii.

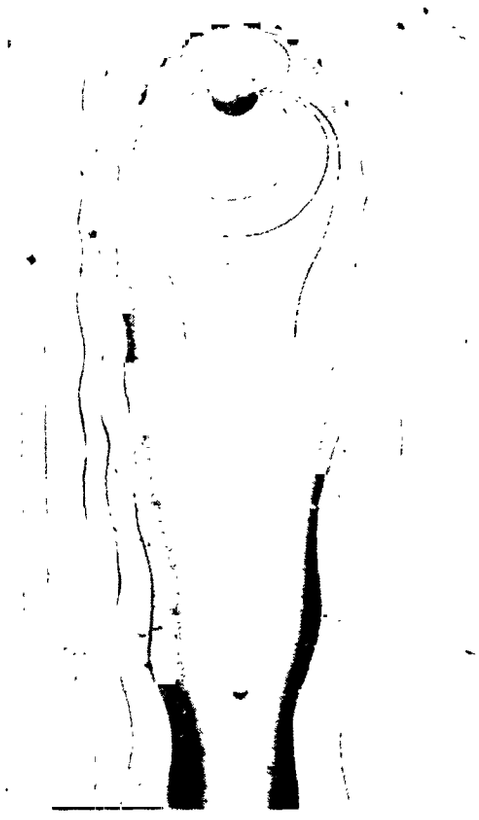


FIGURE 8.—The geomagnetic cavity in the solar wind.

The shadow or cavity carved by the magnetic field of the earth in the solar wind should, in principle, extend back indefinitely far into the solar system behind the earth. However, because the particles of the stream have appreciable transverse velocities associated with their thermal motions, we expect these particles to diffuse together eventually in the shadow of the earth. The ratio of mean transverse to radial velocities is about ¼, hence we expect the geomagnetic cavity to be filled in at a distance of 4 times the diameter of the cavity, or, roughly, the distance of the moon from the earth, as suggested in Fig. 8.

The amount of energy transferred from the solar wind to the atmosphere within the cavity is difficult to estimate. The variable magnetic fields in the solar plasma are believed to glue the particles together and give their motion the properties of fluid flow, in spite of the low density; thus turbulence would be expected at the region of impact of the solar wind on the magnetopause. The bufleting of the magnetospheric boundary associated with this turbulent impact may generate disturbances in the field just within the magnetopause. These disturbances propagate hydromagnetically down or across the field lines into the atmosphere, where they transfer energy which may appear as atmospheric heating, ionization, auroral disturbances, and magnetic storms —that is, the whole complex of disturbances produced in the atmosphere at high geomagnetic latitudes at times of solar activity.

The interaction of the solar plasma with the magnetosphere also leads to the expectation that a "shock wave" will be formed some distance beyond the actual magnetopause. This expectation arises from the fact that the flow of the plasma is supersonic and must become subsonic in the vicinity of the earth. Such a transition would set up a shock wave which would stand off some distance from the magnetopause and would have a thickness determined by the ability of the magnetic field to change the bulk motion of the plasma particles. Evidence of the perturbed magnetic fields corresponding to the shock wave and the intervening transition region has been detected by N. F. Ness of the Goddard Space Center with magnetometers flown on the Interplanetary Monitoring Platform (IMP) spacecraft launched on 27 November 1963.

ATMOSPHERE OF VENUS

Venus is the third brightest object in the sky, the planet nearest the earth, and the planet most closely resembling the earth in size and mass. It has been studied with the telescope since Galileo's time, and yet it remains an enigma, because its surface is permanently shrouded by a layer of clouds. In this state of ignorance, hope has flourished that Venus offers a hospital environment for the development of advanced forms of life.

Some information regarding the surface of the planet has been obtained in recent years through study of the microwave radiation emitted in the longwave region of its thermal spectrum. This radiation, with wavelengths in the region of 1 to 10 centimeters, penetrates the clouds without significant attenuation; its intensity is proportional to the temperature of the emitting surface.

The first attempts to measure the microwave radiation from Venus were made in 1956 with the Naval Research Laboratory radio telescope. The

temperature inferred from the measured radiation intensity was approximately 600°K (700°F)—certainly too high to permit any terrestrial forms of life. Repeated measurements have confirmed the Naval Research Laboratory results and have a revision of our ideas regarding the surface and lower atmosphere of Venus.

It is difficult to understand why the temperature of Venus should be so much higher than that of the earth. The path to an explanation would seem to lie in the assumption of an extremely dense atmosphere which absorbs strongly in the infrared region of the spectrum but is transparent with respect to visible radiation. The part of the incident sunlight which is not reflected back by the clouds will therefore penetrate through the atmosphere and heat the surface of the planet. But when the surface layers reradiate this energy at infrared wavelengths, the radiation is absorbed by the atmosphere and returned in large measure to the surface, thus giving an additional flux of energy into the ground and raising its temperature. This atmospheric phenomenon is the greenhouse effect described earlier.

The clouds of Venus reflect three-quarters of the incident sunlight; the remaining quarter of the incident radiation would bring the surface of the planet to a temperature of 235°K if there were no atmospheric greenhouse effect. If the greenhouse effect is to raise the ground temperature to 600°K, the optical thickness of the atmosphere must be 50 mean free paths throughout the far-infrared region. In an atmosphere with this degree of opacity, only one photon in 10^{21} escapes directly, without absorption. This condition is so severe that alternative suggestions have been made, among them the hypothesis that the apparently high radio temperature of Venus is the result of microwave emission from its ionosphere rather than from the surface.

The Mariner II Venus flyby launched on 27 August 1962 included experiments designed to test this hypothesis. This spacecraft passed Venus at a distance of 33,440 kilometers (20,900 miles) on 14 December 1962, and made crucial measurements of the temperature across the disc. The spacecraft was equipped with two sets of radiation detectors, one in the infrared and one in the microwave region. Measurements of the radiation emitted by the planet in the microwave region included measurements of radiation of 19-millimeter wavelength, which passes through the atmosphere with little attenuation and hence provides a measure of the temperature at the ground, provided there is no additional emission from the ionosphere.

A modest degree of atmospheric attenuation is, however, to be expected, and in the scan of Mariner II across the disc of the planet this slight degree of attenuation should show up as a lower intensity of measured radiation at the edge or limb of Venus, where the thickness of the intervening atmosphere is greater. However, if the high microwave intensities and apparent temperatures result from emission by electrons in the atmosphere of Venus, then the readings at the limb should indicate an enhancement or brightening because of the greater thickness of the ionosphere in the line of sight.

The Mariner II results showed a conclusive darkening of the limb of Venus at 19 millimeter wavelength, thus eliminating the possibility of ionospheric emission and confirming the supposition that the measured radio temperature of 600°K is associated with the surface of the planet.

EXPLORATION OF THE MOON

The moon is a uniquely important body in the study of the history of the solar system because its surface has preserved the record of its history remarkably well. The moon has a negligible atmosphere and no oceans. It is, therefore, unchanged by the processes of erosion which erase the history of the earth's surface in a relatively short time—between 10 and 30 million years.

This is evidenced, in part, by the tens of thousands of craters on the lunar surface, produced by the impact of meteorites which, presumably, have been colliding with the moon since its formation. This is perhaps the only physical record which we have of events in the development of the solar system going back to that early time.

Because of the antiquity of the moon's surface, there is another remarkable record preserved—a layer of cosmic dust which is believed to have rained on it from the solar system since its formation. This dust may be as much as 30 centimeters or more in depth and may contain organic molecules and the precursors of life on earth.

The most important measurements of lunar properties from spacecraft have resulted from flights of the Russian Lunik II and Lunik III. From the Lunik II magnetometer data Soviet scientists concluded that an upper limit of approxmately 100 gammas could be placed on the moon's magnetic field. In future flights, refinement of this limiting value for the moon's magnetic field may provide information on the presence or absence of a liquid core within that body (on the earth the magnetic field is supposed to be associated with currents in the liquid core of the planet). This in turn could have a bearing on our understanding of the formation of the moon and of similar bodies in the solar system.

Lunik III has provided us with the first pictures of the remote side of the moon. In spite of some blurring, the photographs are of great interest, for it is possible to distinguish a large number of features resembling the craters and maria on the front face. Perhaps the most interesting feature is the Soviet Mountain Range, a chain extending across the center of the moon's hidden face. It resembles the great ranges on the earth and is unlike the mountain formations characteristic of the moon's front face, which seem to be circular crater walls and deposits of debris formed by the impact of large meteorites on the lunar surface.

According to our present ideas, terrestrial mountains result from the combined effects of erosion and wrinkling of the earth's crust, but such mountain-building forces have been much less effective on the moon. In the photograph, the markings designated the Soviet Mountain Range could have resulted from the running together of several obscured but independent markings. However, if these features continue to appear as a single range in later, more detailed pictures, we may have to revise our theories of lunar structure.

SOLAR PHYSICS

One of the most interesting questions in solar physics concerns the manner in which energy is transported above the surface of the sun to heat the chromosphere and corona.

We know that near the center of the sun, where the temperature is approximately 15 million degrees Kelvin, hydrogen is converted into helium by a variety of nuclear reactions. We also know that the sun is a self-adjusting system which expands or contracts in order to maintain a precise balance between the energy generation at the center and the energy emission from the surface.

All regular mechanisms of energy transport can carry heat only from a region of high temperature to a region of low temperature. Therefore, in order for the heat generated by nuclear reactions to be carried away from the center of the sun, the temperature must fall continuously from the center to the edge. This is in fact the case; the temperature falls from 15 million degrees at the center of the sun to 5800 degrees at the visible edge.

However, above the visible edge, which is called the photosphere, there lies a relatively tenuous region of gas which constitutes the atmosphere of the sun. This region is divided into the chromosphere and, above that, the corona.

The puzzling fact is that the temperature of the sun *rises* again above the photosphere, reaching a value of 1.5 to 2 million degrees in the corona. One of the burning questions of solar physics is what constitutes the source of the energy which produces the very high temperatures in the solar corona. Also, what is the mechanism of energy transport by which energy can be carried without appreciable losses through the dense gases of the photosphere and yet undergo great losses in the tenuous regions of the corona?

A current belief is that a wave motion—either a sound wave, a hydromagnetic wave, or a gravity wave—carries energy upward from the photosphere and deposits it in the corona. When a sound wave propagates into a region of decreasing density, its amplitude increases and it steepens into a shock wave. This is a mechanism in which considerable dissipation of energy takes place. It appears that hydromagnetic waves are rapidly damped out below the photosphere, but if they could be generated in the region of the chromosphere, then they would tend not to be dissipated until the waves had reached the corona. Magnetic disturbances above the photosphere may be particularly effective in generating these waves. Gravity waves consist of a kind of rolling motion similar to the waves on the surface of the ocean. These may, like sound waves, be generated by the motions of convecting material in the transition layer; they would have a vertical component of

propagation and would be dissipated in the corona.

It may be that all three of these mechanisms are operating in the heating of the chromosphere and corona. If this is the case, there may be a steady heating of the corona upon which is superimposed a localized heating associated with magnetic activity. Thus, the heating of the corona is expected to depend upon the magnetic structure in the outer layers of the sun. This dependence is observed in many phenomena. In particular, in sunspot regions where the strengths of the magnetic field are higher than is normal on the sun's surface, both the chromosphere and the corona have a higher than normal temperature.

The behavior of the chromosphere and the corona is most easily observed by studying the ultraviolet emission from the sun, since in the ultraviolet region the amount of light emitted from the photosphere greatly decreases, whereas the higher temperatures in the chromosphere and corona are responsible for the presence of large numbers of emission lines. The most important emission lines are attributable to hydrogen and helium. In order to understand solar-surface physics in more detail, it is essential to obtain observations of the time variations of these emission lines as indicators of the time variations of behavior in the chromosphere and corona.

The first experiments in this direction were very successfully accomplished by the flight of the first Orbiting Solar Observatory, which was launched on 7 March 1962. It gave data over several months, continuously monitoring a number of different wavelength regions for emission from the sun.

Particularly interesting are the data for the interval 11 through 22 March 1962. At the beginning of this period the sun was in an exceptionally quiet condition, but as the period progressed the sun became more and more active, until on 22 March there was a flare of intensity 3. Experiments revealed that the Lyman alpha line of helium II at 304 angstroms increased in intensity by some 33 percent during the interval, and during the flare itself the intensity of the line increased by an additional 14 percent. The lines of iron XV at 284 angstroms and iron XVI at 335 angstroms also increased in intensity by a factor of 4. At longer wavelengths the Lyman alpha line of hydrogen was observed to increase in intensity by 6.8 percent during the flare.

Very interesting results were also obtained in the x-ray region, 1 to 10 angstroms. During the quiet period a flux was observed which was 360 times the theoretical background radiation which would be obtained from a corona at a temperature of 1.8×10^6 degrees Kelvin. This indicates that nonthermal processes are present and important in the corona under even the quietest solar conditions.

A continuing series of Orbiting Solar Observatories is planned in which these interesting phenomena can be monitored continuously during future years.

X-RAYS AND GAMMA RAYS

The space research program is not confined to the discovery of new facts about the solar system. It also gives the astrophysicist an important opportunity to extend his knowledge of more distant parts of space through observations at wavelengths for which photons do not penetrate through the atmosphere. The principal regions involved are the x-ray and gamma ray region, the ultraviolet, the infrared, and long-wavelength radio waves. The early rocket and satellite measurements of x-rays and gamma rays have been particularly interesting to physicists because they suggest several possible new types of phenomena in space.

X-rays and gamma rays can be produced by a variety of high-energy processes. These processes include collisions between high-energy nucleons, which can create neutral pions, which in turn decay to give gamma rays exceeding 50 Mev in energy. Fast electrons can produce X-rays of bremsstrahlung when they pass close to a nucleus. Fast electrons can also collide with photons of visible starlight and increase the energy of the photons up to the x-ray and gamma ray regions. If radioactive nuclei are produced and dispersed in space between the stars, then some of them should emit characteristic gamma ray energies which might be detected. If positrons are produced in dense regions of matter, such as stellar surfaces, then, upon being slowed down and annihilated, they will emit the characteristic gamma rays of 0.51-Mev energy. If neutrons are produced near stellar surfaces and are slowed down

and captured by the overwhelmingly abundant hydrogen that is present, then these will provide characteristic capture gamma rays with an energy of 2.31 Mev. Finally, if objects should exist in space with surface temperatures of some millions of degrees Kelvin, then photons in the x-ray region would be emitted by thermal processes from their surfaces.

Preliminary measurements now exist of the fluxes of x-rays and gamma rays in a number of different energy intervals. A general background of x-rays of energy of a few thousand electron volts was observed in a rocket flight by R. Giacconi, H. Gursky, and F. R. Paolini, of American Science and Engineering, Inc., and by B. B. Rossi of the Massachusetts Institute of Technology. A general background radiation of gamma rays in the region near 1-Mev energy was measured in the Ranger III flight by J. R. Arnold of the University of California (La Jolla), A. E. Metzger, of the California Institute of Technology, and E. C. Anderson and M. A. Van Dilla, of Los Alamos. A small but still significant flux of gamma rays with energies exceeding 50 Mev was observed with the Explorer XI satellite by W. L. Kraushaar and G. W. Clark of the Massachusetts Institute of Technology.

A number of attempts have been made to explain the presence of these background x-rays and gamma rays. Most mechanisms thus far examined appear quantitatively inadequate to explain the observed fluxes. One promising explanation was proposed by J. E. Felten and P. Morrison of Cornell University, who suggested the importance of the inverse Compton effect in which the high-energy electrons present in the cosmic rays collide with photons with energies of the order of 1 electron volt which are emitted from stars. Following such a collision, the energies of the photons can easily be raised to the observed x-ray or gamma ray energies, depending upon the energies of the electrons with which they collide. Calculations by Felten and Morrison were based on this effect (Fig. 9). A flux will be emitted by the outer halo region of our galaxy if the observed flux of high-energy electrons at the earth exists throughout this large outer region of the galaxy. Electrons in the halo fail to account for the observed x-ray and gamma ray fluxes by some 2½ orders of magnitude. However, if the flux of high-energy electrons is the same throughout space as near the earth, then a background radiation some 30,000 times that which would be produced within the galactic halo would be observed. Evidently, such high fluxes of electrons cannot exist throughout space. One percent of such a flux of electrons can be expected to give a background of x-rays and gamma rays which fits the observations very nicely.

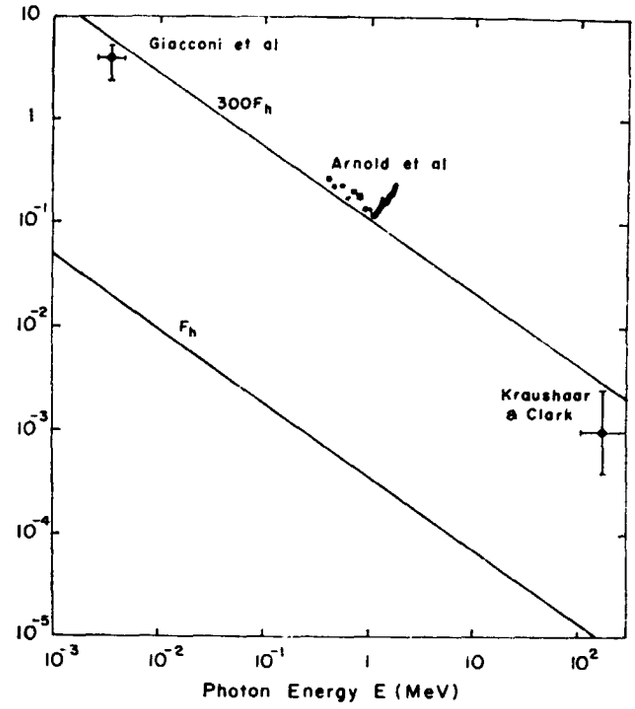


FIGURE 9.—Data on hard photon fluxes in space, compared with theoretical recoil spectra. The data points are as plotted by Felton and Morrison. F_h is the expected contribution from scattering in our galactic halo. The upper curve shows a flux 300 times greater, but still less by two orders of magnitude than the flux that would be obtained if the electrons in our galactic halo extended throughout intergalactic space. [From J. E. Felten and P. Morrison, *Phys. Rev. Letters* 10, 455 (1963)].

Perhaps the most interesting questions concerning the celestial x-rays have been raised through the discovery of discrete sources by Rossi and his colleagues and by H. Friedman. S. Bowyer, T. A. Chubb, and E. T. Byram of the Naval Research Laboratory. Both groups have observed a strong x-ray source in Scorpius which is not coincident with any conspicuous object. Friedman has suggested that this object is a neutron star having a surface temperature of several million degrees, and that the x-rays are due to

thermal emission from the surface layers. Rossi and his colleagues have determined from atmospheric absorption measurements that if the Scorpius source has a thermal spectrum its temperature is approximately 8×10^6 degrees Kelvin. Friedman and his colleagues have also observed x-rays from the direction of the Crab Nebula, the remnant of the supernova explosion of A.D. 1054 (Fig. 10).

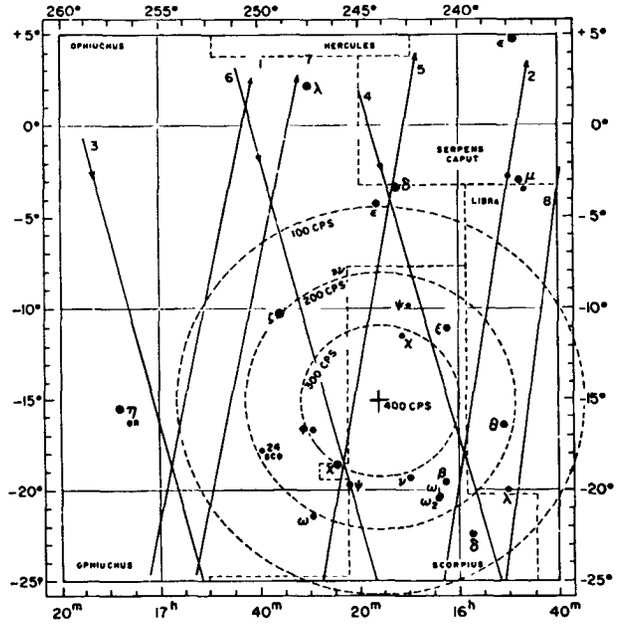


FIGURE 10.—Tracks of eight scans across the Scorpius region. The numbers along the tracks are counts per 0.09-second interval. The dashed circles are best fits to equal-intensity contours and indicate a central intensity peak of 400 count/sec at $\alpha = 16$ hours 15 minutes, $\delta = -15°$. [From S. Bowyer, E. T. Bryam, T. A. Chubb, H Friedman, *Nature* 201, 1307 (1964)].

Neutron stars are hypothetical objects which form one class of degenerate stars, the other class being the degenerate white dwarf stars, which are observed. A typical density for matter in a white dwarf star is 10^6 g/cm^3, and the electrons form a degenerate gas which exerts sufficient pressure to maintain the stars against further contraction. If mass were to be added to such a star, the central region would have to become denser in order to supply the additional pressure required to support the additional mass. There is a relativistic upper limit to the mass of white dwarf stars, but before this limit is reached, the energies of the degenerate electrons have become so high that the nuclei are forced to undergo multiple electron capture reactions, and the nuclei dissolve mainly into neutrons, with only enough protons and electrons left to prevent the neutrons from undergoing their usual mode of decay into electrons and protons.

At 10^{15} g/cm^3 or more, densities comparable to those in the atomic nucleus, this neutron-rich nuclear matter itself becomes degenerate, and it is expected that stable stars could be constructed of it. Such stars may be formed in the central regions of more massive stars when these stars undergo supernova explosions and blow off most of their mass. Recent work by D. Morton of Princeton, E. E. Salpeter of Cornell, and H. Y. Chiu, S. Tsuruta, and A. G. W. Cameron of the Goddard Institute for Space Studies indicates that the surface temperature of a neutron star is likely to lie below its central temperature by between one and two orders of magnitude. Thus, if such stars are formed with central temperatures over 10^9 degrees Kelvin, such as would probably be produced in a supernova explosion, then their surface temperatures are likely to be many millions of degrees for several thousand years.

BIBLIOGRAPHY

Geodesy

W. M. KAULA, *J. Geophys. Res.* **68**, 5183 (1963).
G. J. F. MACDONALD, *Science* **143**, 921 (1964).

Meteorology

K. TELEGADS and J. LONDON, *A Physical Model for the Northern Hemisphere for Winter and Summer* (New York Univ. College of Engineering, New York, 1954).
J. LONDON, *A Study of the Atmospheric Heat Balance* (New York Univ. College of Engineering, New York, 1957).
S. FRITZ and H. WEXLER, "Planet Earth as seen from Space," in *Planets and Satellites*, G. P. Kuiper and B. M. Middlehurst Eds. (Univ. of Chicago Press, Chicago, 1961), p. 1.
W. R. BANDEEN, R. A. HANEL, J. LICHT, R. A. Stampfl, W. G. Stroud, *J. Geophys. Res.* **65**, 3169 (1961).
W. NORDBERG, W. R. BANDEEN, B. J. CONRATH V. KUNDE, I. PERSANO, *J. Atmospheric Sci.* **19**, 20 (1962).
D. Q. WARK, G. YAMAMOTO, J. H. LIENESCH, *ibid.*, p. 369.
C. PRABHAKARA and S. I. RASOOL, *International Symposium on Rocket and Satellite Meteorology*, H. Wexler and J. E. Caskey, Jr., Eds. (North-Holland, Amsterdam, 1963), p. 234.
J. S. WINSTON and P. K. RAO, *Monthly Weather Rev.* **91**, 641 (1963).
A. ARKING, *Science* **143**, 569 (1964).
S. I. RASOOL, *ibid.*, p. 567.

Upper Atmosphere

M. NICOLET, *J. Geophys. Res.* **66**, 2263 (1961).
R. E. BOURDEAU, E. C. WHIPPLE, JR., J. L. DONLEY, S. J. BAUER, *ibid*, **67**, 467 (1962).
I. HARRIS and W. PRIESTER, *J. Atmospheric Sci.* **19**, 286 (1962).
———, *J. Geophys. Res.* **67**, 4585 (1962).
———, *ibid.* **68**, 5891 (1963).
L. G. JACCHIA, in *Space Research* (vol. 3 of *International Space Science Symposium*), W. Priester Ed. (North-Holland, Amsterdam, 1963), p. 3.
L. G. JACCHIA and J. SLOWEY, *J. Geophys. Res.* **69**, 905 (1964).

Magnetosphere

S. F. SINGER, *Trans. Am. Geophys. Union* **38**, 175 (1957).
J. A. VAN ALLEN, G. H. LUDWIG, E. C. RAY, C. E. MCILWAIN, *Jet Propulsion* **28**, 588 (1958).
B. J. O'BRIEN, *J. Geophys. Res.* **67**, 1227 (1962).
———, *ibid.*, p. 3687.
———, *Space Sci. Rev.* **1**, 415 (1962).
S. AKASOFU and S. CHAPMAN, *J. Geophys. Res.* **66**, 1321 (1963).

Magnetopause

L. A. FRANK, J. A. VAN ALLEN, E. MACAGNO, *ibid.* **68**, 3543 (1963).
L. J. CAHILL and P. G. AMAZEEN, *ibid.*, p. 1835.
A. BONETTI, H. S. BRIDGE, A. J. LAZARUS, B. ROSSI, F. SCHERB, *ibid.*, p. 4017.
H. R. ANDERSON, *Science* **139**, 42 (1963).
E. PARKER, *Interplanetary Dynamical Processes* (Interscience, New York, 1963).

Atmosphere of Venus

C. H. MAYER, T. P. MCCULLOUGH, R. M. SLOANAKER, *Astrophys. J.* **127**, 1 (1958).
D. E. JONES, *Planetary Space Sci.* **5**, 166 (1961).
S. I. RASOOL, *Am. Inst. Aeron. Astronaut, J.* **1**, 6 (1963).
S. C. CHASE, L. D. KAPLAN, G. NEUGEBAUER, *J. Geophys. Res.* **68**, 6157 (1963).
R. T. BARATH, A. H. BARRETT, J. COPELAND, D. E. JONES, E. A. LILLY, *Astron. J.* **69**, 49 (1964).

Exploration of the Moon

N. P. BARABASHOR, A. A. MIKHAILOV, Y. N. LIPSKIY, *Atlas of the Other Side of the Moon* (Pergamon, London, 1961).
S. S. DOLGINOV, E. G. EROSHENKO, L. I. ZHUZGOV, N. V. PUSHKOV, "A study of the magnetic field of the Moon," in *The Moon*, Z. Kopal and Z. V. Mikhailov, Eds. (Academic Press, New York, 1962).

Solar Physics

D. E. OSTERBROCK, *Astrophys. J.* **134**, 347 (1961).
W. A. WHITTAKER, *ibid.* **137**, 914 (1963).
J. C. LINDSAY, *Trans. Am. Geophys. Union* **44**, 722 (1963).
D. W. MOORE and E. A. SPIEGAL, *Astrophys. J.* **138**, 48 (1964).

X-rays and Gamma Rays

W. L. KRAUSHAAR and G. W. CLARK, *Phys. Rev. Letters* **8**, 106 (1962).
R. GIACCONI, H. GURSKY, F. R. PAOLINI, B. R. ROSSI, *ibid.* **9**, 439 (1962).
J. ARNOLD, A. E. METZGER, E. C. ANDERSON, M. A. VAN DILLA, *J. Geophys. Res.* **67**, 4878 (1962).
J. E. FELTEN and P. MORRISON, *Phys. Rev. Letters* **10**, 453 (1963).
S. BOWYER, E. T. BYRAM, T. A. CHUBB, H. FRIEDMAN, *Nature* **201**, 1307 (1964).
D. MORTON, *ibid.*, p. 1308.
H. Y. CHIU and E. E. SALPETER, *Phys. Rev. Letters* **12**, 413 (1964).
———, *Ann. Phys. N.Y.* **26**, 364 (1964).

g_J OF THE $(5s5p)\ ^3P_1$ LEVEL OF Cd AND THE $(6s6p)\ ^3P_1$ LEVEL OF Hg BY HIGH-FIELD DOUBLE RESONANCE*

R. KOHLER† AND P. THADDEUS

An optical double resonance experiment at high magnetic fields has determined g_J for the $(5s5p)\ ^3P_1$ level of the even cadmium isotopes to be 1.499846(13), and g_J for the $(6s6p)\ ^3P_1$ level of the even mercury isotopes to be 1.486094(8). Resonance was observed in all cases at a frequency of exactly 24 Gc/sec, and magnetic fields near 11 430 G. At these high fields the Zeeman energy is of the order of 10^{-3} of the fine structure separations of the triplet terms of the $(nsnp)$ configuration, and the $\Delta m = \pm 1$ transitions are split by 9.51(7) G for Cd, and by 2.99(7) G for Hg. This splitting represents several linewidths in the case of Hg, and many linewidths in that of Cd. The average field of the two $\Delta m = \pm 1$ transitions, however, determines g_J to high precision independent of second-order corrections.

I. INTRODUCTION

The double resonance[1,2] and level crossing techniques[3,4] have been used in extensive studies of hyperfine structure (hfs) in the lowest lying $(nsnp)\ ^3P_1$ level of mercury,[5–11] cadmium,[12–17] and zinc.[18–20] Sufficient sensitivity has been obtained to allow the determination of the hfs, and consequently the nuclear moments, of radioactive isotopes with half-lives as short as a few hours.[15,16]

While the level crossing method has the great advantage in practice of doing away with all radio-frequency and microwave apparatus, aside from nuclear magnetic resonance (NMR) magnetometers, it allows a determination of only the ratio of the hyperfine interaction constants to g_J, and an independent measurement of g_J must therefore be made to determine these constants to high precision.[16] While g_J can of course be measured by double resonance in the odd isotopes at various fields, or by observing both double resonance and level crossings,[9,13] it is most directly found from double resonance in the even isotopes of the element.

In principle, the higher the frequency and the greater the magnetic field, the more accurate the determination of g_J. In a series of experiments designed to measure g_J of the lowest lying $^3P_1$ level of zinc, cadmium, and mercury to high precision (it is $\frac{3}{2}$ on the basis of LS coupling), we therefore chose to work at a frequency of 24 Gc/sec, and a field near 11 400 G—approaching the limit attainable with the available 12-in. electromagnet when operating with a gap of 2 in.

The $\Delta m = \pm 1$ transitions for these states coincide at low magnetic fields, but at high fields interaction with the neighboring fine structure terms of the configuration (and to a negligible degree with levels of other configurations) splits the resonances into two well-resolved components, as shown in Fig. 1. Precise measurement of the splitting of the two resonances therefore serves also as a check on the theory of the second-order fine structure interactions.

II. THEORY

Signal Strengths

It is well known that double resonance can be observed between hyperfine or Zeeman states excited by resonance radiation in a variety of ways. Population differences can be produced by resonance radiation that is polarized, nonisotropic, or confined to a narrow wavelength interval,

*Published in *The Physical Review*, 134 (5A): A 1204-A1209, June 1, 1964 work supported by the Joint Services (the U.S. Army, the U.S. Navy, Officers of Naval Research, and the U.S. Air Force Office of Scientific Research), and by the U.S. Office of Naval Research under, contract Nonr-266645).

†Columbia University, New York, and New York University, New York.

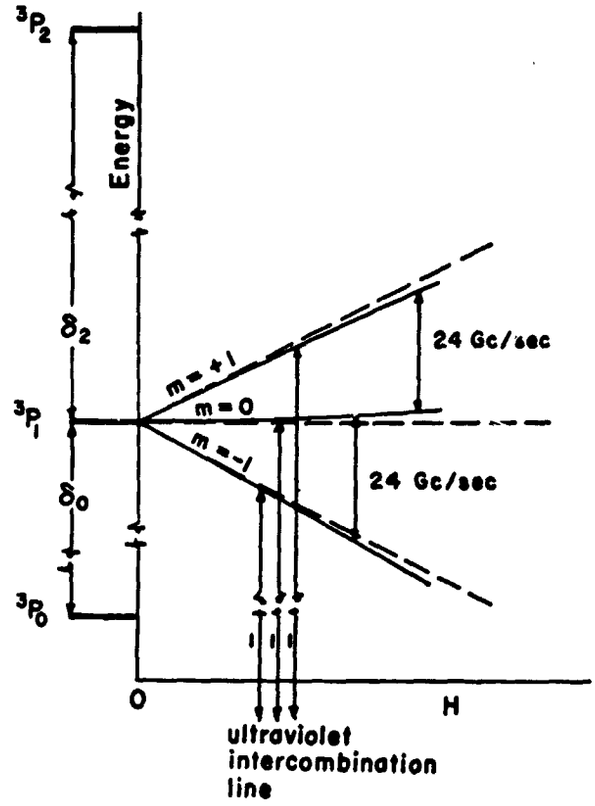


FIGURE 1.—Zeeman effect of the $^3P_1$ state of the even isotopes of zinc, cadmium and mercury, showing the effect of second-order interaction with the other fine structure levels of the configuration.

and all of these methods have been put to practical use.[1,2,6]

To dispense with ultraviolet polarizers, we relied on the latter two effects and used the exeprimental geometry shown in Fig. 2. Only lamps containing the natural metal were used. In the case of cadmium, the Zeeman shift of the $m=\pm 1$ states at 11 400 G is somewhat greater than the isotope shifts and the hyperfine splittings of the natural cadmium isotopes (and the optical Doppler width),[14] and only the $m=0$ state is therefore illuminated. It is then readily calculated in terms of the transition probabilities that at double resonance the resonance fluorescence at 90° is reduced by a factor of $\frac{1}{4}$.

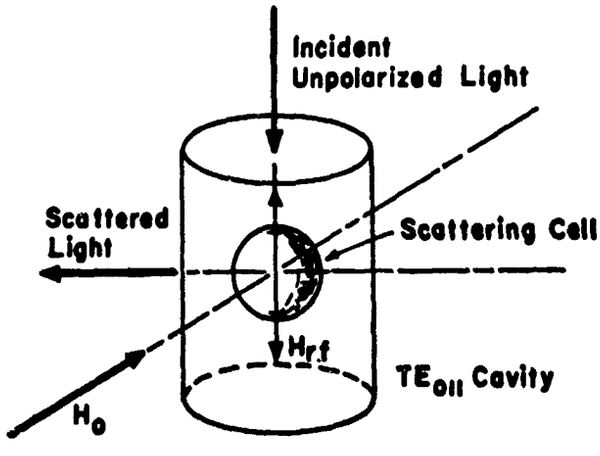


FIGURE 2.—Geometry of the experiment.

For mercury, on the other hand, the isotope shifts and hyperfine splittings are comparable to the Zeeman shift of the $\Delta m=\pm 1$ states,[21] and we must rely mainly on the nonisotropic nature of the resonance radiation. In the least favorable case, on the assumption that all three states are equally illuminated, it is calculated that the resonance fluorescence at 90° is reduced by a factor of 1/12 at double resonance.

Linewidths

Several effects determine the linewidth and shape of the double resonance effect, and have all been considered in detail elsewhere. Most fundamentally, the natural lifetime τ of the state gives a Lorentzian line shape with full width at half-maximum of $g_J\mu_0\Delta H_n=\Delta\nu_n=1/\pi\tau$.[1,6] The natural linewidths for zinc, cadmium, and mercury calculated from the lifetimes in Table I are listed in Table II.

If the rf magnetic field is strong enough to induce many transitions during the lifetime τ, the resonance will be power broadened.[1,6] In practice the rf power is always attenuated to the point where this effect is small, and it will therefore not be considered further.

Double resonance owes its high resolution to the fact that Doppler broadening is determined by the radio rather than the optical frequency. At microwave frequencies, however, the Doppler linewidth $\Delta\nu_d=(\nu/c)\times(8NkT\ \ln 2/M)^{1/2}$, where N is Avogadro's number and M is the molecular weight, may not be altogether negligible, as is shown in Table II.

Inhomogeneity of the magnetic field over the scattering sample contributes a linewidth ΔH_m that can often be reduced by careful alignment of the pole pieces. We found, however, that as our magnetic field approached 11 400 G, its homogeneity fell off rapidly due presumably to uneven

TABLE I.—*Atomic constants of Zn, Cd, and Hg.* $\lambda(^3P_1)$ *and* $\lambda(^1P_1)$, *the wavelengths of the resonance lines of the lowest lying* ($n$snp) *configuration, and* δ_2 *and* δ_0, *the fine structure separations of the triplet terms, are taken from C. E. Moore.*[a]

Atom	τ ($^3P_1$) (sec)	τ ($^1P_1$) (sec)	λ ($^3P_1$) (Å)	λ ($^1P_1$) (Å)	δ_2 (cm^{-1})	δ_0 (cm^{-1})
Zn	$2.00(20)\times10^{-8}$ [b]	$1.38(5)\times10^{-9}$ [c]	3077	2139	388.927	190.082
Cd	$2.39(4)\times10^{-6}$ [d]	$1.66(5)\times10^{-9}$ [e]	3261	2288	1170.866	542.113
Hg	1.18×10^{-7} [f]	0.3–1.6×10^{-9} [g]	2537	1849	4630.677	1767.220

a. C. E. Moore, *Atomic Energy Levels* (U.S. Government Printing Office, 1949–58).
b. See Ref. 19.
c. See Ref. 20.
d. See Ref. 23.
e. A. Lurio and R. Novick, Phys. Rev. 134, A608 (1964).
f. See the last paper of Ref. 22.
g. A. C. G. Mitchell and M. W. Zemansky, *Resonance Radiation and Excited Atoms* (Cambridge University Press, Cambridge, 1961), 2nd ed., p. 147

TABLE II.—*Theoretical and observed linewidths in milligauss.* ΔH_{calc} *is taken as the sum of the natural, Doppler, magnetic field, and wall collision widths.*

Atom	ΔH_n	ΔH_d	ΔH_m	ΔH_w	ΔH_{calc}	ΔH_{obs}
Zn	8	29	50	37	124	-------
Cd	68	18	50	26	162	175
Hg	1295	10	50	15	1370	1300

saturation. We estimate that $\Delta H_m = 50$ mG over a 5-mm sample.

Collisions with the walls of the scattering cell effectively shorten the lifetime of the excited state, and contribute a wall-quenching linewidth $\Delta\nu_w$ which depends mainly on the size of the cell. We have made estimates of $\Delta\nu_w$ in Table II on the assumption that $\Delta\nu_w = 1/\pi\tau_w = (3NkT/M)^{1/2}/\pi a$, where τ_w is the time for an atom at the temperature of the optimum vapor pressure to travel the radius a of the scattering cell.

Collisions between the scattering atoms will broaden the resonance at high vapor densities, while multiple scattering of the optical photons may actually decrease the linewidth (coherence narrowing).[22,23] Both of these effects may be avoided by working at low enough temperatures.

Quenching by atomic or molecular contaminants in the scattering cell, on the other hand, may contribute a linewidth $\Delta\nu_c$ which cannot be easily estimated in advance, and cannot be reduced once the cell is made. We shall see that possibly except for zinc we found no evidence of quenching of this kind.

Zeeman Effect

Interaction with the two neighboring triplet levels of the ($nsnp$) configuration contributes a term which is of the order of the (Zeeman energy)^2/(fine structure energy) $\approx 10^{-3}\times$(Zeeman energy) to the energy of the $m=\pm1$, 0 states. More precisely, from perturbation theory,[24]

$$W_{+1} = g_J\mu_0 H - \frac{\alpha^2(g_S-g_L)^2\mu_0^2H^2}{4\delta_2}, \qquad (1a)$$

$$W_0 = \alpha^2(g_S-g_L)^2\mu_0^2H^2\left(\frac{2}{3\delta_0}-\frac{1}{3\delta_2}\right), \qquad (1b)$$

$$W_{-1} = -g_J\mu_0 H - \frac{\alpha^2(g_S-g_L)^2\mu_0^2H^2}{4\delta_2}, \qquad (1c)$$

where μ_0 is the Bohr magneton, α is the mixing coefficient which indicates the departure from Russell-Saunders coupling, and δ_0 and δ_2 are the magnitude of the fine structure separations. We adopt the convention that g_J is positive. Interaction with the singlet term of the configuration, and with terms of other configurations, contributes less than one part in 10^6 to the energy, and third-order terms vanish.

For double resonance at a fixed frequency ν_{dr}, the field H^+, for which $m=0\rightarrow+1$, is greater than H^-, for which $m=0\rightarrow-1$. To within one part in 10^6 we find from Eq. (1) that in terms of the average field $\bar{H}=(H^++H^-)/2$

$$g_J = h\nu_{dr}/\mu_0\bar{H}. \qquad (2a)$$

This becomes

$$g_J = g_p(\nu_{dr}/\nu_p), \qquad (2b)$$

if the field is measured in terms of the NMR frequency ν_p of protons. g_p is defined such that the proton magnetic dipole-moment operator is $\mathbf{u}_p = g_p \mu_0 \mathbf{I}_p$. For protons in mineral oil we will take $g_p = 3.041978(3) \times 10^{-3}$, from the average of the measurements of g_s/g_p for protons in mineral oil of Koenig, Prodell, and Kusch,[25] and Beringer and Heald,[26] and the recent measurement of g_s of Wilkinson and Crane.[27]

For the separation of the two resonances we find that

$$\Delta H = H^+ - H^- = \frac{\alpha^2 (g_S - g_L)^2 \mu_0 H^2}{g_J} \left(\frac{4}{3\delta_0} - \frac{1}{6\delta_2} \right), \quad (3a)$$

where δ_0 and δ_2 are in ergs. When the field is measured in terms of the proton frequency,

$$\Delta \nu_p = \frac{\alpha^2 (g_S - g_L)^2 \bar{\nu}_p^2}{c g_p g_J} \left(\frac{4}{3\delta_0} - \frac{1}{6\delta_2} \right), \quad (3b)$$

where δ_0 and δ_2 are in cm^{-1}.

III. EXPERIMENT

A schematic illustration of the experiment is shown in Fig. 3. Resonance radiation from an rf lamp was focused with a quartz condensing lens onto a small cell containing cadmium or mercury vapor. The cell was located within a microwave cavity between the poles of a 12-in. electromagnet. This cavity in the case of cadmium also served as an oven to control the vapor density in the scattering cell.

Resonance fluorescence of the atomic vapor was observed with a photomultiplier at an angle of 90°. A light pipe allowed the photomultiplier to be removed from the magnetic field without loss of solid angle. A glass color filter effectively removed all scattered light except the resonance radiation.

The microwave cavity was tuned to resonance with a klystron kept at a frequency of exactly 24 Gc/sec, the field being modulated at a low audio frequency, typically 37 cps, by small rectangular coils attached to the pole faces. At resonance, the light scattered at 90° changed in intensity, and the signal from the photomultiplier was displayed on an oscilloscope.

Field measurements were made with an NMR magnetometer, whose small mineral-oil probe was located as close to the scattering cell as possible. The offset correction was measured with a second probe and oscillator.

Optics

The cadmium lamp used consisted of a small quartz cylinder, 1.5 cm diameter and 2.5 cm long, carefully baked out, purged with an rf discharge in argon and then filled with a small amount of distilled natural cadmium and about 1 torr of spectroscopically pure argon. It was identical to the lamps previously used in this laboratory for several level crossing and double resonance experiments.[13,14] It was excited by being placed, without cooling, in the tank coil of a Hartley oscillator operating in the vicinity of 30 Mc/sec. Experience has shown that these lamps furnish about 1 W (total output over 4π solid angle) in the 3261-Å intercombination line.

For mercury an Osram lamp was used that had the outer jacket removed, and was cooled with a gentle air stream to eliminate self-reversal. It too was excited with a radiofrequency oscillator operating near 30 Mc/sec and was run at the same

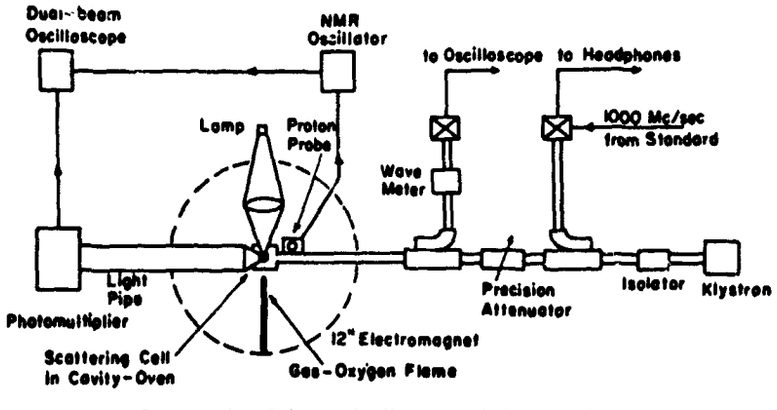


FIGURE 3.—Schematic diagram of the experiment.

current specified by the manufacturer for operation at 60 cps.

The scattering cells were designed to have as small an effect as possible on the electrical properties of the microwave cavity. They were very thin walled quartz bubbles, 5 mm in diameter, blown at the end of a length of 2 mm outer diameter quartz tubing, which, when cut off at a length of 5 cm, served as the tail of the cell. These cells were baked out at 1000°C for a day or longer, purged with a Tessla coil for several minutes, and filled with a small quantity of distilled metal, all under very high vacuum (less than 10^{-8} Torr for the last stage). The light pipe used to remove the RCA-1P28 photomultiplier from the vicinity of the magnet was a glass tube aluminized on the inside, 1 in. in diameter, and about 2½ ft. long. To pass the 3261-Å line of cadmium, or the 2537-Å line of mercury, and exclude other scattered light, a Schott UG-11 or UG-5 colored glass filter was cemented to the end of the light pipe.

Microwave Cavity and Oven

The TE_{011} cylindrical aluminum cavity oven used for cadmium is shown in Fig. 4. It had an internal diameter of 1.4 cm, a Q of about 4000 when loaded with the cell, and was coupled about 50 percent to the microwave line. A hole 6 mm in diameter at the top of the cavity, below cutoff at 24 Gc/sec, allowed illumination of the scattering cell. Small slits in the side of the cavity, also below cutoff, allowed the light scattered at 90° to be observed.

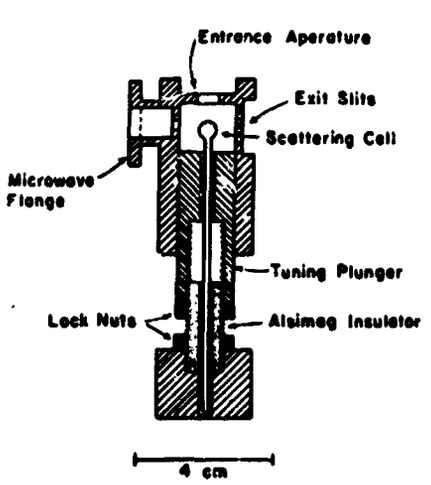


FIGURE 4.—The cavity oven used for cadmium.

The waveguide connecting the cavity to the microwave line was thinned to provide thermal insulation. The cavity was heated with a small gas-oxygen flame, the temperature being monitored by a small copper-constantin thermocouple. Thermal regulation was provided by a Honeywell R7086A thermocouple-activated regulator that played a stream of cool air against the cavity when the thermocouple voltage exceeded a preset figure.

The long tail of the scattering cell passed through an aluminum tuning plunger, and an insulating refractory section, and terminated in the aluminum cylinder shown in the bottom of Fig. 4. This cylinder was heated with another gas-oxygen flame, and its temperature regulated with a separate regulator unit. It was always kept at least 40°C cooler than the oven, and consequently determined the cadmium vapor pressure.

For mercury, in an attempt to gain a higher Q, a thin-walled copper cavity was used that was kept at room temperature. With a brass tuning plunger, a loaded cavity Q of about 8000 was obtained. The tail of the cell was usually kept at room temperature, but could also be cooled in an ice bath.

Microwave Measurements

The Varian VA-98E klystron used in this experiment is an extremely stable tube that oscillates only over the region from 23.6 to 24.4 Gc/sec, with an average output of about 30 mW. To achieve maximum stability, we used batteries for both the heater and repeller voltages, and a well-regulated power supply for the beam voltage. No electronic stabilization of the tube was employed. During a run it was kept manually to within an audiobeat not of the 24th harmonic of the 1000 Mc/sec output of the laboratory's frequency standard. This represents a stability of about one part in 10^6, and is therefore quite adequate for the purpose of this experiment.

Field Measurements

The proton probe of the NMR magnetometer used for field measurements consisted of a small spherical cavity filled with mineral oil drilled in the center of a teflon spool. The inductance for a transistorized marginal oscillator was wound on the spool which was then placed in a brass or copper shield. It was located as close as possible

to the scattering cell (about 4 cm away), and the output signal of the marginal oscillator was displayed with the double resonance signal on a dual beam oscilloscope. The NMR frequency, near 48 Mc/sec, was measured with a Hewlett-Packard direct reading frequency counter whose reference frequency was taken from the laboratory standard. The offset correction was measured by removing the oven, placing a second probe driven by another identical oscillator at the scattering cell location, and measuring the beat frequency (≈ 4 kc/sec) of the two oscillators with the frequency counter. A pentode mixer gave the difference frequency without excessive cross coupling of the two oscillators.

The offset correction was found to depend on the field history, and a correction was therefore made after each field measurement.

IV. RESULTS

For cadmium the onset of power broadening was observed when about 5 mW of microwave power were incident on the cavity, in rough accord with our preliminary estimates based on the lifetime of the $(5s5p)\ ^3P_1$ level, and the electrical properties of the cavity. With oscilloscope display giving a detector bandwidth of about 1 kc/sec we succeeded in obtaining a signal-to-noise ratio of about 30 for both the $\Delta m = \pm 1$ double resonance transitions [Fig. 5(a)]. The observed linewidth (Table II) shows no evidence of contamination broadening.

The shorter lifetime of the $(6s6p)\ ^3P_1$ level of mercury was found to require for maximum signal strength somewhat higher rf fields than the VA-98E klystron and aluminum cavity could provide. A higher Q copper cavity subsequently used for mercury just allowed the level of power broadening to be reached. The observed resonances [Fig. 5(b)] had a signal-to-noise ratio of about 5. It should be recalled that because of the illumination of the $m = \pm 1$ states the mercury signals are expected to be about 3 times weaker than those of cadmium. The additional discrepancy in signal strength can be attributed to the somewhat weaker output of the mercury lamp. The linewidth was in agreement with expectations (Table II).

We had originally hoped to extend our measurements to the $(4s4p)\ ^3P_1$ level of the even zinc isotopes, although we were aware that the long lifetime of this level would produce serious quenching of the excited state atoms against the wall of the scattering cell. We, in fact, made several runs using narrow-band phase-sensitive detection with time constants as long as 5 sec, but failed to detect either double resonance or fluorescence. Beyond the quenching and broadening mechanisms listed in Table II, we are inclined to attribute this to contaminants produced by the high cell temperatures required for zinc.

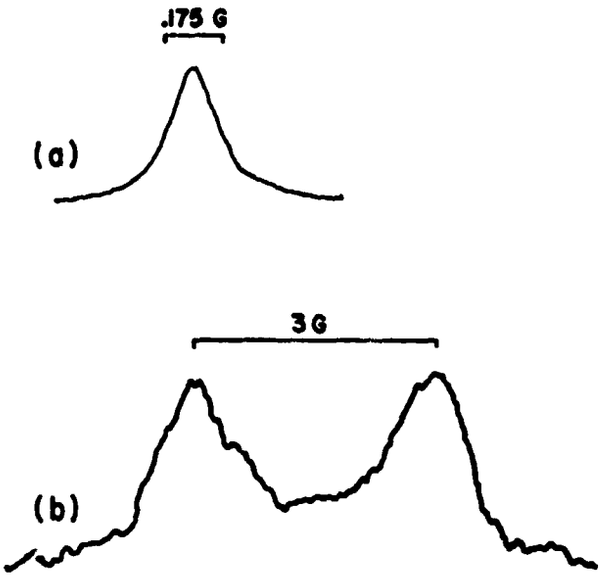


FIGURE 5.—Oscilloscope traces of the double resonance signals: (a) is one of the $\Delta m = \pm 1$ cadmium transitions, (b) is the sum of these two transitions for mercury, split by the second-order interactions.

The magnetic fields at which double resonance was observed for cadmium and mercury are listed in Table III. From Eq. (2) we then find for the $(5s5p)\ ^3P_1$ level of cadmium that

$$g_J = 1.499846(13) = \tfrac{3}{2} - 154(13) \times 10^{-6},$$

in good agreement with the values $g_J^{111} = \frac{3}{2} - 171(9) \times 10^{-6}$ and $g_J^{113} = \frac{3}{2} - 157(9) \times 10^{-6}$ found for the odd isotopes from level crossings[13] and zero-field double resonance[12] (we have here recalculated g_J^{111} and g_J^{113} taking $g_p = 3.041978 \times 10^{-3}$).

Likewise, we find for the $(6s6p)\ ^3P_1$ level of mercury that

$$g_J = 1.486094(8) = \tfrac{3}{2} - 13\ 906(8) \times 10^{-6}.$$

This is in reasonably good agreement with a recent preliminary value of $g_J^{199} = 1.486118(16)$ found by

Smith,[28] who considered both his own level-crossing study of Hg^{199}, and the results of low-field double resonance.[11] It is also in agreement with the less precise value 1.486350(300) found by Dodd[29] from double resonance in the even isotopes. It is in disagreement, however, with the self-consistent values $g_J^{199}=1.486147(10)$ and $g_J^{201}=1.486156(18)$ calculated by Kaul in his thesis[9] from his study of level crossings, and the low-field double resonance results.[6,11]

For cadmium the known lifetimes of the $^3P_1$ and $^1P_1$ levels allow α to be calculated directly from the relation

$$\frac{\beta^2}{\alpha^2}=\frac{\tau(^1P_1)\lambda^3(^3P_1)}{\tau(^3P_1)\lambda^3(^1P_1)}, \tag{4}$$

and the normalization condition $\alpha^2+\beta^2=1$. Using the lifetimes and wavelengths of Table I we find that $\alpha=0.99900(5)$ and $\beta=-0.0448(11)$. The phase convention of Condon and Shortley is used to determine the sign of β.[24] From Eq. (3) the splitting of the resonances is then calculated to be $\Delta\nu_p=40.245(4)$ kc/sec, or $\Delta H=9.452(1)$ G. This is seen from Table III to be 0.255 kc/sec lower than the experimental value, but just within the experimental uncertainty.

Knowing α and β independently from the lifetimes, we are in a position to determine the sum of the relativistic and diamagnetic corrections to g_J. Theoretically

$$g_J=\tfrac{1}{2}\alpha^2(g_L+g_S)+\beta^2 g_L+\Delta g_{\rm rel}+\Delta g_{\rm diam}. \tag{5}$$

Taking $g_L=1$,[30] and $g_S=2.002319$,[27] we find that[31]

$$\Delta g_{\rm rel}+\Delta g_{\rm diam}=-318(40)\times 10^{-6}.$$

Since for mercury the lifetime of the $^1P_1$ level is poorly known, the sum of the relativistic and diamagnetic corrections cannot be found. Instead, however, we may use Eq. (5) neglecting these corrections to obtain $\alpha=0.9849(4)$ and $\beta=-0.1733(24)$. We then calculate from Eq. (3) the splitting of the mercury resonances be to $\Delta\nu_p=12.468(8)$ kc/sec or $\Delta H=2.928(2)$ G. It is interesting to note that this is lower than the experimental value by 0.252 kc/sec, nearly the same as in the case of cadmium. Since the measurement of the splitting ΔH is not susceptible to the same systematic errors as the absolute field measurements, the uncertainties in ΔH given in Table III must be considered quite conservative, and this discrepancy may be significant.

V. DISCUSSION

While an accuracy of about one part in 10^5 has been attained in the field measurements reported in this experiment, for cadmium this represents an uncertainty of one linewidth, and can certainly be improved upon. In retrospect we believe that higher precision could have been realized at somewhat lower frequencies and fields. Most of our uncertainty can be ascribed to the inhomogeneities which crept in as the field approached 11 400 G. Had we worked at half the present field and frequency, for example, the field would have been

TABLE III.—*Measured fields for double resonance in cadmium and mercury, in terms of the NMR frequency of protons in mineral oil, and in G, taking for $1/2\pi\times$ the gyromagnetic ratio 4.25760(4) kc/sec-G.*[a] *A series of runs were made for each atom, care being taken to randomize systematic errors from run to run by slightly relocating the cell and probe within the field, reversing the polarity of the magnet, varying the modulation frequency, changing the vapor pressure, reassembling the optics, etc. The uncertainties quoted are 3 times the standard deviation of the means of the individual runs.*

Atom	H^+	H^-	ΔH
Cd	48 696.91(66) kc/sec	48 656.40(45) kc/sec	40.50(28) kc/sec
	11 437.64(19) G	11 428.13(15) G	9.51(7) G
Hg	49 133.45(35) kc/sec	49 120.73(38) kc/sec	12.72(26) kc/sec
	11 540.17(14) G	11 537.19(14) G	2.99(7) G

[a] R. L. Driscoll and P. L. Bender, Phys. Rev. Letters 1, 413 (1958).

considerably more uniform, the scattering cell would have been twice as large (and easier to construct and more durable), the Doppler and wall collision linewidths halved, and the exit and entrance apertures increased in area by four. It is probable that under these circumstances we would have had sufficient signal strength to observe double resonance in zinc, and that the cadmium resonance could have been measured to a small fraction of its width.

ACKNOWLEDGMENTS

We wish to express our thanks to the staff of the Columbia Radiation Laboratory for their invaluable assistance, and to Professor R. Novick for his advice and support. We would like especially to acknowledge the assistance of H. Feldman, who collaborated with us during a part of the cadmium experiment. Finally, Raymond Eisenstark assisted us liberally during all phases of this work as data taker and calculator, and T. Psaropulos kindly prepared the drawings.

REFERENCES

1. J. Brossel and F. Bitter, Phys. Rev. 86, 308 (1952).
2. G. W. Series, Rept. Progr. Phys. 22, 280 (1959).
3. F. D. Colegrove, P. A. Franken, R. R. Lewis, and R. H. Sands, Phys. Rev. Letters 3, 420 (1959).
4. P. A. Franken, Phys. Rev. 121, 508 (1961).
5. J. N. Dodd, Proc. Phys. Soc. (London) 77, 669 (1961).
6. R. H. Kohler, Phys. Rev. 121, 1104 (1961).
7. H. R. Hirsch, J. Opt. Soc. Am. 51, 1192 (1961).
8. C. Brot, J. Phys. Radium 22, 412 (1961).
9. R. D. Kaul, Ph.D. thesis, Case Institute of Technology, 1963 (unpublished).
10. W. Smith, Ph.D. thesis, Massachusetts Institute of Technology, 1963 (unpublished).
11. C. V. Stager, Phys. Rev. 132, 275 (1963).
12. R. F. Lacey, Ph.D. thesis, Massachusetts Institute of Technology, 1959 (unpublished).
13. P. Thaddeus and R. Novick, Phys. Rev. 126, 1774 (1962).
14. M. N. McDermott and R. Novick, Phys. Rev. 131, 707 (1963).
15. F. W. Byron, M. N. McDermott, and R. Novick, Phys. Rev. 132, 1181 (1963).
16. P. Thaddeus and M. N. McDermott, Phys. Rev. 132, 1186 (1963).
17. M. N. McDermott, R. Novick, B. W. Perry, and E. Saloman, Phys. Rev. 134, B25 (1964).
18. V. K. Böckman, H. Krüger, and E. Rechnagel, Ann. Physik 20, 250 (1957).
19. F. W. Byron, M. N. McDermott, R. Novick, B. W. Perry, and E. Saloman, Phys. Rev. 134, A47 (1964).
20. A. Landman and R. Novick, Phys. Rev. 134, A56 (1964).
21. F. Bitter, S. P. Davis, B. Richter, and J. E. R. Young, Phys. Rev. 96, 1531 (1954).
22. J. P. Barrat, J. Phys. Radium 20, 541, 633, 657 (1959).
23. F. W. Byron, M. N. McDermott, and R. Novick, Phys. Rev. 134, A615 (1964).
24. A. Lurio, M. Mandel, and R. Novick, Phys. Rev. 126, 1758 (1962).
25. S. H. Koenig, A. G. Prodell, and P. Kusch, Phys. Rev. 88, 191 (1952).
26. R. Beringer and M. A. Heald, Phys. Rev. 95, 1474 (1954).
27. D. T. Wilkinson and H. R. Crane, Phys. Rev. 130, 852 (1963).
28. W. Smith (private communication).
29. J. N. Dodd, Proc. Phys. Soc. (London) 78, 65 (1961).
30. This neglects effects due to nuclear motion, which are expected to contribute a correction to g_L of the order of a few ppm. See A. Abragam and J. H. Van Vleck, Phys. Rev. 92, 1448 (1953).
31. The theoretical values of $g_J=\frac{3}{2}-18(80)\times10^{-6}$ and $\Delta g_{\text{rel}}+\Delta g_{\text{diam}}=-135(81)\times10^{-6}$ given in Ref. 13 are in error. These values were calculated from Eq. (13), Ref. 13 on the basis of outdated lifetimes—not those given in Table I, Ref. 13. To the best of our knowledge, the remaining results in that work are self-consistent.

N64-24084

ORIGIN OF DIAMONDS IN THE URELITES*

MICHAEL E. LIPSCHUTZ†

Diamonds in two of the three diamantiferous stony meteorites (ureilites) show a pronounced crystallographic orientation. This suggests that the diamonds were produced by shock during breakup of the meteorite's parent body rather than by high gravitational pressure.

In previous papers (*1*, *2*) on the origin of meteoritic diamonds (Table 1) I concluded that the diamonds in the Canyon Diablo iron were formed by conversion of graphite as a result of the shock of impact of the meteoroid with the earth. The subsequent finding of coesite and stishovite (*3*) in the sandstone around the Canyon Diablo crater indicates shock pressures of at least 100 kbar. The successful shock-induced conversion of graphite to diamond in the laboratory (*4*) also lends strong support to the shock origin of Canyon Diablo diamonds.

Diamonds also occur in a rare type of stony meteorite, the ureilites (*5*). These meteorites, however, are too small (Table 1) to have hit the earth at speeds much greater than terminal velocity, and this implies that the diamonds could not have been formed during the impact of the ureilites with the earth. I therefore suggested (*1*) that the diamonds in these meteorites were formed by some preterrestrial shock—probably the one occurring during catastrophic breakup of the ureilites' parent body. In this paper I report on an x-ray study of meteoritic and synthetic diamonds which was undertaken to examine this suggestion.

The direct conversion of graphite to diamond is not a complex one. It is necessary only to increase the interatomic distance within individual carbon planes by 0.12 Å (Fig. 1) and to decrease the interplanar spacing (*c*-axis) by about 1.86 Å. If a randomly oriented infinite distribution of graphite crystallites is assumed, it should be possible to determine whether a given diamond aggregate was produced by shock, inasmuch as only those graphite crystallites with *c*-axes parallel to the direction of motion of the shock will be converted to diamond (*6*). The resulting diamonds should exhibit a preferred crystallographic orientation. Diamonds formed by hydrostatic (gravitational) compression, on the other hand, will show no such structure since hydrostatic pressure acts isotropically in a medium. In fact, the situation may not be quite so simple because of

TABLE 1.—*Identification of diamonds in ureilites and the phases identified in individual diamonds.*

Meteorite	Recovered mass (kg)	Phases identified
Dyalpur (*21*)	0·26	Diamond, graphite, α-iron
Goalpara (*10*)	2·7	Diamond, graphite, α-iron, trolite, chromite
Novo Urei (*22*)	1·9	Diamond, graphite, α-iron, clinopyroxene

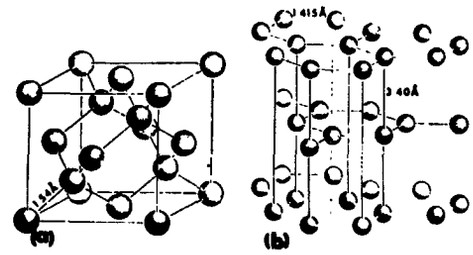


FIGURE 1.—Crystal structures of diamond (a) and graphite (b) [after Taylor (9)].

*Published in *Science*, 143 (3613): 1431-1434, March 27, 1964.

†Goddard Space Flight Center and Physikalisches Institut, Universität Bern, Sidlerstrasse 5, Bern, Switzerland.

complicating effects due to secondary and rarefaction (reflection) waves. Therefore, a preferred orientation should be apparent in at least some instances of diamonds formed by shock. The absence of such an orientation, however, cannot rule out the formation of diamonds by shock.

Specimens from all three ureilites were prepared for x-ray analysis by carefully carving or chipping them out so as not to affect the possible orientation of the crystallites within each fragment. The specimens, each of which weighed less than about 0.1 mg, were then examined by previously described diffraction techniques and equipment (*2*) to determine their composition. Fortunately the other phases in the specimens did not possess diffraction lines which would interfere with those of diamond.

After verification of the presence of diamond in the fragment, the specimen was studied with a Unicam model S.25 single crystal goniometer (*7*). In this experimental arrangement, a polycrystalline aggregate would be indicated by uniform ellipses centered on the x-ray outlet port, while preferred orientation would be evidenced by enhancement of segments of these ellipses. Large single crystals would be indicated by single spots, while deformed single crystals would be evidenced by asterism of these spots in directions approximately perpendicular to the ellipses.

Stationary x-ray photographs were taken every 10° with iron-filtered CoKα [for the (111) diamond planes] and zirconium-filtered MoKα [for the (220) and (311) planes of Goalpara diamond]. The angles ρ and Φ, corresponding to the ends of the preferred orientation zones in each set of normals to the reflecting planes, were measured (*8*) and then plotted on a polar stereographic net graduated in 2° increments (*9*).

The composition and sizes of the diamonds in the ureilites vary. All of the discrete diamond grains that I have analyzed are polycrystalline aggregates consisting of diamond, randomly oriented polycrystalline graphite, and other phases (Table 1).

The α^1–α^2 doublet of the diamond (331) plane provides a useful measure of the average crystallite size. From the lack of resolution of this doublet in Goalpara, Zachariasen (*10*) has estimated that the diamonds are about 100 Å. The diamonds from Dyalpur and Novo Urei on the other hand

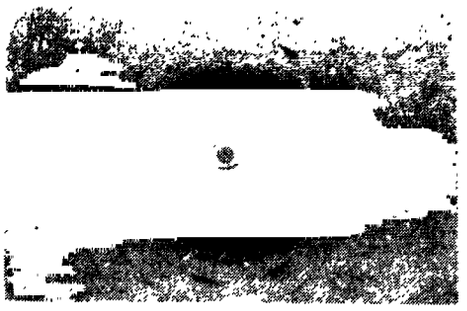


FIGURE 2.—Zirconium-filtered MoKα x-ray pattern of nonrotated diamond grain from Goalpara (×2). The graphite [002] plane shows that the graphite is polycrystalline and randomly oriented. The diamond [111], [220], [311], [400], and [331], on the other hand, shows preferred orientation.

are relatively large; those from Dyalpur are very much larger than 300 Å.

The diffraction pattern of a Goalpara diamond is shown in Fig. 2. The diamond reflections indicate pronounced preferred orientation whereas the graphite is randomly oriented. Stereographic projections of the diamond [111], [220], and [311] directions (Fig. 3 *a*, *b*, and *c*) reveal very striking patterns. The angles listed on each projection are those which that simple cubic plane makes with its (311) plane while the roman numerals indicate the multiplicity of these angles (*11*). Agreement is quite good and indicates that the oriented planes in the diamond are probably those parallel to (311).

It is important to note that the preferred orientation of diamond, which is indicative of formation by an anisotropic process, is present in Goalpara and (as will be demonstrated) also in Novo Urei. Natural terrestrial diamonds have never been reported to contain preferred orientation and in an extensive monograph (*12*) on these structures in terrestrial rocks and minerals there is no mention of this "fabric" in diamond.

Previous x-ray studies of hydrostatically produced synthetic diamond (*13*) do not indicate preferred orientation. The two samples which I studied do not show such orientation.

Synthetic diamonds produced by the anisotropic application of temperature or pressure or both, that is by flash-heating (*14*) or by a 300-kbar shock (Fig. 4), have oriented structures. The stereographic projections of these data show a grouping at about $\rho = 70°$ for the former and a

single "node" (compare with Fig. 5) at about 40° for the latter. Diamonds produced by shock at pressures significantly lower than 300 kbar are very polycrystalline and thus do not have detectable orientation. Thus, in accord with theory, the only diamonds that exhibit preferred orientation were produced by some anisotropic process.

The x-ray pattern of a diamond from Novo Urei (Fig. 6) is similar to that of a diamond produced

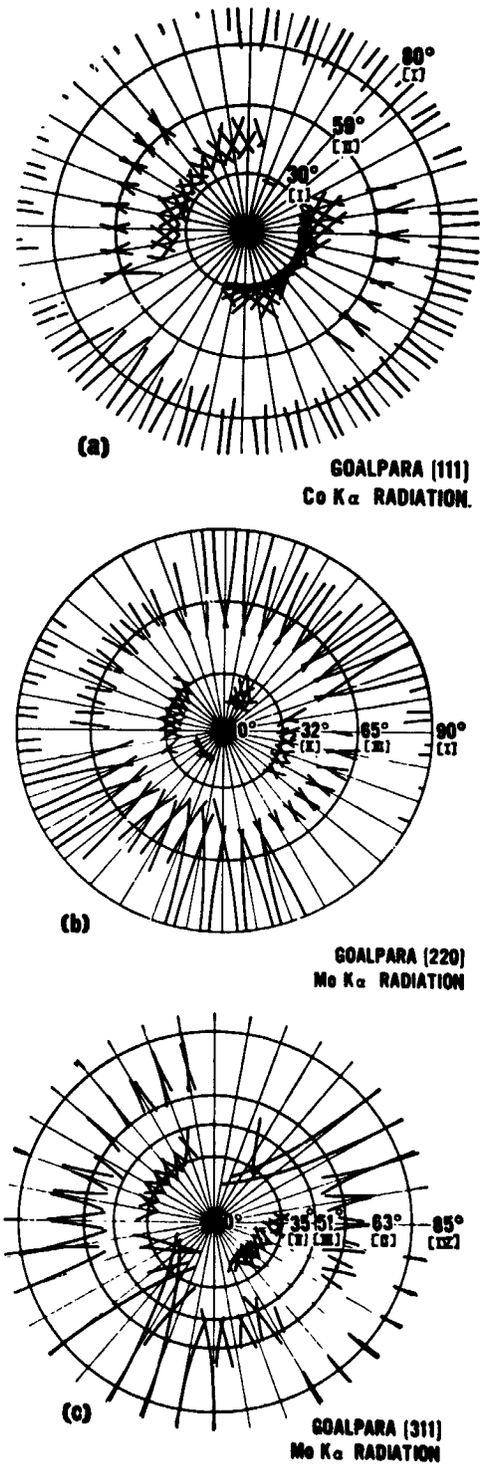


FIGURE 3.—Stereographic projections of Goalpara diamond (a) [111], (b) [220], and (c) [311]. The theoretical angles shown between each direction and the [311] are in agreement with the observed groupings.



FIGURE 4.—Iron filtered *CoKα* x-ray pattern of graphite partly converted to diamond by a 300-kbar shock. While the graphite is polycrystalline and randomly oriented, the diamond shows preferred orientation.

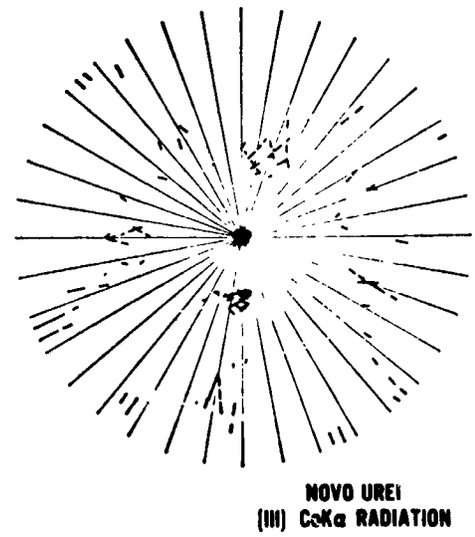


FIGURE 5.—Stereographic projection of Novo Urei diamond [111] directions. Although a pattern is evident, it is somewhat different in nature from that in Figure 3a.



FIGURE 6.—X-ray pattern (iron-filtered *CoKα*) of Novo Urei diamond. The diamond crystalline orientation is similar to that shown in Figure 4.

by a 300-kbar shock (Fig. 4). The stereographic projection of the Novo Urei data (Fig. 5) is different in character from that of Goalpara (Fig. 3*a*), possibly arising from a different alignment of the Novo Urei diamond relative to the x-ray beam or from actual differences in their pressure history.

Other phases in these two meteorites can yield information on the ureilites' pressure-temperature history. Olivine grains from Novo Urei (Fig. 7*a*) still appear to be reasonably good single crystals, although fractured to some extent. Goalpara olivine, on the other hand, is heavily fractured (Fig. 7*b*) and in some regions shows signs of recrystallization along grain boundaries, indicating excursion into temperatures well above its melting point. Since the only known natural process that can give rise to significant short-term anisotropic pressure is shock, it appears reasonable that the diamonds in at least these two ureilites were formed in this manner. It does not seem too extreme to suppose that the diamond formation and recrystallization of the olivine in Goalpara were contemporaneous. The data of Hughes and McQueen (*15*) on shocked olivine (dunite) indicate that even at the maximum pressure studied (700 kbar), the temperature reached only 720°. At pressures of 400 kbar, which is well in the diamond-forming region, the temperature rises to less than 200°. Thus, Goalpara, which was shocked more severely than Novo Urei, was probably exposed to pressures considerably in excess of 700 kbar.

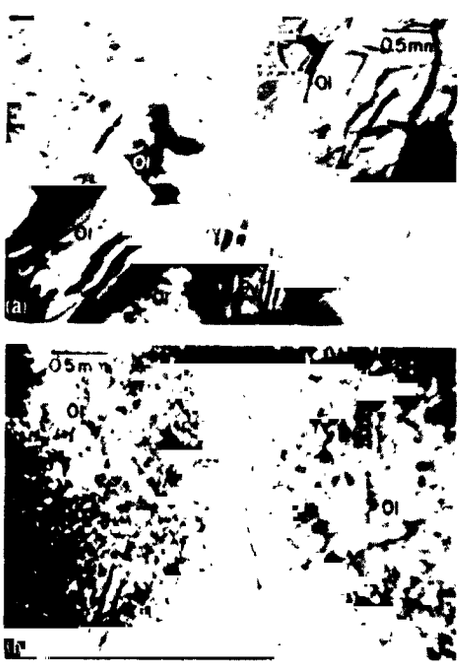


FIGURE 7.—Thin section of (a) Novo Urei ureilite and (b) Goalpara ureilite (both × 50). In (a), olivine (*Ol*) apparently still exists as single crystals although somewhat fractured. The clinopyroxene (*Py*) grains show cleavages. The opaque material is a graphite-diamond-kamacite mixture. In (b), the clinopyroxene grains are still apparently single crystals although some show an undulatory character under polarized light. Olivine grains (*Ol*) are badly shattered and in some locations are apparently recrystallized.

Diamond from Dyalpur shows no preferred orientation. One could suggest that the diamonds in this meteorite were formed by gravitational pressure rather than by shock as in Goalpara and Novo Urei. However, Dyalpur's textural similarity to the other two ureilites (particularly to Novo Urei (*16*) would argue against different modes of diamond formation by Occam's principle (see *17*).

Figure 3 (*a*, *b*, and *c*) shows certain features which may have bearing on the details of shock formation of diamond in Goalpara. First, some degree of symmetry seems evident in the distribution of groupings in each set of normals. Second, there are two "nodes" in the [220] that occur at about 20°. The former effect may be explained by lattice distortion of the diamond crystallites while the latter may be due to the presence of some diamonds formed by rarefaction waves or by secondary shocks. Another feature which may be of significance is that there appear to be only a few groupings at about 51°. These angles occur three times within the crystal (multiplicity III) and should be more prominent than the 63° angle which occurs twice (multiplicity II). This effect may only be an apparent one inasmuch as the zones in the Goalpara [311] are quite extended and markedly overlap the 51° line. The only other reasonable match for the groupings would be by planes parallel to the (220). Were these the oriented planes, however, we would not expect the group at about 30° in the [220] or at about 60° in the [111]. It is possible that the observed distribution of axes is due to some combination of oriented (110) and (311) planes, although the available evidence favors the interpretation of the oriented planes being parallel to (311) only.

From an examination of meteoritic and artificially shock-produced diamonds some sugges-

tions can be made concerning the mechanism of conversion of graphite to diamond by shock. At pressures well below 300 kbar, polycrystalline fine-grained diamond is apparently formed from polycrystalline graphite, possibly by the compression of rhombohedral graphite only (*4*). From pressures of about 300 kbar to well above 700 kbar the solid-state reaction involves conversion of the basal (001) plane of graphite to diamond (311), or possibly some combination of (311) and (110). Only at very high pressures may the conversion involve the hypothetical "metallic carbon" postulated by Libby (*18*) or possibly formation from shock-melted graphite.

The simultaneous presence of graphite, diamond, and kamacite in all of the ureilitic "diamonds" that I have investigated shows that these grains are not equilibrium assemblages. This raises the question of whether they represented an arrested stage in the conversion of graphite to diamond, or vice versa. Studies at 0 kbar (*19*) and 100 kbar (*20*) show that graphite formed from diamond shows preferred orientation. In at least two meteorites, however, it is the diamonds which are oriented, and in all diamantiferous meteorites the graphite is polycrystalline and randomly oriented. Thus, the graphite cannot have been formed from diamond. On the other hand, the theory predicts that, under favorable circumstances, shock-formed diamond would show preferential orientation. It thus appears that all meteoritic diamonds were formed by shock rather than by gravitational compression of graphite.

REFERENCE AND NOTES

1. M. E. Lipschutz and E. Anders. *Geochim. Cosmochim. Acta* **24**, 83 (1961).
2. ———, *Science* **134**, 2095 (1961).
3. E. C. T. Chao, E. M. Shoemaker, B. M. Madsen, *ibid.* **132**, 220 (1960); E. C. T. Chao, J. Fahey, J. Littler, D. Milton, *J. Geophys. Res.* **67**, 419 (1962).
4. P. DeCarli and J. Jamieson, *Science* **133**, 1821 (1961).
5. Major minerals in these meteorites are olivine, $(Mg, Fe)_2SiO_4$, and clinopyroxene, (Ca. Mg, Fe) SiO_3. The meteorites contain, in addition, minor amounts of diamond, graphite, kamacite (α-iron), troilite (FeS), and chromite ($FeCr_2O_4$). No other stony meteorites are known to contain diamond although well over 260 others have been examined for diamond by various investigators.
6. E. Anders, *Astrophys. J.* **134**, 1006 (1961).
7. Unicam Instruments Ltd., Cambridge, England.
8. N. F. M. Henry, H. Lipson, W. A. Wooster, *The Interpretation of X-Ray Diffraction Photographs* (Macmillan, New York, ed. 2, 1961), Fig. 108.
9. A. Taylor, *X-Ray Metallography* (Wiley, New York, 1961), Fig. 258.
10. H. C. Urey, A. Mele, T. Mayeda, *Geochim. Cosmochim. Acta* **13**, 1 (1957).
11. *International Tables for X-Ray Crystallography* (Kynock, Birmingham, England, 1959), vol. 2.
12. H. W. Fairbairn, *Structural Petrology of Deformed Rocks* (Addison-Wesley, Cambridge, Mass., ed. 2, 1949).
13. H. L. Pugh, J. Lees, J. A. Bland, *Nature* **191**, 865 (1961); J. A. Kohn and D. W. Eckart, *Am. Mineralogist* **47**, 1422 (1962); H. O. A. Meyer and H. J. Milledge, *Nature* **199**, 167 (1963).
14. F. P. Bundy, *J. Chem. Phys.* **38**, 631 (1963).
15. D. S. Hughes and R. G. McQueen, *Trans. Am. Geophys. Union* **39**, 959 (1958).
16. L. G. Kwascha, *Chem. Erde* **19**, 249 (1958).
17. Since this paper was submitted, A. P. Vinogradov and G. P. Vdovykin (*Geokhimiya* No. 8, 715 (1963)) independently confirmed the presence of diamond in the three ureilites. In addition, they have announced the discovery of diamond in yet another stony meteorite, Ghubara. When samples of this meteorite can be obtained, the properties of Ghubara's diamonds and olivine can be compared with those of the ureilites.
18. W. F. Libby, *Proc. Natl. Acad. Sci. U.S.* **48**, 1475 (1962).
19. H. J. Grenville-Wells, *Mineral. Mag.* **29**, 803 (1952).
20. V. M. Titova and S. I. Futergendler, *Soviet Phys. Cryst.* (*English Transl.*) **7**, 749 (1963).
21. M. E. Lipschutz, *Science* **138**, 1266 (1962).
22. M. Yerofeyev and P. Lachinov, *Compt. Rend.* **196**, 1679 (1888).

23. I am very grateful to Edward Anders for his many suggestions and comments during the course of this investigation. I also thank John Jamieson and Paul S. DeCarli, for many stimulating discussions, and E. C. T. Chao for his comments on the thin sections of the meteorites. I gratefully acknowledge the loan of the Dyalpur meteorite (E. Olsen, Chicago Natural History Museum) and the Goalpara and Novo Urei meteorites and thin sections (E. P. Henderson, U.S. National Museum). Gifts of synthetic diamonds produced by various techniques were received from R. H. Wentorf, Jr. (General Electric Research Laboratory), M. Waxman (National Bureau of Standards), and P. S. DeCarli (Stanford Research Institute). Miss Vida Milanovic and Miss Donna Reichel provided valuable assistance. The Materials Research and Development Section, NASA, lent equipment during the course of this research, which was supported in part by National Science Foundation grant G 14298.

N69-17124

THE BACKWARD RECURRENCE METHOD FOR COMPUTING THE REGULAR BESSEL FUNCTION*

THOMAS E. MICHELS

The Backward Recurrence Method (suggested by Dr. J. C. P. Miller) for computing the regular Bessel function is investigated for real, imaginary, and complex arguments. From this, one can obtain the functions $J_n(z)$, $I_n(z)$, $ber_n(z)$, and $bei_n(z)$. A section has also been devoted to show the behavior of the functions for all arguments, and graphs are presented which show the magnitudes of the functions for all orders and real arguments between 0 and 6000.

INTRODUCTION

Bessel's Equation is a second order differential equation with a singular point at $z=0$ and is of the form

$$\frac{d^2y}{dz^2} + \frac{1}{z}\frac{dy}{dz} + \left(1 - \frac{n^2}{z^2}\right)y = 0 .$$

Its two independent solutions are referred to as the regular and irregular Bessel functions (see Appendix A). Computation of the irregular function, $Y_n(z)$, is fairly easy, since one can use a straightforward recurrence relationship once he has two starting values (Reference 1) and obtain good accuracy for all n and z values. However, this method cannot be used to compute the regular function, $J_n(z)$, due to error accumulation in the recurrence. For large order and argument, asymptotic formulas prove to be too cumbersome; and in computer computation they are also inaccurate. But a method was devised by Dr. J. C. P. Miller (Reference 2) whereby an ingenious use of the recurrence relationship yields good values for all orders and arguments which fall within the limits of any particular computer.

It is the purpose of this report to explain the method and the need for it, as well as give details on the behavior and use of the recurrence formula and the accuracies obtained. (Hopefully it will also give insight into the behavior of the functions themselves.) This has been done for the argument z real, imaginary, and complex. In order to keep to this primary purpose, not much will be said about the irregular function. In general, however, the irregular functions behave just oppositely to the regular function in the range where $J_n(z)$ decreases or increases exponentially, while they behave similarly in the rest of the range. That is, if the magnitude of $J_n(z)$ *decreases exponentially*, the irregular function *increases;* and when $J_n(z)$ oscillates, the irregular function does the same.

It should be noted that the recurrence formula computes a sequence of Bessel functions as a function of n, the order, rather than the usual tabulations which are a function of z. Throughout this discussion, where the order is referred to as n in $J_n(z)$, it should be taken to mean any value of n (integer, half orders, one-third orders, etc.), unless otherwise specified. Also m is always greater than n, and $z(=x+iy)$ is regarded as a complex number. If just the real numbers are being referred to, x will be used, and for the imaginary iy will be used.

Reference is made from time to time to single and double precision. All tests were made on an IBM 7090 computer which has a word length of 36 binary bits. In the floating point mode, single precision arithmetic is done with 9 decimal digits and double precision with 18 decimal digits. This is implied throughout.

EXPLANATION OF THE METHOD FOR COMPUTING THE REGULAR BESSEL FUNCTION

Importance of the Backward Recurrence

It is well known that both the regular $J_n(z)$ and the irregular $Y_n(z)$, Bessel functions (see Appen-

*Published in *NASA Technical Note* D 2141, March 1964.

dix A) obey the recurrence relationship

$$F_{n+1}(z) + F_{n-1}(z) = \frac{2n}{z} F_n(z) .$$

If the recurrence is started with either $J_0(z)$, $J_1(z)$ or $Y_0(z)$, $Y_1(z)$, in the computer where there are a fixed number of digits, good accuracy is obtained from $n \leq z$ for both functions. This accuracy continues for the irregular function since $Y_n(z)$ *increases exponentially* for $n > z$, and the number of accurate digits is maintained or increased; but for the regular function at this point accuracy is lost very rapidly, since $J_n(z)$ *decreases* exponentially for $n > z$.

Actually, throughout the recurrence, by recurring in either direction, we obtain numbers which represent solutions to Bessel's equation, but they are linear combinations of the regular and irregular solution. Therefore, the error introduced in recurring for the regular function is actually some constant times the irregular function.

In using the formula for recurring Bessel functions, and recurring in the direction of increasing n, the number of digits of accuracy is lost very rapidly. This is not even taking into account loss of accuracy due to round off.

Let us take an extreme case and look at it this way: Assume we are in a computer which produces 8 significant digits and we are using the recurrence formula below with a small error in each term on the right, say in the last decimal place. We ask what the error ϵ_{n+1} is in our new term F_{n+1}:

$$F_{n+1}(z) + \epsilon_{n+1} = \frac{2n}{z} \left[F_n(z) + \epsilon_n\right] - \left[F_{n-1}(z) + \epsilon_{n-1}\right] .$$

Let us assume reasonable values—say for the second term on the right $F_{n-1}(z) \approx 10^{-1}$—and make the error the best we could have, and therefore in the last significant digit, or $\epsilon_{n-1} \approx 10^{-8}$. Let the magnitude of $F_n(z)$ be smaller, say 10^{-2}, again with the error being in the last place, or $\epsilon_n \approx 10^{-9}$. Now we recur to a smaller number $F_{n+1}(z)$, say of the magnitude 10^{-3}. Still, there was an error at 10^{-8}, and therefore $\epsilon_{n+1} \approx 10^{-8}$. The $(n-1)$ term had 7 accurate digits and two terms away, at $(n+1)$, the number of accurate digits is reduced by two and our error is now in the sixth significant digit.

A similar argument can be made to show that when we recur to numbers of increasing magnitude, the number of significant digits of accuracy is actually increased, or at least stays equal to the maximum number of digits available in the computer. Therefore, the use of the foregoing recurrence formula to compute the regular Bessel functions in the region where they decrease in magnitude is not acceptable as it stands.

Statement of Method

Miller's Method

A scheme was devised by Miller (Reference 2) whereby the recurrence formula could be used and accuracy still maintained throughout the whole range of n values. We start at some m, with

$$F_{m+1}(z) = 0 ,$$

$$F_m(z) = a ,$$

where a is any constant, and use the recurrence formula, but in *decreasing* magnitude in n (where we are recurring to larger numbers and therefore increasing accuracy) and generate a series of functions $F_n(z)$, $F_{n-1}(z), \cdots F_0(z)$ at some $n < m$, by

$$F_{n-1}(z) = \frac{2n}{z} F_n(z) - F_{n+1}(z)$$

which are all a constant multiple of the regular Bessel function. That is,

$$F_n(z) = \alpha J_n(z) .$$

Proof

Since the recurrence formula follows for both the regular and irregular function, we can say that it follows for a linear combination of them also, or

$$F_n(z) = \alpha J_n(z) + \beta Y_n(z) .$$

Since

$$F_{m+1}(z) = 0 = \alpha J_{m+1}(z) + \beta Y_{m+1}(z) .$$

we have

$$F_n(z) = \alpha \left[J_n(z) - J_{m+1}(z) \frac{Y_n(z)}{Y_{m+1}(z)} \right] .$$

Therefore at some $n < m$

$$J_{m+1}(z) \frac{Y_n(z)}{Y_{m+1}(z)} \approx 0$$

and hence

$$F_n(z) = \alpha J_n(z)$$

to any desired degree of accuracy. Therefore all that remains to obtain the regular Bessel function is to determine α.

Determination of α

There are a number of ways in which α can be determined. One method is to use an addition formula such as the following (Reference 3):

$$2^{\nu} \sum_{k=0}^{\infty} (\nu + 2k) J_{\nu+2k}(z) z^{-\nu} \frac{\Gamma(\nu + k)}{k!} = 1 .$$

which, of course, for integral orders or $\nu = 0$, becomes

$$J_0(z) + 2 \sum_{k=1}^{\infty} J_{2k}(z) = 1 .$$

or, for $F_k(z) = \alpha J_k(z)$,

$$F_0(z) + 2 \sum_{k=1}^{\infty} F_{2k}(z) = \alpha .$$

The summation can be carried out to $k = n$ and will give α the particular accuracy desired. Here n implies the number of the term $< m$ where $F_n(z) = \alpha J_n = \alpha J_n(z)$ (see page 1251).

Of course, if an expression is known for a particular $J_n(z)$ value, say

$$J_{1/2}(z) = \left(\frac{2}{\pi z}\right)^{1/2} \sin z .$$

then α is more easily determined by

$$\alpha = \frac{F_{1/2}(z)}{J_{1/2}(z)} .$$

For z real, the foregoing addition formula proves to be fairly accurate in the determination of α; however, for z complex or imaginary, using the addition formula in the comput r proves to be inaccurate (see page 1252).

The value of the Bessel function is only as accurate as α, since each term is divided by α; therefore the computation of α is extremely critical.

Error Introduced from Additional Formulas for Complex or Imaginary Argument

In looking at the regular Bessel function of complex or imaginary argument, we observe that

$$|J_0(z)| \approx 0.1 e^{y}$$

and

$$\lim_{n \to \infty} J_n(z) = 0 .$$

As n increases, $J_n(x+iy)$ decreases at an exponential rate almost immediately for $|z| < 10$, and for larger z it takes a few more terms before the exponential decrease. It is almost a logarithmic spiraling into the origin of the complex plane.

With this in mind, assume we are performing the addition formula for integral orders. We want to investigate the error that will exist in the determination of α, assuming an error in some $F_n(z)$.

In our downward recurrence we have the following:

$$F_0(z) + \epsilon_0 + 2 \sum_{n=1}^{\infty} \left(F_{2n}(z) + \epsilon_{2n} \right) = (1 + \epsilon) \alpha .$$

Assume the error to be in the last digit in all $F_n(z)$, which of course is the best we could expect in the

computer. Upon factoring out α, we have

$$\alpha(J_0(z) + \epsilon_0) + 2\alpha \sum_{n=1}^{\infty} (J_{2n}(z) + \epsilon_{2n}) = \alpha(1 + \epsilon) ;$$

and since

$$J_0(z) + 2 \sum_{n=1}^{\infty} J_{2n}(z) = 1 ,$$

we have

$$\epsilon_0 + 2 \sum_{n=1}^{\infty} \epsilon_{2n} = \epsilon ,$$

where ϵ is now the error we introduce in computing the number 1.

Since the errors decrease at an exponential rate as n increases,

$$\epsilon \approx \epsilon_0 ,$$

which says the error we introduce in computing the digit 1 is of the order of the error in $J_0(z)$. Therefore α will be in error by this factor. It can be seen that if $J_0(z)$ had five figure accuracy and was of the order of magnitude 10^7, the summation would actually become meaningless.

Thus, to determine α one of the Bessel function values must be computed. This can be done by an asymptotic formula which simplifies for $J_0(z)$, or the following integral relationship can be used:

$$J_n(z) = \frac{1}{\pi} \int_0^{\pi} \cos(n\theta - z \sin\theta)\, d\theta ,$$

or

$$J_0(z) = \frac{1}{\pi} \int_0^{\pi} \cos(z \sin\theta)\, d\theta .$$

Integration of the expression for $J_0(z)$ by the trapezoidal rule generally yields sufficient accuracy, especially for large z values for which the slopes of the functions become large. For larger z values, of course, $\Delta\theta$ must be decreased. Integration performed by using the trapezoidal rule on various $|z|$ values (less than 75) and with $\Delta\theta = \pi/1000$ gave accuracy in the seventh and eighth digit.

When using the recurrence formula for $I_n(y)$

$$G_{n-1}(y) - G_{n+1}(y) = \frac{2n}{y} G_n(y) ,$$

the summation

$$e^y = I_0(y) + 2 \sum_{n=1}^{\infty} I_n(y)$$

can be used to determine α. This formula is of course used in the same manner as are the $F_n(z)$ functions. That is,

$$G_{m+1}(y) = 0 ,$$

$$G_m(y) = a ,$$

and, at the same n,

$$G_n(y) = aI_n(y)$$

to any desired degree of accuracy, and

$$ae^y = G_0(y) + \sum_{n=1}^{\infty} G_n(y).$$

No tests were made with this recurrence formula.

Determination of n versus m Values

We cannot give a definite value to the number of terms necessary in the recurrence to obtain good values for $J_n(z)$. It depends, of course, on z as well as the value given to a in $F_m(z) = a$, and also upon the particular accuracy desired.

The following empirical formulas will give values for m which generate a sequence $J_0(z), \cdots, J_n(z)$, where $J_n(z)$ is of the magnitude 10^{-21}, that have the accuracies described in a later section. This is assuming computation is done in single

precision and $a = 10^{-20}$.

$$m = 5.0\,|z| + 15 \quad \text{for} \quad 0.1 \le |z| < 10 ; \tag{1}$$

$$m = 1.48\,|z| + 48 \quad \text{for} \quad 10 \le |z| < 150 ; \tag{2}$$

$$m = 1.05\,|z| + 112 \quad \text{for} \quad 150 \le |z| \le 6000 . \tag{3}$$

It is felt that Equation 3 can be extended indefinitely for all values of $z \ge 150$; however, tests did not go beyond the range given.

For all values of z and n the following general rule can be employed: *For single precision accuracy, a safe* m *vs* n *is obtained when the following condition is met:*

$$\frac{J_m(z)}{J_n(z)} < 10^{-5} .$$

Table 1 gives various n vs m values for real arguments in the range from 0.01 to 6000.

Details on Accuracy

z Real

It has been said that accuracy is gained in the downward recurrence until the point where oscillations begin to occur. If the correct n vs m has been used for a particular x, then at this n value $F_n(x) = \alpha J_n(x)$ to any desired degree of accuracy. From here down to $J_0(x)$, accuracy loss is due only to round off and truncation in the computer. But, in this region, the functions oscillate usually between +0.3 and the number of accurate digits

Table 1

The integral order n of the regular Bessel function having a particular magnitude in the range where $n > x$ is given for various x values, and the recurrence term m which will yield eight place accuracy for all orders from zero to n. Also tabulated is the value of m which yields overflow in the computer for the downward recurrence at $F_0(x)$. These values were obtained using double precision with $F_m(x) = 10^{-20}$.

Argument x	Magnitude of $J_n(x)$, where $n > x$, for particular orders n and recurrence terms m										m value yielding overflow
	10^{-2}		10^{-8}		10^{-16}		10^{-24}		10^{-32}		
	n	m	n	m	n	m	n	m	n	m	
.01	1	4	4	6	6	8	9	11	11	13	19
.05	1	5	5	7	8	10	11	13	14	16	22
.10	2	5	5	7	9	11	12	14	16	18	26
.50	3	7	8	11	13	15	17	19	21	23	34
2.0	6	12	12	16	19	22	25	28	30	33	47
5.0	9	13	18	23	26	30	34	38	40	43	60
8.0	14	16	22	28	32	37	41	45	48	51	70
12.0	17	20	28	34	39	44	48	53	57	61	81
20.0	25	29	39	46	51	57	62	67	72	76	100
40.0	46	55	63	72	79	86	92	98	104	109	138
100.0	109	119	130	143	152	163	170	179	186	193	229
300.0	310	321	334	361	374	389	398	411	420	431	485
500.0	511	530	555	583	587	604	615	631	642	655	715
1000.	1012	1047	1063	1088	1107	1128	1145	1164	1177	1192	1270
2000.	2015	2055	2080	2115	2135	2165	2185	2210	2225	2245	2340
3000.	3015	3065	3090	3127	3154	3208	3208	3235	3255	3277	3385
4000.	4015	4070	4095	4140	4170	4205	4230	4260	4280	4305	4425
6000.	6015	6080	6110	6160	6195	6235	6260	6295	6320	6350	6485

becomes more or less as the recurrence goes from larger to smaller numbers. Therefore, accuracy loss due to round off is not quite as bad as it might seem. In fact, when recurring for $x=6000$, and $m=6400$, the error only crept over to the fifth significant digit; this was using single precision in the computer.

Particular values were checked in already existing tables (References 3, 4) for $0.05 \leq x \leq 100$ and $0 \leq n \leq 100$. Values having greater x and n were checked against values obtained using the recurrence in double precision. The accuracies obtained were the following:

Argument	Accuracy
	within a few figures
$0.05 < x \leq 30$	in the 8th digit
$30 < x \leq 100$	at least 7 digits
$100 < x \leq 1000$	at least 5 digits
$1000 < x \leq 6000$	at least 4 digits

These accuracies are the worst obtained over the whole range of orders out to where $J_n(x) \approx 10^{-38}$.

Using double precision, of course, would give approximately eight more digits of accuracy for all arguments tabulated. It seemed unnecessary to check the recurrence for real arguments greater than 6000, in view of their sparse usage.

The accuracy stated above is obtained only if α in

$$F_n(x) = \alpha J_n(x)$$

is computed to at least the number of digits stated. Of course, if α is only good to a few digits then the Bessel function can be no better.

z Complex or Imaginary

In the downward recurrence for imaginary and complex argument $(x+iy)$, where $|y/x|$ is not small, the magnitudes of the numbers increase over the whole range of n values. Therefore, accuracy is gained or maintained over the whole range. These functions were checked in tables of Thomson functions for ber(x) and bei(x) and in formulas for $J_{1/2}(z)$, $J_{3/2}(z)$, $J_{5/2}(z)$. These all proved to be accurate to seven digits for various $|z|$ values ranging from .05 to 100.

For argument $(x+iy)$ where $|y/x|$ is small, approximately, less than 0.1) the function behaves similarly to the real function in that the magnitudes oscillate and then continually decrease when n becomes greater than z. In general the accuracies should behave similarly to the accuracies of the real function.

Details on Use of the Recurrence
Choosing the Value for α

When using the above method for computing the regular Bessel function, the value we give to a is completely arbitrary. In other words the method "works" for any value of a. However, as a general rule, we should use as small a value, in magnitude, as is practical for the computer. This enables us to start out farther in the downward recurrence or use a larger m value and thereby obtaining higher order Bessel functions—and, of course, with less chance of overflow in recurring to numbers which are too large for the particular computer.

Consider a sequence of functions, say for real z,

$$J_0(x),\ J_1(x) \cdots J_n(x),\ J_{n+1}(x)\ .$$

The magnitudes of the functions may vary from approximately 0.1 for $J_0(x)$ down to say 10^{-38} for $J_n(x)$. That is, the values cover a range equal to approximately 10^{38}. Since the same range would be covered in our downward recurrence, the magnitude of $F_0(x)$ depends completely on the value of a at any particular m. If $a=10^{-12}$, then $F_0(x)$ would be 38 orders of magnitude away or 10^{25}, and so on. Therefore, to compute larger order Bessel functions, a smaller value for

$$F_m(x) = a\ ,$$

should be used. It should be noted that the value given to a will not increase or decrease accuracy.

z Real

As has been stated, $J_0(x)$ does not become greater than 1 for all x, and

$$\lim_{n \to \infty} J_n(x) = 0\ ,$$

the values oscillate usually between ± 0.3 until n becomes greater than x, then they stay positive

and begin to decrease at exponential rate to zero at $n = \infty$. The recurrence must start beyond this point for the method to succeed.

As a general rule, for all x, recurrence can begin at an m value for a particular order n where

$$\frac{F_m(x)}{F_n(x)} = \frac{J_m(x)}{J_n(x)} < 10^{-s} ,$$

and still obtain eight figure accuracy for $J_n(x)$.

z Complex

In contrast to the function of real argument,

$$|J_0(z)| \approx 0.1\, e^{y} ,$$

but still

$$\lim_{n\to\infty} J_n(x) = 0 ,$$

and the magnitude of the complex function decreases continually, slowly at first for low orders

Table 2

Values obtained for $J_n(10)$ from the forward recurrence in double and single precision and the backward recurrence in double precision.

n	$J_n(10)$		
	Forward Recurrence Single Precision	Forward Recurrence Double Precision	Backward Recurrence Double Precision
0	-0.24593576E-00	-0.24593576E-00	-0.24593576E-00
1	0.43472745E-01	0.43472745E-01	0.43472745E-01
2	0.25463031E-00	0.25463031E-00	0.25463031E-00
3	0.58379377E-01	0.58379379E-01	0.58379379E-01
4	-0.21960268E-00	-0.21960268E-00	-0.21960268E-00
5	-0.23406152E-00	-0.23406152E-00	-0.23406152E-00
6	-0.14458837E-01	-0.14458842E-01	-0.14458842E-01
7	0.21671092E-00	0.21671092E-00	0.21671092E-00
8	0.31785411E-00	0.31785412E-00	0.31785412E-00
9	0.29185566E-00	0.29185568E-00	0.29185568E-00
10	0.20748607E-00	0.20748610E-00	0.20748610E-00
11	0.12311649E-00	0.12311652E-00	0.12311652E-00
12	0.63370191E-01	0.63370255E-01	0.63370255E-01
13	0.28971968E-01	0.28972083E-01	0.28972083E-01
14	0.11956926E-01	0.11957163E-01	0.11957163E-01
15	0.45074252E-02	0.45079730E-02	0.45079730E-02
16	0.15653493E-02	0.15667561E-02	0.15667561E-02
17	0.50169241E-03	0.50564667E-03	0.50564667E-03
18	0.14040489E-03	0.15244248E-03	0.15244248E-03
19	0.37652134E-05	0.43146276E-04	0.43146277E-04
20	-0.12609708E-03	0.11513367E-04	0.11513369E-04
21	-0.50815354E-03	0.29071915E-05	0.29071994E-05
22	-0.20081478E-02	0.69683732E-06	0.69686851E-06
23	-0.83276965E-02	0.15889271E-06	0.15902198E-06
24	-0.36299255E-01	0.34069179E-07	0.34632629E-07
25	-0.16590872E-00	0.46393399E-08	0.72146349E-08
26	-0.79324437E-00	-0.10872479E-07	0.14405452E-08
27	-0.39589619E 01	-0.61176230E-07	0.27620052E-09
28	-0.20585150E 02	-0.31947916E-06	0.50937552E-10
29	-0.11131787E 03	-0.17279071E-05	0.90497669E-11
30	-0.62505851E 03	-0.97023620E-05	0.15510961E-11
31	-0.36390331E 04	-0.56486385E-04	0.25680948E-12
32	-0.21936946E 05	-0.34051320E-03	0.41122714E-13
33	-0.13675742E 06	-0.21227981E-02	0.63758926E-14
34	-0.88066203E 06	-0.13669954E-01	0.95817661E-15
35	-0.58517443E 07	-0.90832891E-01	0.13970838E-15
36	-0.40081548E 08	-0.62216029E 00	0.19782068E-16
37	-0.28273539E 09	-0.43887211E 01	0.27225057E-17
38	-0.20521604E 10	-0.31854376E 02	0.36447453E-18
39	-0.15313683E 11	-0.23770464E 03	0.47500695E-19
40	-0.11739457E 12	-0.18222410E 04	0.60308953E-20

until, approximately when n becomes greater than x/2, they decrease at an exponential rate. The role that x plays will be discussed in a later section. Therefore the downward recurrence can begin anywhere in the range of n values as long as the condition

$$\frac{F_m(z)}{F_n(z)} < 10^{-5} ,$$

is met to give eight figure accuracy in $J_n(z)$.

z Imaginary

The function of imaginary argument behaves similarly to the function of complex argument in that they continually decrease in magnitude and, approximately at z/2, they decrease exponentially. For integral orders, the functions oscillate between being pure real and pure imaginary; but for orders that are not integral, $J_n(iy)$ possess both real and imaginary parts. The integral orders are usually multiplied by $e-in\ \pi/2$ to give the real function, $I_n(y)$, which is widely used. That is,

$$I_n(y) = e^{-in\pi/2} J_n(iy) .$$

More will be discussed on the function $I_n(y)$ in the next section. An example of the method will be found in Table 2, which gives $J_n(10)$ by the forward recurrence technique in both single and double precision, and by the backward recurrence technique in double precision.

NUMERICAL BEHAVIOR OF THE REGULAR BESSEL FUNCTION

In seeking to present a complete picture of the regular solutions to Bessel's equation

$$y'' + \frac{1}{z} y' + \left(1 - \frac{n^2}{z^2}\right) y = 0$$

it seemed that a better understanding of the solutions was brought about when first they were observed as functions of n, and then as functions z. Then they could be brought together and a better picture could be visualized for all n and z.

z Real

Two-dimensional graphs have been drawn (Appendix B) showing various lines of constant magnitude for all real z and n for $0<x\leq6000$. These can be extended to larger x, since the lines are fairly linear.

It n be seen from a plot of $J_n(x)$ vs. n (Figure 1) that the functions oscillate between ±1, with

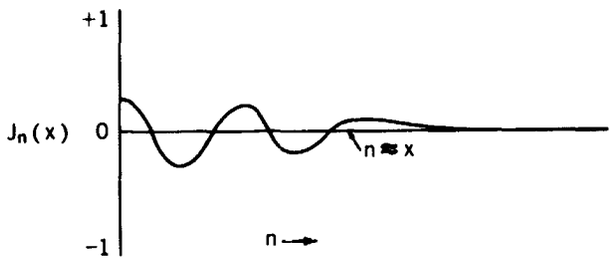


FIGURE 1.—Plot of $J_n(x)$ vs. n, n continuous.

the oscillations dampening in the range $0\leq n\leq x$. Then for $n\geq x$ the values stay positive and begin to decrease very rapidly (here is where forward recurrence begins to fail), until at $n=\infty$, $J_n(x)=0$. This is true for all values of x with the exception $J_n(0)=0$, and $J_0(0)=1$. Let us now consider the function more closely.

For $5<x<n$, $J_n(x)$ oscillates between ±0.3 and for $0<x\leq5$, $n\leq3$, $J_n(x)$ oscillates between ±0.2 and for all n, $J_n(n)\leq\cdot1$. It can be seen that dampening of the oscillations takes longer for larger x, since it takes more terms for n to become similar to x.

On the other hand, it can be seen from a plot of $J_n(x)$ vs. x (Figure 2) that the functions start

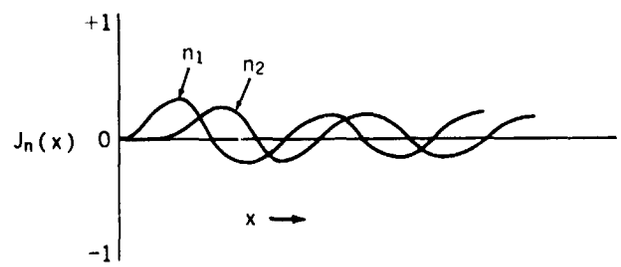


FIGURE 2.—Plot of $J_n(x)$ vs. x, $n_1<n_2$.

at $x=0$, (except that $J_0(0)=1$) and increase very slowly; the larger the n value the slower it increases, staying positive and never exceeding 0.1 until $n\approx x$. At this point they continue to increase, but as x increases oscillations begin. For $n>3$ and all x, the oscillations are between ±0.3.

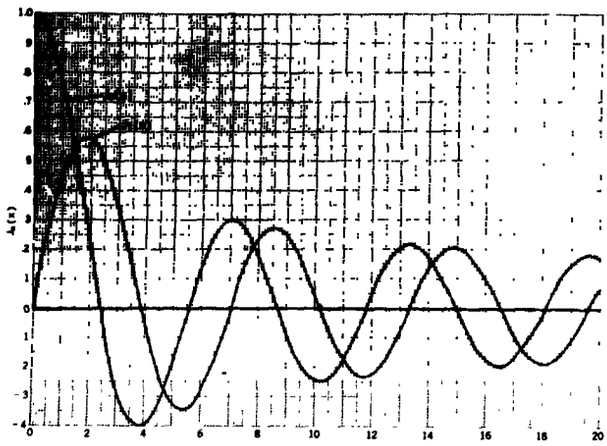


FIGURE 3.—Detailed plot of $J_0(x)$ and $J_1(x)$ vs. x, $0 \leq x \leq 20$.

Figure 3 shows a detailed plot of $J_0(x)$ and $J_1(x)$ versus x.

z Complex

The magnitude of the function for complex argument $z = x+iy$ begins at

$$|J_0(x+iy)| \approx 0.1\, e^y$$

for any z and, where $|y/x|$ is not small (see page 9)

$$\lim_{n\to\infty} J_n(x+iy) = 0 .$$

The functions continually decrease in magnitude as n becomes larger and, at approximately where $n > |z|/2$, they decrease at an exponential rate.

If we plot the modulus of $J_n(z)$ vs. n, we get the exponential curve shown in Figure 4. The

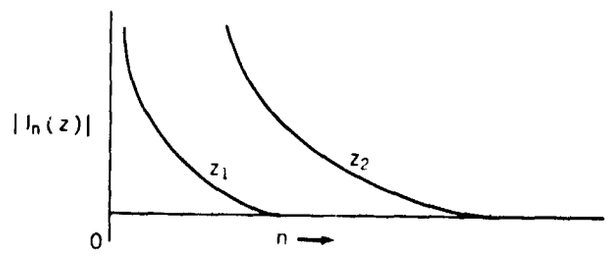


FIGURE 4.—Plot of $|J_n(x+iy)|$ vs. n.

effect of x in the argument is that for larger x, the gradient of the slope becomes less.

In considering a picture of the function of complex argument, it seems that the easiest and perhaps the best approach is set

$$J_n(x+iy) = u+iv$$

and plot the functions in both planes.

In Figure 5, the spiral is started at $J_0(x+iy)$, and as n increases the function spirals into the

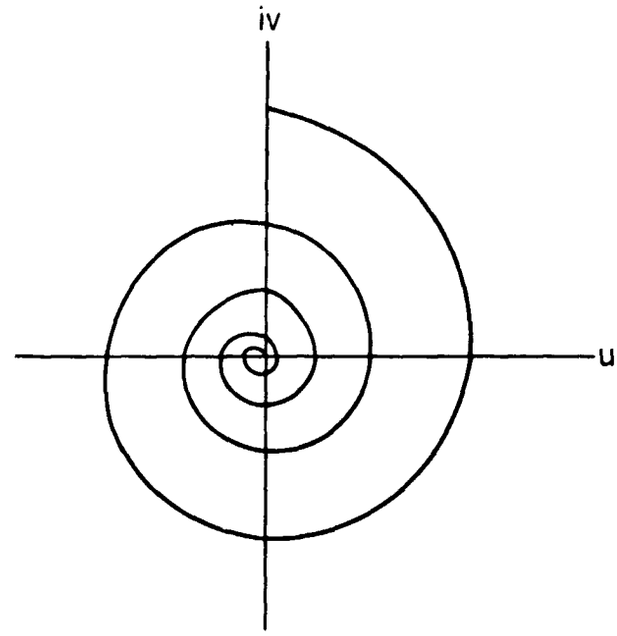


FIGURE 5.—Projection of $J_n(x+iy)$ on the u, iv plane as a function of n, n continuous.

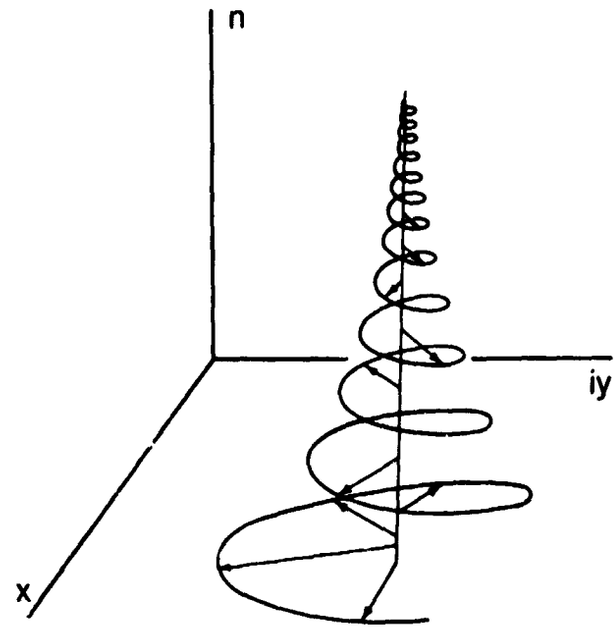


FIGURE 6.—Plot of $J_n(x+iy)$ in the x, iy, n space as function of n, n continuous.

origin as a limit. Figure 6 is again a plot of $J_n(x+iy)$ as a function of n but now it is the three dimensional picture in x, iy, and n.

If the spiral is moved around in the x, iy plane, the base of the spiral becomes larger as y increases, and becomes more spread out or flatter as x increases. If movement was directed along one of the 45 degree lines, the slope and the width would change at approximately the same rate. It should be noted that for movement along the 45 degree lines the plot would be of the Thomson Functions, $ber_n(z)$, $bei_n(z)$.

The direction of the spiraling is not significant, for it changes direction as the argument of z changes. The functions of arguments (x+iy) and (x−iy) would spiral into the origin in opposite directions.

If a three dimensional plot is made of $J_n(x+iy)$ in the x, iy plane, contour lines of constant magnitude would give the effect of a crater (Reference 5, page 127).

For argument (x+iy) where |y/x| is small, the modulus of $J_n(x+iy)$ oscillates until n becomes greater than z and at this point, it stays in the first quadrant of the u, iv plane, continually decreasing to the origin at $n = \infty$.

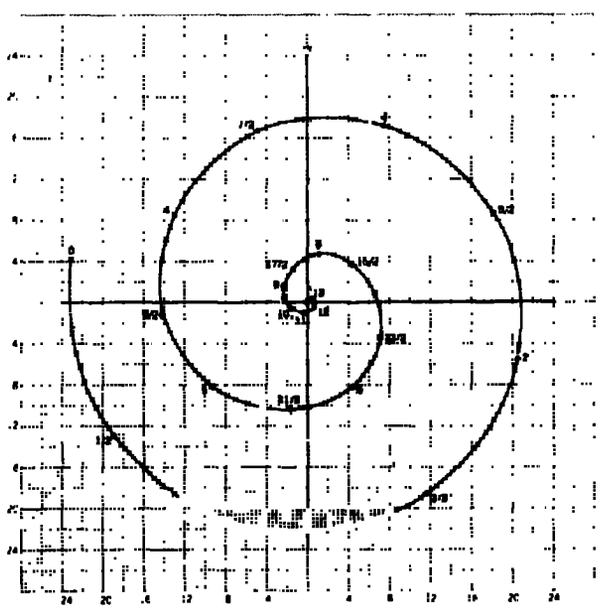


FIGURE 7.—Detailed plot of the projection of $J_n(10+i10)$ on the u,iv plane as a function of n.

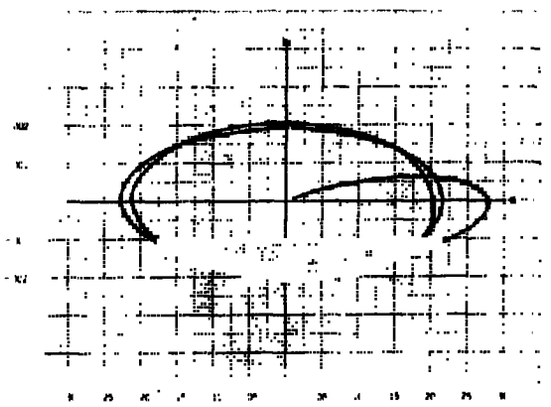


FIGURE 8.—Detailed plot of the projection of $J_n(15+6.01)$ on the u,iv plane vs. n.

z Imaginary

The regular Bessel function of imaginary argument $J_n(iy)$ behaves similarly to the function of complex argument. $|J_0(iy)|$ is of the order of magnitude $0.1e^y$ with

$$\lim_{n\to\infty} J_n(iy) = 0 ,$$

and the function spirals in toward the origin of the u, iv plane. The figures for imaginary argument would be similar to the complex argument with the integral orders lying on the axis and half orders displaced by an angle $\pi/4$, etc. In the u, iv plane they would appear as in Figure 9. Here again the direction of the spiral is determined by the argument of z, or $\pm\pi/2$. These functions

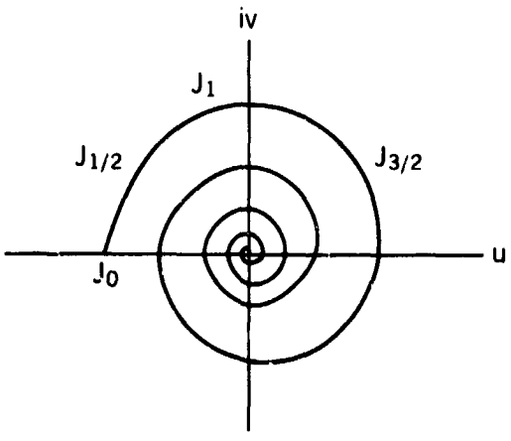


FIGURE 9.—Projection of $J_n(iy)$ on the u,iv plane.

are really solutions of the equation

$$y'' + \frac{1}{z} y' - \left(1 + \frac{n^2}{z^2}\right) y = 0 ,$$

which occurs frequently in problems of mathematical physics; however, it is usually desirable to express the solutions in real form.

The solutions $J_n(iy)$, $Y_n(iy)$ are not always real, as was shown for the regular function; however the function $e^{-in\pi/2}J_n(iy)$ is always real and is a solution of the equation. In effect the imaginary values are rotated thru an angle of $\pi/2$. These functions are the $I_n(y)$ functions and are defined by

$$I_n(y) = e^{-in\pi/2} J_n(iy) ;$$

they obey the recurrence formula

$$I_{n-1}(y) - I_{n+1}(y) = \frac{2n}{y} I_n(y)$$

the use of which was discussed in a previous section.

CONCLUDING REMARKS

It has been shown that the backwa.d recurrence is an accurate and especially easy method for computing the regular Bessel Function. Indirectly, this report came about through a need for and investigation of computing the 1/2 order Bessel functions for the intensity functions of light scattering by spherical particles. Since the calculations required the Bessel functions varying in the order, the recurrence was especially suitable. The backward recurrence method should be a very useful tool to those requiring Bessel functions of any order or argument:

The magnitude graphs (Appendix B, Figures B2, B3, B4) are significant in that they give an overall picture of the functions varying in both the order and the argument. This could be useful in many computations, e.g., an integral of the type

$$A = \int_a^b J_n(x)\,dx .$$

Insight might be obtained as to the range of x that would add to the value of the integral and so on.

REFERENCES

1. Goldstein, M., and Thaler, R. M., "Recurrence Techniques for the Calculation of Bessel Functions," *Math. Tables & Aids to Comput.* 13:102–108, 1959.
2. British Association for the Advancement of Science, "Bessel Functions. pt. II. Functions of Positive Integer Order," Vol. X of series *Mathematical Tables*, Cambridge, England: The University Press, 1931–48.
3. Watson, G. N., "A Treatise on the Theory of Bessel Functions," 2nd. ed., Cambridge, England: The University Press, 1944.
4. Mathematical Tables Project, National Bureau of Standards, "Tables of Spherical Bessel Functions," 2 Vols., New York: Columbia University Press, 1947.
5. Jahnke, E., and Emde, F., "Tables of Functions with Formulae and Curves," 4th ed., New York: Dover Publications, 1945.

ADDITIONAL BIBLIOGRAPHY

Harvard University, Computation Laboratory, "Tables of the Bessel Functions of the First Kind of Orders," *Ann. Comput, Lab. Harvard Univ.* Vol. 3–14, Cambridge, Mass.: Harvard University Press, 1947–49.

Nosova, L. N., "Tables of Thomson Functions and Their First Derivatives." Translated from Russian by Prasenjit Basu, New York: Pergamon Press, 1961.

Stegun, I. A., and Abramowitz, M., "Generation of Bessel Functions on High Speed Computers," *Math. Tables & Aids to Comput.* 11:255–257, 1957.

Appendix A

SOLUTIONS OF BESSEL'S DIFFERENTIAL EQUATION

If α is any real constant, Bessel's equation of order α is

$$y'' + \frac{1}{x} y' + \left(1 - \frac{\alpha^2}{x^2}\right) y = 0 .$$

This is a linear differential equation of second order, and therefore has two and only two independent solutions. If these two solutions are $y_1(x)$ and $y_2(x)$, then there cannot exist two non-zero constants c_1 and c_2 such that

$$c_1 y_1(x) + c_2 y_2(x) = 0 .$$

That is to say that any other solution, say $y_3(x)$, is a linear combination of the two independent solutions, or

$$y_3(x) = c_1 y_1(x) + c_2 y_2(x) .$$

Note that Bessel's equation has a singular point at $x=0$, but a solution can be obtained by the method of Frobenius. That is, assume a solution of the type

$$y = \sum_{k=0}^{\infty} C_k x^{k+\alpha}$$

and it turns out that the two solution types are completely dependent upon α. They are referred to by order α and argument x. If α is an integer, the two independent solutions are $J_\alpha(x)$ and $Y_\alpha(x)$, α = integer. If α is not an integer the two solutions are $J_\alpha(x)$ and $J_{-\alpha}(x)$, $\alpha \neq$ integer. If α is half an odd integer they are usually referred to as the *plus and minus half orders*. $J_\alpha(x)$ is the regular solution; and $J_{-\alpha}(x)$, α not an integer, and $Y_\alpha(x)$, α = integer, are the irregular solutions. The two solutions are also referred to as Bessel functions of the first and second kind respectively.

If the argument x is purely imaginary, the solution is referred to as

$$I_\alpha(x) \quad \text{and} \quad K_\alpha(x) .$$

$I_\alpha(x)$ is analogous to the regular Bessel function and $K_\alpha(x)$ to the irregular Bessel function.

$H_\alpha^{(1)}(x)$ and $H_\alpha^{(2)}(x)$ are known as Bessel functions of the third kind, or first and second Hankel Functions respectively. They are defined as

$$H_\alpha^{(1)}(x) = J_\alpha(x) + i Y_\alpha(x) . \qquad (A1)$$

$$H_\alpha^{(2)}(x) = J_\alpha(x) - i Y_\alpha(x) . \qquad (A2)$$

These are obtained from

$$F_\alpha(x) = A J_\alpha(x) + B Y_\alpha(x)$$

with $A=1$ and $B=\pm i$. If we add and subtract Equations A1 and A2, we arrive at other identities which are sometimes useful:

$$J_\alpha(x) = \frac{1}{2}\left(H_\alpha^{(1)}(x) + H_\alpha^{(2)}(x)\right) .$$

$$Y_\alpha(x) = \frac{-i}{2}\left(H_\alpha^{(1)}(x) - H_\alpha^{(2)}(x)\right) .$$

For complex argument $i^{3/2}x$, one obtains the Thomson function $ber_\alpha(x)$, $bei_\alpha(x)$:

$$J_\alpha\left(i^{3/2}x\right) = ber_\alpha(x) + i\, bei_\alpha(x) .$$

If $\alpha = 0$ the subscripts are usually omitted. Listed below are the two general types of differential equations and their associated functions as

solutions:

$$y'' + \frac{1}{x} y' + k^2 y = 0 \qquad \left\{ \begin{array}{l} y = AJ_0(kx) + BY_0(kx) \\ y = AH_0^{(1)}(kx) + BH_0^{(2)}(kx) \end{array} \right\};$$

$$y'' + \frac{1}{x} y' + \left(k^2 - \frac{\alpha^2}{x^2}\right) y = 0 \qquad \left\{ \begin{array}{l} y = AJ_\alpha(kx) + BY_\alpha(kx) \\ y = AJ_\alpha(kx) + BJ_{-\alpha}(kx) \\ y = AH_\alpha^{(1)}(kx) + BH_\alpha^{(2)}(kx) \end{array} \right\};$$

$$y'' + \frac{1}{x} y' - k^2 y = 0 \qquad \left\{ y = AI_0(kx) + BK_0(kx) \right\};$$

$$y'' + \frac{1}{x} y' - \left(k^2 - \frac{\alpha^2}{x^2}\right) y = 0 \qquad \left\{ \begin{array}{l} y = AI_\alpha(kx) + BK_\alpha(kx) \\ y = AI_\alpha(kx) + BI_{-\alpha}(kx) \end{array} \right\}.$$

Appendix B

THE MAGNITUDE GRAPHS

The Magnitude graphs are plots of constant $J_n(x)$ in the x, n plane. The plot of $J_n(x)$ is continuous in this plane both as a function of n and as a function of x. It is a rolling surface and slopes down into the surface of the graph as x and n become larger. Between the zeros, lines of constant magnitude can be drawn, of which there would be two of the same value that connect at some larger x and n value; however, it was felt these would be superfluous here and they were left out. Since the surface is continuous, peak values either plus or minus would fall between two zeros and the magnitude of the function would decrease as we go toward the zero. All lines below the 10^{-2} line represent the approximate n and x value for the zeros of the Bessel function with the sign of the function indicated between the zeros.

In looking at the magnitude graphs that follow, pick a particular case, say $J_n(x)$ varying in n for $x=40$; it can be seen that $J_0(40)$ is positive and there are six oscillations before the functions stay positive. It can be seen that

$$-0.3 \leq J_n(x) \leq 0.3 \quad , \qquad 0 \leq n \leq 46 \quad ,$$

$$0 < J_n(x) < 10^{-2} \quad , \qquad n > 46 \quad ,$$

$$0 \leq J_n(x) \leq 10^{-16} \quad , \qquad n > 79 \quad ,$$

and so on. A graph of $J_n(40)$ versus n would appear as in Figure B1.

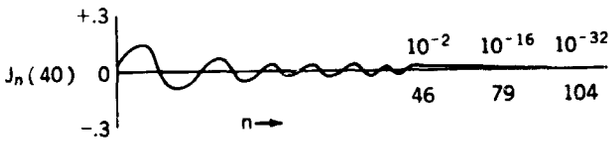


FIGURE B1.—Plot of $J_n(40)$ vs. n.

On the other hand, looking at $J_n(x)$ varying in x, say for $n=50$, the function increases from 0 to 10^{-32} for x ranging from 0 to approximately 8.8, and oscillations do not begin until x is approximately 55. It can be seen that

$$J_{50}(x) < 10^{-32} \quad , \qquad 0 \leq x < 8.8 \quad ,$$

$$10^{-32} \leq J_{50}(x) < 10^{-2} \quad , \qquad 8.8 \leq x \leq 44 \quad ,$$

$$-0.3 \leq J_{50}(x) \leq 0.3 \quad , \qquad x > 44 \quad .$$

Therefore, an idea of how the function behaves can be obtained for any order or argument within the limits of the graph.

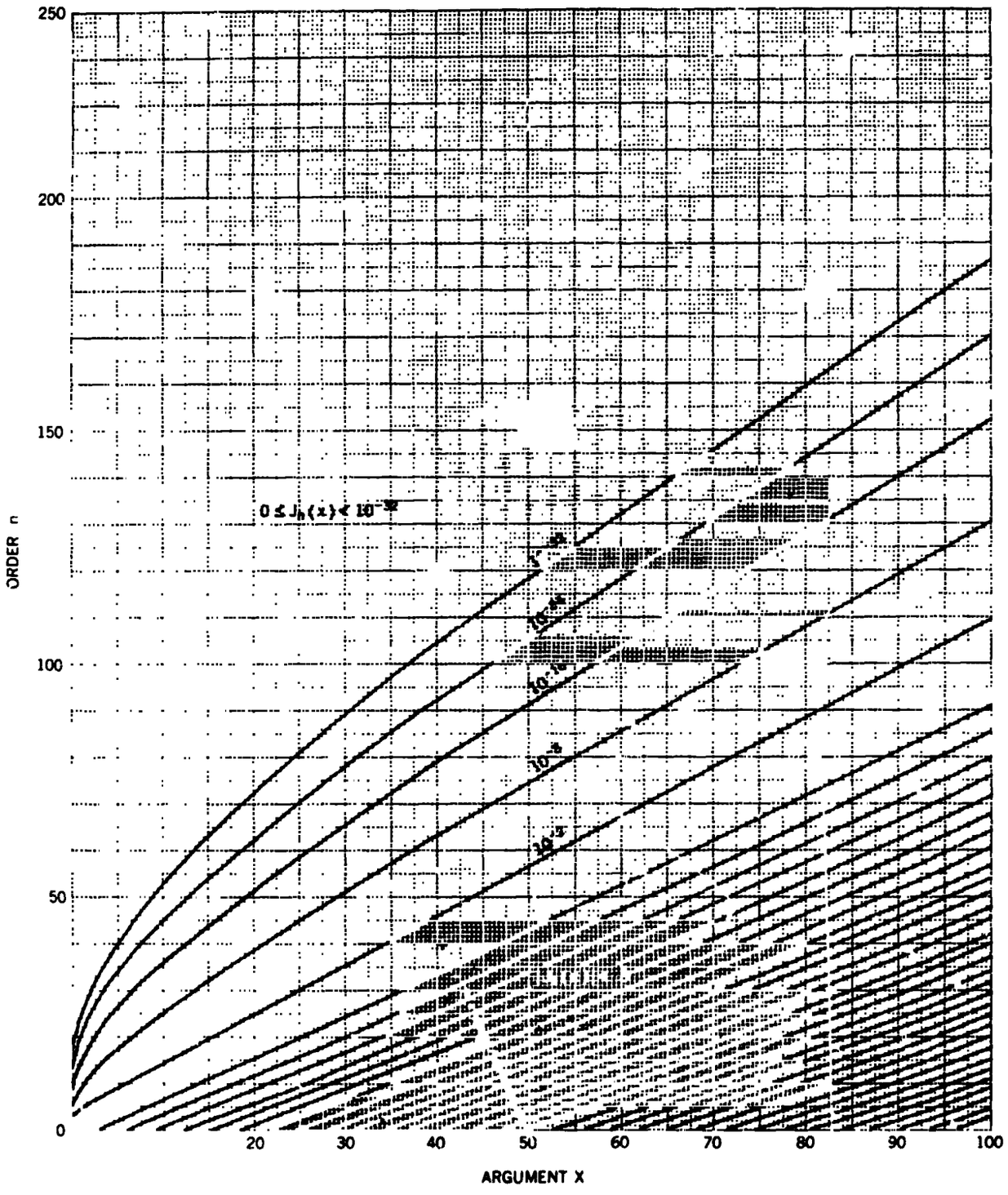

250
200
150
100
50
0
ORDER n
20
30
40
50
60
70
80
90
100
ARGUMENT X
$0 \leq J_n(x) < 10^{-38}$
10^{-8}
10^{-6}
10^{-4}
10^{-2}

FIGURE B2

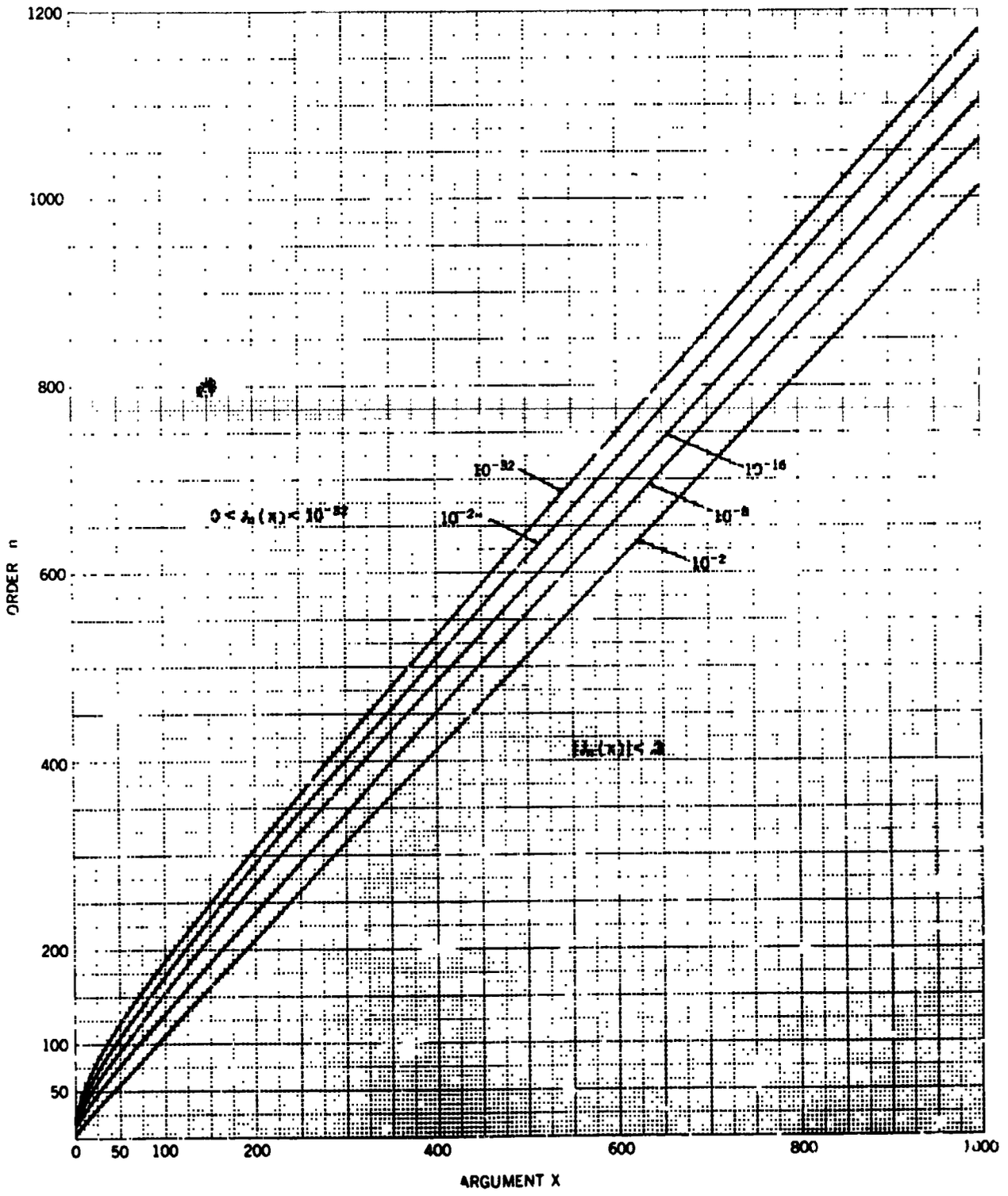

1200
1000
800
600
400
200
100
50
ORDER n
0
50
100
200
400
600
800
1000
ARGUMENT X
10^{-32}
10^{-16}
10^{-8}
10^{-2}
$0 < J_n(x) < 10^{-32}$

FIGURE B3

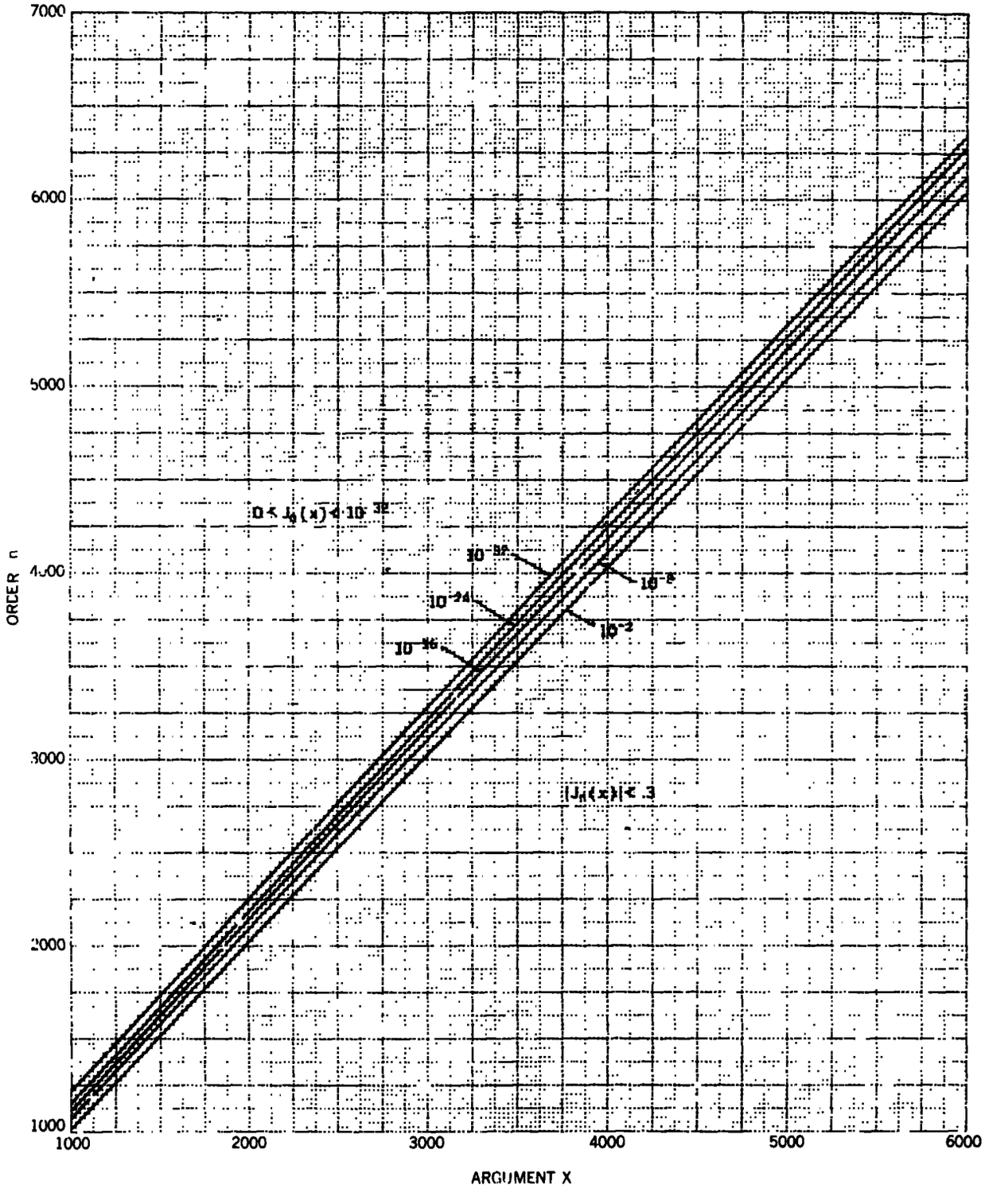

7000
6000
5000
4000
3000
2000
1000
ORDER n
1000
2000
3000
4000
5000
6000
ARGUMENT X
$0 \leq J_n(x) < 10^{-32}$
10^{-32}
10^{-8}
10^{-24}
10^{-2}
10^{-16}
$|J_n(x)| < .3$

FIGURE B4

N65 37971

THEORY OF THE 2s AND 2p EXCITATION OF THE HYDROGEN ATOM INDUCED BY ELECTRON IMPACT*

KAZEM OMIDVAR

A numerical calculation has been carried out to evaluate the 3×3 cross-section matrix involved in the electron impact excitation of the ground state of H atom to the 2s and 2p levels. The method of solution is that of atomic eigenstates expansion. In this paper, instead of the iterative technique used by other authors, the definite integral terms in the coupled radial differential equations are eliminated through some linear transformation of the radial functions, thus avoiding iteration of these equations. The accuracy of the numerical integration is tested by satisfying the equation of reciprocity and the equation of continuity of currents with an error-to-value ratio less than 1 per 1000 on the average; and the maximum of this ratio, except for a few cases, has been kept below 5 percent. The results are in agreement with the results of an iterative technique.

To evaluate the effect of the long-range and the centrifugal potential, a simple perturbation theory is developed. The six cross sections 1s→2s, 1s→2p, 1s→1s, 2s→2s, 2s→2p, and 2p→2p are tabulated. The 2p→2p cross section requires the solution of the sets of differential equations with different parities. With the validity of the eigenstates expansion assumed, it is found by comparison with the eigenstates expansion calculation that the Born approximation, despite its simplicity, gives meaningful results for low and close-to-the-threshold energies of the bombarding electrons. The effect of the exchange potentials on the cross sections is also investigated. Finally, an interesting structure of the 1s→2s excitation cross section above threshold is found.

INTRODUCTION

Calculation of the excitation cross sections in atomic hydrogen by electron impact corresponds to the solution of the problem of three interacting bodies: one proton, and two electrons. By taking the position of the proton as the center of mass, the problem will reduce to the task of finding the nonseparable wave function of the system of the two electrons with an attractive center of force. Such a solution has not been found. However, if this wave function is expanded in terms of the eigenstates of the hydrogen atom, the coefficients of the expansion, which are functions of the position vector of the free electron, can be found through numerical integration. When an infinite number of terms is included in the expansion, the solution to the problem is exact. Furthermore, the expansion has the advantage that the asymptotic form of its coefficients is automatically the asymptotic form of the free electron wave function scattered from different atomic states, which are simply related to the excitation cross sections.

In this paper, atomic states 1s, 2s, 2p are included in the expansion and, by antisymmetrizing the two electron wave functions according to the exclusion principle, some contribution from the continuum in the expansion is also taken into account. The first calculation of this type was performed by Marriot (Reference 1), whose expansion consisted of the 1s and 2s states in order to calculate the 1s→2s transition cross section. This calculation was extended by Smith (Reference 2) to higher total orbital angular momenta of the system. Percival and Seaton (Reference 3) have formulated the eigenstate expansion technique in general and have tabulated the coefficients of the integro-differential equations for s, p, and d atomic electrons. Burge, Smith, and Schey (References 4 and 5†), using the equations of Percival and Seaton for the three states 1s, 2s, 2p, have inte-

*Published as *NASA Technical Note D-2145*, June 1964. This work has also appeared in a condensed form in the *Physical Review*, 133, February 17, 1964.

†A similar calculation has been performed by R. Damburg and R. Peterkop; this will appear in the *USSR Journal of Experimental and Theoretical Physics*. A different method to calculate the 1s–2s electron impact transition cross section in hydrogen is being considered by L. Kyle and A. Temkin, adopting the nonadiabatic theory of electron scattering developed by A. Temkin (see References 21, 22); the calculation is in progress.

IN D-

grated the resulting integro-differential equations. In this paper we solve the same differential equations by a linear transformation of the differential equations in order to avoid the need for iteration of these equations (Reference 6).*

The numerical integrations were carried out for all partial waves, while in higher partial waves the Born approximation was used. The transition between the eigenstates expansion calculation and the Born approximation takes place when the results of the two calculations agree closely.

FORMULATION

Derivation of the Differential Equations

Since spin orbit interaction of the electrons is neglected, the total orbital angular momentum L and the total spin angular momentum S are separately conserved. We can then divide the interactions into antiparallel spin states, where $S=0$, and parallel spin states, where $S=1$. We deal with spatial wave functions of the electrons only, and for brevity we call the orbital angular momentum the *angular momentum.*

Neglecting the motion of the proton of the hydrogen atom and taking its position as the origin of the coordinate system, the Schroedinger equation for the system can be written

$$[H-E]\,\psi(\mathbf{r}_1, \mathbf{r}_2) = 0 \,. \tag{1}$$

where $\mathbf{r}_1$ and $\mathbf{r}_2$ are the position vectors of the bound and free electrons; and in atomic units

$$H - E = -\frac{1}{2}\nabla_1^2 - \frac{1}{2}\nabla_2^2 - \frac{1}{r_1} - \frac{1}{r_2} + \frac{1}{r_{12}} - E \,, \tag{2}$$

where E is the total energy of the system and r_{12} is the distance between the two electrons. We expand the total wave function $\psi(\mathbf{r}_1, \mathbf{r}_2)$ in terms of the eigenfunctions of the total angular momentum L,

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \sum_{L=0}^{\infty} \psi_L(\mathbf{r}_1, \mathbf{r}_2) \,. \tag{3}$$

Since these eigenfunctions are orthogonal and distinct, substitution of Equation 3 in Equation 1 gives

$$[H-E]\,\psi_L(\mathbf{r}_1, \mathbf{r}_2) = 0 \tag{4}$$

The explicit form of $\psi_L(\mathbf{r}_1, \mathbf{r}_2)$ is given by

$$\psi_L(\mathbf{r}_1, \mathbf{r}_2) = (1+\beta P_{12}) \sum_{n_1 \ell_1 \ell_2} \sum_{m_1 m_2} C^{\ell_1 \ell_2 L}_{m_1 m_2 M}\, \phi(n_1 \ell_1 m_1, \mathbf{r}_1)$$

$$\times\, r_2^{-1}\, u(k_{n_1} \ell_2, r_2)\, Y_{\ell_2 m_2}(\Omega_2) \,, \tag{5}$$

$$\phi(n_1 \ell_1 m_1, \mathbf{r}_1) = r_1^{-1} P(n_1 \ell_1, r_1)\, Y_{\ell_1 m_1}(\Omega_1) \,. \tag{6}$$

Here $\phi(n_1\, \ell_1\, m_1, \mathbf{r}_1)$ is the hydrogen atom wave function with radial part $r_1^{-1}P(n_1\, \ell_1, r_1)$ and angular part $Y_{\ell_1 m_1}(\Omega_1)$ and quantum numbers $n\ell_1\, m_1$; $r_2^{-1}u(k_{n_1}\, \ell_2, r_2)$ is the radial part and $Y_{\ell_2 m_2}(\Omega_2)$ is the angular part of the free electron wave function with quantum numbers $k_{n_1}\, \ell_2\, m_2$. The relation between the wave number k_{n_1} and n_1 is given by

$$k_{n_1}^2 = 2\left(E + \frac{1}{2n_1^2}\right) \,. \tag{7}$$

Finally the constants $C^{\ell_1 \ell_2 L}_{m_1 m_2 M} = (\ell_1\, \ell_2\, m_1\, m_2 | LM)$, with M representing the total magnetic quantum number, are vector coupling coefficients which make the linear combination of the products of the one electron wave function in Equation 5 the eigenfunction of L. In the problem under consideration, $n_1 = 1, 2$; $\ell_1 = 0, 1$; $\ell_2 = |L-\ell_1|, \cdots, |L+\ell_1|$; $m_1 = -\ell_1, \cdots, \ell_1$; and $m_2 = -\ell_2, \cdots, \ell_2$. To make the total wave function symmetric for antiparallel spins or antisymmetric for parallel spins, the operator P_{12} interchanges $\mathbf{r}_1$ and $\mathbf{r}_2$ while β is $+1$ for the first case and is -1 for the second.

*A similar calculation has been carried out in Reference 6. Here the L=0,1 cases have been solved by noniterative, and all other cases by iterative, methods.

By taking **L** perpendicular to the z-axis, $M = 0$ and $m_2 = m_1$. Equation 5 can then be written

$$\psi_L(\mathbf{r}_1, \mathbf{r}_2) = (1 + \beta P_{12}) \sum_{n_1 \ell_1 \ell_2} \sum_{m_1} C^{\ell_1 \ell_2 L}_{m_1 - m_1 0} \phi(n_1 \ell_1 m_1, \mathbf{r}_1) \times r_2^{-1} u(k_{n_1 \ell_2}, r_2) Y_{\ell_2 m_2}(\Omega_2) . \quad (8)$$

In order that $\psi(\mathbf{r}_1, \mathbf{r}_2)$ closely approximates the exact wave function, we minimize the expectation value of the energy operator with respect to the radial parts of the free electron wave functions,

$$\delta \int \psi_L^*(\mathbf{r}_1, \mathbf{r}_2) [H - E] \psi_L(\mathbf{r}_1, \mathbf{r}_2) d^3 r_1 d^3 r_2 = 0 . \quad (9)$$

It has been shown by Kohn (Reference 7) that the differences between the scattering amplitudes obtained from these equations and the exact scattering amplitudes are quadratic in the difference between $\psi_L(\mathbf{r}_1, \mathbf{r}_2)$ and the exact wave function. When the variation is carried out inside the integral, we obtain

$$\sum_{m_1} C^{\ell_1 \ell_2 L}_{m_1 - m_1 0} \int \phi^*(n_1 \ell_1 m_1, \mathbf{r}_1) Y^*_{\ell_2 m_2}(\Omega_2) [H - E] \psi_L(\mathbf{r}_1, \mathbf{r}_2) d^3 r_1 d\Omega_2 = 0 . \quad (10)$$

By means of Equations 2 and 8, the Schroedinger equation for the hydrogen atom,

$$\left[\nabla_1^2 + \frac{2}{r_1}\right] \phi(n_1 \ell_1 m_1, \mathbf{r}_1) = \frac{1}{n_1^2} \phi(n_1 \ell_1 m_1, \mathbf{r}_1) , \quad (11)$$

and Equation 7, Equation 10 reduces to

$$\sum_{m_1} \sum_{n_1' \ell_1' \ell_2'} \sum_{m_1'} C^{\ell_1 \ell_2 L}_{m_1 - m_1 0} C^{\ell_1' \ell_2' L}_{m_1' - m_1' 0} \int \phi^*(n_1 \ell_1 m_1, \mathbf{r}_1) Y^*_{\ell_2 m_2}(\Omega_2) (1 + \beta P_{12}) \times \left[\nabla^2_{r_2} - \frac{\ell_2'(\ell_2' + 1)}{r_2^2} + k^2_{n_1'} + 2\left(\frac{1}{r_2} - \frac{1}{r_{12}}\right)\right] \phi(n_1' \ell_1' m_1', \mathbf{r}_1) \times r_2^{-1} u(k_{n_1'} \ell_2', r_2) Y_{\ell_2' m_2'}(\Omega_2) d^3 r_1 d\Omega_2 = 0 , \quad (12)$$

where ∇^2_{r2} is the radial part of ∇_2^2. By orthogonality of the hydrogen atom and spherical harmonics wave functions, the relation (Reference 8)

$$\sum_{m_1} \left[C^{\ell_1 \ell_2 L}_{m_1 - m_1 0}\right]^2 = 1 , \quad (13)$$

the integration by parts of the exchange terms, and the relation*

$$C^{\ell_2 \ell_1 L}_{-m_1 m_1 0} = (-)^{L - \ell_1 - \ell_2} C^{\ell_1 \ell_2 L}_{m_1 - m_1 0} . \quad (14)$$

*Reference 8, Equation 3.16b.

Equation 12 leads to

$$\left[\nabla^2_{r_2} - \frac{l_2(l_2+1)}{r_2^2} + k_{n_1}^2 + \frac{2}{r_2}\right]\frac{u(k_{n_1} l_2, r_2)}{r_2} - 2\sum_{m_1}\sum_{n_1' l_1' l_2'}\sum_{m_1'} C^{l_1 l_2 L}_{m_1 - m_1' 0}\, C^{l_1' l_2' L}_{m_1 - m_1' 0}$$

$$\int \frac{\phi^*(n_1 l_1 m_1, r_1)\, Y^*_{l_2 m_2}(\Omega_2)}{r_{12}} \times \left[\phi(n_1' l_1' m_1', r_1)\frac{u(k_{n_1'} l_2', r_2)}{r_2} Y_{l_2' m_2'}(\Omega_2)\right.$$

$$\left. + \beta\phi(n_1' l_1' m_1', r_2) \times \frac{u(k_{n_1'} l_2', r_1)}{r_1} Y_{l_2' m_2'}(\Omega_1)\right] d^3 r_1\, d\Omega_2$$

$$+ \beta \sum_{n_1' l_1' l_2'} (-)^{L - l_1 - l_2}\, \delta(l_1' l_2', l_2 l_1)\left(\frac{1}{n_1^2} + k_{n_1'}^2\right)$$

$$\int_0^\infty P(n_1 l_1, r_1)\, r_2^{-1} P(n_1' l_1', r_2) \times u(k_{n_1'} l_2', r_1)\, dr_1 = 0. \qquad (15)$$

If $1/r_{12}$ is expanded in terms of the Legendre polynomials and use is made of the addition theorem,* we obtain

$$\frac{1}{r_{12}} = \sum_{\lambda=0}^{\infty} \frac{r_<^\lambda}{r_>^{\lambda+1}} P_\lambda(\cos\theta_{12})$$

$$= \sum_{\lambda=0}^{\infty}\sum_{\mu=-\lambda}^{+\lambda} \frac{4\pi}{2\lambda+1}\frac{r_<^\lambda}{r_>^{\lambda+1}} Y_{\lambda\mu}(\Omega_1)\, Y^*_{\lambda\mu}(\Omega_2). \qquad (16)$$

In this expression θ_{12} is the angle between the position vectors $\mathbf{r}_1$ and $\mathbf{r}_2$ at the origin, and $r<$ is the smaller and $r>$ is the larger of $|\mathbf{r}_1|$ and $|\mathbf{r}_2|$. We also introduce

$$y_\lambda(n l\, n' l', r_2) = r_2^{-(\lambda+1)} \int_0^{r_2} P(n l, r_1)\, P(n' l', r_1)\, r_1^\lambda\, dr_1$$

$$+ r_2^\lambda \int_{r_2}^\infty P(n l, r_1)\, P(n' l', r_1)\, r_1^{-(\lambda+1)}\, dr_1. \qquad (17)$$

Then it follows that

$$\int_0^\infty \frac{P(n l, r_1)\, P(n' l', r_1)}{r_{12}}\, dr_1 = \sum_{\lambda=0}^{\infty}\sum_{m=-\lambda}^{+\lambda} Y_{\lambda\mu}(\Omega_1)\, Y^*_{\lambda\mu}(\Omega_2)\, y_\lambda(n l\, n' l', r_2). \qquad (18)$$

*Reference 8, Equation 4.28.

By means of Equation 18, the relation*

$$\int Y^*_{\ell_3 m_3} Y_{\ell_2 m_2} Y_{\ell_1 m_1} d\Omega = \left[\frac{(2\ell_1+1)(2\ell_2+1)}{4\pi(2\ell_3+1)}\right]^{1/2} C^{\ell_1\ell_2\ell_3}_{m_1 m_2 m_3} C^{\ell_1\ell_2\ell_3}_{000} , \tag{19}$$

and the definition

$$\left(n\ell | k_{n_1}\ell'\right) = \int_0^\infty P(n\ell, r)\, u\left(k_{n_1}\ell', r\right) dr , \tag{20}$$

Equation 15 when multiplied by r_2 gives

$$\left[\frac{d^2}{dr_2^2} - \frac{\ell_2(\ell_2+1)}{r_2^2} + k_{n_1}^2 + \frac{2}{r_2}\right] u\left(k_{n_1}\ell_2, r_2\right)$$

$$- 2\left(\frac{2\ell_2+1}{2\ell_1+1}\right)^{1/2} \sum_{n_1'\ell_1'\ell_2'} \sum_{m_1 m_1'} \sum_{\lambda=0}^{\infty} \sum_{\mu=-\lambda}^{+\lambda} C^{\ell_1\ell_2 L}_{m_1 - m_1 0} C^{\ell_1'\ell_2' L}_{m_1' - m_1' 0}$$

$$\times \left\{ \left(\frac{2\ell_1'+1}{2\ell_2'+1}\right)^{1/2} C^{\lambda\ell_1'\ell_1}_{\mu m_1' m_1} C^{\lambda\ell_1'\ell_1}_{000} C^{\lambda\ell_2\ell_2'}_{\mu m_2 m_2'} C^{\lambda\ell_2\ell_2'}_{000} g_\lambda\left(n_1\ell_1 n_1'\ell_1', r_2\right) u\left(k_{n_1'}\ell_2', r_2\right) \right.$$

$$\left. + \beta\left(\frac{2\ell_2'+1}{2\ell_1'+1}\right)^{1/2} C^{\lambda\ell_2'\ell_1}_{\mu m_2 m_1} C^{\lambda\ell_2'\ell_1}_{000} C^{\lambda\ell_2\ell_1'}_{\mu m_2 m_1'} C^{\lambda\ell_2\ell_1'}_{000} P\left(n_1'\ell_1', r_2\right) g_\lambda\left(n_1\ell_1 k_{n_1'}\ell_2', r_2\right) \right\}$$

$$+ \beta \sum_{n_1'\ell_1'\ell_2'} (-)^{L-\ell_1-\ell_2} \delta\left(\ell_1'\ell_2', \ell_2\ell_1\right)\left(\frac{1}{n_1^2} + k_{n_1'}^2\right) P\left(n_1'\ell_1', r_2\right)\left(n_1\ell_1 | k_{n_1'}\ell_2'\right) = 0 . \tag{21}$$

In the exchange integrals above we have defined $P(k_{n_1}\ell_2, r)$ as $u(k_{n_1}\ell_2, r)$.

The vectorial equations

$$\mathbf{L} = \boldsymbol{\ell}_1 + \boldsymbol{\ell}_2 , \tag{22}$$

where $\mathbf{L}$ is constant but $\boldsymbol{\ell}_1$ and $\boldsymbol{\ell}_2$ take the values given before, can be divided into two groups, one with $L-\ell_1-\ell_2$ even and the other with $L-\ell_1-\ell_2$ odd. Since the total spatial wave function has the parity of $\ell_1+\ell_2$, in the first group the wave function has the parity of L and in the second a parity opposite to L. By conservation of parity we have two distinct groups of interactions. In this problem, where 1s, 2s, and 2p states of atomic hydrogen are taken into account, it is easy to see that, when $L-\ell_1-\ell_2$ is even, the set of quantum numbers $k_{n_1}\ell_2$ has four values: one for each of the 1s and 2s states, and two for the 2p state. When $L-\ell_1-\ell_2$ is odd, $k_{n_1}\ell_2$ has one value which corresponds to the elastic scattering of electrons by the 2p state of the hydrogen atom. Equation 21 is evaluated for these cases, and the resulting differential equations are listed in Appendix A. In evaluating Equation 21, it should be noted that the $C^{\ell_1\ell_2\ell_3}_{m_1 m_2 m_3}$ coefficients are subject to the condition that $\ell_1\ \ell_2\ \ell_3$ form a closed triangle and $m_3 = m_1 + m_2$.† This limits the summation over λ and μ considerably to few terms only. Summation over m_1, m_1', λ, and μ is carried out using the numerical values of the C coefficients given by Condon and Shortley (Reference 9).

*Reference 8, Equation 4.34.

†Reference 8, Equation 3.14.

Percival and Seaton (Reference 3) have derived the same differential equations for the scattering of free electrons by atomic s, p, and d electrons in the hydrogen atom using the theory of irreducible tensor operators to evaluate the interaction terms between the two electrons in the differential equations. The calculation becomes considerably simpler in this way. The results of the two methods are identical.

In the rest of the paper, except the section on page 14, we discuss the solutions to the four coupled differential equations given in Appendix A and which arise when $L-\ell_1-\ell_2$ is even. The single differential equations for $L-l_1-l_2$ odd are derived in the excepted section (page 14). Its numerical integration can be treated as a special case of the four coupled differential equations.

When the integrals representing the direct potentials in the four differential equations are evaluated and some change is made in the limits of the exchange potential integrals, these equations can be written in the following matrix form:

$$\left[\frac{d^2}{dr^2} + k_n^2 - \frac{l_n(l_n+1)}{r^2}\right] u(k_n l_n, r) = 2V u(k_n l_n, r) . \tag{23}$$

The four components of **u** are the four radial functions of the free electron. **V** is a 4x4 symmetric matrix that is the sum of three matrices,

$$\left.\begin{aligned} V_{ij} &= D_{ij} + E_{ij} , \\ E_{ij} &= F_{ij} + \sum_{\nu=1}^{\sigma} g_{ij}^{\nu} \int_0^{\infty} h_{ij}^{\nu}\, dr , \end{aligned}\right\} \tag{24}$$

where D_{ij} is the direct potential, E_{ij} is the exchange potential, and where both are functions of r. The matrix E_{ij} contains in addition integrals with respect to r, and for the purpose of numerical integration it can be written as the sum of two matrices. The explicit forms of D_{ij}, F_{ij}, g_{ij}^{ν}, and h_{ij}^{ν} are given in Appendix B. The value of σ is 2 for i=j=3 and i=j=4, and is 1 for all other values of i and j. It is understood that for the exchange terms the components of u on the right-hand side of Equation 23 are inside the integrals of the exchange terms.

Derivation of the Transmission Matrix from Solutions of the Differential Equations

The method is similar to that used by Bransden and McKee (Reference 10), and by Marriot (Reference 1). Equation 23 constitutes a set of four coupled second-order differential equations. Three components of **u** can be eliminated from these equations, resulting in an 8th order differential equation for the remaining component. Therefore there are eight sets of solutions to Equation 23. However, only half of these solutions are regular at the origin. Each of the four regular solutions corresponds to a definite vector **u**. The four vectors can properly be represented by a 4x4 matrix u_{nj}, n, j=1, 2, 3, 4, where n corresponds to the particular component and j corresponds to the particular solution of **u**. The four solutions are carried out numerically in the next section.

From the explicit form of **V** it can be seen that **V** vanishes at infinity. The asymptotic solution of **u** as given by Equation 23 is therefore

$$u_{nj}(r) \approx a_{nj} \sin\left(k_n r - \frac{l_n \pi}{2} + \delta_{nj}\right) \quad (n, j = 1, 2, 3, 4) , \tag{25}$$

where a_{nj} is the amplitude and δ_{nj} is the phase shift of the j^{th} solution of the n^{th} component of **u**.

Corresponding to the four components of **u**, there are 4 channels open to the reaction. If the incident wave is in the m^{th} channel (m=1, 2, 3, 4), the traveling wave in the n^{th} channel will be given by

$$u_n(r) \approx \exp\left[-i\left(k_n r - \tfrac{1}{2} l_n \pi\right)\right] \delta(m, n) - S_{mn} \exp\left[i\left(k_n r - \tfrac{1}{2} l_n \pi\right)\right] \quad (n = 1, 2, 3, 4) . \tag{26}$$

The constants S_{mn} are the amplitudes of the scattered waves. Since Equation 26 is also the asymptotic solution of Equation 23, they must be equal to linear combinations of Equation 25. If we call the coefficients of the linear combinations P_j, we must have

$$\sum_{j=1}^{4} P_j a_{nj} \sin\left(k_n r - \frac{l_n \pi}{2} + \delta_{nj}\right) =$$

$$(k_n)^{-1/2}\left\{\exp\left[-i\left(k_n r - \frac{1}{2} l_n \pi\right)\right]\delta(m, n) - S_{mn}\exp\left[i\left(k_n r - \frac{1}{2} l_n \pi\right)\right]\right\}$$

$$(n, m = 1, 2, 3, 4) . \quad (27)$$

On the right-hand side, we have used the normalization of Blatt and Weisskopf (Reference 11). If we equate the coefficients of $\exp[-i(k_n r - 1/2 l_n \pi)]$ and $\exp[i(k_n r - 1/2 l_n \pi)]$ in Equation 27, we obtain

$$\left.\begin{aligned} \sum_{j=1}^{4} P_j a_{nj} \exp - [i\delta_{nj}] &= \frac{-2i}{\sqrt{k_n}}\delta(m, n) . \\ \sum_{j=1}^{4} P_j a_{nj} \exp [i\delta_{nj}] &= \frac{-2i}{\sqrt{k_n}} S_{mn} . \end{aligned}\right\} \quad (28)$$

Separation of Equations 28 into real and imaginary parts gives

$$\left.\begin{aligned} \sum_{j=1}^{4}\left[(\Re P_j)\sin\delta_{nj} - (\Im P_j)\cos\delta_{nj}\right] a_{nj} &= \frac{2}{\sqrt{k_n}}\delta(m, n) , \\ \sum_{j=1}^{4}\left[(\Re P_j)\cos\delta_{nj} + (\Im P_j)\sin\delta_{nj}\right] a_{nj} &= 0 , \\ \sum_{j=1}^{4}\left[(\Re P_j)\sin\delta_{nj} + (\Im P_j)\cos\delta_{nj}\right] a_{nj} &= \frac{-2}{\sqrt{k_n}}\Re S_{mn} , \\ \sum_{j=1}^{4}\left[(\Re P_j)\cos\delta_{nj} - (\Im P_j)\sin\delta_{nj}\right] a_{nj} &= \frac{2}{\sqrt{k_n}}\Im S_{mn} . \end{aligned}\right\} \quad (29)$$

In the above R or I represent the real or the imaginary part of the quantity that follows them. Equations 29 are a set of 16 linear equations for 16 unknowns RP_j, IP_j, RS_{mn} and IS_{mn}. Once these unknowns are found,* the magnitude of S_{mn} will be given by

$$|S_{mn}|^2 = (RS_{mn})^2 + (IS_{mn})^2 . \quad (30)$$

The cross section is obtained by asymptotic expansion in spherical harmonics of the incident plane wave†

$$\exp[ikz] \approx \frac{\pi^{1/2}}{kr} \sum_{l=0}^{\infty} (2l+1)^{1/2} i^{l+1} \left\{ \exp\left[-i\left(kr - \frac{1}{2} l\pi\right)\right] - \exp\left[i\left(kr - \frac{1}{2} l\pi\right)\right] \right\} Y_{l,0} . \quad (31)$$

The magnitude of the ingoing wave on the right-hand side of Equation 27 for $n=m$ is $[k_m/\pi(2l_m+1)]^{1/2}$ times the magnitude of the partial wave of the expansion of $r \exp[ik_m z]$. The plane wave has a flux of v which, in atomic units, is equal to k. The ingoing flux of the right-hand side of Equation 27 is therefore $k_m^2/[\pi(2l_m+1)]$. The outgoing flux in the channel $n \neq m$ is $|S_{mn}|^2$. The cross section is obtained when we average the ratio of the outgoing flux to the ingoing flux over the initial states and sum over the final states. For a particular spin state of the two electrons, unpolarized electron beam, and unoriented atoms, the multiplicity of the initial states is $(2l_1+1)(2l_2+1)$, where l_1 and l_2 are the angular momentum of the bound and free electrons. For a polarized beam, $m_2 = 0$, where m_2 is the magnetic quantum number of the free electron. Then $m_1 = M$, where m_1 and M are the bound electron and the total magnetic quantum numbers. Since M is constant, there is only one initial state for a polarized beam. The multiplicity of the final states is $2L+1$, where L is the total orbital angular momentum. Since $\ell_m = \ell_2$, the cross section for $m \neq n$ is

$$Q_{mn} = \frac{\pi(2L+1)}{k_m^2(2l_1+1)} |S_{mn}|^2 \quad (m \neq n) . \quad (32)$$

The outgoing partial wave in the incident channel m consists of the scattered wave plus the outgoing wave given in the expansion of the plane wave. Then, according to Equation 27 for $n=m$, the magnitude of the amplitude of the scattered wave is $|1-S_{mn}|$. The elastic scattering cross section is therefore given by

$$Q_{mm} = \frac{\pi(2L+1)}{k_m^2(2l_1+1)} |1-S_{mm}|^2 . \quad (33)$$

If we define a matrix T by the relation

$$T = 1 - S , \quad (34)$$

Equations 32 and 33 can then be combined into a single equation,

$$Q_{mn} = \frac{\pi(2L+1)}{k_m^2(2l_1+1)} |T_{mn}|^2 . \quad (35)$$

T_{mn} is the transmitted amplitude in the n^{th} channel due to an incident wave in the m^{th} channel. The elements of T_{mn} constitute the transmission matrix.

The matrix S has two properties that are useful as tests on the accuracy of numerical integration. From Equation 26 it can be seen that S transforms the ingoing wave into the outgoing waves. The continuity of the electronic current requires that S be a unitary matrix.

$$\sum_{n=1}^{4} |S_{mn}|^2 = 1 \quad (m = 1, 2, 3, 4) . \quad (36)$$

Furthermore, since the Hamiltonian is Hermetian,

*Equations 29 with their present form and without further simplifications are solvable by the computer.

†Reference 11, Ch. VIII, Equation 2.7.

S must be symmetric (Reference 11):

$$S_{mn} = S_{nm} . \tag{37}$$

Equations 36 and 37 are used as tests on the accuracy of numerical integration.

A Useful Relation

A relation based on the symmetry of the interaction potentials, which serves as another test on the accuracy of the solutions, can be derived. The l^{th} and the k^{th} solutions of the i^{th} component of $\mathbf{u}$ by Equation 23 are given by

$$\left.\begin{aligned}\left[\frac{d^2}{dr^2} + k_i^2 - l_i \frac{(l_i+1)}{r^2}\right] u_{il} &= \sum_j V_{ij} u_{jl} ,\\ \left[\frac{d^2}{dr^2} + k_i^2 - l_i \frac{(l_i+1)}{r^2}\right] u_{ik} &= \sum_j V_{ij} u_{jk} .\end{aligned}\right\} \tag{38}$$

Multiplying the first by u_{ik} and the second by u_{il}, subtracting the two expressions, and summing over i gives

$$\sum_i \left[u_{ik} \frac{d^2}{dr^2} u_{il} - u_{il} \frac{d^2}{dr^2} u_{ik}\right] = \sum_{i,j} V_{ij} \left[u_{ik} u_{jl} - u_{il} u_{jk}\right] . \tag{39}$$

Since $V_{ij} = V_{ji}$, the interchange of the summation indices changes the sign on the right-hand side of the equation; the right-hand side must therefore be zero. Integrating the left-hand side from zero to infinity, we obtain

$$\sum_i \int_0^\infty \left[u_{ik} \frac{d^2}{dr^2} u_{il} - u_{il} \frac{d^2}{dr^2} u_{ik}\right] dr = 0 . \tag{40}$$

Integrating the above equation by parts, and applying Equation 25, we obtain

$$\sum_{i=1}^{4} k_i a_{ik} a_{il} \sin(\delta_{ik} - \delta_{il}) = 0 \qquad (k, l = 1, 2, 3, 4, k \neq l) . \tag{41}$$

Although the terms containing the exchange potentials do not cancel out on the right-hand side of Equation 39, the cancellation does take place after the integration is carried out in Equation 40.

Transmission Matrix According to Born Approximation

The Born approximation consists of neglecting the exchange potential terms appearing in the V matrix of Equation 23, and also of neglecting all the direct potential terms in this matrix except those terms that connect the incident channel to all other channels (Reference 12). Equation 23, when the incident wave is in the m^{th} channel, reduces to

$$\left[\frac{d^2}{dr^2} + k_n^2 - \frac{l_n (l_n+1)}{r^2}\right] u_n = 2 D_{nm} u_m \qquad (n = 1, 2, 3, 4) ; \tag{42}$$

u_m and u_n are given asymptotically by

$$u_m \approx k_m^{-1/2} \sin\left(k_m r - l_m \frac{\pi}{2}\right) , \tag{43}$$

$$u_n \approx k_n^{-1/2} B_{nm} \cos\left(k_n r - l_n \frac{\pi}{2}\right) . \tag{44}$$

We have chosen the constants of proportionality of u_m and u_n such that B_{nm} is the Born approximation of the reactance matrix R (Reference 13).* Equation 43 shows that u_m must have the following

*Also Reference 11, Ch. V, Sec. 4.

form (Reference 14):

$$u_m = k_m^{1/2} r j_{l_m}(k_m r) , \qquad (45)$$

where $jl_m(k_m r)$ are spherical Bessel functions. Furthermore, if y_n represents the homogenous solution of Equation 42, it must have the following forms:

$$y_n = a_n k_n r j_{l_n}(k_n r) , \qquad (46)$$

$$y_n \approx a_n \sin\left(k_n r - l_n \frac{\pi}{2}\right) , \qquad (47)$$

with a_n some unknown constant. Multiplying Equation 42 on the left by y_n and integrating the result from zero to infinity, we obtain by partial integration

$$2\int_0^\infty y_n D_{nm} u_m \, dr = \int_0^\infty y_n \left[\frac{d^2}{dr^2} + k_n^2 - \frac{l_n(l_n+1)}{r^2}\right] u_n \, dr$$

$$= \left[y_n \frac{d}{dr} u_n - u_n \frac{d}{dr} y_n\right]_0^\infty$$

$$= -k_n^{1/2} a_n B_{nm} .$$

The last equality has been obtained by noticing that y_n and u_n vanish at the origin, and by using their asymptotic forms as given by Equations 44 and 47. We therefore have

$$B_{nm} = -2(k_n k_m)^{1/2} \int_0^\infty j_{l_n}(k_n r) D_{nm} j_{l_m}(k_m r) r^2 \, dr . \qquad (48)$$

This is identical to the expression given for B by Seaton.*

The transmission and the reactance matrices are related by $\mathbf{T} = -2i\mathbf{R}/(1-i\mathbf{R})$. Since in the Born approximation $\mathbf{R} = \mathbf{B} \ll 1$, the transmission matrix according to the Born approximation is given by

$$T_{nm}^B = 4i(k_n k_m)^{1/2} \int_0^\infty j_{l_n}(k_n r) D_{nm} j_{l_m}(k_m r) r^2 \, dr . \qquad (49)$$

Substitution of Equation 49 into 35 would give the cross section according to the Born approximation. It should be noted that the symmetry of T insures Equation 37 to be satisfied while Equation 36 is no longer satisfied.

Elastic Scattering of Electrons by the 2p States of the Hydrogen Atom

The angular momentum of the free electron l_2 in the 2p channel has the values L−1, L, L+1, where L is the total angular momentum of the system. The first and the last values were considered in previous sections. The case $l_2 = L$ corresponds to a wave function in the 2p channel with a parity different from all channel wave functions considered previously. It therefore corresponds to elastic scattering. The wave function

*Reference 13, Equation 3.10.

in this case is given by

$$\psi_L(r_1, r_2) = (1 + \beta P_{12}) \sum_{m_1=-1}^{+1} C^{1LL}_{m_1 - m_1 0}\, \phi_{2p m_1}(r_1)\, \frac{u(k_2 L, r_2)}{r_2}\, Y_{L-m_1}(\Omega_2) \quad (50)$$

When Equation 9 is formed with this wave function, and minimized with respect to $u(k_2L, r_2)$, treatment which led to the derivation of the four differential equations will give the following differential equation:

$$\left[\frac{d^2}{dr^2} + k_2^2 - \frac{L(L+1)}{r^2} + \frac{2}{r}\right] u_L(r)$$

$$- \beta\left(\frac{1}{4} + k_2^2\right) \delta(L, 1)\, r\, R_{21}(r)\, \left(2p | k_2 L\right)$$

$$+ 2\left[y_0(2p\,2p, r) - \frac{1}{5} y_2(2p\,2p, r)\right] u_L(r)$$

$$+ 2\beta r R_{21}(r)\left[-\frac{3y_{L-1}(2p k_2 L, r)}{(2L+1)(2L-1)} + \frac{3y_{L+1}(2p k_2 L, r)}{(2L+1)(2L+3)}\right] = 0 . \quad (51)$$

The asymptotic solution of this equation is given by

$$u_L \approx a_L \sin(k_2 r - L\pi/2 + \delta_L) . \quad (52)$$

If the scattering amplitude is designated by T_{55}, it can be shown from the section on page 8 that for a particular L

$$T_{55} = 1 - \exp 2i\delta = -2i \exp(i\delta) \sin\delta , \quad (53)$$

where for simplicity we have suppressed the subscript L. The cross section, according to Equation 35, is given by

$$Q_{55} = \frac{4\pi(2L+1)}{3k_2^2} \sin^2\delta . \quad (54)$$

The total elastic scattering cross section by the 2p states is the sum of this cross section and the cross section corresponding to $l_2 = L-1$ and $l_2 = L+1$ (which were considered previously).

The Born amplitude, Equation 48, in this case is given by

$$B_{55} = -2k_2 \int_0^\infty j_L(k_2 r)\, D_{55}\, j_L(k_2 r)\, r^2\, dr , \quad (55)$$

where, by Equation 51,

$$D_{55} = -\frac{1}{r} + y_0(2p\,2p, r) - \frac{1}{5} y_2(2p\,2p, r) . \quad (56)$$

NUMERICAL INTERGRATION

Decomposition of the Differential Equations

If it were not for the definite integrals appearing in the potential matrix **V**, the set of the four coupled differential equations (23) could be integrated by any standard technique. The presence of these unknown constants whose integrand involves the unknown functions makes it necessary to solve these equations by iteration or by transformation of **u** into other vectors, whose differential equations do not contain definite integrals. Since the terms containing definite integrals are small as compared with the direct potentials, the iteration method can be used by assuming that the values of these integrals are zero. The differential equations are then integrated, the values of the definite integrals that are subsequently obtained are substituted in the differential equations, and the integration is repeated. The process is repeated

until sufficiently consistent values of these integrals are obtained. This method is useful if the convergences of the constants are fast enough, and the cross section is not very sensitive to the values of these constants.

In the second method, the transformation of **u** fixes the values of the constants and thus avoids iteration, whereby the computation is reduced considerably. The description of the method will be given here (Reference 15).*

By making use of Equations 24, Equation 23 can be written

$$\left[\frac{d^2}{dr^2} + k_i^2 - \frac{l_i(l_i+1)}{r^2}\right] u_i = 2\sum_{j=1}^{4}\left[(D_{ij}+F_{ij})\,u_j + \sum_{\mu=1}^{\sigma} g_{ij}^{\mu} C_{ij}^{\mu}\right], \tag{57}$$

where

$$C_{ij}^{\mu} = \int_0^{\infty} h_{ij}^{\mu}(r)\,u_j(r)\,dr . \tag{58}$$

We introduce the functions v_i and u_i^{kl} that are solutions of the following differential equations:

$$\left[\frac{d^2}{dr^2} + k_i^2 - \frac{l_i(l_i+1)}{r^2}\right] v_i = 2\sum_{j=1}^{4}\left[D_{ij}+F_{ij}\right] v_j , \tag{59}$$

$$\left[\frac{d^2}{dr^2} + k_i^2 - \frac{l_i(l_i+1)}{r^2}\right] u_i^{kl} = 2\sum_{j=1}^{4}\left[D_{ij}+F_{ij}\right] u_j^{kl} + 2\delta(i,k)\,g_{kl}^{\nu} . \tag{60}$$

Then u_i is given by the following expression:

$$u_i = v_i + \sum_{k=1}^{4}\sum_{l=1}^{4}\sum_{\nu=1}^{\sigma} C_{kl}^{\nu}\,u_i^{kl} . \tag{61}$$

Equation 61 can be verified by multiplying Equation 60 by C_{kl}^{ν}; summing over k, l, and ν; and adding to Equation 59—whereupon Equation 57 results. Substitution of Equation 61 in Equation 58 gives.

$$\sum_{k=1}^{4}\sum_{l=1}^{4}\sum_{\nu=1}^{\sigma}\left[\delta(ij\mu, kl\nu) - A_{ij}^{\mu kl}\right] C_{kl}^{\nu} = B_{ij}^{\mu}$$

$$(i, j = 1, 2, 3, 4;\ \mu = 1, 2 \text{ for } i = j = 3 \text{ and } i = j = 4;\ \mu = 1 \text{ otherwise}) , \tag{62}$$

where $A_{ij}^{\mu kl}$ and B_{ij}^{μ} are defined by

$$\left.\begin{aligned} A_{ij}^{\mu kl} &= \int_0^{\infty} h_{ij}^{\mu}\,u_j^{kl}\,dr , \\ B_{ij}^{\mu} &= \int_0^{\infty} h_{ij}^{\mu}\,v_j\,dr . \end{aligned}\right\} \tag{63}$$

The numerical integration is carried out by integrating Equations 59 and 60 by any standard method, calculating $A_{ij}^{\mu kl}$ and B_{ij}^{μ} by Equations 63

*Also, Reference 1. This description differs from the description of Reference 15 and the present paper. In Reference 1, v_i, in Equation 61 is set to 0; this makes $B_{ij}^{\mu}=0$. Equations 62 then reduce to a set of homogeneous equations whose determinant must be 0. Since the amplitude of any of the four components of **u** can be left arbitrary, one of the C_{kl} is set equal to 1 and the rest of the constants are found subsequently.

and, finally, solving the system of 18 algebraic equations given by Equation 62 to find C_{kl}^{ν}. With the known values of these constants the integration of Equation 57 is straightforward.

The determinant of Equation 62 becomes singular for L = 0 and 1. This is shown in Appendix C. To remove the singularity, some of the C_{kl}^{ν} are chosen arbitrarily, and the rest of the C_{kl}^{ν} are found in terms of the chosen ones.

Solution at the Origin

In order that the four solutions of **u** be independent of each other, we must have

$$\sum_{j=1}^{4} C_j u_{ij} \neq 0 \quad (i = 1, 2, 3, 4) , \tag{64}$$

where C_j are some constants. A necessary condition for this to be satisfied is that the determinant of Equation 64 be nonzero:

$$||u_{ij}|| \neq 0 . \tag{65}$$

It is not difficult to see that this also is a sufficient condition. At the origin the solution u_{ij} can be expressed as power series in r,

$$u_{ij} = \sum_{\nu=0}^{\infty} a_{ij}^{\nu} r^{s_i+\nu} , \tag{66}$$

where a_{ij}^{ν} are the coefficients of expansion, and s_i are given integers for each component of **u** and are fixed by the behavior of Equation 23 at the origin. We can satisfy Equation 65 at the origin by having

$$||a_{ij}^{0}|| \neq 0 . \tag{67}$$

By choosing suitable values of a_{ij}^{0}, subject to the restriction of Equation 67, four independent solutions are obtained.

Solution at Large r

With given initial values, the solution of Equation 23 can be extended from origin to any desired value of r. To obtain the asymptotic amplitudes and the phase shifts, the presence of the centrifugal and long-range potentials makes it necessary to extend the solutions to infinity. This is undesirable because of the time consumption on the computer, and the accumulated error due to the long-range integration. Seaton (Reference 13) has solved the problem of r^{-2} long-range potentials occurring in the off-diagonal terms of the potential matrix **V** by diagonalizing the asymptotic form of the differential Equation 23 and the corresponding S-matrix. By an inverse transformation the elements of the original S-matrix are found.

Instead, we develop here a perturbation theory which is based on the method described by Mott and Massey (Reference 16). The error in the resulting solution is inversely proportional to the square of the distance from the origin.

Equation 23 for large distances of r can be written

$$\left[\frac{d^2}{dr^2} + k_n^2\right] u(k_n l_n, r) = 2\mathbf{U}u(k_n l_n, r) , \tag{68}$$

where **U** is the sum of the centrifugal potential matrix and the asymptotic form of the **V** matrix. The elements of **U** are given in Appendix D. A component of Equation 68 is of the following form:

$$\left.\begin{aligned} \left[\frac{d^2}{dr^2} + k^2\right] u(r) &= g(r) , \\ g(r) &\ll k^2 u(r) \quad (g(r) \to 0 \text{ as } r \to \infty) . \end{aligned}\right\} \tag{69}$$

The perturbation theory is applied between some large distance R and infinity. Suppose u vanishes at R; then we have the following boundary condition:

$$u(R) = 0 . \tag{70}$$

If we represent the solution of the homogenous equation by y (r), at infinity, we must have

$$\left.\begin{aligned} y(r) &= a \sin(kr - kR) , \\ u(r) &= (a + \Delta a) \sin(kr - kR + \eta) , \end{aligned}\right\} \tag{71}$$

where a is the amplitude of u(r) if g(r) were identically zero and ΔA and η are generated by g(r). Since g(r) is small, we can write

$$u = y(1+\zeta) , \tag{72}$$

where ζ is a small function. Substitution of Equation 72 in Equation 69 gives

$$\frac{d}{dr}\left(y^2 \frac{d\zeta}{dr}\right) = g(r)\, y , \tag{73}$$

where, upon double integration, we obtain

$$\zeta = \int_R^r \frac{dr}{y^2} \int_R^r g(r')\, y\, dr' . \tag{74}$$

The constants of integrations are fixed by the condition (70) and the fact that $u'(R) = y'(R)$.

We now integrate Equation 74 by parts,

$$\zeta = \left[\int_R^r g(r)\, y\, dr\right]\left[\int_R^r \frac{dr}{y^2}\right] - \int_R^r g(r)\, y\, dr \int_R^r \frac{dr}{y^2} . \tag{75}$$

When the integration with respect to y is carried out, and the result is substituted in Equation 73, we obtain

$$u(r) = \sin(kr - kR)\left[a + \frac{1}{k}\int_R^r g(r)\cos(kr - kR)\, dr\right] + \cos(kr - kR)\left[-\frac{1}{k}\int_R^r g(r)\sin(kr - kR)\, dr\right] . \tag{76}$$

Comparison of the second of Equations 71 and 76 shows that

$$\left.\begin{aligned} \Delta a &= \frac{1}{k}\int_R^\infty g(r)\cos(kr - kR)\, dr , \\ \eta &= -\frac{1}{ak}\int_R^\infty g(r)\sin(kr - kR)\, dr \end{aligned}\right\} \tag{77}$$

to first order. The functions g(r) in the four differential Equations 68 are given by

$$g_i(r) = 2\sum_j U_{ij}\, u_j . \tag{78}$$

To first order this can be written by

$$g_i(r) = 2\sum_j a_j\, U_{ij} \sin\left(k_i\, r - k_j\, R_j\right) , \tag{79}$$

where R_j is the last zero of u_j with positive slope. Substitution of this equation in Equation 77 gives

$$\left.\begin{aligned} \Delta a_i &= -\sum_j \frac{a_j}{k_i}\int^{R_i} \cos\left(k_i\, r - k_i\, R_i\right) U_{ij} \sin\left(k_j\, r - k_j\, R_j\right) dr , \\ \eta_i &= \sum_j \frac{a_j}{a_i\, k_i}\int^{R_i} \sin\left(k_i\, r - k_i\, R_i\right) U_{ij} \sin\left(k_j\, r - k_j\, R_j\right) dr . \end{aligned}\right\} \tag{80}$$

Δa_i and η_i can easily be calculated by substituting the values of U_{ij} from Appendix D, integrating the resulting integrals by parts, and retaining the leading terms.

The asymptotic amplitudes and phase shifts are given by

$$\left.\begin{aligned} a_i(\infty) &= a_i(R_i) + \Delta a_i , \\ \delta_i(\infty) &= \delta_i(R_i) + \eta_i + \left[L - \delta(i,3) + \delta(i,4)\right]\frac{\pi}{2} , \end{aligned}\right\} \tag{81}$$

where $a_i(R_i)$ and $\delta_i(R_i)$ are the amplitudes and total phase shifts calculated at R_i by the machine, and where $\delta(i, 3)$ and $\delta(i, 4)$ are the δ functions.

Details of the Numerical Integration

Milne's (Reference 17) method with variable mesh size and Simpson's* rule were used for the integration of the differential equations and evaluations of the integrals respectively. As the solution advances from the origin, the differential equations become less sensitive to the size of the increment, and the error of integration falls below certain small number ϵ. At each value of r the value of the function is found, first with the given value of the increment, and second with the value of increment divided in half. The error of integration is defined as the difference between these two solutions. When the error becomes small, the increment is doubled until a maximum value is reached. At some distance R_1 all the exchange potentials and, similarly, all the direct potentials except those representing optically allowed transitions and the 2p−2p elastic scattering potential become vanishingly small (see Appendix D). At this distance the set of differential equations is replaced by the simpler set containing only these potentials. The integration is continued until some distance R_2, where the first-order solution of the rest of the range of integration is obtained by the method developed in the previous section. No attempt was made to solve any set of linear equations or any matrix equations, as these equations are solvable by the computer in their original form.

The values of the constants of the numerical integration are given below; h_i and h_f are the initial and the final increments of integration. In some exceptional cases, different values were used.

$$h_i = 1 \times 10^{-5}$$

$$h_f = 0.05$$

$$\epsilon = 1 \times 10^{-4}$$

$$R_1 = 30$$

$$R_2 = 200$$

(All quantities are in units of Bohr radius except ϵ, which is dimensionless.)

RESULTS AND DISCUSSION

The differential equations (23) with the known values of the elements of the potential matrix **V** as given in Appendix B were integrated numerically by the methods described in the last section. By choosing different values for the determinant (Equation 67) different sets of independent solutions can be generated. The cross sections reported in this paper have been obtained by averaging the cross sections obtained from two independent sets of solutions. To test the accuracy of the numerical integration, we define the three quantities D_{mn}, D'_{mn}, and D''_m given by

$$D_{mn} = \frac{\left|\sum_{i=1}^{4} k_i a_{im} a_{in} \sin(\delta_{im} - \delta_{in})\right|}{\sum_{i=1}^{4} k_i a_{im} a_{in} \left|\sin(\delta_{im} - \delta_{in})\right|} \quad (m, n = 1, 2, 3, 4, m \neq n), \tag{82}$$

$$D'_{mn} = \frac{|S_{mn} - S_{nm}|}{|S_{mn}| + |S_{nm}|} \quad (m, n = 1, 2, 3, 4, m \neq n), \tag{83}$$

$$D''_m = \frac{\left|\sum_{n=1}^{4} |S_{mn}|^2 - 1\right|}{\sum_{n=1}^{4} |S_{mn}|^2 + 1} \quad (m = 1, 2, 3, 4). \tag{84}$$

*Reference 17, Sec. 33.

Based on Equations 36, 37, and 41 in an exact solution of the four differential equations, the right-hand side of these equations would vanish; they can therefore be used to test the accuracy of the numerical integration. As an illustration the numerical values of D_{mn}, D'_{mn}, and D''_{m} for the case of $1s-2s-2p$ coupling, $\beta=+1$, $k_1=2.0$, and $L=3$, are given below:

$$D_{12} = 1.4 \times 10^{-3}, \quad D_{13} = 2.6 \times 10^{-4}, \quad D_{14} = 1.3 \times 10^{-3},$$

$$D_{23} = 5.1 \times 10^{-4}, \quad D_{24} = 2.2 \times 10^{-3}, \quad D_{34} = 1.8 \times 10^{-3},$$

$$D'_{12} = 7.6 \times 10^{-4}, \quad D'_{13} = 5.1 \times 10^{-3}, \quad D'_{14} = 5.6 \times 10^{-3},$$

$$D'_{23} = 5.4 \times 10^{-3}, \quad D'_{24} = 5.7 \times 10^{-3}, \quad D'_{34} = 1.3 \times 10^{-3},$$

$$D''_{1} = 1.8 \times 10^{-4}, \quad D''_{2} = 7.6 \times 10^{-5}, \quad D''_{3} = 2.5 \times 10^{-4}, \quad D''_{4} = 4.4 \times 10^{-6}.$$

To compare the results of the numerical integration by the noniterative method we have carried out here with those of the iterative method of References 2, 4, and 5, we have provided Table 1.* The $1s \rightarrow 2s$ excitation cross section is given by the two methods; I and II refer to the iterative and noniterative methods, respectively; and E_{max} is the maximum of the error to value ratios in the reciprocity relations (Equation 83). In the $1s-2s$ eigenstates coupling approximation the non-iterative method is far more accurate than the iterative method and, as is seen, the cross

Table 1

Comparison of the Iterative and the Noniterative Results for the Singlet ($L = 0, 1$; $k_1 = 0.9, 1.0$), $1s - 2s$ Excitation Cross Section.*

(a) $1s - 2s$ coupling

k_1	L	$Q_{1s \rightarrow 2s}$		E_{max} (percent)	
		†I	II	I	II
0.90	0	0.0384	0.0375	7.1	0.72
1.00	0	0.0714	0.0725	unknown	0.53
0.90	1	0.008	0.0017	386	0.91
1.00	1	0.051	0.0583	55	0.75

(b) $1s - 2s - 2p$ coupling

k_1	L	$Q_{1s \rightarrow 2s}$		E_{max} (percent)	
		‡I	II	I	II
0.90	0	0.0529	0.523	0.40	0.40
1.00	0	0.0766	0.0768	0.12	0.60
0.90	1	0.0045	0.0048	2.3	10
1.00	1	0.0145	0.0147	0.33	1.3

*I and II refer to iterative and noniterative methods, respectively; E_{max} is the maximum of the error to value ratios in the reciprocity relation.
†See Reference 2.
‡See References 4 and 5.

*Author is indebted to Dr. K. Smith for sending some of the data used in Table 1.

sections by the two methods differ from each other sometimes in their first significant figure. In the 1s−2s−2p eigenstates coupling approximation, on the other hand, the results by the iterative method seems to be somewhat more accurate. The reason is contributed to the effect of the r^{-2} long-range potential, which appears in the differential equations when the 2p state is included in the eigenstates coupling approximation. Two different methods are used in References 4 and 5 and in the present paper to estimate the effect of this potential for large distances; and it may be that in References 4 and 5 this effect is better taken into account. Nevertheless the cross sections are the same in their first three decimal places.

In Figure 1 we present the theoretical and the experimental estimate of the 1s−2s excitation cross section. The calculated curves are Born, 1s—2s coupling, 1s−2s−2p coupling exchange neglected, and 1s−2s−2p coupling exchange included approximation. The first three of these curves are the same as References 4 and 5. The experimental curves are those of Lichten and Schultz (Reference 18) and Stebbings, Fite, and Hummer (Reference 19). The various calculated results agree better with the results of Lichten and Schultz. However, recent calculations of Taylor and Burke (Reference 20) have shown that, in an eigenstates expansion calculation where 1s, 2s, 2p, 3s, and 3p are included, the cross section at the peak of the 1s−2s−2p curve is reduced by 30 percent. This suggests that, within the eigenstates expansion approximation, more states should be included to insure that the convergence has been achieved; and the discrepancy between the two experimental results is still an unresolved problem. As another theoretical approach to the problem, H. L. Kyle and A. Temkin (Reference 21) have extended the nonadiabatic theory of scattering developed by A. Temkin (Reference 22) to the L=0, 1s→2s inelastic scattering of electrons by the hydrogen atom. They find a 30 percent decrease in the 1s→2s cross section as calculated by the 1s−2s close coupling approximation.

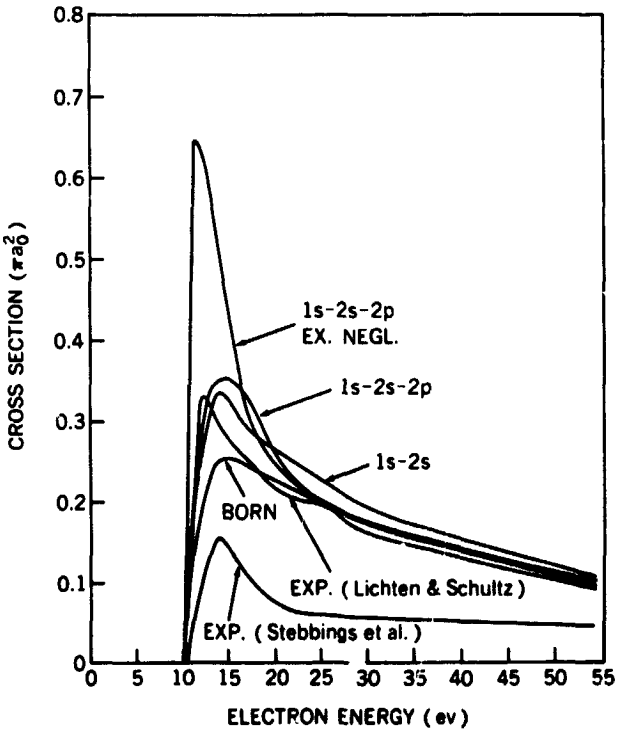


FIGURE 1.—1s→2s total excitation cross section. (1s–2s refers to 1s–2s eigenstates coupling approximation; 1s–2s–2p has similar meaning. EX.NEGL. refers to exchange neglected case, BORN is the Born approximation, and EXP. refers to experiment.)

Comparison of the exchange neglected and exchange included 1s−2s−2p coupling shows that exchange is mostly important at threshold, and its effect does not extend beyond 20 electron volts.

Table 2 gives the numerical values of the 1s→2s cross section in different approximations. The 1s→2s excitation cross section in the singlet state has an interesting behavior immediately above threshold. In Figure 2 this cross section for a range of 600 milli electron volts (mev) above threshold is plotted. In the 1s−2s coupling approximation a maximum appears at 34 mev, while in the 1s−2s−2p coupling approximation there are three maxima of approximately the same magnitudes at 17, 34, and 87 mev, respectively. In the singlet case the cross section rises sharply within a range of 17 mev above threshold to a value of about $0.04\pi a_0^2$; it then rises with an approximately constant and small slope. The contribution of the triplet case is seen to be almost negligible at the threshold, and it has no maximum in this region. It should be noted that the principal maximum in the 12→2s excitation cross section appears at about 3 ev with a value of about 0.35, and has a contribution from a higher angular momentum than L=0. Although no study has been made to relate the existence of the maxima above threshold to any physical phenom-

Table 2
1s - 2s Excitation Cross Section.

(a) Born approximation

k_i	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ	Q_T
0.9	0.16370	0.00981	0.00021	0.00000	0.00000	0.00000	0.00000	0.00000	0.17378	0.17379
1.0	0.19578	0.04795	0.00428	0.00026	0.00000	0.00000	0.00000	0.00000	0.24827	0.24827
1.1	0.16272	0.07073	0.01141	0.00125	0.00011	0.00001	0.00000	0.00000	0.24622	0.24623
1.2	0.12704	0.07896	0.01858	0.00299	0.00039	0.00004	0.00000	0.00000	0.22800	0.22800
1.5	0.05872	0.06606	0.02979	0.00939	0.00242	0.00053	0.00010	0.00002	0.16703	0.16706
2.0	0.01946	0.03363	0.02521	0.01365	0.00614	0.00236	0.00081	0.00025	0.10151	0.10187
3.0	0.00388	0.00909	0.01019	0.00866	0.00628	0.00394	0.00226	0.00116	0.04546	0.04758
4.0	0.00123	0.00320	0.00420	0.00431	0.00385	0.00301	0.00216	0.00141	0.02337	0.02720

(b) Exchange neglected 1s - 2s - 2p eigenstates coupling approximation

k_i	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ	Q_T
0.9	0.2202	0.0749	0.3535	--	--	--	--	--	0.6486	0.6486
1.0	0.1685	0.1427	0.1598	0.0517	--	--	--	--	0.5227	0.5227
1.1	0.0951	0.1142	0.0298	0.0616	0.0231	--	--	--	0.3238	0.3238
1.2	0.0594	0.1137	0.0032	0.0360	0.0244	0.0135	--	--	0.2502	0.2502
1.5	0.0249	0.0861	0.0201	0.0068	0.0112	0.0118	0.0074	--	0.1683	0.1683
2.0	0.0101	0.0373	0.0255	0.0107	0.0046	0.0034	0.0033	--	0.0949	0.0953

(c) 1s - 2s eigenstates coupling approximation

Singlet										
k_i	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_S	---
0.9	0.0375	0.0017	0.0000	0.0000	0.0000	0.0000	0.0000	--	0.0392	---
1.0	0.0725	0.0583	0.0002	0.0000	0.0000	0.0000	0.0000	--	0.1310	
1.1	0.0701	0.0525	0.0023	0.0000	0.0000	0.0000	0.0000	--	0.1249	
1.2	0.0547	0.0534	0.0054	0.0002	0.0000	0.0000	0.0000	--	0.1137	
1.5	0.0241	0.0384	0.0110	0.0022	0.0004	0.0001	0.0000	--	0.0762	
2.0	0.0072	0.0157	0.0093	0.0041	0.0015	0.0005	0.0002	--	0.0385	

Triplet										$\Sigma_S + \Sigma_T$	Q_T
k_i	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_T		
0.9	0.0004	0.1686	0.0060	0.0000	0.0000	0.0000	0.0000	--	0.1750	0.2142	0.2142
1.0	0.0021	0.1528	0.0446	0.0021	0.0001	0.0000	0.0000	--	0.2017	0.3327	0.3327
1.1	0.0044	0.1052	0.0568	0.0068	0.0005	0.0000	0.0000	--	0.1737	0.2987	0.2987
1.2	0.0061	0.0737	0.0576	0.0114	0.0015	0.0002	0.0000	--	0.1505	0.2642	0.2642
1.5	0.0073	0.0355	0.0406	0.0174	0.0050	0.0012	0.0002	--	0.1072	0.1833	0.1833
2.0	0.0049	0.0162	0.0205	0.0143	0.0074	0.0032	0.0012	--	0.0677	0.1062	0.1068

(d) 1s - 2s - 2p eigenstates coupling approximation

Singlet										
k_i	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_S	---
0.9	0.0523	0.0048	0.0620	--	--	--	--	--	0.1191	---
1.0	0.0768	0.0147	0.0833	0.0092	--	--	--	--	0.1840	
1.1	0.0585	0.0245	0.0647	0.0236	0.0055	--	--	--	0.1768	
1.2	0.0382	0.0251	0.0246	0.0252	0.0081	0.0028	--	--	0.1240	
1.5	0.0123	0.0308	0.0015	0.0041	0.0051	0.0034	0.0023	0.0026	0.0621	
2.0	0.0049	0.0152	0.0068	0.0021	0.0010	0.0010	0.0011	0.0008	0.0329	
3.0	0.0010	0.0031	0.0031	0.0023	0.0015	0.0009	0.0005	0.0006	0.0130	
4.0	0.0003	0.0010	0.0012	0.0012	0.0010	0.0008	--	--	0.0055	

Triplet										$\Sigma_S + \Sigma_T$	Q_T
k_i	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_T		
0.9	0.0013	0.0748	0.0019	--	--	--	--	--	0.0780	0.1971	0.1971
1.0	0.0040	0.1224	0.0195	0.0214	--	--	--	--	0.1673	0.3513	0.3513
1.1	0.0050	0.1013	0.0326	0.0077	0.0131	--	--	--	0.1597	0.3366	0.3366
1.2	0.0055	0.0724	0.0359	0.0036	0.0105	0.0070	--	--	0.1355	0.2596	0.2596
1.5	0.0045	0.0333	0.0309	0.0072	0.0046	0.0054	0.0049	0.0043	0.0951	0.1573	0.1573
2.0	0.0031	0.0155	0.0176	0.0101	0.0044	0.0025	0.0023	0.0019	0.0574	0.0903	0.0907
3.0	0.0013	0.0048	0.0065	0.0059	0.0044	0.0029	0.0018	0.0012	0.0288	0.0418	0.0439
4.0	0.0006	0.0019	0.0028	0.0030	0.0027	0.0022	--	--	0.0132	0.0187	0.0261

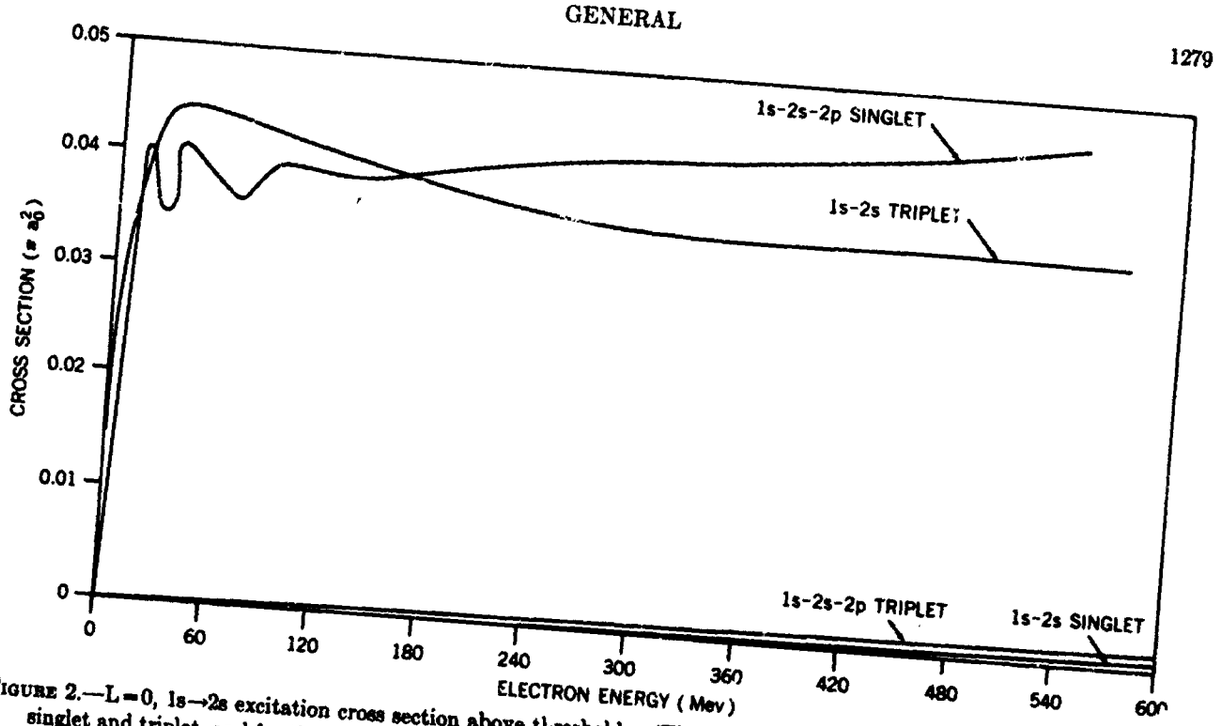


FIGURE 2.—L=0, 1s→2s excitation cross section above threshold. (The cross sections are given for the two spin states singlet and triplet, and for the two approximations 1s−2s and 1s−2s−2p. The total cross section is the sum of the singlet and the triplet cross sections.)*

Table 3

The Singlet L = 0, 1s - 2s Excitation Cross Section Near Threshold.*

k_2	0	0.01	0.02	0.025	0.030	0.035	0.04
E(mev)	0	1.36	5.44	8.50	12.2	16.7	21.8
Q_1	0	0.0168	0.0298	--	0.0377	--	0.0423
Q_2	0	--	0.0149	0.0259	0.0349	0.0405	0.0353
k_2	0.045	0.050	0.060	0.070	0.080	0.090	0.100
E(mev)	27.5	34.0	49.0	66.6	87.0	110.	136.
Q_1	--	0.0446	0.0441	0.0435	0.0423	0.0412	0.0405
Q_2	0.0346	0.0405	0.0391	0.0361	0.0395	0.0392	0.0385

*k_2 is the wave number of the inelastically scattered wave, and E is the corresponding energy in mev; Q_1 and Q_2 are the cross sections according to the 1s - 2s and the 1s - 2s - 2p couplings, respectively.

ena, it may be said that, similar to resonances below threshold in the elastic scattering of electrons by the hydrogen atom, these maxima are due to formation of some unstable states of the negative hydrogen ion. The numerical values of the 1s→2s cross section at threshold are given in Table 3. Damburg and Peterkop (Reference 6), and Gailitis and Damburg (Reference 23) have made an extensive study of the behavior of different cross sections near threshold in the 1s−2s

*According to Gailitis and Damburg, when the energy difference between the 2s and 2p states is neglected in the 1s-2s-2p couplings, the 1s→2s excitation cross section does not go to zero at threshold (see Reference 23, and Figure 2). Figures 2 and 4 show that in the 1s-2s-2p couplings, if E_{n-1} and E_n represent the energy with respect to the threshold of the two neighboring maxima or minima then $E_n/E_{n-1} \approx$ constant. This may be attributed to the r^{-4} potential, which is due to the coupling between the 2s and the 2p states. For further details, see Reference 23.

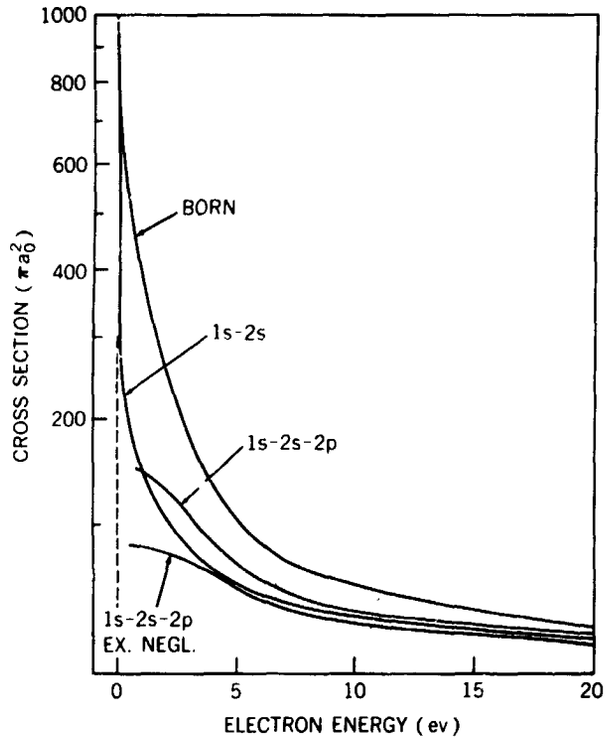


FIGURE 3.—2s→2s total elastic cross section. (Curves are designated as in Figure 1).

and the 1s−2s−2p eigenstates coupling approximations.

In Figure 3 we have shown the 2s→2s elastic cross section. The 1s−2s coupling approximation gives a value of $944\pi\ a_0^2$ at zero incident energy, while the corresponding value in the Born approximation is $768\pi\ a_0^2$. The high value of this cross section at zero energy is in sharp contrast with its geometrical cross section. The zero energy 2s→2s cross section in the 1s−2s−2p coupling approximation, because of the r^{-2} potential, is difficult to find. The 2s→2s cross section has certain maxima and minima at low energy which are not found in the 1s→1s cross section. Figure 4 shows the L=0 singlet and triplet 2s→2s cross section in the two approximations. While there is one minimum in the 1s−2s coupling approximation, there are three minima in the 1s−2s−2p coupling approximation. It is thought that the existence of these minima is due to a wider potential range in the 2s→2s scattering, a case which does not exist in the 1s→1s scattering. In Table 4 we have listed the numerical

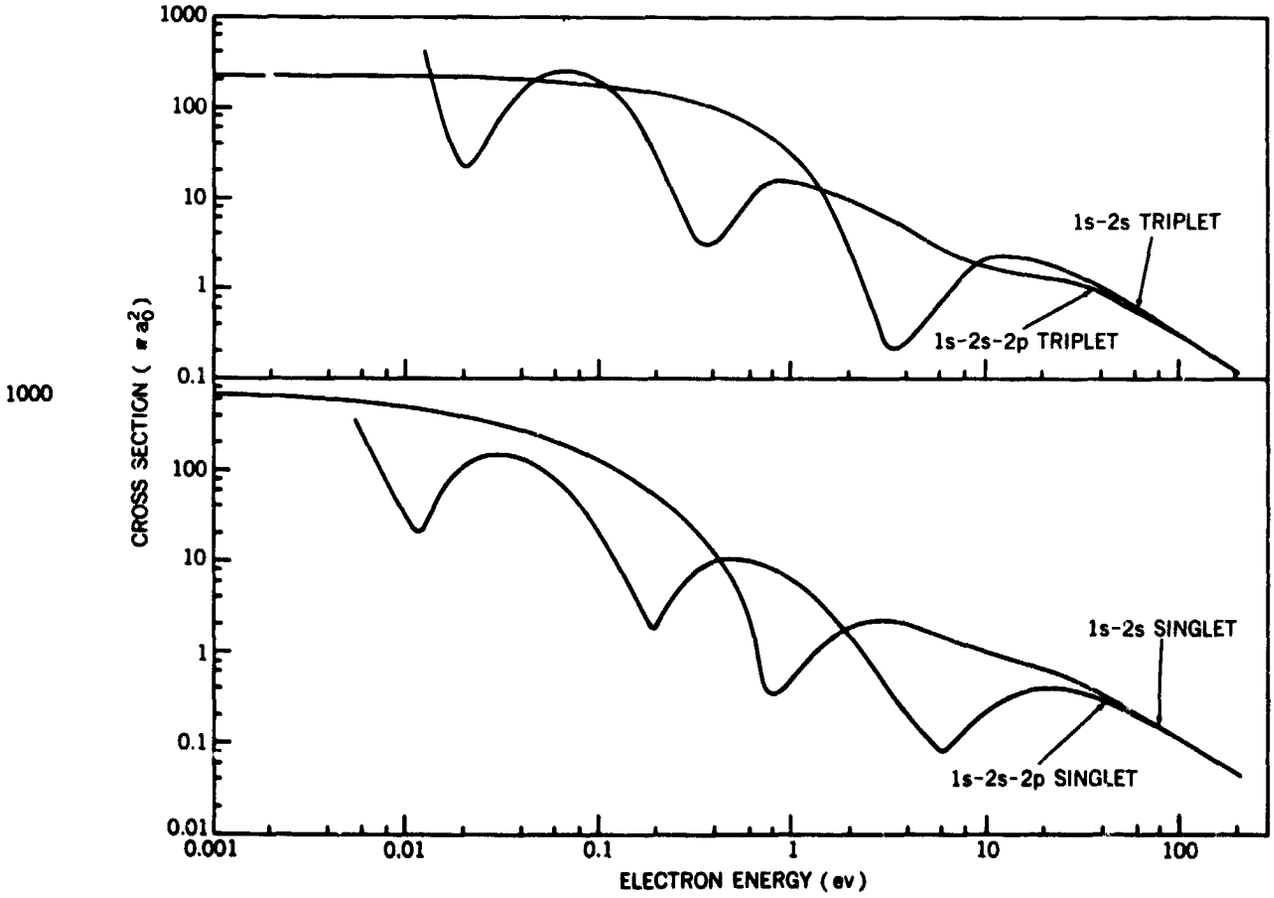


FIGURE 4.—L=0, 2s→2s elastic cross section. (Curves are designated as in Figure 2).

Table 4
2s - 2s Elastic Cross Section.

(a) Born approximation

k_2	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ	Q_T
0.24	389.97	15.910	0.27680	0.00319	0.00003	0.00000	0.00000	0.00000	406.16	406.17
0.50	105.52	32.129	5.0152	0.55306	0.04846	0.00347	0.00022	0.00001	143.269	143.276
0.68	47.381	24.746	7.4981	1.7023	0.31602	0.04869	0.00658	0.00078	81.700	81.703
0.83	26.562	17.895	7.5601	2.5080	0.69820	0.16365	0.03393	0.00621	55.427	55.440
1.23	8.2461	7.6846	4.9702	2.7483	1.3500	0.57745	0.22452	0.07803	25.879	25.990
1.80	2.4331	2.7773	2.2763	1.6805	1.1541	0.71994	0.41776	0.22120	11.680	11.105
2.87	0.5245	0.71381	0.68903	0.61123	0.51876	0.4104	0.31076	0.21860	4.0011	.8280
3.91	0.18794	0.27611	0.28658	0.27339	0.25098	0.21580	0.17946	0.14029	1.8106	2.7417

(b) Exchange neglected 1s - 2s - 2p eigenstates coupling approximation

k_2	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ	Q_T
0.24	32.02	8.489	42.26	--	--	--	--	--	82.77	82.78
0.50	2.041	7.710	23.35	9.004	9.368	7.702	5.870	--	65.045	65.052
0.68	2.255	8.247	15.16	5.547	4.373	3.755	2.975	--	42.312	42.316
0.83	1.861	8.206	10.67	4.637	2.792	2.223	1.818	--	32.207	32.226
1.23	1.716	4.987	5.030	3.321	1.797	1.058	0.7179	--	18.537	18.726
1.80	1.020	2.134	2.150	1.750	1.275	0.8320	0.5457	--	9.707	10.352

(c) 1s - 2s eigenstates coupling approximation

Singlet										
k_2	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_S	---
0.24	0.3303	3.196	0.2628	0.0028	--	--	--	--	8.792	---
0.50	1.532	10.38	0.0275	0.0048	0.0008	0.0002	0.0249	--	11.97	
0.68	1.115	5.536	1.502	0.1150	0.0087	0.0010	0.0017	--	8.279	
0.83	0.8980	3.512	1.997	0.4303	0.0747	0.0129	0.0032	--	6.928	
1.23	0.5702	1.413	1.236	0.7010	0.3129	0.1228	0.0450	--	4.401	
1.80	0.2825	0.5370	0.5285	0.4193	0.2931	0.1863	0.1110	--	2.358	

Triplet										$\Sigma_S + \Sigma_T$	Q_T
k_2	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_T		
0.24	45.94	118.8	7.713	0.0540	--	--	--	--	172.51	181.30	181.31
0.50	0.2102	34.44	21.05	2.776	0.2521	0.0316	0.0994	--	58.86	70.83	70.84
0.68	1.366	18.13	12.74	4.059	0.8282	0.1463	0.0442	--	37.31	45.60	45.60
0.83	2.112	11.65	8.725	3.887	1.230	0.3225	0.0850	--	28.01	34.94	34.96
1.23	1.811	4.691	4.008	2.585	1.399	0.6637	0.2862	--	15.44	19.84	20.03
1.80	0.8989	1.735	1.652	1.316	0.9510	0.6359	0.4008	--	7.590	9.947	10.592

(d) 1s - 2s - 2p eigenstates coupling approximation

Singlet										
k_2	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_S	---
0.24	7.800	14.79	22.42	--	--	--	--	--	45.01	---
0.50	0.2858	0.6960	3.149	4.447	2.925	2.063	1.491	--	15.057	
0.68	0.0661	1.044	2.455	1.884	1.480	1.098	0.7928	--	8.820	
0.83	0.1675	1.088	2.105	0.9905	0.8282	0.6606	0.5071	--	6.347	
1.23	0.3739	1.075	1.196	0.7196	0.3940	0.2496	0.1827	0.1409	4.332	
1.80	0.2416	0.4974	0.5133	0.4133	0.2923	0.1928	0.1247	0.0799	2.3553	
2.87	0.0847	0.1489	0.1574	0.1466	0.1280	0.1057	0.0852	0.0658	0.9223	
3.91	0.0365	0.0635	0.0702	0.0698	0.0660	0.0599	--	--	0.3659	

Triplet										$\Sigma_S + \Sigma_T$	Q_T
k_2	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_T		
0.24	16.52	1.236	89.65	--	--	--	--	--	107.41	152.42	152.43
0.50	6.172	17.19	30.69	12.90	5.559	5.373	4.257	--	82.14	97.206	97.213
0.68	2.346	12.88	14.52	7.937	3.046	2.459	2.090	--	45.28	54.099	54.103
0.83	1.709	9.166	8.976	5.367	2.365	1.514	1.246	--	30.343	36.686	36.705
1.23	1.391	4.199	3.880	2.681	1.585	0.9110	0.5658	0.4094	15.622	19.955	20.066
1.80	0.7898	1.656	1.603	1.291	0.9466	0.6520	0.4305	0.2786	7.648	10.002	10.427
2.87	0.2622	0.4652	0.4857	0.4479	0.3901	0.3234	0.2623	0.2045	2.8413	3.7636	4.5905
3.91	0.1113	0.1945	0.2146	0.2122	0.2000	0.1812	--	--	1.1138	1.4797	2.7306

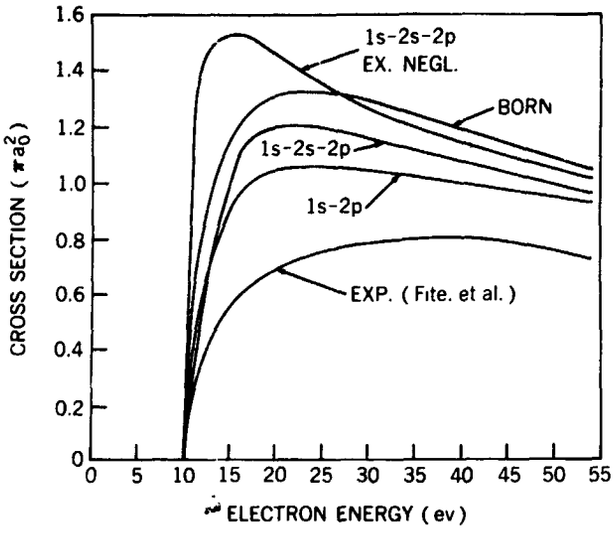


FIGURE 5.—1s→2p total excitation cross section. (1s–2p refers to 1s–2p eigenstates coupling approximation; 1s–2s–2p has similar meaning. EX.NEGL. refers to exchange case, BORN is the Born approximation, and EXP. refers to experiment.

values of the 2s→2s cross section in different approximations.

In Figure 5 the four calculated curves for the 1s→2p excitation cross section are compared with the measurement of Fite, Stebbings, and Brackmann (References 24 and 25). The 1s−2s−2p and the Born curves are the same as in References 4 and 5, but the 1s−2s−2p exchange neglected and the 1s−2p curves are not calculated in these references. As concluded before, the calculated curves are higher than the experimental. Moreover, we notice that, similar to the 1s→2s excitation cross section, the inclusion of the exchange lowers the value of the cross section at threshold. Table 5 gives the numerical values of the 1s→2p cross section in different approximations.

The calculation of the 2p→2p elastic cross section is more complicated than the cases so far considered. For a given total angular momentum L, the angular momentum of the partial wave which is scattered from the 2p state may be L−1, L, and L+1. The first and the third values correspond to a wave function which has the same parity as the wave functions in the 1s and the 2s channels; in this case, $L-\ell_1-\ell_2$ is even. The second value corresponds to a wave function with a different parity, and the only process that occurs with this parity is the 2p elastic scattering; in this case, $L-\ell_1-\ell_2$ is odd. We have calculated the

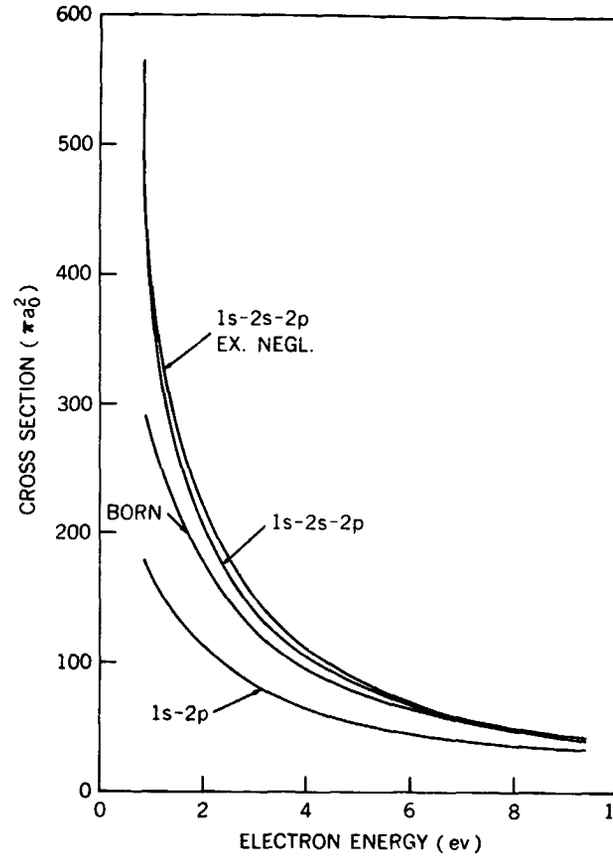


FIGURE 6.—2p→2p total elastic cross sections. (Curves are designated as in Figure 5. The cross section at zero energy is finite but is not found here.)

2p→2p cross sections for the two cases, and they are listed in Table 6. The total cross section is shown in Figure 6 (on page 33). Because of the r^{-2} potential it is difficult to find the zero energy value of this cross section.

The 2s→2p transition cross section has application in some plasma and stellar atmosphere calculations. The total cross section using the Born approximation is given by Seaton (Reference 26). In Table 7 (on page 32) we list the partial cross section using the close coupling approximation. This table may be found useful in problems in which plasma shielding occurs, where only electrons with an impact parameter with a given range can induce the 2s→2p transition.

It may be noted that the cross sections for the inverse processes 2s→1s, 2p→1s, and 2p→2s may be calculated by Equation 35 and the symmetry of the T-matrix.

In all the tables listed here, k_1 is the wave number in the 1s, and k_2 is the wave number in

Table 5
1s → 2p Excitation Cross Section.

(a) Born approximation

k_1	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ	Q_T
0.9	0.00107	0.46700	0.09607	0.01025	0.00087	0.00007	0.00001	0.00000	0.56534	0.57535
1.0	0.00499	0.48867	0.35645	0.13467	0.03964	0.01032	0.00254	0.00059	1.03787	1.03851
1.1	0.00702	0.36207	0.41698	0.24992	0.11533	0.04649	0.01753	0.00624	1.22158	1.22859
1.2	0.00747	0.25540	0.38137	0.29903	0.17908	0.09269	0.04453	0.02004	1.27961	1.30741
1.5	0.00550	0.09184	0.19986	0.23775	0.21386	0.16292	0.11291	0.07232	1.09696	1.28101
2.0	0.00234	0.02222	0.05938	0.09386	0.11287	0.11376	0.10248	0.08374	0.59065	1.04055
3.0	0.00048	0.00285	0.00806	0.01537	0.02309	0.02920	0.03271	0.03273	0.14449	0.66256
4.0	0.00013	0.00066	0.00179	0.00359	0.00586	0.00816	0.01009	0.01114	0.04142	0.45252

(b) Exchange neglected 1s - 2s - 2p eigenstates coupling approximation

k_1	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ	Q_T
0.9	0.1600	0.3985	0.6497	--	--	--	--	--	1.2082	1.2194
1.0	0.1007	0.2917	0.8190	0.2190	0.0476	0.0224	--	--	1.5004	1.5041
1.1	0.0980	0.2008	0.6201	0.3696	0.1380	0.0586	0.0166	--	1.5017	1.5150
1.2	0.0822	0.1251	0.4481	0.3922	0.2044	0.1014	0.0403	--	1.3937	1.4416
1.5	0.0372	0.0334	0.1671	0.2568	0.2295	0.1706	0.1136	--	1.0082	1.2645
2.0	0.0105	0.0068	0.0394	0.0837	0.1089	0.1165	0.1042	--	0.4700	1.0036

(c) 1s - 2p eigenstates coupling approximation

Singlet

k_1	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_S	---
0.9	0.0044	0.1216	0.1422	--	--	--	--	--	0.2682	---
1.0	0.0168	0.0655	0.3011	0.0206	0.0057	0.0059	--	--	0.4156	
1.1	0.0299	0.0366	0.3948	0.0851	0.0260	0.0099	0.0060	--	0.5883	
1.2	0.0296	0.0169	0.3088	0.1421	0.0517	0.0236	0.0103	--	0.5830	
1.5	0.0059	0.0037	0.0821	0.0989	0.0718	0.0458	0.0304	--	0.3386	
2.0	0.0010	0.0006	0.0131	0.0271	0.0327	0.0335	0.0296	--	0.1376	

Triplet

k_1	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_1	$\Sigma_S + \Sigma_T$	Q_T
0.9	0.0002	0.2066	0.0005	--	--	--	--	--	0.2073	0.4755	0.4867
1.0	0.0016	0.1078	0.0020	0.1651	0.0363	0.0187	--	--	0.3315	0.7471	0.7508
1.1	0.0037	0.0540	0.0060	0.1599	0.1002	0.0365	0.0217	--	0.3820	0.9703	0.9836
1.2	0.0055	0.0249	0.0098	0.1446	0.1336	0.0695	0.0341	--	0.4220	1.0050	1.0529
1.5	0.0059	0.0027	0.0133	0.0868	0.1231	0.1100	0.0849	--	0.4267	0.7652	1.0215
2.0	0.0028	0.0002	0.0073	0.0311	0.0559	0.0700	0.0772	--	0.2445	0.3820	0.9156

(d) 1s - 2s - 2p eigenstates coupling approximation

Singlet

k_1	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_S	---
0.9	0.0390	0.0745	0.1027	--	--	--	--	--	0.2162	---
1.0	0.0360	0.1123	0.2575	0.0317	--	--	--	--	0.4375	
1.1	0.0358	0.1094	0.3405	0.0886	0.0308	0.0113	0.0075	--	0.6239	
1.2	0.0345	0.0806	0.2912	0.1278	0.0506	0.0237	0.0105	--	0.6189	
1.5	0.0172	0.0175	0.0953	0.1003	0.0693	0.0440	0.0290	0.0238	0.3964	
2.0	0.0036	0.0023	0.0170	0.0303	0.0344	0.0333	0.0293	0.0229	0.1731	
3.0	0.0004	0.0002	0.0015	0.0037	0.0060	0.0078	0.0094	0.0106	0.0396	
4.0	0.0001	0.0001	0.0003	0.0008	0.0014	0.0020	--	--	0.0047	

Triplet

k_1	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_T	$\Sigma_S + \Sigma_T$	Q_T
0.9	0.0007	0.0682	0.0112	--	--	--	--	--	0.0801	0.2963	0.3075
1.0	0.0033	0.0801	0.0500	0.1730	--	--	--	--	0.3064	0.7439	0.7976
1.1	0.0070	0.0626	0.0567	0.1841	0.1082	0.0404	0.0209	--	0.4799	1.1039	1.1172
1.2	0.0096	0.0418	0.0537	0.1761	0.1409	0.0729	0.0332	--	0.5282	1.1471	1.1950
1.5	0.0107	0.0131	0.0351	0.1081	0.1343	0.1157	0.0842	0.0596	0.5608	0.9570	1.1410
2.0	0.0053	0.0038	0.0143	0.0393	0.0625	0.0731	0.0738	0.0642	0.3363	0.5094	0.9593
3.0	0.0010	0.0006	0.0025	0.0066	0.0122	0.0175	0.0220	0.0246	0.0870	0.1266	0.6095
4.0	0.0002	0.0002	0.0006	0.0016	0.0031	0.0048	--	--	0.0105	0.0152	0.4475

Table 6
2p - 2p Elastic Cross Section.

(a) L - ℓ_1 - ℓ_2 odd, Born approximation

k_2	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_0
0.24	26.562	6.1961	2.1260	0.92680	0.44497	0.22694	0.08105	36.56
0.50	14.210	5.2190	2.1768	1.0526	0.55433	0.31476	0.18383	23.71
0.68	8.8346	4.0761	1.9476	1.0100	0.54983	0.31728	0.18718	16.923
0.83	6.0065	3.2062	1.6993	0.94122	0.53204	0.31331	0.18680	12.8854
1.23	2.5053	1.7048	1.1013	0.70920	0.44518	0.28212	0.17577	6.9237
1.80	0.91169	0.74615	0.56949	0.42475	0.30158	0.21122	0.14245	3.30733
2.87	0.23647	0.22647	0.20035	0.17178	0.13878	0.10926	0.08177	1.16488

(b) L - ℓ_1 - ℓ_2 odd, exchange neglected 2p eigenstates coupling approximations

k_2	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_0
0.24	61.12	8.444	2.5808	1.1408	0.6200	0.3720	0.2516	74.52
0.50	15.292	6.884	2.5436	1.1580	0.6216	0.3640	0.2352	27.10
0.68	8.008	4.940	2.2548	1.1160	0.6160	0.3640	0.2352	17.54
0.83	5.108	3.6364	1.9292	1.0380	0.5968	0.3592	0.2340	12.90
1.23	2.0484	1.7392	1.1812	0.7652	0.4972	0.3236	0.2208	6.776
1.80	0.7640	0.7220	0.5816	0.4436	0.3312	0.2400	0.1784	3.261
2.87	0.2132	0.2180	0.2004	0.1776	0.1520	0.1272	0.1064	1.195

(c) L - ℓ_1 - ℓ_2 odd, 2p eigenstates coupling approximation

Singlet								
k_2	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_{0S}
0.24	2.963	4.161	0.6725	0.2861	0.1552	0.0930	0.0629	8.394
0.50	3.735	3.182	0.7915	0.3066	0.1576	0.0913	0.0589	8.323
0.68	2.165	1.728	0.6851	0.3025	0.1587	0.0920	0.0590	5.190
0.83	1.371	1.107	0.5562	0.2799	0.1547	0.0913	0.0589	3.619
1.23	0.5280	0.4603	0.3115	0.1991	0.1277	0.0823	0.0558	1.7647
1.80	0.1928	0.1831	0.1475	0.1124	0.0837	0.0605	0.0449	0.8249
2.87	0.0534	0.0547	0.0503	0.0445	0.0381	0.0319	0.0267	0.2996
Triplet								
k_2	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_{0T}
0.24	49.22	3.265	1.850	0.8528	0.4650	0.2788	0.1886	56.12
0.50	7.791	1.851	1.504	0.8193	0.4598	0.2722	0.1764	12.87
0.68	4.720	2.219	1.361	0.7700	0.4481	0.2700	0.1756	9.964
0.83	3.373	2.099	1.237	0.7198	0.4318	0.2652	0.1744	8.300
1.23	1.484	1.225	0.8376	0.5509	0.3630	0.2384	0.1639	4.863
1.80	0.5674	0.5337	0.4295	0.3283	0.2456	0.1784	0.1330	2.4159
2.87	0.1596	0.1632	0.1500	0.1328	0.1136	0.0951	0.0797	0.8940

(d) L - ℓ_1 - ℓ_2 even, Born approximation

k_2	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_E	Q_T
0.24	12.488	230.42	4.0427	1.9014	0.30797	0.57480	0.29675	0.06128	250.09	290.69
0.50	0.1758	74.475	7.0963	0.92172	0.48094	0.45745	0.30098	0.05695	83.965	109.714
0.68	0.07386	36.478	6.5573	1.0156	0.29902	0.27895	0.23113	0.04364	44.978	63.742
0.83	0.22032	21.559	5.4143	1.1653	0.26124	0.15835	0.15225	0.02533	28.956	43.572
1.23	0.25024	7.3349	2.9703	1.1453	0.37750	0.11091	0.04377	0.03572	12.269	20.571
1.80	0.13477	2.3518	1.3155	0.74676	0.39290	0.18754	0.07731	0.02824	5.2348	9.6066
2.87	0.04395	0.55809	0.40443	0.30267	0.21866	0.15188	0.09753	0.05832	1.8355	3.8272
3.91	0.01868	0.20876	0.17048	0.14307	0.11656	0.09255	0.06896	0.04864	0.8677	2.1455

(e) L - ℓ_1 - ℓ_2 even, exchange neglected 1s - 2s - 2p eigenstates coupling approximation

k_2	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_E	Q_T
0.24	31.96	91.21	154.6	92.86	55.86	36.79	24.37	--	487.7	566.32
0.50	9.371	12.65	30.88	20.53	12.99	8.451	5.695	--	100.6	129.8
0.68	4.156	5.953	13.19	9.628	6.563	4.508	3.130	--	47.13	66.56
0.83	2.542	4.560	7.424	5.249	3.811	2.793	2.032	--	28.41	43.07
1.23	1.208	2.879	2.906	1.734	1.106	0.8438	0.6912	--	11.37	19.56
1.80	0.5612	1.385	1.238	0.8299	0.5109	0.3005	0.2022	--	5.028	9.382

Table 6 (Concluded)
2p - 2p Elastic Cross Section.

(f) L - ℓ_1 - ℓ_2 even, 1s - 2p eigenstates coupling approximation

Singlet										
k_2	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_{ES}	---
0.24	1.964	5.238	17.34	1.260	0.4896	0.2460	--	--	26.54	---
0.50	0.5131	3.159	1.544	1.439	0.4063	0.1906	0.1133	--	7.365	
0.68	0.2346	2.091	1.039	0.4944	0.2539	0.1520	0.0984	--	4.363	
0.83	0.1227	1.252	0.9930	0.1532	0.1191	0.0990	0.0766	--	2.816	
1.23	0.1071	0.6796	0.6167	0.2249	0.0635	0.0249	0.0217	--	1.7384	
1.80	0.0571	0.3366	0.2926	0.1818	0.0954	0.0426	0.0172	--	1.0233	
Triplet										
k_2	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_{ET}	Q_T
0.24	15.62	26.72	32.44	6.674	1.252	0.7359	--	--	83.44	178.89
0.50	4.844	3.384	19.97	12.52	0.8285	0.3266	0.2754	--	42.15	72.80
0.68	2.247	3.886	10.33	5.922	1.035	0.2270	0.1764	--	23.82	45.22
0.83	1.293	3.650	6.287	3.476	0.9248	0.2097	0.1162	--	15.957	32.45
1.23	0.4842	2.287	2.424	1.418	0.6072	0.2079	0.0717	--	7.500	17.28
1.80	0.1909	1.063	0.9673	0.6659	0.4001	0.2024	0.0962	--	3.586	8.943

(g) L - ℓ_1 - ℓ_2 even, 1s - 2s - 2p eigenstates coupling approximation

Singlet										
k_2	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_{ES}	---
0.24	7.852	13.45	38.56	21.41	15.44	8.610	6.432	--	111.75	---
0.50	2.470	5.026	7.433	4.900	3.201	2.101	1.417	--	26.548	
0.68	1.344	2.283	3.025	2.683	1.756	1.166	0.7903	--	13.047	
0.83	0.7424	1.316	1.579	1.518	1.079	0.7580	0.5313	--	7.524	
1.23	0.2813	0.6752	0.6254	0.3663	0.2726	0.2289	0.1916	--	2.641	
1.80	0.1357	0.3339	0.2862	0.1816	0.1065	0.0659	0.0471	0.0383	1.1952	
2.87	0.0513	0.1053	0.0930	0.0739	0.0553	0.0391	0.0268	0.0181	0.4628	
3.91	0.0267	0.0451	0.0422	0.0373	0.0318	0.0261	0.0212	0.0166	0.2470	
Triplet										
k_2	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	L = 7	Σ_{ET}	Q_T
0.24	27.90	75.79	63.10	87.21	41.31	26.12	17.13	--	338.56	518.92
0.50	4.337	4.540	20.66	21.62	10.38	6.373	4.255	--	72.17	122.01
0.68	2.823	4.018	10.75	8.986	5.149	3.337	2.302	--	37.37	67.46
0.83	2.032	3.674	6.400	4.796	2.950	2.013	1.468	--	23.333	44.54
1.23	0.9741	2.290	2.465	1.599	0.9568	0.6421	0.4924	--	9.419	20.11
1.80	0.4291	1.055	0.9303	0.6614	0.4150	0.2540	0.1639	0.1180	4.027	9.531
2.87	0.1564	0.3217	0.2907	0.2330	0.1780	0.1291	0.0907	0.0623	1.4619	3.945
3.91	0.0805	0.1366	0.1285	0.1143	0.0982	0.0810	0.0655	0.0524	0.7570	--

the 2s or the 2p channels. The energy, in electron volts, of the incident electron in each channel is given by $E = 13.6\ k^2$, where k could be k_1 or k_2. All cross sections are in units of πa_0^2. In the different tables, Σ is the sum of the partial cross sections calculated. The total cross section Q_T is obtained by adding the contribution of higher partial waves than those calculated using the regular Born approximation; this could easily be done with the help of the table of the Born approximation.

CONCLUSION

The noniterative technique employed here can be applied to a large class of problems containing exchange integrals. The method is particularly useful when the exchange potential is comparable to the direct potential, in which case the convergence of iteration is slow.

ACKNOWLEDGMENTS

The programming of the numerical integration of the radial differential equations on the IBM 7090 computer has been performed by Mr. Edward Sullivan; through his meticulous and systematic programming the solution of the present problem has become available.

I am indebted to Dr. A. Temkin for many fruitful and illuminating discussions. I also wish to thank Prof. Myers of the University of Maryland for a clarifying discussion.

Table 7
2s - 2p Excitation Cross Section.

(a) Born approximation

k_2	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	Σ	$^*Q_T$
0.245	210.45	465.14	449.77	343.68	148.60	209.74	158.31	1985.69	13560
0.500	6.2469	30.729	57.561	64.948	54.469	49.590	40.391	303.935	3465.0
0.678	0.92003	6.6619	16.736	24.356	25.702	24.598	21.122	120.096	1930.9
0.831	0.24580	2.1480	6.4066	11.011	13.641	14.135	12.473	60.560	1308.4
1.225	0.02039	0.20887	0.78437	1.7481	2.7859	3.6015	3.9501	13.099	620.51
1.803	0.00187	0.01900	0.07823	0.20201	0.38645	0.59807	0.78354	2.0692	294.95
2.872	0.00011	0.00107	0.00444	0.01211	0.02545	0.04450	0.06688	0.15456	120.26
3.905	0.00002	0.00017	0.00067	0.00182	0.00389	0.00699	0.01096	0.02452	66.509

(b) Exchange neglected 1s - 2s - 2p eigenstates coupling approximation

k_2	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	Σ	Q_T
0.245	5.311	12.59	23.21	--	--	--	--	41.11	12476
0.500	0.8651	10.55	1.143	18.92	26.25	--	--	57.73	3308.8
0.678	1.150	5.907	0.5760	7.168	12.28	14.46	15.08	56.62	1867.4
0.831	1.249	2.859	0.3831	3.158	6.664	8.591	9.391	32.295	1280.1
0.225	0.3847	0.4266	0.0881	0.4560	1.403	2.328	3.022	8.108	615.52
1.803	0.0654	0.0516	0.0142	0.0553	0.1846	0.3930	0.6183	1.3824	294.26

(c) 1s - 2s - 2p eigenstates coupling approximation

Singlet

k_2	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	Σ_S	---
0.245	2.243	4.424	3.276	--	--	--	--	9.943	---
0.500	0.1241	1.605	1.348	6.360	7.159	--	--	16.596	
0.678	0.0362	1.446	0.5518	3.056	3.693	3.911	--	12.694	
0.831	0.1866	0.9881	0.2615	1.488	2.144	2.433	2.505	10.006	
1.225	0.1048	0.1516	0.0384	0.1639	0.4584	0.7031	0.8529	2.473	
1.803	0.0175	0.0157	0.0046	0.0148	0.0530	0.1108	0.1726	0.3890	
2.872	0.0014	0.0010	0.0004	0.0010	0.0032	0.0075	0.0133	0.0278	
3.905	0.0002	0.0002	0.0001	0.0002	0.0005	0.0012	0.0019	0.0043	

Triplet

k_2	L = 0	L = 1	L = 2	L = 3	L = 4	L = 5	L = 6	Σ_T	$\Sigma_S + \Sigma_T$	Q_T
0.245	0.0000	10.40	56.74	--	--	--	--	67.14	77.083	12512
0.500	2.322	7.125	2.357	3.838	17.01	--	--	32.652	49.248	3300.3
0.678	1.590	2.363	0.3333	1.442	6.810	9.868	--	22.406	35.100	1867.0
0.831	0.9885	1.125	0.1226	0.8444	3.411	5.518	6.556	18.566	28.572	1276.4
1.225	0.2648	0.2173	0.0402	0.2297	0.7623	1.423	1.987	4.924	7.397	614.8
1.803	0.0455	0.0310	0.0094	0.0383	0.1223	0.2544	0.4070	0.9079	1.297	294.18
2.872	0.0040	0.0024	0.0009	0.0031	0.0095	0.0208	0.0359	0.0766	0.1044	120.21
3.905	0.0007	0.0004	0.0002	0.0005	0.0015	0.0035	0.0063	0.0131	0.0174	66.502

REFERENCES

1. Marriott, R., "Calculation of the 1s–2s Electron Excitation Cross Section of Hydrogen," *Proc. Phys. Soc.* (*London*) 72(1):121–129, July 1958.
2. Smith, K., "Elastic and Inelastic Scattering of Electrons from the S States of Atomic Hydrogen," *Phys. Rev.* 120(3):845–847, November 1, 1960.
3. Percival, I. C., and Seaton, M. J., "The Partial Wave Theory of Electron-Hydrogen Atomic Collisions," *Proc. Cambridge Phil. Soc.* 53(3):654–662, July 1957.
4. Burke, P. G., and Smith, K., "The Low Energy Scattering of Electrons and Positrons by Hydrogen Atoms," *Rev. Mod. Phys.* 34(3):458–502, July 1962.
5. Burke, P. G., Schey, H. M., and Smith, K., "Collisions of Slow Electrons and Positrons with Atomic Hydrogen," *Phys. Rev.* 129(3):1258–1274, February 1, 1963.
6. Damburg, R., and Peterkop, R., "Application of the Multichannel Effective Range Theory to Electron-Hydrogen Scattering," *Proc. Phys. Soc.* 80:563 and 1073 (1962).
7. Kohn, W., "Variational Methods in Nuclear Collision Problems," *Phys. Rev.* 74:1763–1772, December 15, 1948.
8. Rose, M. E., "Elementary Theory of Angular Momentum," New York: John Wiley, 1957: Eq. 3.7.
9. Condon, E. U., and Shortley, G. H., "The Theory of Atomic Spectra," Section 14³, Oxford: Cambridge, 1953.
10. Brandsen, B. H., and McKee, J. S. C., "The 1s–2s Excitation of Hydrogen by Electron Impact," *Proc. Phys. Soc.* 69A(5):422–425, May 1956.
11. Blatt, J. M., and Weisskopf, V. F., "Theoretical Nuclear Physics," New York: Wiley, 1960: Ch. X.
12. Massey, H. S. W., "Theory of the Scattering of Slow Electrons," *Rev. Mod. Phys.* 28(3):199–213, July 1956.

13. SEATON, M. J., "Strong Coupling in Optically Allowed Atomic Transitions Produced by Electron Impact," *Proc. Phys. Soc.* 77A(1):174–183, January 1961.
14. SCHIFF, L. I., "Quantum Mechanics," 1st ed., New York: McGraw-Hill Book Company, 1949: Ch. 4.
15. OMIDVAR, K., "Non-Adiabatic Treatment of Elastic Scattering of Electrons by Atomic Hydrogen," Institute of Mathematical Sciences, New York University, New York, CX–37, January 1959: 22.
16. MOTT, N. F., and MASSEY, H. S. W., "The Theory of Atomic Collisions," 2nd ed., Oxford: Clarendon Press, 1949: Ch. 2.
17. MILNE, W. E., "Numerical Calculus," Princeton: Princeton University Press, 1949: Sec. 40.
18. LICHTEN, W., and SCHULTZ, S., "Cross Sections for the Excitation of the Metastable 2s State of Atomic Hydrogen by Electron Collision," *Phys. Rev.* 116(5): 1132–1139, December 1, 1959.
19. STEBBINGS, R. F., FITE, W. L., HUMMER, D. G., and BRACKMANN, R. T., "Collisions of Electrons with Hydrogen Atoms. V. Excitation of Metastable 2s Hydrogen Atoms," *Phys. Rev.* 119(6):1939–1945, September 15, 1960.
20. TAYLOR, A. J., and BURKE, P. G., "Bulletin of the Third International Conference on the Physics of Electronic and Atomic Collisions," University College London, 22–26 July 1963.
21. KYLE, H. L., and TEMKIN, A., "Bulletin of the Third International Conference on the Physics of Electronic and Atomic Collisions," University College London, 22–26 July 1963.
22. TEMKIN, A., "Nonadiabatic Theory of Electron-Hydrogen Scattering," *Phys. Rev.* 126(1):130–142, April 1,1962.
23. GAILITIS, M., and DAMBURG, R., "The Influence of Close Coupling on the Threshold Behavior of Cross Sections of Electron-Hydrogen Scattering," *Proc. Phys. Soc.*, 82 Pt 2(526):192, 1963.
24. FITE, W. L., and BRACKMANN, R. T., "Collisions of Electrons with Hydrogen Atoms. II. Excitation of Lyman-Alpha Radiation," *Phys. Rev.* 112(4):1151–1156, November 15, 1958.
25. FITE, W. L., STEBBINGS, R. F., and BRACKMANN, R. T., "Collisions of Electrons with Hydrogen Atoms. IV. Excitation of Lyman-Alpha Radiation Near Threshold," *Phys. Rev.* 116(2):356, October 15, 1959.
26. SEATON, M. J., "Cross Sections for 2s–2p Transitions in H and 3s–3p Transitions in Na Produced by Electron and by Proton Impact," *Proc. Phys. Soc.* 68A(6):457–473, June 1955.

Appendix A

THE FOUR DIFFERENTIAL EQUATIONS

$$\left[\frac{d^2}{dr^2} + k_1^2 - \frac{L(L+1)}{r^2} + \frac{2}{r}\right] u(k_1 L, r)$$

$$= 2y_0 (1s\,1s, r)\, u(k_1 L, r) + 2y_0 (1s\,2s, r)\, u(k_2 L, r)$$

$$+ 2\left[\frac{L}{3(2L+1)}\right]^{1/2} y_1 (1s\,2p, r)\; u(k_2 L - 1, r)$$

$$- 2\left[\frac{L+1}{3(2L+1)}\right]^{1/2} y_1 (1s\,2p, r)\; u(k_2 L + 1, r)$$

$$+ \frac{2\beta}{2L+1} rR_{10}(r)\, y_L (1s\,k_1 L, r) + \frac{2\beta}{2L+1} rR_{20}(r)\, y_L (1s\,k_2 L, r)$$

$$+ 2\beta\left[\frac{3L}{(2L+1)(2L-1)^2}\right]^{1/2} rR_{21}(r)\, y_{L-1} (1s\,k_2 L - 1, r)$$

$$- 2\beta\left[\frac{3(L+1)}{(2L+1)(2L+3)^2}\right]^{1/2} rR_{21}(r)\, y_{L+1} (1s\,k_2 L + 1, r)$$

$$- \beta\delta(L, 0)\left(1 + k_1^2\right)\left(1s|k_1 L\right) rR_{10}(r)$$

$$- \beta\delta(L, 0)\left(1 + k_2^2\right)\left(1s|k_2 L\right) rR_{20}(r)$$

$$- \beta\delta(L-1, 0)\left(1 + k_2^2\right)\left(1s|k_2 L - 1\right) rR_{21}(r) \qquad \text{(A1)}$$

$$\left[\frac{d^2}{dr^2} + k_2^2 - \frac{L(L+1)}{r^2} + \frac{2}{r}\right] u(k_2 L, r)$$

$$= 2y_0\ (1s\,2s, r)\, u(k_1 L, r) + 2y_0\ (2s\,2s, r)\, u(k_2 L, r)$$

$$+ 2\left[\frac{L}{3(2L+1)}\right]^{1/2} y_1 (2s\,2p, r)\ u(k_2 L - 1, r)$$

$$- 2\left[\frac{L+1}{3(2L+1)}\right]^{1/2} y_1 (2s\,2p, r)\ u(k_2 L + 1, r)$$

$$+ \frac{2\beta}{2L+1}\, rR_{10}(r)\, y_L (2s\, k_1 L, r) + \frac{2\beta}{2L+1}\, rR_{20}(r)\, y_L (2s\, k_2 L, r)$$

$$+ 2\beta\left[\frac{3L}{(2L+1)(2L-1)^2}\right]^{1/2} rR_{21}(r)\, y_{L-1} (2s\, k_2 L - 1, r)$$

$$- 2\beta\left[\frac{3(L+1)}{(2L+1)(2L+3)^2}\right]^{1/2} rR_{21}(r)\, y_{L+1} (2s\, k_2 L + 1, r)$$

$$- \beta\delta(L, 0)\left(\frac{1}{4} + k_1^2\right) (2s|k_1 L)\, rR_{10}(r)$$

$$- \beta\delta(L, 0)\left(\frac{1}{4} + k_2^2\right) (2s|k_2 L)\, rR_{20}(r)$$

$$- \beta\delta(L-1, 0)\left(\frac{1}{4} + k_2^2\right) (2s|k_2 L - 1)\, rR_{21}(r)\ . \qquad \text{(A2)}$$

$$\left[\frac{d^2}{dr^2} + k_2^2 - \frac{(L-1)L}{r^2} + \frac{2}{r}\right] u(k_2 L - 1, r)$$

$$= 2\left[\frac{L}{3(2L+1)}\right]^{1/2} y_1(1s\,2p, r)\, u(k_1 L, r) + 2\left[\frac{L}{3(2L+1)}\right]^{1/2} y_1(2s\,2p, r)\; u(k_2 L, r)$$

$$+ 2\left[y_0(2p\,2p, r) + \frac{L-1}{5(2L+1)} y_2(2p\,2p, r)\right] u(k_2 L - 1, r)$$

$$- \frac{6}{5}\frac{\sqrt{L(L-1)}}{2L+1} y_2(2p\,2p, r)\, u(k_2 L + 1, r)$$

$$+ 2\beta\left[\frac{3L}{(2L+1)(2L-1)^2}\right]^{1/2} rR_{10}(r)\, y_{L-1}(2p\,k_1 L, r)$$

$$+ 2\beta\left[\frac{3L}{(2L+1)(2L-1)^2}\right]^{1/2} rR_{20}(r)\, y_{L-1}(2p\,k_2 L, r)$$

$$+ \frac{6\beta}{2L-1} rR_{21}(r)\left[\frac{y_L(2p\,k_2 L - 1, r)}{(2L+1)^2} + \frac{L-1}{2L-3} y_{L-2}(2p\,k_2 L - 1, r)\right]$$

$$- \frac{6\sqrt{L(L-1)}}{(2L-1)^2} y_L(2p\,k_2 L + 1, r)\, rR_{-1}(r)$$

$$- \beta\delta(L, 1)\left(\frac{1}{4} + k_1^2\right)(2p|k_1 L)\, rR_{10}(r)$$

$$- \beta\delta(L, 1)\left(\frac{1}{4} + \kappa_2^2\right)(2p|k_2 L)\, rR_{10}(r)$$

$$- \beta\delta(L, 2)\left(\frac{1}{4} + k_2^2\right)(2p|k_2 L - 1)\, rR_{21}(r)\,. \qquad \text{(A3)}$$

$$\left[\frac{d^2}{dr^2} + k_2^2 - \frac{(L+1)(L+2)}{r^2} + \frac{2}{r}\right] u(k_2 L+1, r)$$

$$= -2\left[\frac{L+1}{3(2L+1)}\right]^{1/2} y_1(1s\,2p, r)\, u(k_1 L, r) - 2\left[\frac{L+1}{3(2L+1)}\right]^{1/2} y_1(2s\,2p, r)\, u(k_2 L, r)$$

$$- \frac{6}{5}\frac{\sqrt{L(L+1)}}{2L+1} y_2(2p\,2p, r)\, u(k_2 L-1, r)$$

$$+ 2\left[y_0(2p\,2p, r) + \frac{(L+2)\, y_2(2p\,2p, r)}{5(2L+1)}\right] u(k_2 L+1, r)$$

$$- 2\beta\left[\frac{3(L+1)}{(2L+1)(2L+3)^2}\right]^{1/2} rR_{10}(r)\, y_{L+1}(2p\,k_1 L, r)$$

$$- 2\beta\left[\frac{3(L+1)}{(2L+1)(2L+3)^2}\right]^{1/2} rR_{20}(r)\, y_{L+1}(2p\,k_2 L, r)$$

$$- \frac{6\beta\sqrt{L(L+1)}}{(2L+1)^2} rR_{21}(r)\, y_L(2p\,k_2 L-1, r)$$

$$+ \frac{6\beta}{2L+3} rR_{21}(r)\left[\frac{y_L(2p\,k_2 L+1, r)}{(2L+1)^2} + \frac{(L+2)\, y_{L+2}(2p\,k_2 L+1, r)}{(2L+5)}\right]$$

$$- \beta\delta(L, 0)\left(\frac{1}{4} + k_2^2\right)(2p|k_2 L+1)\, rR_{21}(r)\,. \qquad \text{(A4)}$$

Appendix B

ELEMENTS OF THE POTENTIAL MATRIX

Elements of D_{ij}

$$D_{11} = -\left(1+\frac{1}{r}\right)e^{-2r}, \quad D_{22} = -\left(\frac{1}{r}+\frac{3}{4}+\frac{r}{4}+\frac{r^2}{8}\right)e^{-r},$$

$$D_{33} = -\left[\frac{1}{r}+\frac{3}{4}+\frac{r}{4}+\frac{r^2}{24}\right]e^{-r} + \frac{6(L-1)}{2L+1}\left[\frac{1}{r^3}-\left(\frac{1}{r^3}+\frac{1}{r^2}+\frac{1}{2r}+\frac{1}{6}+\frac{r}{24}+\frac{r^2}{144}\right)e^{-r}\right],$$

$$D_{44} = -\left[\frac{1}{r}+\frac{3}{4}+\frac{r}{4}+\frac{r^2}{24}\right]e^{-r} + \frac{6(L+2)}{2L+1}\left[\frac{1}{r^3}-\left(\frac{1}{r^3}+\frac{1}{r^2}+\frac{1}{2r}+\frac{1}{6}+\frac{r}{24}+\frac{r^2}{144}\right)e^{-r}\right],$$

$$D_{12} = D_{21} = \frac{2\sqrt{2}}{9}\left(r+\frac{2}{3}\right)e^{-3/2\,r},$$

$$D_{13} = D_{31} = \frac{128\sqrt{2}}{243}\times\left(\frac{L}{2L+1}\right)^{1/2}\left[\frac{1}{r^2}-\left(\frac{1}{r^2}+\frac{3}{2r}+\frac{9}{8}+\frac{27r}{64}\right)e^{-3r/2}\right],$$

$$D_{14} = D_{41} = -\frac{128\sqrt{2}}{243}\left(\frac{L+1}{2L+1}\right)^{1/2}\left[\frac{1}{r^2}-\left(\frac{1}{r^2}+\frac{3}{2r}+\frac{9}{8}+\frac{27r}{64}\right)e^{-3r/2}\right],$$

$$D_{23} = D_{32} = -3\left(\frac{L}{2L+1}\right)^{1/2}\left[\frac{1}{r^2}-\left(\frac{1}{r^2}+\frac{1}{r}+\frac{1}{2}+\frac{r}{6}+\frac{r^2}{24}\right)e^{-r}\right],$$

$$D_{24} = D_{42} = 3\left(\frac{L+1}{2L+1}\right)^{1/2}\left[\frac{1}{r^2}-\left(\frac{1}{r^2}+\frac{1}{r}+\frac{1}{2}+\frac{r}{6}+\frac{r^2}{24}\right)e^{-r}\right],$$

$$D_{34} = D_{43} = -18\left[\frac{L(L+1)}{(2L+1)^2}\right]^{1/2}\left[\frac{1}{r^3}-\left(\frac{1}{r^3}+\frac{1}{r^2}+\frac{1}{2r}+\frac{1}{6}+\frac{r}{24}+\frac{r^2}{144}\right)e^{-r}\right].$$

Elements of F_{ij}

$$F_{11} = \frac{\beta}{2L+1}\left[\frac{R_{10}}{r^L}\int_0^r R_{10}\,r'^{L+1}\,dr' - R_{10}\,r^{L+1}\int_0^r \frac{R_{10}}{r'^L}\,dr'\right],$$

$$F_{22} = \frac{\beta}{2L+1}\left[\frac{R_{20}}{r^L}\int_0^r R_{20}\, r'^{L+1}\, dr' - R_{20}\, r^{L+1}\int_0^r \frac{R_{20}}{r'^L}\, dr'\right],$$

$$F_{33} = \frac{3\beta}{2L-1}\left[\frac{1}{(2L+1)^2}\left(\frac{R_{21}}{r^L}\int_0^r R_{21}\, r'^{L+1}\, dr' - R_{21}\, r^{L+1}\int_0^r \frac{R_{21}}{r'^L}\, dr'\right)\right.$$

$$\left. + \frac{L-1}{2L-3}\left(\frac{R_{21}}{r^{L-2}}\int_0^r R_{21}\, r'^{L-1}\, dr' - R_{21}\, r^{L-1}\int_0^r \frac{R_{21}}{r'^{L-2}}\, dr'\right)\right],$$

$$F_{44} = \frac{3\beta}{2L+3}\left[\frac{1}{(2L+1)^2}\left(\frac{R_{21}}{r^L}\int_0^r R_{21}\, r'^{L+1}\, dr' - R_{21}\, r^{L+1}\int_0^r \frac{R_{21}}{r'^L}\, dr'\right)\right.$$

$$\left. + \frac{L+2}{2L+5}\left(\frac{R_{21}}{r^{L+2}}\int_0^r R_{21}\, r'^{L+3}\, dr' - R_{21}\, r^{L+3}\int_0^r \frac{R_{21}}{r'^{L+2}}\, dr'\right)\right],$$

$$F_{12} = \frac{\beta}{2L+1}\left[\frac{R_{20}}{r^L}\int_0^r R_{10}\, r'^{L+1}\, dr' - R_{20}\, r^{L+1}\int_0^r \frac{R_{10}}{r'^L}\, dr'\right],$$

$$F_{21} = F_{12}\left[R_{10} \rightleftarrows R_{20}\right],$$

$$F_{13} = \sqrt{3}\beta\left[\frac{L}{(2L+1)(2L-1)^2}\right]^{1/2} \times \left[\frac{R_{21}}{r^{L-1}}\int_0^r R_{10}\, r'^L\, dr - R_{21}\, r^L\int_0^r \frac{R_{10}}{r'^{L-1}}\, dr'\right],$$

$$F_{31} = F_{13}\left[R_{10} \rightleftarrows R_{21}\right],$$

$$F_{14} = -\sqrt{3}\beta\left[\frac{L+1}{(2L+1)(2L+3)^2}\right]^{1/2} \times \left[\frac{R_{21}}{r^{L+1}}\int_0^r R_{10}\, r'^{L+2}\, dr' - R_{21}\, r^{L+2}\int_0^r \frac{R_{10}}{r'^{L+1}}\, dr'\right],$$

$$F_{41} = F_{14}\left[R_{10} \rightleftarrows R_{21}\right],$$

$$F_{23} = \sqrt{3}\beta\left[\frac{L}{(2L+1)(2L-1)^2}\right]^{1/2} \times \left[\frac{R_{21}}{r^{L-1}}\int_0^r R_{20}\, r'^L\, dr' - R_{21}\, r^L\int_0^r \frac{R_{20}}{r'^{L-1}}\, dr'\right],$$

$$F_{32} = F_{23}\left[R_{20} \rightleftarrows R_{21}\right],$$

$$F_{24} = -\sqrt{3}\beta\left[\frac{L+1}{(2L+1)(2L+3)^2}\right]^{1/2} \times \left[\frac{R_{21}}{r^{L+1}}\int_0^r R_{20}\, r'^{L+2}\, dr' - R_{21}\, r^{L+2}\int_0^r \frac{R_{20}}{r'^{L+1}}\, dr'\right],$$

$$F_{42} = F_{24}\left[R_{20} \rightleftarrows R_{21}\right] ,$$

$$F_{34} = -3\beta\left[\frac{L(L+1)}{(2L+1)^4}\right]^{1/2} \times \left[\frac{R_{21}}{r^L}\int_0^r R_{21}\, r'^{L+1}\, dr' - R_{21}\, r^{L+1}\int_0^r \frac{R_{21}}{r'^L}\, dr'\right] ,$$

$$F_{34} = F_{43} .$$

Elements of g_{ij} and h_{ij}

$$g_{11} = \frac{\beta R_{10}\, r^{L+1}}{2L+1} , \quad h_{11} = R_{10}\left[\frac{1}{r^L} - \frac{1+k_1^2}{2}\,\delta(L, 0)\, r\right] ,$$

$$g_{22} = \frac{\beta R_{20}\, r^{L+1}}{2L+1} , \quad h_{22} = R_{20}\left[\frac{1}{r^L} - \frac{\frac{1}{4}+k_2^2}{2}\,\delta(L, 0)\, r\right] ,$$

$$g_{33}^1 = \frac{3\beta R_{21}\, r^{L+1}}{(2L-1)(2L+1)^2} , \quad h_{33}^1 = \frac{R_{21}}{r^L} ,$$

$$g_{33}^2 = \frac{3\beta(L-1) R_{21}\, r^{L-1}}{(2L-1)(2L-3)} , \quad h_{33}^2 = R_{21}\left[\frac{1}{r^{L-2}} - \frac{\frac{1}{4}+k_2^2}{2}\,\delta(L, 2)\, r\right] ,$$

$$g_{44}^1 = \frac{3\beta R_{21}\, r^{L+1}}{(2L+3)(2L+1)^2} , \quad h_{44}^1 = R_{21}\left[\frac{1}{r^L} - \frac{\frac{1}{4}+k_2^2}{2}\,\delta(L, 0)\, r\right] ,$$

$$g_{44}^2 = \frac{3\beta(L+2) R_{21}\, r^{L+3}}{(2L+3)(2L+5)} , \quad h_{44}^2 = \frac{R_{21}}{r^{L+2}} ,$$

$$g_{12} = \frac{\beta}{2L+1} R_{20}\, r^{L+1} , \quad h_{12} = R_{10}\left[\frac{1}{r^L} - \frac{1+k_2^2}{2}\,\delta(L, 0)\, r\right] ,$$

$$g_{21} = g_{12}\left[R_{20} \longrightarrow R_{10}\right] , \quad h_{21} = h_{12}\left[R_{10} \longrightarrow R_{20}\right] ,$$

$$g_{13} = \sqrt{3}\beta\left[\frac{L}{(2L+1)(2L-1)^2}\right]^{1/2} R_{21}\, r^L , \quad h_{13} = R_{10}\left[\frac{1}{r^{L-1}} - \frac{1+k_2^2}{2}\,\delta(L, 1)\, r\right] ,$$

$$g_{31} = g_{13}\left[R_{21} \longrightarrow R_{10}\right] , \quad h_{31} = h_{13}\left[R_{10} \longrightarrow R_{21}\right] ,$$

$$g_{14} = -\sqrt{3}\beta\left[\frac{L+1}{(2L+1)(2L+3)^2}\right]^{1/2} R_{21}\, r^{L+2} , \quad h_{14} = \frac{R_{10}}{r^{L+1}} ,$$

$$g_{41} = g_{14}\left[R_{21} \longrightarrow R_{10}\right] , \quad h_{41} = h_{14}\left[R_{10} \longrightarrow R_{21}\right] ,$$

$$g_{23} = \sqrt{3}\beta \left[\frac{L}{(2L+1)(2L-1)^2}\right]^{1/2} R_{21} r^L , \quad h_{23} = R_{20}\left[\frac{1}{r^{L-1}} - \frac{\frac{1}{4} + k_2^2}{2}\delta(L\ 1)\, r\right] ,$$

$$g_{32} = g_{23}\left[R_{21} \longrightarrow R_{20}\right] , \quad h_{32} = h_{23}\left[R_{20} \longrightarrow R_{21}\right] ,$$

$$g_{24} = -\sqrt{3}\beta\left[\frac{L+1}{(2L+1)(2L+3)^2}\right]^{1/2} R_{21} r^{L+2} , \quad h_{24} = \frac{R_{20}}{r^{L+1}} ,$$

$$g_{42} = g_{24}\left[R_{21} \longrightarrow R_{20}\right] , \quad h_{42} = h_{24}\left[R_{20} \longrightarrow R_{21}\right] ,$$

$$g_{34} = -3\beta\left[\frac{L(L+1)}{(2L+1)^4}\right]^{1/2} R_{21} r^{L+1} , \quad h_{34} = \frac{R_{21}}{r^L} ,$$

$$g_{43} = g_{34} .$$

In the F_{ij} matrix the interchange of the functions R_{10}, R_{20}, and R_{21} accompanies the interchange of their arguments too.

Appendix C

REMOVAL OF THE SINGULARITY IN THE DETERMINANT OF THE COEFFICIENTS OF THE LINEAR TRANSFORMATION FOR L=0, 1

Case I: L=0

By using the definition of D_{ij} and F_{ij} and Equations 63 in the text, the following relation can be derived from Equation 59 (text):

$$\int_0^\infty \left[rR_{20}\left(\frac{d^2}{dr^2} + k_1^2\right) v_1 - \beta\, rR_{10}\left(\frac{d^2}{dr^2} + k_2^2\right) v_2 \right] dr = -\frac{2}{\sqrt{3}} \left[a_{13}\, B_{24} - \beta\, a_{23}\, B_{14} \right] , \tag{C1}$$

where the superscript μ is suppressed when there is only one value for μ, and

$$a_{13} = \int_0^\infty R_{10}\, R_{21}\, r^3\, dr = \left[2^{15} \times 3^{-9} \right]^{1/2} ,$$

$$a_{23} = \int_0^\infty R_{20}\, R_{21}\, r^3\, dr = -3\sqrt{3} .$$

Integrating the left-hand side of Equation C1 by parts, and making use of Equations 11 and 63 (text), we obtain

$$\int_0^\infty \left[rR_{20}\left(\frac{d^2}{dr^2} + k_1^2\right) v_1 - \beta\, rR_{10}\left(-\frac{d^2}{dr^2} + k_2^2\right) v_2 \right] dr = -2 \left[B_{21} - \beta B_{12} \right] . \tag{C2}$$

We conclude that

$$B_{21} - \beta B_{12} = \frac{1}{\sqrt{3}} \left[a_{13} B_{24} - \beta\, a_{23} B_{14} \right] . \tag{C3}$$

Equation C3 connects the right-hand sides of the four equations of Equation 62 specified by ij = 21, 12, 24, 14. A similar relation should hold among the left-hand sides of these equations. This in fact is the case and, by making use of the first of Equations 63, it can be shown directly that equations similar to Equation C3 hold among the elements of each column k1ν of the left-hand side of Equation 62 specified by ij = 21, 12, 24, 14. We conclude that one equation of Equation 62 is linearly dependent on others and the determinant of Equation 62 is singular.

Case II: L = 1

Similar to the previous case, the following relation can be derived from Equation 59 in the text:

$$\int_0^\infty \left[rR_{21}\left(\frac{d^2}{dr^2} + k_1^2 - \frac{2}{r^2}\right) v_1 - \beta\, rR_{10}\left(\frac{d^2}{dr^2} + k_2^2\right) v_3 \right] dr$$

$$= -\frac{2}{3}\left[\beta a_{13} B_{11} + \beta a_{23} B_{12} - a_{13} B_{33}^1 + \sqrt{2}\left(a_{13} B_{34} - \frac{3}{5}\beta a_{33} B_{14}\right)\right] , \qquad \text{(C4)}$$

where

$$a_{33} = \int_0^\infty R_{21}^2\, r^4\, dr = 30 .$$

Integrating the left-hand side of Equation C4 by parts, and making use of Equations 11 and 63, we obtain

$$\int_0^\infty \left[rR_{21}\left(\frac{d^2}{dr^2} + k_1^2 - \frac{2}{r^2}\right) v_1 - \beta\, rR_{10}\left(\frac{d^2}{dr^2} + k_2^2\right) v_3 \right] dr = -2\left[B_{31} - \beta B_{13}\right] . \qquad \text{(C5)}$$

Combining Equations C4 and C5, we get

$$B_{31} - \beta B_{13} = \frac{1}{3}\left[\beta a_{13} B_{11} + \beta a_{23} B_{12} - a_{13} B_{33}^1 + \sqrt{2}\left(a_{13} B_{34} - \frac{3}{5}\beta a_{33} B_{14}\right)\right] . \qquad \text{(C6)}$$

Finally, Equation 59 (text) gives the following relation:

$$\int_0^\infty \left[rR_{21}\left(\frac{d^2}{dr^2} + k_2^2 - \frac{2}{r^2}\right) v_2 - \beta\, rR_{20}\left(\frac{d^2}{dr^2} + k_2^2\right) v_3 \right] dr$$

$$= -\frac{2}{3}\left[\beta a_{23} B_{22} + \beta a_{13} B_{21} - a_{23} B_{33}^1 + \sqrt{2}\left(a_{23} B_{34} - \frac{3}{5}\beta a_{33} B_{24}\right)\right] . \qquad \text{(C7)}$$

Integration by parts of the left-hand side gives, as before,

$$\int_0^\infty \left[rR_{21}\left(\frac{d^2}{dr^2} + k_2^2 - \frac{2}{r^2}\right) v_2 - \beta\, rR_{20}\left(\frac{d^2}{dr^2} + k_2^2\right) v_3 \right] dr = -2\left[B_{32} - \beta B_{23}\right] , \qquad \text{(C8)}$$

whereupon we get

$$B_{32} - \beta B_{23} = \frac{1}{3}\left[\beta a_{23} B_{22} + \beta a_{13} B_{21} - a_{23} B_{33}^{1} + \sqrt{2}\left(a_{23} B_{34} - \frac{3}{5}\beta a_{33} B_{24}\right)\right] \tag{C9}$$

Similar to the case L=0, Equations C6 and C9 indicate that two equations of Equation 62 are linearly dependent on others and the determinant of Equation 62 is singular.

To remove the singularity in the L=0 case, one of the C_{kl}^{ν} is chosen to be arbitrary; and a degenerate equation is removed from Equation 62(text). Similarly, in the L=1 case two of the C_{kl}^{ν} are chosen arbitrary; and two degenerate equations are removed from Equation 62.

Appendix D

ELEMENTS OF THE MATRIX OF THE SUM OF THE ASYMPTOTIC COULOMB AND CENTRIFUGAL POTENTIALS

$$U_{11} = L(L+1)\,r^{-2}, \quad U_{22} = L(L+1)\,r^{-2},$$

$$U_{33} = (L-1)\,Lr^{-2} + 12(L-1)\,(2L+1)^{-1}\,r^{-3}, \quad U_{44} = (L+1)\,(L+2)\,r^{-2} + 12(L+2)\,(2L+1)^{-1}\,r^{-3},$$

$$U_{12} = U_{21} = 0, \quad U_{13} = U_{31} = \left[\frac{256\sqrt{2}}{243}\right]\left[\frac{L}{(2L+1)}\right]^{1/2} r^{-2},$$

$$U_{14} = U_{41} = -\left[\frac{256\sqrt{2}}{243}\right]\left[\frac{(L+1)}{(2L+1)}\right]^{1/2} r^{-2}, \quad U_{23} = U_{32} = -6\left[\frac{L}{(2L+1)}\right]^{1/2} r^{-2},$$

$$U_{24} = U_{42} = 6\left[\frac{(L+1)}{(2L+1)}\right]^{1/2} r^{-2}, \quad U_{34} = U_{43} = -36\left[L(L+1)\right]^{1/2}\,(2L+1)^{-1}\,r^{-3}.$$

see N64-22748

EFFECTS OF NON-PHOTOCHEMICAL PROCESSES ON THE MERIDIONAL DISTRIBUTION AND TOTAL AMOUNT OF OZONE IN THE ATMOSPHERE*

CUDDAPAH PRABHAKARA

The photochemical equilibrium distribution of ozone below 41 km is computed for the solstices and equinoxes from recent rocket observations of the solar ultraviolet energy received at the top of the atmosphere.

In this study an attempt is made to investigate the important part played by the mean seasonal transport processes in distributing the ozone below 41 km in the stratosphere. To make such an investigation it is assumed that a steady state meridional distribution of ozone is produced when photochemical and transport processes are acting together. Relevant theory and numerical procedure to compute such a steady state meridional distribution of ozone have been developed.

In order to make a model of the transport processes that is needed for the calculation of the distribution of ozone, theoretical and observational information available at present on the various transport processes present in the stratosphere is reviewed. Further, in the calculations of ozone an attempt is made to combine the mean meridional motions and the large scale mixing processes within the limits acceptable to the current ideas on these transport processes.

The seasonal distribution of total ozone thus computed from the model of the transport processes and with an assumption of steady state condition shows a satisfactory agreement with the observed seasonal and latitudinal variation of total ozone. The mean annual total ozone variation with latitude deduced from the computed seasonal distribution of total ozone demonstrates the ability of the method developed here to account for the mean annual total ozone variation with latitude observed in the earth's atmosphere.

The model of the transport processes developed in this study to calculate the distribution of ozone in the atmosphere emphasizes the importance of the large scale eddy diffusion processes. The theoretical and observational findings from the spread of W^{185} (a radioactive isotope of tungsten introduced into the stratosphere by nuclear explosions) have been very useful in making this model of the transport processes.

INTRODUCTION

Our knowledge of the distribution of ozone in the atmosphere is derived largely from the observations of the vertical distribution of ozone at comparatively few and widely scattered stations on the earth.

The Umkehr method (Götz, Reference 1) of obtaining the distribution of ozone was widely used until recently. From the Umkehr method standard curves of the vertical distribution of ozone† at a given station for different amounts of total ozone contained in a vertical column of unit area were constructed. Then, using the observed total amount of ozone at the station for any given day, the vertical distribution of ozone for the day was read off these standard curves. Ramanathan and Kulkarni (Reference 2) have constructed the meridional distribution of ozone using such a technique on the data obtained at Kodaikanal, Poona, Mount Abu, Delhi, Srinagar, Tetano, Arosa, Tromsö, and College, Alaska. The models of the distribution of ozone given by Ramanathan and Kulkarni from the equator to 80°N latitude for November and July are combined as shown in Figure 1 to yield a distribution of ozone from one pole to the other.

Besides the Umkehr method there have been other methods developed to measure the vertical distribution of ozone. Regener (Reference 3), using an ozone-sonde based on the principle of

*Published as *NASA Technical Note D-2106*, June 1964. This work was also published in the *Monthly Weather Review*, 91(9):411-431, September 1963.

†Vertical distribution of ozone is usually expressed in Dobson Units: D.U. $=10^{-3}$cm of ozone at standard temperature and pressure. Also D.U. $=2.69\times10^{16}$ molecules of ozone/cm^2.

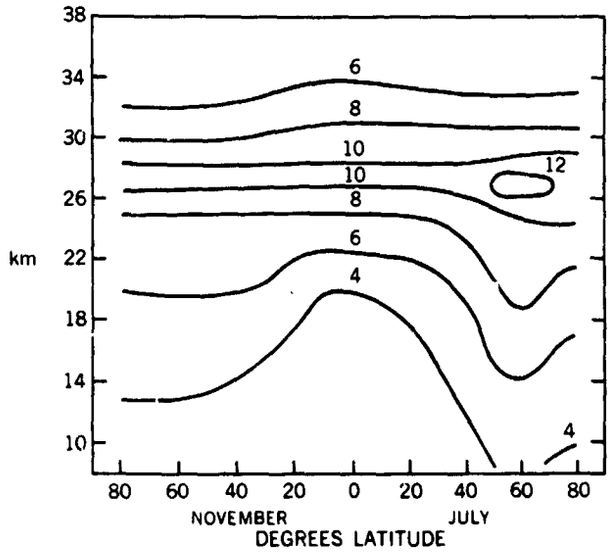


FIGURE 1.— Model of the "observed" meridional distribution of ozone (1.33xD.U./km), constructed after Ramanathan and Kulkarni (Reference 2).

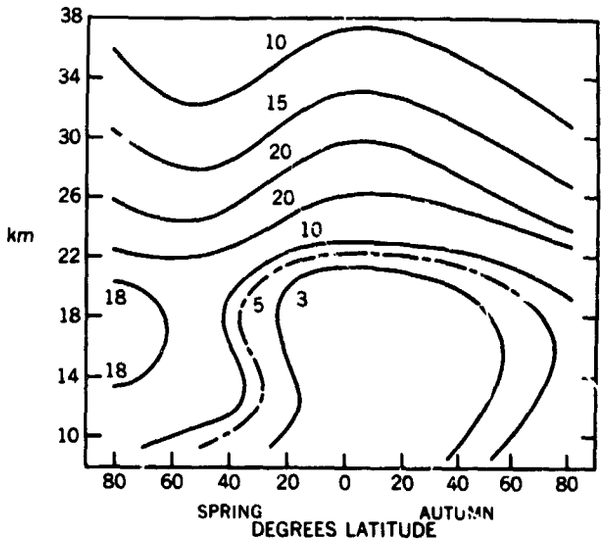


FIGURE 2.—Model of the "observed" meridional distribution of ozone (D.U./km) after Paetzold and Piscalar (Reference 7).

photographic photometry, has measured the vertical distribution of ozone at Albuquerque. Using the observations of the emission and absorption of infrared radiation by ozone in the 9.6μ band, Epstein et al. (Reference 4) at Flagstaff, Arizona, and Goody and Roach (Reference 5) at Oxford, England, have derived the vertical distribution of ozone. Brewer (Reference 6) has made measurements of the vertical distribution of ozone at Liverpool and Malta with a chemical ozone-sonde. It is very difficult to put together such measurements obtained by different methods and to make a model of the meridional distribution of ozone.

Very recently Paetzold and Piscalar (Reference 7) using the observations of the vertical distribution of ozone made during the I.G.Y. with an optical radio ozone-sonde at four stations—Tromsö, Weissenau, Leopoldville, and Denver—have constructed a meridional distribution of ozone for spring and fall, which is presented here as Figure 2.

Both these models of the "observed" meridional distribution of ozone, the construction of which could not be attempted until recently, are still based on data collected from a few stations. Nevertheless, these models bring out the salient features of the atmospheric ozone.

Ozone is produced in the earth's atmosphere by the photochemical reactions induced by the solar ultraviolet radiation (Chapman, Reference 8). Although the photochemical reactions (referred to hereafter as photochemical theory) lead to the formation of ozone in the atmosphere, the observed global distribution of ozone cannot be completely explained on the basis of photochemical theory alone.

Total ozone computed from the photochemical theory is short of the observed total amounts at higher latitudes, and it is in excess of the amounts observed at the equatorial latitudes (Craig, Reference 9). Photochemical theory would lead to summer maximum and winter minimum of total ozone at all latitudes, while the observations show a maximum in spring and a minimum in autumn at high latitudes as shown in Figure 3 (London, Reference 10). Small seasonal variations are observed at the equatorial latitudes (Ramanathan, Reference 11).

The vertical distribution of ozone concentration according to photochemical theory has a maximum at about 30 km height in the atmosphere, and the position of maximum is higher, the higher the latitude (Figure 12). On the contrary, observations on the vertical distribution of ozone show that the position of the maximum is usually lower when the latitude is higher, leading to a lowering of the center of gravity of the ozone layer with increase in latitude (Figure 4).

Photochemical studies made by previous investigators (see, e.g., Wulf and Deming, Reference

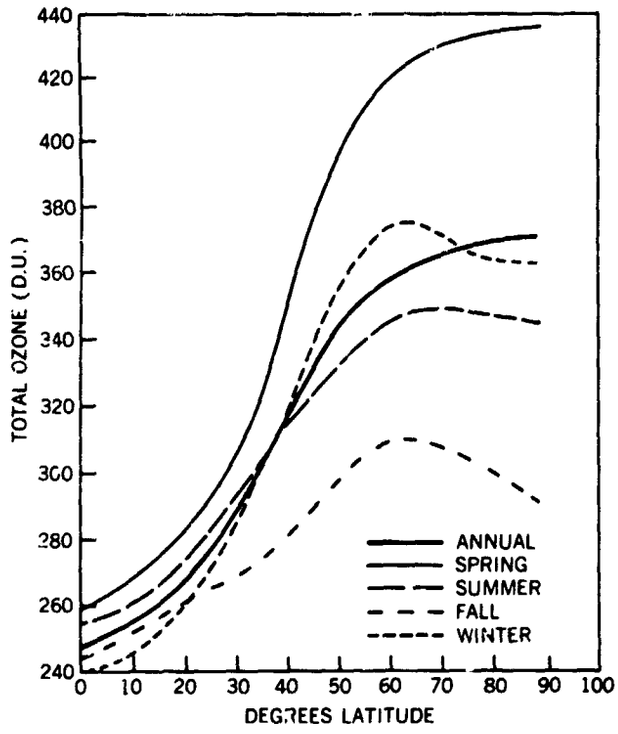


FIGURE 3.—Total ozone variation with latitude and season, after London (Reference 10).

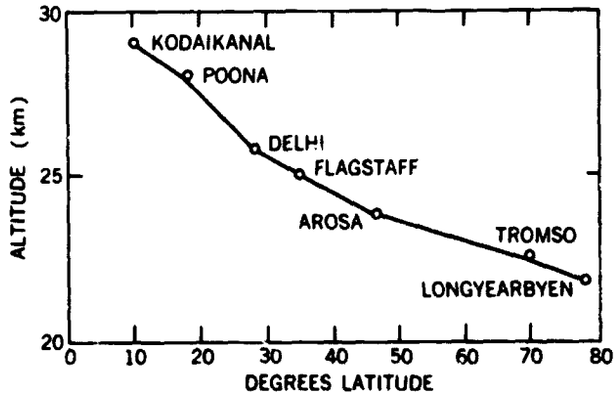


FIGURE 4.—Latitudinal variation of ozone center of gravity after Miller (Reference 12).

13; Craig, Reference 9; Dütsch, Reference 14) showed that the half-life of ozone was of the order of a day at heights above about 30 km, and that photochemical conditions could prevail above that level. However, below that level the production of ozone by solar ultraviolet rapidly decreased, and any departures from photochemical equilibrium conditions could persist for long periods of time. Arguing on such grounds to explain the seasonal and latitudinal variations of ozone, Craig suggested that there should be "some atmospheric process which transports ozone from regions where it can be quickly reformed (high levels or equatorial regions) to lower protected levels where it can persist." He further commented that"... this process must be most effective in middle and high latitudes in winter, when the total amount of ozone increases most rapidly."

Dütsch (Reference 15) first made an attempt to introduce changes of ozone due to Austausch and circulation and to obtain a seasonal variation of ozone as a function of latitude. Craig (Reference 9) remarked that the seasonal and latitudinal variations obtained by Dütsch were qualitatively similar to the observed ones, but of extremely small amplitude (about 20 percent of the observed latitudinal amplitude and about 5 to 10 percent of the observed seasonal amplitude at high latitudes).

Objective of the Present Study

From the earlier discussion it may be seen that transport by advection and diffusion on different space and time scales plays a significant role in producing the geographical distribution of ozone. At present it is difficult to consider elaborately the effects of small scale phenomena owing to the lack of information on nature of these transport processes on small space and time scales in the stratosphere.

In this study it is intended to compute the photochemical distribution of ozone, using the best available data needed for such computations, as a function of latitude and season. This is done as a basic step before proceeding to examine the influence of transport processes on the distribution of ozone.

The measurements of total ozone in the atmosphere show clearly an annual cycle of the ozone variation with the largest and the least amounts in spring and fall respectively. The striking contrast between the mean seasonal ozone amounts in the atmosphere has been demonstrated by Ramanthan (Reference 11) and London (Reference 10). Such mean seasonal features of total ozone suggest that the part played by the mean seasonal transport processes in distributing the ozone in the lower stratosphere, where the bulk of

ozone present is under weak solar control, could be of paramount importance. This influence of the mean seasonal transport processes on ozone in the lower stratosphere is probably capable of masking the direct influence of solar radiation, the ultimate source of ozone. Hence in this study it is intended to make an experiment on the calculation of the meridional distribution of ozone produced under steady state condition by transport processes and solar radiation. Further, it is intended to see if such an approach can explain the seasonal variation of ozone in the earth's atmosphere.

In this study an attempt is made to develop the equations necessary for calculating the distribution of ozone in the atmosphere under steady state conditions when photochemistry is coupled with local change of ozone due to mean seasonal meridional motions and large scale mixing processes. Further, the numerical procedure to solve such equations will be developed.

Posing the problem of atmospheric ozone as a steady state or boundary value problem has a unique advantage; namely, it is not necessary to specify the initial distribution of ozone in the atmosphere to get a steady state solution, so that the solution is not subject to the uncertainties of the initial distribution.

The mean seasonal transport processes in the stratosphere needed for this study will be deduced from the recent available observational and theoretical information.

THEORY

Basic Photochemistry of Ozone

The photochemical reactions involved in producing the ozone layer below 50 km are (see, e.g., Craig, Reference 9; Dütsch, Reference 14):

$$O_2 + h\nu \longrightarrow O_1 + O_1 , \tag{1}$$

(λ < 2400 A, oxygen absorption in Schumann-Runge bands and Herzberg continuum)

$$O_1 + O_2 + M \xrightarrow{K_2} O_3 + M , \tag{2}$$

$$O_3 + O_1 \xrightarrow{K_3} 2O_2 , \tag{3}$$

$$O_3 + h\nu \longrightarrow O_1 + O_2 , \tag{4}$$

(λ < 11,000 A, ozone absorption in Hartley, Huggins, and Chappuis bands)

where O_1, O_2, O_3, and M stand for number of atomic oxygen, molecular ozygen, ozone, and total number of molecules per cm³. Symbol h is the Planck's constant and ν, the frequency of radiation. K_2 (cm⁶ molec⁻² sec⁻¹) and K_3 (cm³ molec⁻¹ sec⁻¹) are the rate coefficients of the Reactions 2 and 3.

The above four reactions give the rate of growth of O_1 and O_3 as

$$\frac{\partial O_1}{\partial t} = 2f_2O_2 + f_3O_3 - K_2O_1O_2M - K_3O_1O_3 , \tag{5}$$

$$\frac{\partial O_3}{\partial t} = -f_3O_3 + K_2O_1O_2M - K_3O_1O_3 , \tag{6}$$

where f_2 and f_3 stand for the number of quanta of energy absorbed per second per molecule of oxygen and ozone.

Under equilibrium conditions

$$\frac{\partial O_1}{\partial t} = \frac{\partial O_3}{\partial t} = 0 . \tag{7}$$

With the above condition we have from Equation 5

$$O_1 = \frac{2f_2O_2 + f_3O_3}{K_2O_2M + K_3O_3} . \tag{8}$$

Substituting for O_1 from Equation 8 into (6) (and neglecting O_3 in comparison to (K_2/K_3)

(O_2M) we get the photochemical equilibrium ozone, O_{3e}:

$$O_{3e} = O_2\sqrt{\frac{K_2}{K_3}\frac{f_2}{f_3}M} . \quad (9)$$

Effects of Transport Processes

General Considerations

The photochemical equations are developed by considering only the solar radiation and its direct influence on the atmosphere. However, if transport processes such as advection and diffusion are present in the atmosphere, they tend to redistribute the ozone produced by solar radiation. In such a redistribution, regions of the atmosphere where photochemical production of ozone is rapid act like sources, and regions where the photochemical production is weak constitute sinks of ozone.

Before introducing the effects of transport processes into the growth equation for ozone, it is intended to develop basic ideas necessary for such a study by considering the influence on ozone of an external input at the rate of F molecules of O_3/cm^3–sec. The growth Equation 6 for ozone now becomes

$$\frac{\partial O_3}{\partial t} = -f_3O_3 + K_2O_2O_1M - K_3O_3O_1 + F . \quad (10)$$

Substituting for O_1 from 8 into 10, we have

$$\frac{\partial O_3}{\partial t} = -f_3O_3 + \{K_2O_2M - K_3O_3\}\frac{(2f_2O_2 + f_3O_3)}{(K_2O_2M + K_3O_3)} + F . \quad (11)$$

We divide the above equation by K_3O_3 and write N_3^* for K_2O_2M/K_3 and get

$$\frac{\partial O_3}{\partial t}\left[\frac{N_3^*}{O_3} + 1\right] = -2f_3O_3 + F\left[1 + \frac{N_3^*}{O_3}\right] + 2f_2O_2\left[\frac{N_3^*}{O_3} - 1\right] . \quad (12)$$

Now the external input at the rate of F molecules of O_3/cm^3–sec and the photochemical process tend to a new equilibrium of ozone O_{3E} given by 13, when $\partial O_3/\partial t = 0$:

$$2f_3O_{3E} - F\left\{1 + \frac{N_3^*}{O_{3E}}\right\} - 2f_2O_2\left\{\frac{N_3^*}{O_{3E}} - 1\right\} = 0. \quad (13)$$

Since $N_3^*/O_{3E} \gg 1$, as was stated in arriving at Equation 9, we can write Equation 13 as

$$2f_3O_{3E} - F\frac{N_3^*}{O_{3E}} - 2f_2O_2\frac{N_3^*}{O_{3E}} = 0 \quad (14)$$

or

$$O_{3E} = \sqrt{\frac{(F + 2f_2O_2)}{2f_3}N_3^*} . \quad (15)$$

Dütsch (Reference 15) first arrived at this equation. Equation 15 may be transformed with the help of Equation 9 as

$$O_{3E} = O_{3e}\sqrt{1 + \frac{F}{2f_2O_2}} . \quad (16)$$

Equation 16 shows that an input of F molecules of O_3/cm^3–sec does not indefinitely increase the amount of ozone, for solar radiation exerts a control on the growth of ozone so that a new equilibrium is reached which is related to the photochemical equilibrium O_{3e} and to the ratio $F/2f_2O_2$.

Similarly from 16 it may further be inferred that $O_{3E} < O_{3e}$ when F is negative. However, when F is negative, a scrutiny of Equation 16 shows that there is no real solution for O_{3E} if $(2f_2O_2 + F) < 0$. This condition can be overcome by the following physical considerations.

Until now F was considered only as a positive or negative rate of input for developing the basic equations. However, in the atmosphere local rate of change of ozone due to advective and diffusive processes results from the spatial distribution of ozone and the relative intensities of these transporting agencies. Thus, under equilibrium conditions, the spatial distribution assumed by ozone will be such that at each point in the field, flux convergence or divergence of ozone is in balance with the photochemical destruction or

production of ozone at that point. On physical grounds it becomes untenable to obtain a field of ozone under steady state that has flux divergence at any point in the field over and above photochemical production at that point.

Transport by Large Scale Diffusion and Mean Meridional Motions

We may now proceed to introduce the changes of ozone due to large scale diffusion and mean meridional motions in the growth Equation 12. However, to take into account the variation of density in the atmosphere we have to choose a property of ozone that is a function of the atmospheric density. Such a property of ozone is its mixing ratio

$$\chi = \frac{O_3 \text{ molecules, cm}^3}{\text{air molecules, cm}^3} .$$

Thus

$$O_3 = M\chi . \tag{17}$$

Now the rate at which ozone increases at any given point per unit volume per unit time, due to a velocity $\mathbb{V}$, is given by the equation of continuity for ozone:

$$\left[\frac{\partial (M\chi)}{\partial t}\right]_a = -\nabla \cdot (M\chi \mathbb{V}) . \tag{18}$$

The subscript a refers to air motions.

Now let u, v, and w be the components of $\mathbb{V}$ along x the east, y the north, and z the vertical direction. Expressing these velocity components and χ in terms of their mean and turbulent part we have

$$\begin{aligned} \chi &= \bar{\chi} + \chi' , \\ \mathbb{V} &= \bar{\mathbb{V}} + \mathbb{V}' , \\ u &= \bar{u} + u' , \\ v &= \bar{v} + v' , \\ w &= \bar{w} + w' , \end{aligned} \tag{19}$$

where the time averaged quantities $\bar{\chi}$, $\bar{\mathbb{V}}$, $\bar{u}$, $\bar{v}$, and $\bar{w}$ are defined as

$$\overline{(\quad)} = \frac{1}{\tau} \int_0^\tau (\quad)\, dt . \tag{20}$$

τ is a period comparable to the length of a season. Thus

$$\left[\frac{\partial (M\chi)}{\partial t}\right]_a = -\nabla \cdot \left[M(\bar{\chi} + \chi')(\bar{\mathbb{V}} + \mathbb{V}')\right] ; \tag{21}$$

that is,

$$\left[\frac{\partial (M\chi)}{\partial t}\right]_a = -\nabla \cdot (M\bar{\chi}\bar{\mathbb{V}}) - \nabla \cdot (M\bar{\chi}\mathbb{V}') - \nabla \cdot (M\chi'\bar{\mathbb{V}}) - \nabla \cdot (M\chi'\mathbb{V}') . \tag{22}$$

Let $\bar{M} = M$. Then averaging over time would give

$$\left[\frac{\partial (M\chi)}{\partial t}\right]_a = -\nabla \cdot (M\bar{\chi}\bar{\mathbb{V}}) - \nabla \cdot (M\bar{\chi}'\bar{\mathbb{V}}') = -\bar{\mathbb{V}} \cdot \nabla (M\bar{\chi}) - M\bar{\chi}\nabla \cdot \bar{\mathbb{V}} - \nabla \cdot (M\bar{\chi}'\bar{\mathbb{V}}') ; \tag{23}$$

therefore,

$$\left[M\frac{\partial \bar{\chi}}{\partial t} + \bar{\chi}\frac{\partial M}{\partial t}\right]_a = -M\bar{\mathbb{V}} \cdot \nabla\bar{\chi} - \bar{\chi}\bar{\mathbb{V}} \cdot \nabla M - M\bar{\chi}\nabla \cdot \bar{\mathbb{V}} - \nabla \cdot (M\bar{\chi}'\bar{\mathbb{V}}') . \tag{24}$$

Now, from the equation of continuity we have

$$\frac{1}{M}\frac{dM}{dt} = -\nabla \cdot \bar{\mathbb{V}} . \tag{25}$$

Therefore,

$$\bar{\chi}\left(\frac{\partial M}{\partial t} + \bar{\mathbb{V}} \cdot \nabla M\right) + M\bar{\chi}\nabla \cdot \bar{\mathbb{V}} = 0 ; \tag{26}$$

and so

$$\left[\frac{\partial \bar{\chi}}{\partial t}\right]_a = -\bar{\mathbb{V}} \cdot \nabla\bar{\chi} - \frac{1}{M}\nabla \cdot (M\bar{\chi}'\bar{\mathbb{V}}') . \tag{27}$$

When we expand $\bar{V}$ and $\bar{V}'$ into their components, we get

$$\left[\frac{\partial \bar{\chi}}{\partial t}\right]_a = -\left(\bar{u}\frac{\partial \bar{\chi}}{\partial x} + \bar{v}\frac{\partial \bar{\chi}}{\partial y} + \bar{w}\frac{\partial \bar{\chi}}{\partial z}\right) - \frac{1}{M}\left\{\frac{\partial}{\partial x} M\overline{\chi' u'} + \frac{\partial}{\partial y} M\overline{\chi' v'} + \frac{\partial}{\partial z} M\overline{\chi' w'}\right\} . \quad (28)$$

We further assume zonal symmetry of $\bar{\chi}$ and a steady zonal wind $\bar{u}$, and we can reduce Equation 28 to

$$\left[\frac{\partial \bar{\chi}}{\partial t}\right]_a = -\left(\bar{v}\frac{\partial \bar{\chi}}{\partial y} + \bar{w}\frac{\partial \bar{\chi}}{\partial z}\right) - \frac{1}{M}\left\{\frac{\partial}{\partial y} M\overline{\chi' v'} + \frac{\partial}{\partial z} M\overline{\chi' w'}\right\} . \quad (29)$$

The term $-\left(\bar{v}\frac{\partial \bar{\chi}}{\partial y} + w\frac{\partial \bar{\chi}}{\partial z}\right)$ in 29 expresses the local change of χ at any given point in the meridional plane due to mean meridional and vertical motions.

The terms $-\frac{\partial}{\partial y} M\overline{\chi' v'}$ and $-\frac{\partial}{\partial z} M\overline{\chi' w'}$ in 29 may be transformed with the help of the Austausch assumption (see, e.g., Haurwitz, Reference 16) of diffusion as

$$-\frac{\partial}{\partial y}\left(M\overline{\chi' v'}\right) = \frac{\partial}{\partial y}\left[Mk_y \frac{\partial \bar{\chi}}{\partial y}\right]$$

and

$$-\frac{\partial}{\partial z}\left(M\overline{\chi' w'}\right) = \frac{\partial}{\partial z}\left[Mk_z \frac{\partial \bar{\chi}}{\partial z}\right], \quad (30)$$

where K_Y and K_Z (cm^2/sec) are the eddy diffusion coefficients along the y and z directions. Upon neglecting the variation of density, M along y, the local change of ozone at a given point due to large scale diffusion and mean meridional motions may be obtained from 29, 30, and 17 as

$$\left[M\frac{\partial \bar{\chi}}{\partial t}\right]_a = -M\left(\bar{v}\frac{\partial \bar{\chi}}{\partial y} + \bar{w}\frac{\partial \bar{\chi}}{\partial z}\right) + M\left[K_y\frac{\partial^2 \bar{\chi}}{\partial y^2} + \frac{\partial K_y}{\partial y}\frac{\partial \bar{\chi}}{\partial y} + K_z\frac{\partial^2 \bar{\chi}}{\partial z^2} + \frac{\partial K_z}{\partial z}\frac{\partial \bar{\chi}}{\partial z} + K_z\frac{1}{M}\frac{\partial M}{\partial z}\frac{\partial \bar{\chi}}{\partial z}\right]. \quad (31)$$

If we substitute for F, in Equation 13, the local change of ozone $M\left[\frac{\partial \bar{\chi}}{\partial t}\right]_a$ from 31 and then divide by $\left(\frac{O_{3E}+N_3^*}{O_{3E}}\right)$, we obtain

$$-2f_3\frac{O_{3E}^2}{N_3^* + O_{3E}} + 2f_2 O_2\left[\frac{N_3^* - O_{3E}}{N_3^* + O_{3E}}\right] + \left[M\frac{\partial \bar{\chi}}{\partial t}\right]_a = 0 \quad (32)$$

or

$$-2\frac{f_3}{M}\frac{(M\bar{\chi})^2}{(N_3^* + M\bar{\chi})} + 2f_2\frac{O_2}{M}\left[\frac{N_3^* - M\bar{\chi}}{N_3^* + M\bar{\chi}}\right] + \left[\frac{\partial \bar{\chi}}{\partial t}\right]_a = 0 \quad (33)$$

Equation 33 describes the combined effects of the photochemical and transport processes under steady state conditions. By solving 33 for $\bar{\chi}$ as a function of latitude and height we can get the distribution of $\bar{\chi}$ and hence O_{3E} in the meridional plane.

The numerical methods and boundary conditions necessary to solve Equation 33 are discussed in the next section.

During the polar night in the winter hemisphere there is no discussion of oxygen or ozone by radiation. Hence the local change of ozone at any point under such conditions will be entirely due to transport processes. It is also possible that air motions could transport atomic oxygen into the polar night. Such transported atomic oxygen could form ozone in the polar night through Reaction 2. Kellogg (Reference 17) has demonstrated by his calculations the importance of such transported atomic oxygen in heating the layers of the atmosphere above 80 km. However, at lower levels below about 40km, it is reasonable

to assume, following Craig (Reference 9), that O_1 is used up in Reaction 2 in situ, so that the O_1 transport may be neglected.

PROCEDURE OF THE INVESTIGATION

Discussion of the Data Used

It is necessary to have data on the solar energy received at the top of the atmosphere, the absorption coefficients of oxygen and ozone, the temperature and density distribution in the atmosphere, and the reaction rate coefficients K_2 and K_3 in order to calculate the vertical distribution of ozone in the atmosphere.

From the most recent rocket measurements of the solar spectrum we have now very reliable estimates of the energy in the ultraviolet radiation received at the top of the atmosphere. The two measurements of the solar ultraviolet spectrum, one reported by Detwiler et al. (Reference 18) and the other by Hinteregger (Reference 19), agree with one another quite well. The value of the solar ultraviolet energy received at the top of the atmosphere from 1800 to 2600 A used in this study are taken from measurements of Detwiler et al. Above 2600 A the solar radiation received at the top of the atmosphere tabulated by Gast (Reference 20) is adopted. The solar spectrum thus constructed is shown in Figure 5.

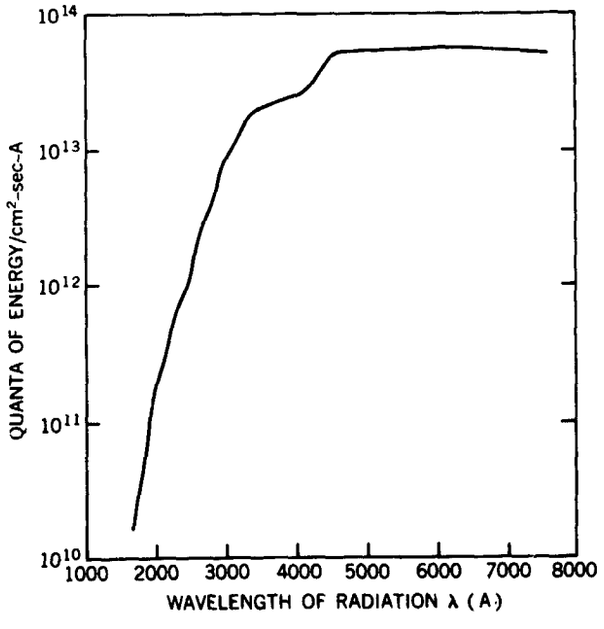


FIGURE 5.—Solar energy received at the top of the atmosphere.

Watanabe et al. (Reference 21), using vacuum ultraviolet technique, measured the absorption coefficients of oxygen between 1050 and 1900 A. The absorption coefficients of oxygen in the Schumann-Runge bands from 1800 to 1900 A are adopted from the measurements reported, independent of pressure, by these investigators. However, they point out that their results show a pressure variation of the coefficients by a factor of 4 to 5 in the absorption peaks, and 2 to 3 in the absorption minima. In view of the apparent pressure dependency of the "observed coefficients" reported by them, they remark that their "data were smoothed somewhat arbitrarily; that is, low pressure data were used for maxima, and high pressure data for minima" to obtain absorption coefficients of oxygen independent of pressure.

Heilpern (Reference 22) showed that the absorption coefficient of oxygen is pressure dependent in the region 2100 to 2400 A. In his measurements he used eight different pressures (between 50 and 127 kgm/cm²) to obtain the functional form of pressure dependency of the extinction coefficients for a mixture of oxygen and nitrogen having a composition similar to air. Samples of air gave the same results within the accuracy of his experimental values. He fitted his results for the entire region to an equation of the form

$$\epsilon_p = \epsilon_1 P_1 + \epsilon_2 P_2 + \epsilon_3 P_1 P_2 ,$$

where $\epsilon_p(cm^{-1})$ is the extinction coefficient, P_1 (kgm/cm²) is the partial pressure of oxygen, P_2(kgm/cm²) is the partial pressure of nitrogen, and ϵ_1, ϵ_2 and ϵ_3 are coefficients evaluated by fitting the data to the above equation.

Heilpern's investigation of the extinction coefficient of a similar mixture in the pressure range 0.2 to 130 kgm/cm² and at 2144 A agreed with the functional form obtained from the whole region 2100 A to 2400 A. On the strength of the observations at 2144 A, which were extended to the lower pressure range, this form of pressure dependency is accepted in this work to hold good at low pressure for the entire region 2100 to 2400 A. The absorption coefficients at pressures below 0.2 kgm/cm², encountered in the atmosphere, are obtained by extrapolating the experimental data with the help of the above functional form of pressure dependency.

The absorption coefficient of oxygen for the region 1900 to 2100 A, lying between the limits of the experimental studies of Watanabe et al. and Heilpern, were obtained by interpolation. The interpolation was made at each desired pressure in the atmosphere.

The absorption coefficient of oxygen thus obtained at sea level pressure is compared in Figure 6 with that of Craig (Reference 9).

Vassy (Reference 23) determined the absorption coefficients of ozone from 2020 A to 2170 A using ozonized oxygen and a photographic photometric method. However, she used Ny and Choong's (Reference 24) value of the absorption coefficient of ozone in the Huggins bands to measure the path length of ozone in her experiment.

Absorption coefficients of ozone in the region 1050 to 2200 A were measured by Tanaka et al. (Reference 25) using a vacuum ultraviolet technique. Their curve for the absorption intensity showed a minimum at about 2015 A. They used pure ozone obtained by vaporizing liquid ozone instead of the ozonized oxygen used by the earlier experimenters. Tanaka et al., however, remarked that "the results of Vassy, and Ny and Choong, who measured down to 2020 A and 2135 A respectively, are somewhat higher than the present data. The difference was about 10 percent at 2200 A and about 40 percent at 2020 A as compared with the data reported by Vassy. The discrepancies are greater than can be explained by ordinary experimental errors, and might be due to their use of the ozonized oxygen and photographic detection as opposed to the use of purer ozone and photoelectric detection."

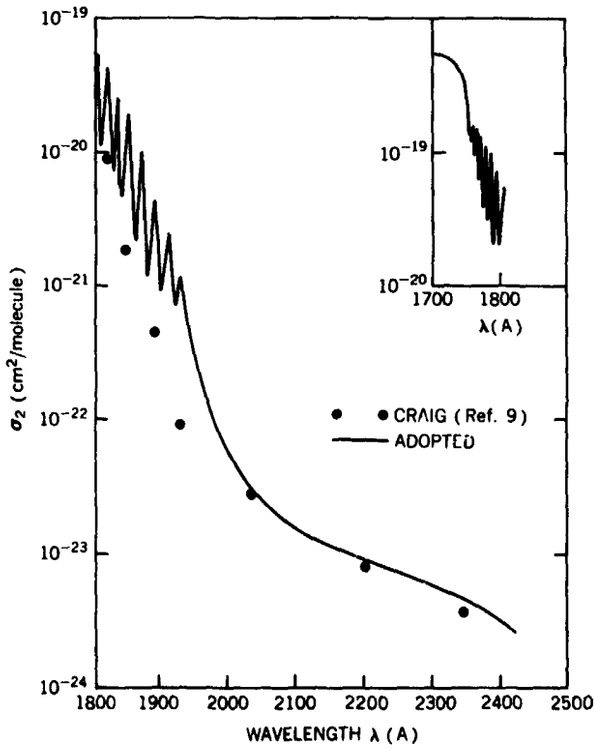


FIGURE 6.—Absorption coefficients of oxygen.

But Tanaka et al. have observed chemiluminescence of pure ozone in their experiment. It is possible that the presence of such chemiluminescence of pure ozone in the investigation of Tanaka et al. may have been responsible for somewhat smaller values of the absorption coefficients of ozone.

As an extension to the measurements made by Tanaka et al., Inn and Tanaka (Reference 26) measured the ozone absorption in the near ultraviolet from 2000 to 3500 A and in the visible region from 4000 to 7000 A.

Vigroux (Reference 27) studied the absorption of ozone in the Hartley, Huggins, and Chappuis bands using photographic photometric method and ozonized oxygen. He also studied the temperature dependency of the absorption coefficient of ozone. However, his observations do not extend to wavelengths shorter than 2300 A in the Hartley bands.

Gast (Reference 20) has tabulated the absorption coefficients of ozone from 2200 to 3000 A determined by Inn and Tanaka. For the region 3000-3500 A he has listed the absorption coefficients measured by Vigroux which he remarks are "recommended" by Inn and Tanaka. Gast has presented the absorption coefficients of ozone at 18°C and also (suitably applying the temperature correction given by Vigroux) at −44°C. The value corresponding to −44°C is better suited to calculate transmission of solar energy in the atmosphere.

In the present study the absorption coefficients of ozone at wavelengths longer than 2200 A are adopted from the tables presented by Gast as shown in Figure 7(a). At wavelengths shorter than 2200 A there is apparently a large difference between the absorption coefficients obtained by Tanaka et al., using pure ozone, and Vassy, who used ozonized oxygen.

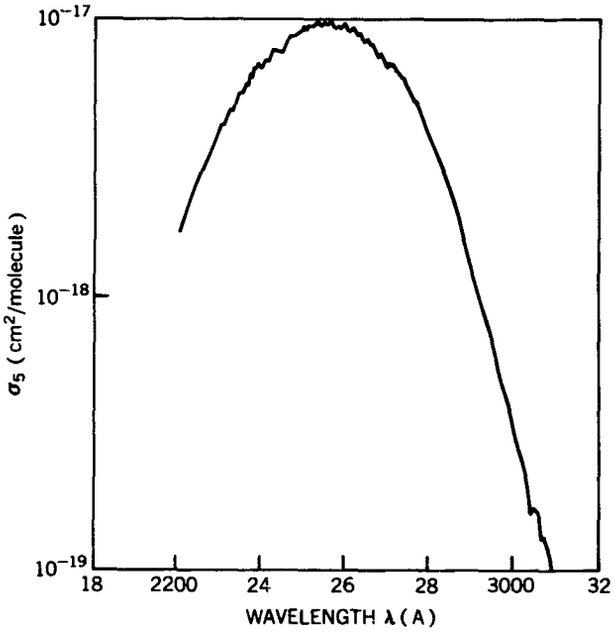


FIGURE 7a.—Absorption coefficients of ozone between 2200–3100A.

In this study the absorption coefficients of Vassy in the spectral region 2020 to 2200 A have been adopted after reducing them by 10 percent to be consistent with the measurements of Inn et al. at 2200 A. The absorption curve thus constructed after Vassy between 2020 and 2200 A is then extrapolated to agree with that of Tanaka et al. at 1800 A as shown in Figure 7(b).

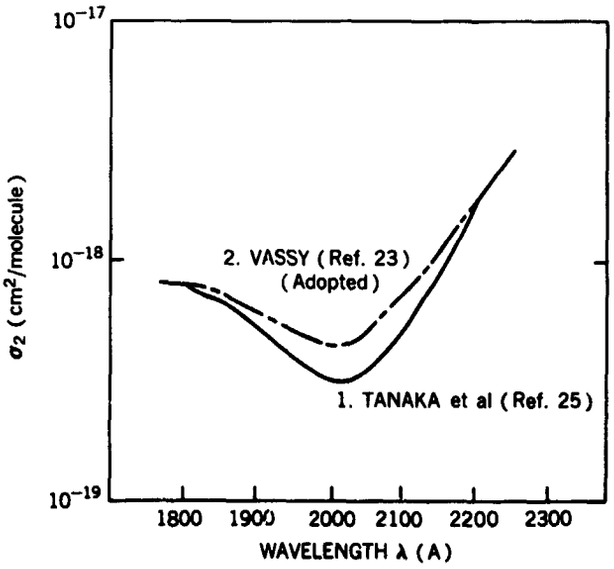


FIGURE 7b.—Absorption coefficients of ozone between 1800–2200A.

Temperature and density distributions for the solstices are taken from Murgatroyd (Reference 28) as a function of altitude and latitude. The value of the temperature and density below 16 km, the lowest level given by Murgatroyd, are obtained from the investigations of London et al. (Reference 29).

The reaction rate coefficients K_2 and K_3 are adopted from Campbell and Nudelman's (Reference 30) least square analysis of Glissmann and Schumacher's (Reference 31) data on thermal decomposition of ozone in the temperature range from 70°–110°C. According to Campbell and Nudelman, the functional form of K_2 and K_3 is given by

$$K_2 = 1.87 \times 10^{14} \text{ Exp } 1.49 \times 10^2/T \quad \text{cm}^6 \text{ gm mol}^{-2} \text{ sec}^{-1},$$

$$K_3 = 5.44 \times 10^{12} \text{ Exp } 2.133 \times 10^3/T \quad \text{cm}^3 \text{ gm mol}^{-1} \text{ sec}^{-1}.$$

The ratio of these two coefficients K_2/L_3 is compared in Figure 8 with that obtained by Eucken and Patat (Reference 32) in their study on the temperature dependence of the photochemical formation of ozone. The logarithm of the ratio K_2 K_3, when plotted against 1/T, where T is the temperature in °K, gives a straight line as shown in Figure 8.

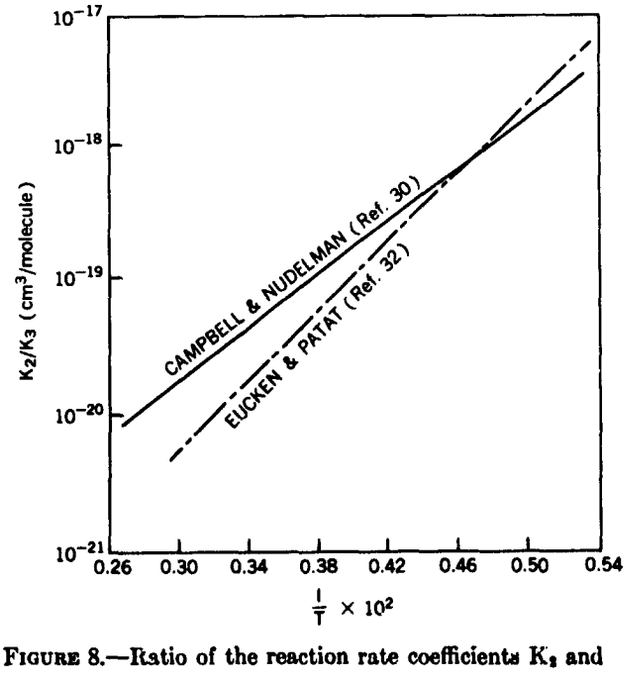


FIGURE 8.—Ratio of the reaction rate coefficients K_2 and K_3 vs. 1/T°K.

Method of Obtaining the Meridional Distribution of Ozone When Transport Processes Are Present

Solution of Equation 33 gives us the meridional distribution of $\bar{\chi}$ and hence of ozone. To solve 33, the following numerical procedure is adopted.

The meridional plane from pole to pole and between 41 km to the height of the tropopause is divided into a grid, the gridpoints in which are spaced 10° of latitude apart horizontally, and 2 km apart vertically as shown in Figure 9. The

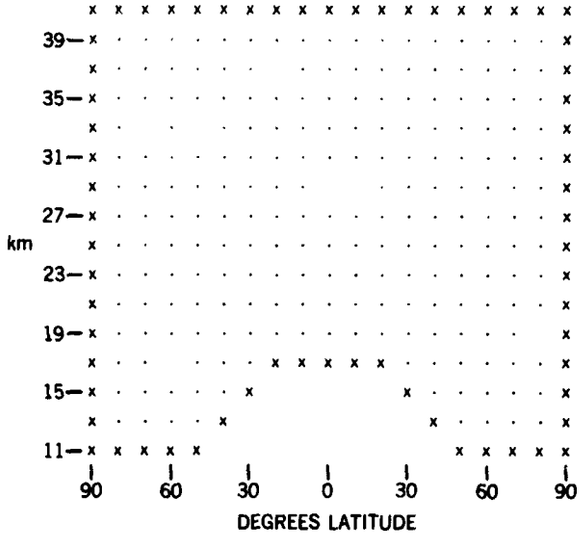


FIGURE 9.—Grid used for the numerical calculations.

boundary at 41 km representing the layer 42 to 40 km is chosen for the reason that ozone is reasonably undisturbed from its photochemical equilibrium at and above this level.

The boundary conditions used to solve 33 are:

(1) The flux of ozone through the vertical boundaries at the poles is zero.

(2) Photochemical equilibrium is assumed at 41 km, the upper boundary. However, during the polar night in the winter hemisphere, it is assumed that from 60° to the pole the photochemical equilibrium is constant at 41 km.

(3) On the lower boundary at the tropopause level, the value of the ozone concentration is specified from the "observed" model of the meridional distribution of ozone given by Paetzold and Piscalar (Reference 7).

The boundary condition (1) is purely a mathematical constraint and does not call for any knowledge of the distribution of ozone along the verticals at the poles. The boundary condition (2) at the 41 km level is quite justified from considerations of the photochemical theory, while the boundary condition (3) lacks rigor of the preceding two conditions, and hence is subject to question. Further discussion of the boundary condition (3) is presented in the results.

The complete form of Equation 33, after expansion of the term $\left[\frac{\partial \bar{\chi}}{\partial t}\right]_a$, is seen to be

$$-\frac{2f_3}{M}\frac{(M\bar{\chi})^2}{(N_3^* + M\bar{\chi})} + 2f_2\frac{O_2}{M}\left[\frac{N_3^* - M\bar{\chi}}{N_3^* + M\bar{\chi}}\right] - \left(\bar{v}\frac{\partial\bar{\chi}}{\partial y} + \bar{w}\frac{\partial\bar{\chi}}{\partial z}\right)$$

$$+\left(K_y\frac{\partial^2\bar{\chi}}{\partial y^2} + \frac{\partial K_y}{\partial y}\frac{\partial\bar{\chi}}{\partial y} + K_z\frac{\partial^2\bar{\chi}}{\partial z^2} + \frac{\partial K_z}{\partial z}\frac{\partial\bar{\chi}}{\partial z} + K_z\frac{1}{M}\frac{\partial M}{\partial z}\frac{\partial\bar{\chi}}{\partial z}\right) = 0 . \quad (34)$$

$\bar{\chi}$, the ozone mixing ratio at each gridpoint, is calculated implicitly from Equation 34. Numerical integration of the above equation is performed at each gridpoint starting from 80° latitude in one hemisphere at the upper boundary, and proceeding downward along the vertical to the lower boundary, and then recommencing at the upper boundary, at the next latitude, and so on until 80° latitude in the other hemisphere is reached.

The parameters f_2 and f_3 at each gridpoint depend on the solar energy reaching that point. However solar energy transmitted to any given height in the atmosphere is a function of the zenith angle of the sun. In order to simplify the calculations it is assumed that the sun radiates at a mean zenith angle [ie., 1/2 (zenith angle at sun rise+zenith angle at noon)] corresponding to the given latitude and season. Further, to take into account the variation of the length of day, f_2 and f_3 at each gridpoint are multiplied by the ratio (daylight hours/24 hours) corresponding to each given latitude and season.

The finite difference scheme adopted here to solve Equation 34 involves the central difference approximation to the second order space derivatives. For the first order space derivatives arising from advective terms, the Lelevier (Richtmyer, Reference 33) method of approximation is used. The basic idea behind Lelevier's method is to get the advection from the windward side of the gridpoint using non-centered differences.

The finite difference approximations to the various terms in 34 are written below with the help of Figure 10.

All of the parameters χ, M, v, w, K_Y, and K_Z are subscripted with the numbers 1, 2, 3, or 4 denoting the value of the parameter at the respective gridpoints in Figure 10. The value of any of these parameters denoted without a subscript refers to the central point.

Further, we denote the horizontal velocity v as positive in the y direction pointing to the north. Similarly, the sign of the vertical velocity w is taken positive when directed upward along z.

$$-\bar{v}\frac{\partial\bar{\chi}}{\partial y} \simeq \begin{cases} -v\dfrac{(\chi-\chi_4)}{\Delta y}, & (v>0) \\ -v\dfrac{(\chi_2-\chi)}{\Delta y}, & (v<0) \end{cases} \tag{35}$$

$$-\bar{w}\frac{\partial\bar{\chi}}{\partial z} \simeq \begin{cases} -w\dfrac{(M+M_2)}{2M}\dfrac{(\chi-\chi_2)}{\Delta z}, & (w>0) \\ -w\dfrac{(M+M_1)}{2M}\dfrac{(\chi_1-\chi)}{\Delta z}, & (w<0) \end{cases} \tag{36}$$

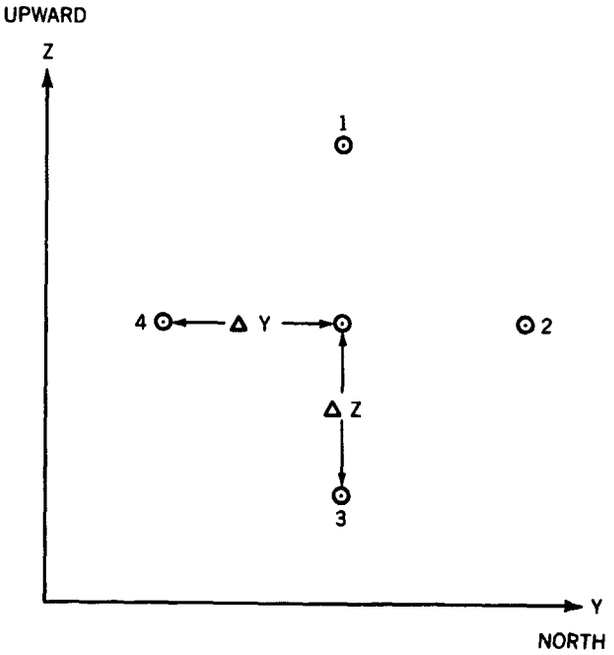


FIGURE 10.—Finite difference scheme.

The coefficients $\frac{M+M_2}{2M}$ and $\frac{M+M_1}{2M}$ in 36 arise due to the variation of density along z.

$$\left.\begin{aligned} K_y\frac{\partial^2\bar{\chi}}{\partial y^2} &\simeq K_y\frac{\chi_2+\chi_4-2\chi}{(\Delta y)^2} \\ \frac{\partial K_y}{\partial y}\frac{\partial\bar{\chi}}{\partial y} &\simeq \frac{(K_{y2}-K_{y4})}{(2\Delta y)}\frac{(\chi_2-\chi_4)}{(2\Delta y)} \end{aligned}\right\}, \tag{37}$$

$$\left.\begin{aligned} K_z\frac{\partial^2\bar{\chi}}{\partial z^2} &\simeq K_z\frac{\chi_1+\chi_2-2\chi}{(\Delta z)^2} \\ \frac{\partial K_z}{\partial z}\frac{\partial\bar{\chi}}{\partial z} &\simeq \frac{(K_{z1}-K_{z3})}{(2\Delta z)}\frac{(\chi_1-\chi_3)}{(2\Delta z)} \\ K_z\frac{1}{M}\frac{\partial M}{\partial z}\frac{\partial\bar{\chi}}{\partial z} &\simeq K_z\frac{M_1-M_3}{2M\Delta z}\frac{(\chi_1-\chi_3)}{2\Delta z} \end{aligned}\right\}. \tag{38}$$

With the finite difference approximations 35 to 38, we can transform Equation 34 into an equation for χ at the central gridpoint:

$$\begin{aligned} \chi = \Bigg\{ & 2f_2\frac{O_2}{M}\left[\frac{N_3^*-M\chi}{N_3^*+M\chi}\right] + K_y\frac{\chi_2+\chi_4}{(\Delta y)^2} + \frac{(K_{y2}-K_{y4})}{2\Delta y}\frac{(\chi_2-\chi_4)}{2\Delta y} \\ & + K_z\frac{\chi_1+\chi_3}{(\Delta z)^2} + \frac{(K_{z1}-K_{z3})}{2\Delta z}\frac{(\chi_1-\chi_3)}{2\Delta z} \\ & + K_z\left[\frac{M_1-M_3}{2M\Delta z}\right]\frac{(\chi_1-\chi_3)}{2\Delta z} + v\frac{\chi_4}{\Delta y} + w\left[\frac{M+M_3}{2M}\right]\frac{\chi_3}{\Delta z}\Bigg\} \\ & \div\left\{2f_3\frac{M\chi}{N_3^*+M\chi} + \frac{v}{\Delta y} + \frac{w}{\Delta z}\left[\frac{M+M_3}{2M}\right] + 2\frac{K_z}{\Delta z^2} + 2\frac{K_y}{\Delta y^2}\right\}. \end{aligned} \tag{39}$$

Equation 39 is solved by successive approximations to the value of χ. For example, the first guess to χ, let us say, is χ'. Then substituting χ' in the right-hand side of 39, we calculate a new value for χ denoted χ''. Now the new guess for χ, namely χ''', is obtained from

$$\chi''' = \frac{\chi' + \chi''}{2} . \tag{40}$$

Successive approximations to χ are made until the criterion

$$\left| 1 - \frac{\chi'''}{\chi''} \right| \leq 0.02 \tag{41}$$

is satisfied. After the criterion 41 is satisfied, the ozone concentration $M\chi$ at the center of the layer is obtained.

The boundary conditions (2) and (3) do not present any problem since they are specified to start with. But the boundary condition (1) has to be satisfied implicitly whenever Equation 39 is being solved along the vertical at 80°N or 80°S latitude. This is achieved by solving Equation 39 with the condition that, in any layer, χ at 80° latitude is equal to χ at the pole. In other words, along the vertical at 80° latitude, we solve the equation

$$\chi = \left\{ 2f_2 \frac{O_2}{M} \left[\frac{N_3^* - M\chi}{N_3^* + M\chi} \right] + K_y \frac{\chi_4}{\Delta y^2} + \frac{(K_{y2} - K_{y4})}{2\Delta y} \frac{(\chi_2 - \chi_4)}{2\Delta y} \right.$$

$$+ K_z \frac{\chi_1 + \chi_3}{\Delta z^2} + \frac{(K_{z1} - K_{z3})}{2\Delta z} \frac{(\chi_1 - \chi_3)}{2\Delta z}$$

$$\left. + K_z \frac{(M_1 - M_3)}{2M\Delta z} \frac{(\chi_1 - \chi_3)}{2\Delta z} - v \frac{\chi_4}{\Delta y} + w \left[\frac{M + M_3}{2M} \right] \frac{\chi_3}{\Delta z} \right\} \div$$

$$\left\{ 2f_3 \frac{M\chi}{N_3^* + M\chi} + \frac{v}{\Delta y} + \frac{w}{\Delta z} \left[\frac{M + M_3}{2M} \right] + 2 \frac{K_z}{\Delta z^2} + \frac{K_y}{\Delta y^2} \right\} . \tag{42}$$

Equation 42 is applicable at 80°N latitude. A similar equation for 80°S can be easily obtained by substituting $K_y \frac{\chi_2}{\Delta y^2}$ for $K_y \frac{\chi_4}{\Delta y^2}$ in the numerator of the right-hand side of (42).

After the entire field from 80°S to 80°N has been scanned by Equations 39 and 41, a fresh iteration of the whole field is started again from 80°S. The whole field is then iterated several times until the values of χ or O_{3E} at any point in the whole field from two successive iterations do not differ by more than 1 percent—that is, if the value of χ is $[\chi]_1$ at any given point from preceding iteration, and $[\chi]_2$ from the succeeding iteration, the convergence criterion requires

$$\left| 1 - \frac{[\chi]_1}{[\chi]_2} \right| < 0.01 . \tag{43}$$

To solve Equation 39 and to obtain the steady state distribution of ozone corresponding to either the time of solstices or equinoxes, we do not have to specify the distribution of ozone in the interior of the field. One may start with the photochemical equilibrium distribution in the interior of the field as the first guess. It is observed that about 50 iterations of the entire field are adequate to get a convergence level of less than 1 percent stipulated by Equation 43.

General Survey of the Investigations on the Transport Processes in the Stratosphere

To make an investigation on the problem of seasonal variations of atmospheric ozone, it is necessary to know about the large scale diffusive and advective phenomena. Studies made on the transport processes in the atmosphere have in general a tendency to empasize one of the two major transport phenomena: diffusion, or advection. Theories constructed on the basis of either one of these two seem to be capable of explaining phenomena such as the spread of radioactive debris in the stratosphere. An attempt is made here to present an account of both of these schools of thought.

Brewer (Reference 34), from his frost point hygrometer measurements of water vapor content in the lower stratosphere over Great Britain, observed a rapid fall of humidity mixing ratio above the tropopause. The humidity mixing ration at a height of 1 km above the tropopause was about one-tenth of its value at the tropopause. The frost points at 15 km were as low as the lowest temperatures measured in the upper troposphere near the equator. Thus he argued that such extremely dry air present in the polar stratosphere must have originated in the equatorial upper troposphere. Brewer explained the vertical distribution of water vapor in the lower polar stratosphere as a result of upward diffusion of water vapor opposed by a slow sinking of very dry air from above. A reverse current from the troposphere into the stratosphere in the equatorial region was postulated to balance such a slow sinking of air in the higher latitudes from the stratosphere into the troposphere.

Brewer estimated from the observed vertical profiles of water vapor distribution a value of 1.8×10^{-5} to 5×10^{-5} cm^{-1} for the ratio w/K_z, where w(cm/sec) is the velocity of the sinking motion, and K_z (cm^2/sec) is the vertical eddy diffusion coefficient, in the lower stratosphere. He also suggested that the effect of horizontal advection could be only one-tenth of sinking. These ideas of Brewer were accepted by Dobson (Reference 35) to explain the seasonal distribution of ozone in the atmosphere. The circulation model thus came to be known as the "Dobson-Brewer" model.

The most important factor, namely the nature of the vertical distribution of water vapor in the stratosphere, which led Brewer to postulate his model of the organized circulation in the stratosphere, does not appear to be a general characteristic of the stratosphere on the whole globe. The water vapor mixing ration is observed sometimes to increase with height in the stratosphere (Mastenbrook et al., Reference 36).

Murgatroyd and Singleton (Reference 37), neglecting the influence of large scale diffusion, have made a calculation of meridional circulation sufficient to transport heat between radiational sources and sinks in the stratosphere (15–55 km) and mesophere (55–80 km). Combining suitably the radiative heating values of Ohring (Reference 38) and Murgatroyd and Goody (Reference 39), and modifying slightly the combined radiative field to get a solution, they computed the meridional circulation. The radiative heating and the atmospheric temperature needed as a function of time in their computations were obtained with an implicit assumption that these two quantities are in phase with the sun.

The circulation model thus obtained has a rising motion of air below 30 km over the tropical latitudes, with an outflow from the tropics to higher latitudes in both hemispheres. The order of magnitude of the vertical velocity is 10^{-1} cm/sec, and below 40 km the horizontal velocity is about 10^2 cm/sec.

Cross sections of meridional and vertical velocities from pole to pole at different times of the year computed by Murgatroyd and Singleton (Reference 37: page 130, Figure 2) are presented in the Figure 11.

Broadly speaking, the circulation model of Murgatroyd and Singleton suggests rising motion in the tropical latitudes somewhat following the sun in its transit from one hemisphere to the other during the year. In the summer hemisphere at higher latitudes most of the stratosphere is covered with rising motion, and in the winter hemisphere there is correspondingly an intense sinking motion. The circulation model for autumn and spring shows an appreciable difference between the circulation on either side of the equator, with relatively intense mean air motions in spring and weak circulation in autumn. The poleward velocity in general in all seasons intensifies as one goes to lower levels in the stratosphere.

This circulation model of Murgatroyd and Singleton supports the ideas of the "Dobson-Brewer" model. However, the authors remarked that systems on the scale of cyclones and anticyclones, and possibly also diffusion, must play a large part as transport processes in the atmosphere, and that the circulation model obtained by them can be no more than a resultant mean effect.

Spar (see Stebbins, Reference 40), from his analysis of the spread of an artificial radioactive nuclide, W^{185} (Tungsten 185), introduced by the U.S. Hardtack nuclear explosions in 1958, derived meridional and vertical eddy diffusion coefficients

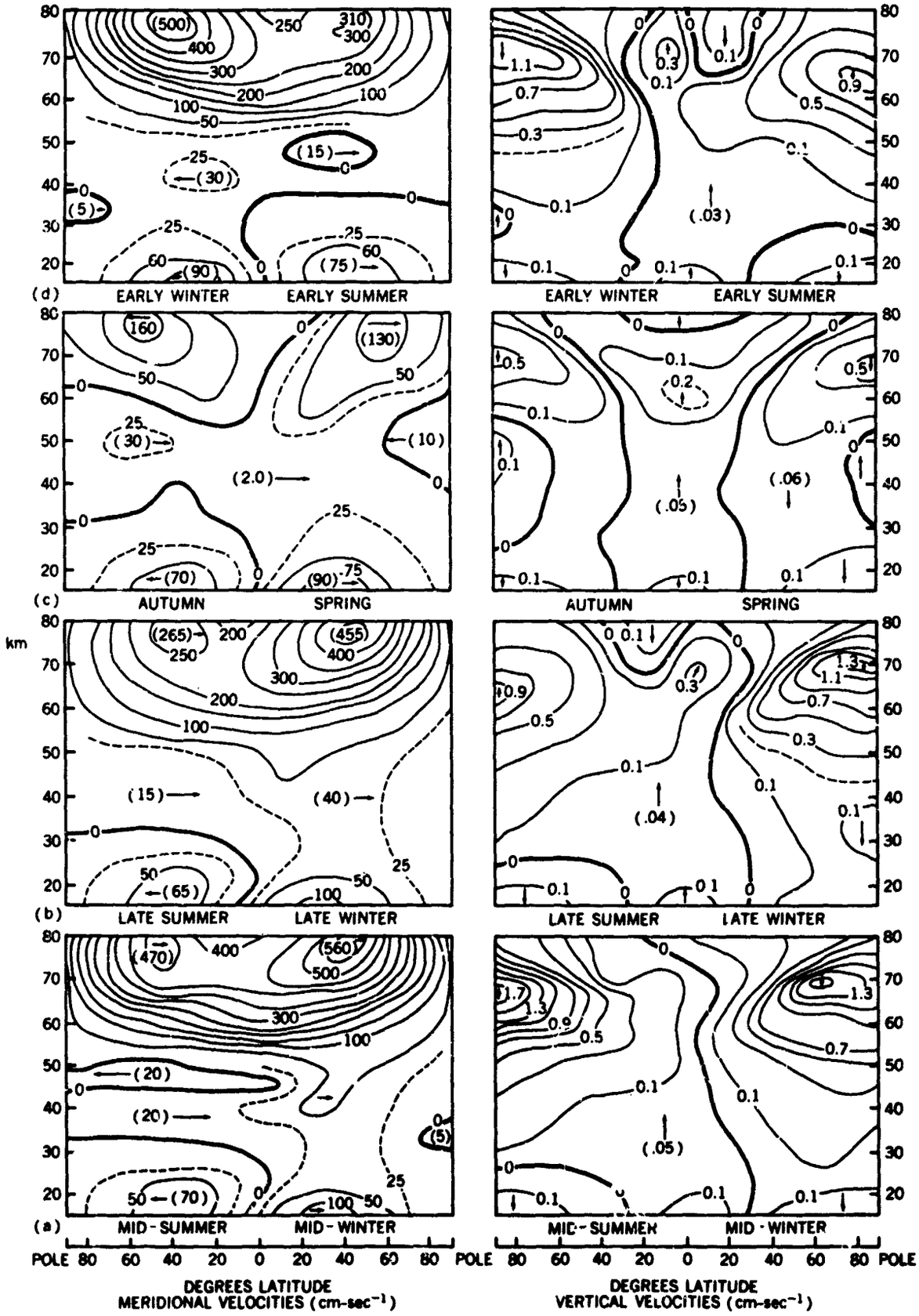


FIGURE 11.—Cross section of meridional and vertical velocities (cm/sec) from pole to pole at different times of the year (after Murgatroyd and Singleton, Reference 37; courtesy of the Royal Meteorological Society Quarterly Journal.)

for the stratosphere. Assuming that the radioactive cloud has a normal distribution of debris about the maximum, he obtained h, the half width of the cloud, as $h=1.67\sqrt{Kt}$, where K(cm²/sec) is the eddy diffusion coefficient and t is the time elapsed after detonation. He thus estimated that the value of the vertical eddy diffusion coefficient was 2×10^4 cm²/sec at the polar latitudes and 4×10^3 cm²/sec at the equatorial latitudes. The estimates of the meridional eddy diffusion coefficient ranged from a low value of about 5×10^5 cm²/sec to a value almost as high as 10^{10} cm²/sec. However, the latitudinal variation of the meridional eddy diffusion coefficient could not be ascertained from this investigation.

W^{185} was observed to have a zone of maximum concentration around 65,000–70,000 ft at the equatorial stratosphere and sloped down toward the poles. Furthermore, the height of the zone of maximum concentration was found to be virtually stationary from September 1958 to April 1960 (Feely and Spar, Reference 41), contradicting the "Dobson-Brewer" calculation model, which hypothesizes an organized meridional circulation with air moving upward through the equatorial tropopause, then horizontally poleward at high levels, and sinking back into the troposphere in the polar region.

Feely and Spar (Reference 41) to explain such an observed distribution of W^{185}, hypothesized a mixing mechanism in which lateral mixing occurs within a sloping layer which decreases in height toward the poles. However, they pointed out "the fact that the mixing surfaces slope more steeply than the isentropic surfaces may indicate the existence of a systematic sinking of the radioactive debris relative to the air as it mixes poleward, possibly by gravitational settling or some other mechanism, or a systematic subsidence of air itself as it mixes into the polar region. There is no indication, however, of a large scale subsidence of polar stratospheric air from very high levels."

Considering the rate at which the half width of the W^{185} cloud, in the north-south direction, increases with time, Spar (see Friend et al., Reference 42) has deduced an effective meridional (poleward) velocity of 8 cm/sec which he states is about one fifth that of Murgatroyd and Singleton's value. Further to account for the inclination, 1:1000, of the zone of W^{185} maximum from equator to pole, he has shown that there should be a rising motion of about 0.01 cm/sec at the equatorial region and a sinking motion of the same magnitude at the higher latitudes. This is about one tenth of the vertical velocity Murgatroyd and Singleton in the lower stratosphere. Thus the terminal velocity of the tungsten-bearing particles, which are apparently held stationary in the rising motion at the equatorial region, must be 0.01 cm/sec. From Stoke's law, one would find the radii of these particles to be about 0.2 micron, which Spar mentions is in good agreement with the size of the observed particles.

Murgatroyd and Singleton conclude from their trajectory calculation that "most of the air in the stratosphere, up to 25 km at least, must have passed through the equatorial tropopause region within the last one year or so." Commenting on this statement, Spar (see Friend et al., Reference 42) points out that if this were so the radioactive debris from nuclear tests would be almost depleted in one year or so. Such a rapid depletion of radioactive debris, however, does not seem to get support from the observed amount of debris contained in the stratosphere.

An interesting feature concerning the meander of the center of W^{185} maximum about the equator was revealed from the investigations of Friend et al. (Reference 42). The W^{185} maximum, although held at a stationary height, was found to migrate to the southern hemisphere during winter, and by early summer was back in the northern hemisphere. This phenomenon suggests a tendency of the peak of the W^{185} cloud to follow the sun.

Spar (see Friend et al., Reference 42) indicated that direct vertical mixing through the tropopause appeared to be too small to account for the stratospheric fallout.

A recent survey by Newell (Reference 43) is very helpful in assessing the relative importance of large scale diffusion processes and mean meridional motions in the stratosphere. Newell has estimated the horizontal eddy transport of ozone by an elegant and simple method. Assuming that the ozone concentration in the layer 12–24 km is correlated with total ozone measured at the gound, he has evaluated the large scale meridional horizontal eddy flux of ozone from the winds in

that layer. This computation was based on data obtained during the I.G.Y. from twenty-five ozone stations that had local upper air wind observations. He had also estimated the meridional transport of ozone due to mean meridional motion. Newell's investigations suggest that the transport due to mean meridional motion. For instance, the values of the eddy transport and the transport due to mean motion at 50°N during spring 1958, according to him, are in the proportion 11.0:1.8. Both these transports of ozone are directed toward the north. Newell further demonstrates quantitatively that these transports are capable of accounting for the observed build-up of ozone at higher latitudes during winter and spring.

DISCUSSION OF THE RESULTS

Computed Models of the Meridional Distribution of Ozone

The photochemical equilibrium distributions of ozone calculated for the solstices and the equinoxes are presented in Figures 12 and 13. Both these distributions indicate a center of ozone high located at about 30 km height at the latitude ovre which the sun is overhead. The concentration of ozone decreases in all directions from this center. A clear symmetry of the ozone distribution about the vertical at the latitude over which the sun is overhead may be noticed.

A comparison of the photochemical equilibrium distributions of ozone shown in Figures 12 and 13 with those of the "observed" models in Figures 1

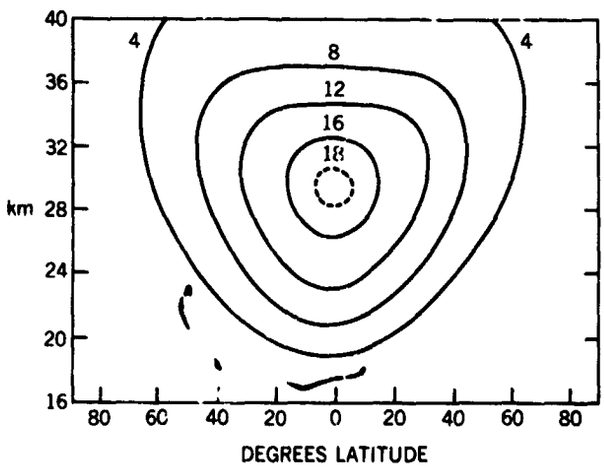


FIGURE 12.—Photochemical equilibrium distribution of ozone (D.U./km) for the equinoxes.

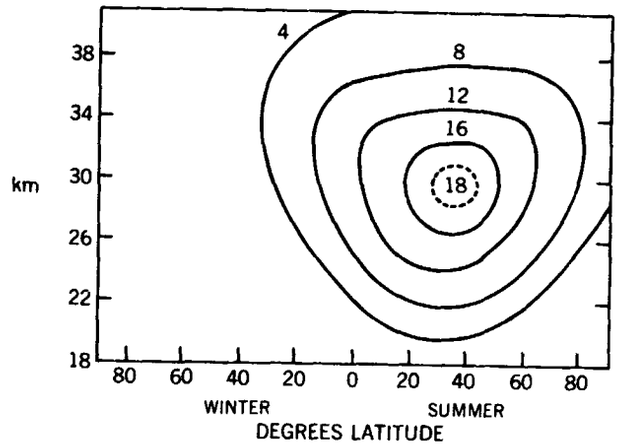


FIGURE 13.—Photochemical equilibrium distribution of ozone (D.U./km) for the solstices.

and 2 reveals all the inadequacies of the pure photochemical theory. Further, we are led to infer from such a comparison the important part played by transport mechanisms present in the atmosphere.

From the earlier discussion (in the section starting on page 18) on the transport processes—namely, large scale diffusion and mean meridional motions—it is not easy to arrive at a comprehensive unified picture of these transport processes. A computation of such a unified model of the transport processes that includes both diffusion and mean meridional motions is beyond the scope of this work. Instead, an attempt is made in this study to combine the mean meridional motions depicted by Murgatroyd and Singleton and the diffusion model of Spar with the help of investigations made by Newell in which an evaluation of the relative importance of each one of these transport processes is made.

The large scale diffusion model of Spar for the stratosphere suggests that the vertical eddy diffusion coefficient increases from a low value of 4×10^3 cm^2/sec in the tropics to 2×10^4 cm^2/sec at the polar latitudes. The variation of the horizontal eddy diffusion coefficient with latitude could not be so readily ascertained. Hence the variation of the horizontal eddy diffusion coefficient has to be elicited from an experimentation with the calculations on the model of the meridional distribution of zone.

The experimentation on the meridional distribution of ozone under steady state condition for the equinoxes is first undertaken.

From a few trial calculations with a constant value of 2×10^9 cm²/sec for the horizontal eddy diffusion coefficient, a rapid increase in the value of K_z, the vertical eddy diffusion coefficient from tropical latitudes to the polar latitudes as shown in Table 1, is chosen.

From similar experimentation an optimum value of about 20 percent of the Murgatroyd and Singleton circulation (comprising the meridional and vertical velocities) for the equinoxes is arrived at. Apparently this estimate of the mean meridional circulation, in a broad sense, is in agreement with the estimate of the mean meridional motion arrived at by Friend et al. (Reference 42) from their investigation on the rate at which the half width of the W^{185} cloud increased in the north-south direction. Newell's (Reference 43) estimate of the mean meridional motion also seems to support this evaluation of the Murgatroyd and Singleton mean meridional circulation.

The model of the distribution of ozone calculated for the equinoxes as explained in the section on page 13, with the above scheme of transport processes, is shown in Figure 14.

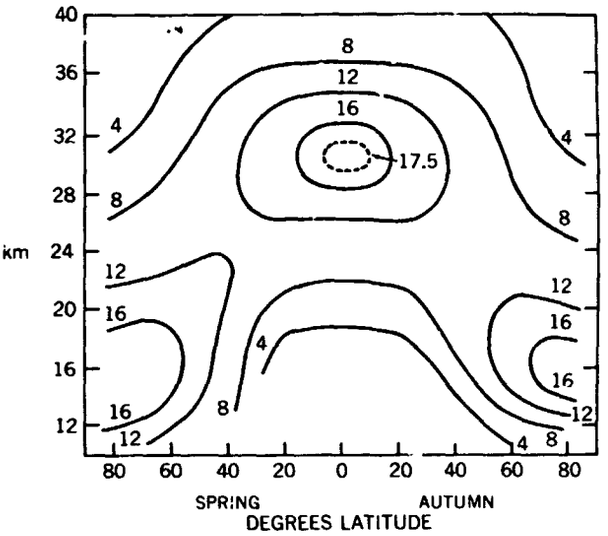


FIGURE 14.—Steady state distribution of ozone (D.U./km) computed for the equinoxes with constant value of K_Y (2×10 cm/sec), K_z (Table 1) increasing toward pole, and 20 percent of Murgatroyd and Singleton's (Reference 37) mean meridional circulation.

The concentration of ozone on the tropopause, the lower boundary of the model, is specified with the help of the "observed" distribution of Paetzold et al. (Figure 4) for spring and autumn.

The computed model thus obtained for the equinoxes still shows the ozone high produced by solar radiation over the tropical latitudes. However, there is a reasonably good lateral spread of ozone that is best illustrated in Figure 15(b) in which the latitudinal variation of total ozone calculated for the equinoxes is plotted. The total ozone obtained from the photochemical equilibrium calculation for equinoxes is shown in Figure 15(a) for comparison.

From the scheme of the transport processes used to calculate this (Figure 14) distribution of ozone, it is obvious that the diffusion pattern is symmetrical about the equator. There is, however, asymmetry of circulation present between the spring and autumn hemispheres. Such superficial argument cannot be carried further to conclude that the difference between the distribution of ozone between spring and autumn is due entirely to the difference in circulation intensity. Such a conclusion is misleading. The large scale diffusion processes in the vertical and horizontal directions interact in a nonlinear fashion with the mean meridional circulation. It is more justified to infer that the difference between the latitudinal distribution of ozone in spring and autumn is a result of the mutual interaction of the circulation and diffusion phenomena.

The increase of total ozone with latitude shown in Figure 15(b) is inadequate to explain the total ozone increase with latitude portrayed by London (Reference 10) for spring and autumn. This immediately leads one to an examination of the horizontal eddy diffusion coefficient. Spar's investigations show that the horizontal eddy diffusion coefficient has a large range of values.

Table 1

Variation of K_z with Latitude for the Equinoxes.

Degrees latitude	0	10	20	30	40	50	60	70	80	90
K_z (10^4 cm²/sec)	0.4	0.4	0.6	0.9	1.4	2	2	2	2	2

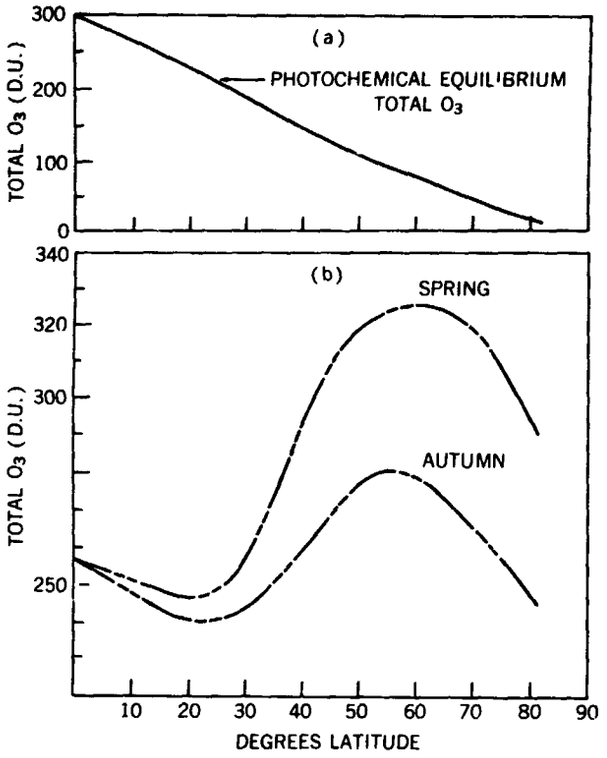


FIGURE 15.—Comparison between the total ozone variation with latitude obtained from pure photochemical theory (Figure 12), and from the combined effects (Figure 14).

Therefore further experimentation on the meridional distribution of ozone has been done, increasing the horizontal eddy diffusion coefficient from tropical latitudes to polar latitudes. Eventually a variation of K_Y somewhat analogous to the variation of K_Z with latitude has been found suitable. The variation of K_Y thus arrived at is shown in Table 2.

The meridional distribution of ozone, calculated with such a variation of K_Y along with the model of the K_Z and mean meridional motion used in the preceding computation, is shown in Figure 16. The variation of the total ozone with latitude given by this distribution is plotted in Figure 17.

The increase of total ozone from tropical latitudes to the polar latitudes shown in Figure 17 is

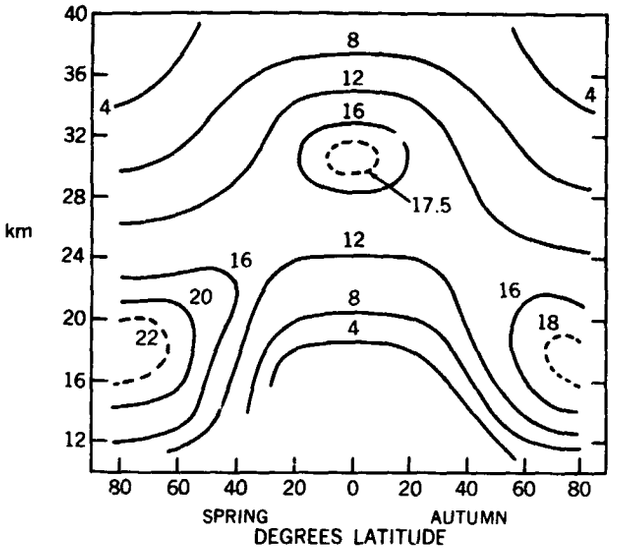


FIGURE 16.—Steady state distribution of ozone (D.U./km) computed for the equinoxes with K_Z (Table 1) increasing with latitude, and 20 percent of Murgatroyd and Singleton's (Reference 37) mean meridional circulation.

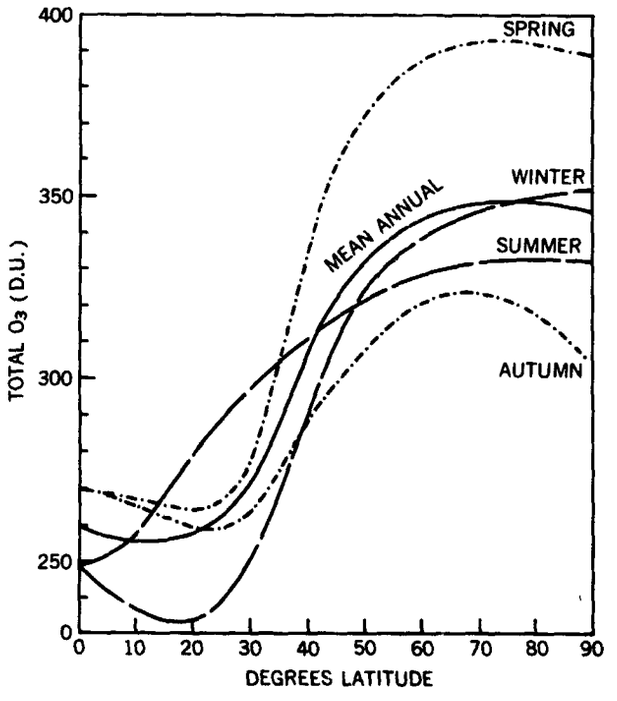


FIGURE 17.—Computed seasonal and latitudinal variation of total ozone corresponding to Figures 16 and 18.

Table 2

Variation of K_Y with Latitude for the Equinoxes.

Degrees latitude	0	10	20	30	40	50	60	70	80	90
$K_Y (10^{10}$ cm^2/sec)	0.2	0.2	0.4	0.6	0.8	1.0	1.0	1.0	1.0	1.0

in better agreement with that of London's curves for spring and autumn. On the basis of reasonable agreement between the observed total ozone curves and the computed ones, it is felt that, in order to explain the spring buildup of ozone at higher latitudes by the steady state approach, increase of the horizontal eddy diffusion coefficient from tropical latitudes to polar latitudes is necessary.

After having studied in detail the behavior of the various transporting mechanisms for the spring and autumn seasons, it is straightforward to proceed to the summer and winter seasons. Furthermore, the quantitative model of the transport mechanisms developed to explain the distribution of ozone for spring and autumn may now be checked in the computation of the steady state distribution of ozone for summer and winter.

From the observations made by Friend et al. (Reference 42) on the W^{185} cloud it is apparent that the center of the cloud has a tendency to move with the sun. This interesting detail suggests a shift of the whole pattern of the diffusion field along with the sun. In other words, the hemisphere tilted toward the sun has a comparatively weak pattern of diffusion, while the hemisphere tilted away from the sun has a correspondingly more intense mixing. Such an idea seems to be quite consistent with the relatively stable lapse rates of temperature in the summer stratosphere and comparatively less stable lapse rates and polar night jet stream (Hare, Reference 44) in the winter stratosphere.

The above ideas have been suitably used to adjust the diffusion pattern for the summer and winter hemispheres as shown in Table 3.

With the pattern of vertical and horizontal eddy diffusion shown in Table 3, and with 20 percent of Murgatroyd and Singleton's mean meridional circulation for the solstices, a steady state distribution of ozone is calculated as explained in the section on page 1313.

The boundary condition on the tropopause is again specified from the Paetzold et al. model. It is tacitly assumed here that the concentration ozone on the tropopause in winter is similar to that in spring, and in summer is similar to that in autumn. Such an assumption may not be unreasonable if one considers the latitudinal distribution of total ozone during these seasons. Incidentally this assumption provides a good ground to test the influence of the lower boundary condition on the distribution of ozone in the interior of the field of computation.

The distribution of ozone resulting under steady state condition for solstices calculated with the above boundary condition on the tropopause is shown in Figure 18. The variation of total ozone with latitude computed for the corresponding seasons is plotted in Figure 18.

The latitudinal increase of total ozone for winter and summer shown in Figure 18 reveals some significant departures from the corresponding curves of London (Reference 10). The solar radiation incident normally at 23.5° latitude on the summer hemisphere in the computed model has apparently contributed to comparatively larger values of the total ozone in the tropical latitudes. In the winter hemisphere the total ozone from the computed model in the tropical latitudes seems to be appreciably smaller than that given by London's curve. The relatively

Table 3
Variation of K_z and K_y with Latitude for the Solstices.

Summer									
Degrees latitude	10	20	30	40	50	60	70	80	90
K_z (10^4 cm²/sec)	0.4	0.4	0.4	0.6	0.9	1.4	2.0	2.0	2.0
K_z (10^{10} cm²/sec)	0.2	0.2	0.2	0.4	0.6	0.8	1.0	1.0	1.0

Winter										
Degrees latitude	0	10	20	30	40	50	60	70	80	90
K_z (10^4 cm²/sec)	0.6	0.9	1.4	2.0	2.0	2.0	2.0	2.0	2.0	2.0
K_z (10^{10} cm²/sec)	0.4	0.6	0.8	1.0	1.0	1.0	1.0	1.0	1.0	1.0

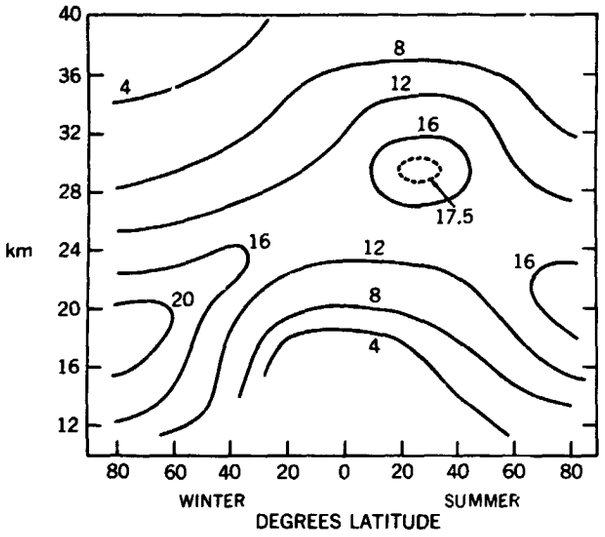


FIGURE 18.—Steady state distribution of ozone (D.U./km) computed for the solstices with K_Y and K_Z (Table 3) increasing from subsolar point, and 20 percent of Murgatroyd and Singleton's mean meridional circulation.

larger total amounts of ozone in summer and smaller amounts in winter in the tropical latitudes shown by London's curves lend a qualitative support to the computed results. London's curves show a maximum difference of about 30 D.U. between the summer and winter at the equator with a mean total ozone of about 250 D.U. The computed total ozone curves show a maximum difference of about 48 D.U. at about 25° latitude with a mean total of about 258 D.U.

Although the total ozone variation with latitude computed for the four seasons shown in Figure 17 qualitatively agrees with the observed curves (London's), there is considerable disagreement between the two. The total ozone computed for spring and winter is still short of the observed increase of ozone for those seasons. On the other hand, the calculated total ozone for summer and autumn is considerably larger than that suggested by the observed values. The difference between the winter and summer total ozone from the observed values, at higher latitudes, is almost double that obtained from the computations.

The seasonal variation of ozone in this study is computed with an assumption of steady state conditions. Such an assumption may not be justified for each season, but should be valid for the mean annual ozone distribution with latitude. The mean annual total ozone distribution is obtained from the computed total ozone curves for the four seasons as shown in Figure 19. The mean annual total ozone distribution constructed in this fashion is in reasonable agreement with that of the observed total ozone curves (London, Reference 10).

Until now the discussion on the computed meridional distributions of ozone has been based on the total ozone variation with latitude. The emphasis laid on the total ozone rather than the details of its distribution in the meridional plane is mainly due to the fact that the total ozone is the only parameter of atmospheric ozone known with a good degree of accuracy.

In general, all of the computed models of the meridional distribution of ozone do not quite resemble the "observed" models. Each one of the computed models clearly show three ozone highs. One of these ozone highs is evidently produced by the solar radiation and is in the tropical stratosphere centered around 30 km height. The other two ozone highs are a result of the transport processes, and their centers are located around 18 km height at the poles in spring, autumn, and winter hemispheres. This center is around 20 km height at the summer pole.

The "observed" models of the meridional distribution of ozone do not indicate the existence of an ozone high in the tropical stratosphere.

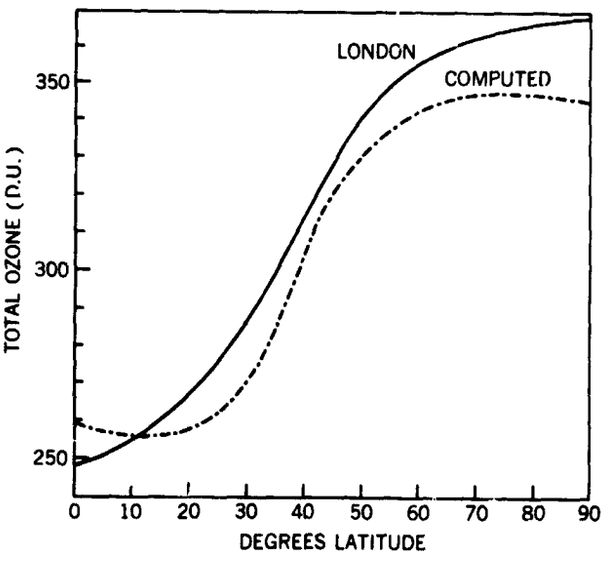


FIGURE 19.—Comparison of the computed mean annual total ozone distribution with that of London (Reference 10).

The ozone high close to the pole is present only in the model of Paetzold in the spring hemisphere.

In the computed models at higher latitudes, large horizontal and vertical gradients of ozone are present close to the tropopause during spring, autumn, and winter, while during summer a rather weak gradient is developed. The distribution of ozone in the lower stratosphere during summer is thus considerably different from the other three seasons. This is primarily a consequence of the weak diffusion pattern and the rising air motions present over large portions of the stratosphere in the summer hemisphere.

The "observed" models suggest the presence of strong horizontal and vertical gradients of ozone in the lower stratosphere at high latitudes during spring and probably during late winter.

The zone of ozone maximum, formed by the three ozone highs in both the equinoctial and the solsticial models of the steady state distribution of ozone calculated in this study, slopes down steeply from the subsolar point to the poles. The "observed" models apparently do not indicate such a steep slope for the zone of ozone maximum.

From the above brief discussion of the meridional distribution of ozone, it may be remarked that the computed models differ from the "observed" models in many details, but the general patterns in both the computed and the "observed" models are similar. Until our knowledge of the "observed" distribution of ozone in the atmosphere is better than it is at present, it will not be possible to judge quantitatively the details of the computed models.

To test the influence of the lower boundary condition on the computed distribution of ozone in the interior of the field, a calculation of the ozone distribution for the solstices is made with radically different boundary condition on the tropopause. In the computed distribution shown in Figure 18 the lower boundary condition has been taken to be exactly the same as that used for spring and autumn. Now in order to test the significance of the lower boundary condition, the distribution of ozone on the tropopause for winter is taken to be half that of spring, and for summer, twice that of autumn. Such an alteration of the boundary condition, however, is done from 40° latitude to the pole on either hemisphere as shown in Table 4.

In order to bring to light the significance of the lower boundary condition, the total ozone variation with latitude corresponding to the two sets of lower boundary condition for solstices is presented in Figure 20. Besides this, in Figure 21

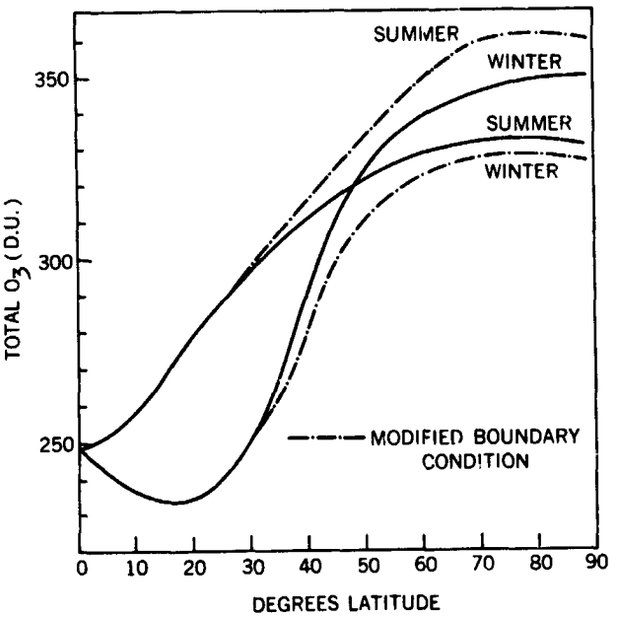


FIGURE 20.—Influence of the boundary condition at the tropopause on the computed total ozone.

the vertical distribution of ozone, taken from the meridional distributions of ozone computed with these two different boundary conditions at 60° latitude in the winter and summer hemispheres, is presented.

From Figure 21 it may be observed that changing the lower boundary condition does not affect

Table 4

Modified Value of the Ozone Concentration (10^{12} molecules/cm^2) at the Tropopause in Winter and Summer.

Degrees latitude	0	10	20	30	40	50	60	70	80	90
Winter	0.54	0.54	0.67	1.08	1.41	1.34	1.48	1.75	2.00	2.00
Summer	0.54	0.54	0.54	0.54	0.08	1.08	2.16	2.64	3.20	3.20

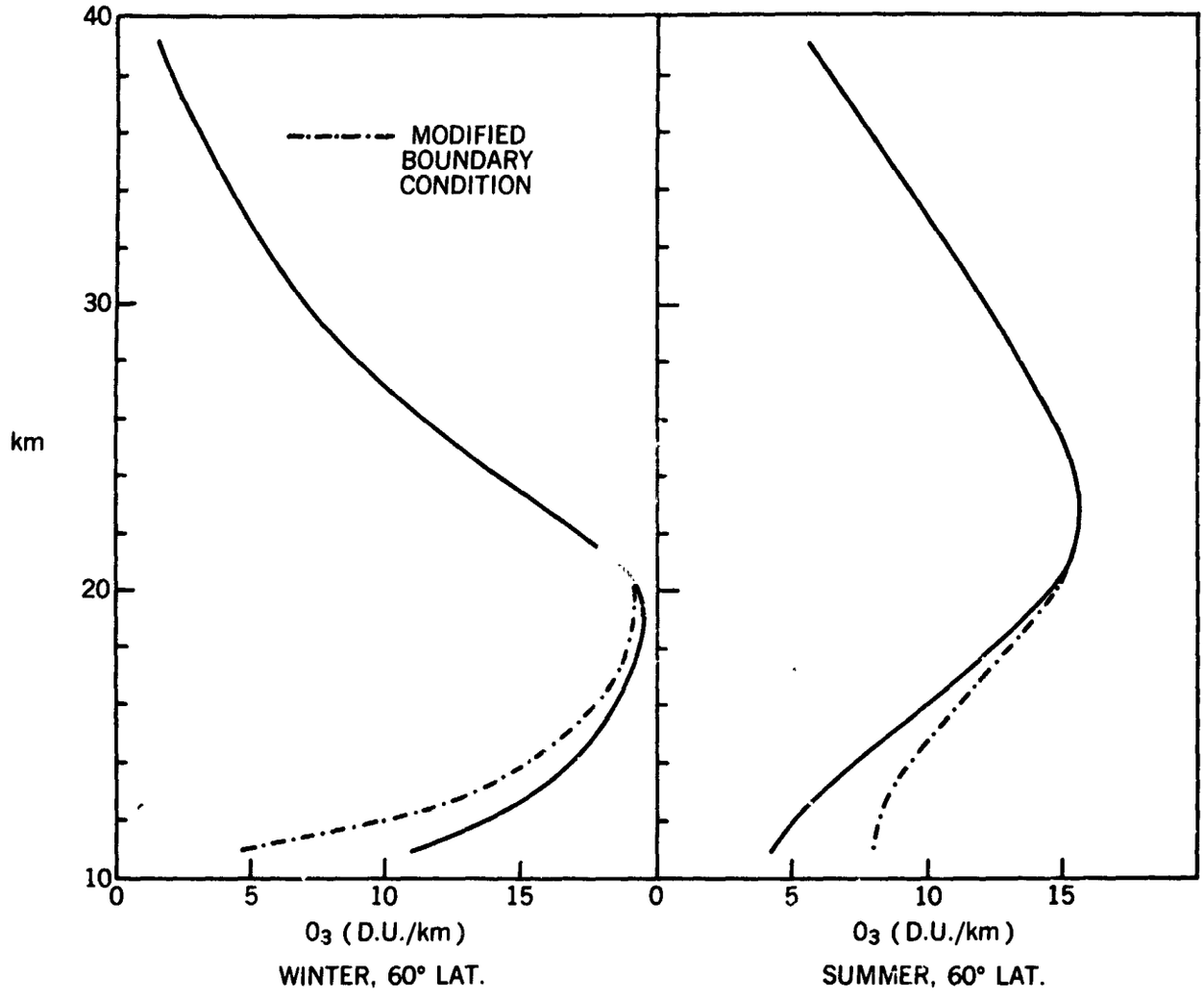


FIGURE 21.—Modification of the vertical distribution of ozone, at 60° latitude in the summer and winter hemispheres produced by changing the boundary condition at the tropopause.

the ozone distribution above about 20 km. The influence of the boundary value rapidly decreases with height. Changing the boundary by a factor of 2 has altered the total ozone by about 8 percent. This comparatively small change is obviously due to the effect of the transport processes that strongly control the distribution of ozone within the interior of the field.

Limitations of the Study

The study of the meridional distribution of ozone made in this investigation is purely a numerical experiment.

In this study the absorption coefficient of ozone in the spectral interval 2200 to 2020 A is adopted, with suitable modification, from Vassy's (Reference 23) photographic photometric measurements. If, however, the absorption coefficient of ozone, in the same spectral interval, reported by Tanaka et al. (Reference 25) is used, one would get much larger photochemical production of ozone, especially at tropical latitudes. Then in order to reduce such large amounts of ozone produced at the tropical latitudes to the amounts observed, it is necessary to use rising motions in the tropical stratosphere of about the same magnitude as shown Murgatroyd and Singleton's (Reference 37) model. So the present study in which Vassy's absorption coefficient of ozone is used cannot unequivocally place the stratospheric meridional

circulation at 20 percent of that of Murgatroyd and Singleton.

The calculation of the meridional distribution of ozone for different seasons has been made as though the mean seasonal transport processes and the mean solar insolation for the season tend to produce a quasi-stationary pattern of ozone. Such an assumption can make the results obtained in this study not applicable into the real atmosphere in which both the solar insolation and transport processes are varying with time. The seasonal variations of ozone computed in this study are especially subject to such a limitation. The mean annual total ozone distribution is probably the least affected quantity obtained in this investigation with the steady state assumption.

A more satisfactory study of atmospheric ozone would be that in which one does not have to invoke the quasi-stationary seasonal distribution of ozone. Such a study would call for an approach to this problem as an initial value problem. This implies that one should have a detailed and reliable initial distribution of ozone, and also good quantitative information on the seasonal variation of large scale transport processes in the atmosphere. In view of the lack of this detailed information demanded by the initial value problem, the present study has been made by taking a less satisfactory but more simple approach. Hence caution must be exercised in interpreting the results obtained from this study with a steady state assumption. The present investigation can be no more than a starting point to more sophisticated and realistic future studies of the atmospheric ozone. The satisfactory manner in which the present investigation has accounted for the observed total ozone in the atmosphere can at best be only a crude but quantitative demonstration of the importance of the transport processes in distributing ozone in the atmosphere.

The models of the transport processes used in this investigation lack rigorous dynamical basis. The efforts made to synthesize the advective and diffusive fields in the atmosphere are purely from the point of view of explaining the seasonal variation of the atmospheric ozone. The ability of such a scheme of transport processes to explain the seasonal variation of temperature or W^{185} in the stratosphere is still questionable. The only justification that can be attributed to this scheme of transport processes is that it is within reasonable limits laid by the current theoretical and observational findings.

The variations of ozone in the east-west direction are by no means negligible. The investigations of the variation of total ozone by Bidner (Reference 45) over Western Europe strongly suggest such zonal variations. The assumption of zonal symmetry of ozone made in this study is a very broad assumption.

The height of the tropopause is assumed to vary smoothly from tropical latitudes to higher latitudes. On the contrary, the observed shape of the tropopause shows a discontinuity between the tropical and middle latitudes. Brewer (Reference 6) and Friend et al. (Reference 42) have discussed at length the important part played by this break in the tropopause and the jet stream present at this break in the exchange of ozone and nuclear debris, respectively, between the troposphere and stratosphere. With the information available at present, it does not seem possible to introduce such features on a quantitative basis in this study.

The influence of the large scale fluctuating vertical velocities in the early spring, which produce explosive warming and a large rise of ozone (Reference 46), has not been included in this study. Also the meridional transport due to the meandering waves and vertical motions associated with the jet stream present in the polar night (Hare, Reference 44) has not been quantitatively incorporated.

SUMMARY AND CONCLUSIONS

In the light of the new data on solar radiation and other parameters necessary for photochemical calculations, a computation of the meridional distribution of photochemical equilibrium of ozone for solstices and equinoxes is made. It is found, in agreement with ideas of the earlier investigators, that the photochemical theory by itself cannot explain the latitudinal and seasonal variation of ozone.

To study the part played by transport processes in the stratosphere, the necessary equation which includes the combined effects of photochemical and transport processes has been developed for steady state conditions. With such an equation and using a scheme of transport process that is

within reasonable agreement with the current ideas on transport processes in the stratosphere, meridional distribution of ozone produced under steady state conditions is calculated for equinoxes and solstices. The seasonal and latitudinal variation of total ozone thus obtained is in qualitative agreement with the observed variations of total ozone in the atmosphere shown by the investigations of London (Reference 10).

Considering the limitations of this study, it is hard to say that all the problems concerning the behavior of ozone in the atmosphere have been answered quantitatively. But it may be remarked that the technique developed in this investigation is potentially good for studying the problem of atmospheric ozone. It is hoped that such a quasi-stationary approach would put the problem of ozone, at least as a first step, on a quantitatively understandable level.

ACKNOWLEDGMENTS

The author is very grateful to Prof. Jerome Spar and Prof. Julius London for their valuable advice. The computational work was carried out on the IBM 7090 computer at the Institute for Space Studies of NASA, with the kind permission of Prof. R. Jastrow.

The research reported in this paper is based on a Ph.D. thesis, New York University, and was supported by Geophysics Research Directorate, Air Force Cambridge Research Division, under contract AF 19(604)-5492.

REFERENCES

1. Götz, F. W. P., "Zum Strahlungsklima des Spitzbergensommers. Strahlungs- und Ozonmessungen in der Königsbucht 1929," *Gerl. Beitr. Geophys.* 31:119–154, 1931.
2. Ramanthan, K. R., and Kulkarni, R. N., "Mean Meridional Distributions of Ozone in Different Seasons Calculated from Umkehr Observations and Probable Vertical Transport Mechanics," *Q. Jour. Roy. Met. Soc.* 86(368):144–155, April 1960.
3. Regener, V. H., "On the Vertical Distribution of Atmospheric Ozone," Contr. AF 19(122)-381, Dept. of Physics, Univ. of New Mexico, Albuquerque, New Mexico, 1954.
4. Epstein, E. S., Osterberg, C., and Adel, A., "A New Method for the Determination of the Vertical Distribution of Ozone from a Ground Station," *J. Met.* 13(4):319–334, August 1956.
5. Goody, R. M., and Roach, W. T., "Determination of the Vertical Distribution of Ozone from Emission Spectra," *Q. Jour. Roy. Met. Soc.* 82(352):217–221, April 1956.
6. Brewer, A. W., "The Transfer of Atmospheric Ozone into Troposphere," Sci. Rep. Planetary Cir. Proj. Dept. of Meteorology, M.I.T., 1960.
7. Paetzold, H. K., and Piscalar, F., "Meridional Ozone Distribution and Stratospheric Circulation," *Naturwissenschaften*, 48(13):474, 1961 (in German).
8. Chapman, S., "Theory of Upper Atmospheric Ozone," *Mem. Roy. Met. Soc.* 3:103–125, June 1930.
9. Craig, R. A., "The Observations and Photochemistry of Atmospheric Ozone and Their Meteorological Significance," Boston: American Meteorological Society, 1950 [Meteorological Monographs, 1(2)].
10. London, J., "Mesosphere Dynamics, Part 3, Final Report," New York Univ., College of Engineering, New York, AFCRL-62-672, May 1962.
11. Ramanathan, K. R., "Atmospheric Ozone and the General Circulation of the Atmosphere," *in:* "Scientific Proceedings of the International Association of Meteorology, Rome 1954," 3–24, London: Butterworth Scientific Publications, 1956.
12. Miller, L. E., "Atmospheric Composition," *in:* "Handbook of Geophysics," (C. F. Campen, Jr., and others, eds.) Chapter 8:1–6, New York: MacMillan Co., 1960.
13. Wulf, O. R., and Deming, L. S., "Distribution of Atmospheric Ozone," *Terr. Mag.* 42:195–202, June 1937.
14. Dütsch, H. U., "Das atmosphärische Ozon als Indikator für Strömungen in der Stratosphäre, I. Die photochemischen Grundlagen," *Archiv Für Meteorl. Geophys. Biokl.* A(9):87–119, 1956.
15. Dütsch, H. U., "Photochemische Theorie des atmosphärischen Ozons unter Berucksichtigung von Nichtgleischgewichtzustanden," Doctoral dissertation, Univ. of Zurich, 1946.
16. Haurwitz, B., "Dynamic Meteorlogy," New York: McGraw-Hill Book Co., 1941.
17. Kellogg, W. W., "Chemical Heating Above the Polar Mesopause in Winter," *J. Meteorol.* 18(3): 373–381, June 1961.
18. Detwiler, C. R., Garrett, D. L., Purcell, J. D., Tousey, R., "The Intensity Distribution in the Ultraviolet Solar Spectrum," *Ann. Geophys.* 17(3): 263–272, 1961.
19. Hinteregger, H. E., "Preliminary Data on Solar Extreme Ultraviolet Radiation in the Upper Atmosphere," 66(8):2367–2380, August 1961.
20. Gast, P. R., "Solar Radiation," *in:* "Handbook of Geophysics," (C. F. Campen, Jr., and others, eds.) chapter 16:14–32, New York: MacMillan Co., 1960.
21. Watanabe, K., Inn, E. C. Y., and Zelikoff, M., "Absorption Coefficients of Gases in the Vacuum Ultraviolet," *J. Chem. Phys.* 20:1969–1970, December 1952.
22. Heilpern, W., "The Effect of Added Nitrogen on the Absorption of Light by Oxygen as a Function of the Pressure," *Helv. Phys. Acta*, 25(7):753–772, 1952 (in German).

23. Vassy, A., "Sur l'Absorption Atmospherique dans l'Ultra-Violet," *Annales de Physique*, 16:145–203, July-September, 1941.
24. Ny, T. Z., and Choong, S. P., "Absorption Band of Ozone Between 3050 Å and 2150 Å," *Comptes Rendus*, 196:916–918, March 27, 1933 (in French).
25. Tanaka, Y., Inn, E. C. Y., and Watanabe, K., "Absorption Coefficients of Gases in the Vacuum Ultraviolet. IV. Ozone." *J. Chem. Phys.* 21:1651–1653, October 1953.
26. Inn, E. C. Y., and Tanaka, Y., "Absorption Coefficient of Ozone in the Ultraviolet and Visible Regions," *J. Opt. Soc. Amer.* 43:870–873, October 1953.
27. Vigroux, E., "Contributions to the Experimental Study of the Absorption of Ozone," *Annales de Physique*, 8:709–762, September-October, 1953 (in French).
28. Murgatroyd, R. J., "Winds and Temperatures Between 20 km and 100 km—A Review," *Q. Jour. Roy. Met. Soc.* 83(358):417–458, October 1957.
29. London, J., Ohring, G., and Ruff, I., "Radiative Properties of the Stratosphere. Final Report," New York Univ. College of Engineering, New York, Contract AF 19(604)–1285, October 1956.
30. Campbell, E. S., and C. Nudelman, "Reaction Kinetics, Thermodynamics and Transport Properties in the Ozone Oxygen System," A FOSR TN–60–502, New York Univ., 1960.
31. Glissmann, A., and Schumacher, H.-J., "Der Thermische Ozonerfall," *Zeits, Physik. Chem.* 21B(4):323–348, May 1933.
32. Eucken, A., and Patat, F., "Effect of Temperature on the Photosynthesis of Ozone," *Zeits. Phys. Chem.* B. 33(6):459–474, 1936 (in German).
33. Richtmyer, R. D., "Difference Methods for Initial-Value Problems," New York: Interscience Publishers, 1957.
34. Brewer, A. W., "Evidence for a World Circulation Provided by the Measurements of Helium and Water Vapor Distribution in the Startosphere," *Q. Jour. Roy. Met. Soc.* 75(326):351–363, October 1949.
35. Dobson, G. M. B., "Origin and Distribution of the Polyatomic Molecules in the Atmosphere," *Proc. Roy. Soc.* A236(1205):187–193, August 2, 1956.
36. Mastenbrook, H. J., and Dinger, J. E., "Distribution of Water Vapor in the Stratosphere," *J. Geophys. Res.* 66(5):1437–1444, May 1961.
37. Murgatroyd, R. J., and Singleton, F., "Possible Meridional Circulations in the Stratosphere and Mesosphere," *Q. Jour. Roy. Met. Soc.* 87(372): 125–135, April 1961.
38. Ohring, G., "Radiation Budget of the Stratosphere," *J. Meteor.* 15:440-451, 1958.
39. Murgatroyd, R. J., and Goody, R. M., "Sources and Sinks of Radiative Energy from 30 to 90 km," *Q. Jour. Roy. Met. Soc.* 84(361):225–234, July 1958.
40. Stebbins, A. K., III, "Special Report on High Altitude Sampling Programs," Defense Atomic Support Agency, Washington, D. C., DASA–532B, June 1, 1960.
41. Feely, H., and Spar, J., "Tungsten 185 from Nuclear Bomb Tests as a Tracer for Stratospheric Meteorology," *Nature*, 188(4756):1062–1064, December 24, 1960.
42. Friend, J. P., Feely, H. W., Krey, P. W., Spar, J., and Walton, A., "The High Altitude Sampling Program, Discussion of Hasp Results. Final Report," Isotopes Inc., Westwood, N. J., DASA 1300, volume 3, August 1961.
43. Newell, E. R., "The Transport of Trace Substances in the Atmosphere and Their Implications for the General Circulation of the Stratosphere," *Geophysica Pura e Applicata* 50:138–158, 1961.
44. Hare, F. K., "The Disturbed Circulation of the Arctic Stratosphere," *J. Met.* 17:36–51, February 1960.
45. Bidner, A., "Synoptic Aspects of the Total Ozone Amount over Europe," M.S. Thesis, New York Univ., Dept. of Meteorology and Oceanography, 1958.
46. Godson, W. L., "Total Ozone and the Middle Stratosphere over Arctic and Sub-Arctic Areas in Winter and Spring," *Q. Jour. Roy. Met. Soc.* 86(369):301–317, July 1960.

N66 37972

See N65-98541

ELECTRON MICROSCOPY OF METEORITIC AND ARTIFICIALLY SHOCKED GRAPHITE*

THOMAS P. SCIACCA AND MICHAEL E. LIPSCHUTZ†

Samples of graphite, partly converted to diamond by shock, and meteoritic graphite were studied by transmission electron microscopy. The shocked graphite remained in the form of single crystals and was mildly deformed while the diamond which had formed was anhedral and polycrystalline. The meteoritic graphite contained minor amounts of troilite (FeS) and elemental sulfur but no detectable carbon phase other than graphite.

As an adjunct to an x-ray crystallographic study of meteoritic diamonds (*1*) we have studied the morphology of the phases present in meteoritic graphite, and in graphite which has been artificially shocked at pressures high enough (300 kbar) to produce diamond (*2*).

Samples of the starting material, high-purity graphite (*3*), as well as the resulting shock-product, and graphite from the interior of a 1.1-g graphite nodule from the Canyon Diablo iron meteorite were crushed to 100 mesh. Slurries in ethanol were then placed in an ultrasonic generator for 5 minutes to disperse the sample uniformly. The suspensions were placed on Formvar, collodion, or carbon substrates for examination in a Hitachi HU-11 electron microscope. Phases were identified by electron diffraction of selected areas with gold and palladium as standards.

Figure 1*a* shows a typical graphite-diamond intergrowth from the artificially shocked graphite. All of the diamonds formed in this manner were anhedral and polycrystalline. The graphite flakes seemed mildly deformed although they remained single crystals. Figure 1*b* shows a typical graphite flake from the starting material.

In both the starting material and shock product the only other phase that we found was in the

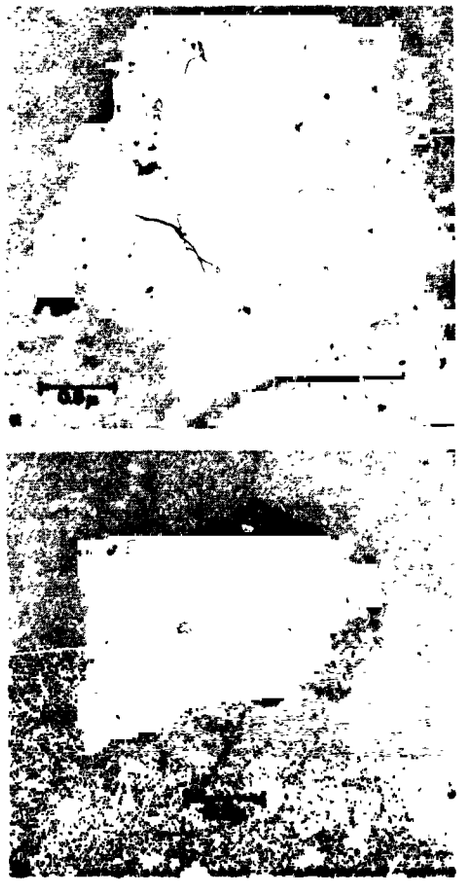


FIGURE 1.—(a) Diamond-graphite intergrowth in artificially shocked graphite. Note the anhedral morphology of the larger diamond grains (D). (b) Typical flake from the graphite starting material. The texture of this flake may be compared with that of the untransformed graphite in (a).

*Published in *Science*, 145 (3636): 1049–1050, September 4, 1964.
*Goddard Space Flight Center.
†Physikalisches Institut and Universitat Bern, Sidlerstrasse 5, Bern, Switzerland.

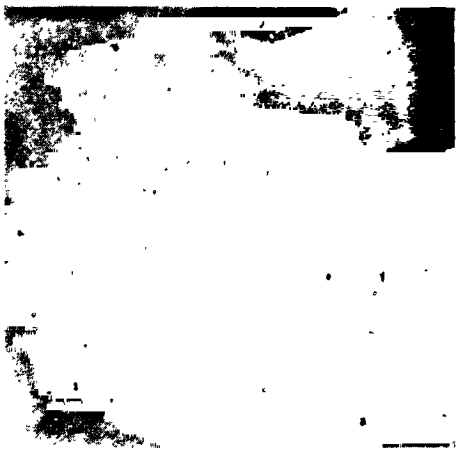


FIGURE 2.—Unidentified species present in both starting material and shock product. The cell dimension of these cubes is similar to that of the "carbon II" phase reported by Aust and Drickamer (4.)

form of euhedral cubes (Fig. 2) which were single crystals of uncertain origin. The cell dimension of these simple cubic crystals ($\alpha_0 = 5.55$ Å) corresponds very closely to that of the "carbon II" phase ($\alpha_0 = 5.545$ Å) reported by Aust and Drickamer (*4*). Inasmuch as the cubes were present in very small numbers ($<<1$ percent) in both the starting material and product we cannot say that they were produced by shock or, in fact, that they were composed of carbon. It seems very unlikely that some extraneous material would possess exactly the same cell dimension and crystallographic habit as "carbon II". Yet the cubes' euhedral appearance, the general perfection of these crystals, and their low abundance suggest that they were contaminants and not shock-formed. It could also be that these cubes were, indeed, "carbon II" which had grown relatively slowly during preparation of the original artificial graphite block. Unfortunately, all three alternatives (contamination, shock-formation, or slow growth) seem equally unlikely and it is impossible to decide among them with the available data.

The anhedral morphology of the artificial diamond might be expected from the short time available for crystal growth under the experimental shock conditions. Shocks generated during asteroidal collision (*1*) on the other hand, no doubt were of longer duration and might well be expected to result in the formation of subhedral diamond grains. These grains would probably be polycrystalline aggregates, since the shock-induced conversion of graphite to diamond apparently proceeds by a mechanism other than the orderly diffusion and addition of carbon atoms to diamond nuclei (*1*, *2*).

X-ray fluorescence analysis of the interior of the Canyon Diablo nodule indicated the presence of only iron and sulfur. The only carbon phase detected was normal graphite. The cubic phase seen in the starting material and shock product was not observed in the Canyon Diablo nodule.

The sulfur in the Canyon Diablo material apparently exists in two forms; as free rhombic sulfur and as troilite (FeS). Figure 3*a* shows anhedral single crystals of sulfur in a graphite matrix at an electron accelerating voltage of 75 kv. At 100 kv, the normal operating voltage



FIGURE 3.—(a) Sulfur (S), graphite mixture from Canyon Diablo (75-kv electrons). (b) The same area examined with 100-kv electrons. The higher accelerating voltage has sufficiently increased the temperature to melt the sulfur grains.

in these studies, the temperature reached was high enough to melt the sulfur (m.p. 113°C), and further aided in its identification (Fig. 3*b*). The troilite in the graphite nodule was identified by both x-ray and electron diffraction. It was present in amounts greater than about 5 percent while the sulfur was present in lesser amounts (about 2 percent).

The coexistence of troilite and elemental sulfur has been observed previously in Type I carbonaceous chondrites (*5*) although in these meteorites there is considerably more sulfur than troilite. Free sulfur has not previously been reported in iron meteorites. It may be that the sulfur was formed by terrestrial weathering of the troilite, the iron thus liberated forming amorphous limonite (*6*). An alternative is that these minerals are preterrestrial. If this was the case it is expected that the S^{32}/S^{34} ratios in the troilite and sulfur may differ.

REFERENCES AND NOTES

1. M. E. LIPSCHUTZ, Science 143, 1431 (1964).
2. P. S. DECARLI and J. JAMIESON, *ibid.* 133, 1821 (1961).
3. National Carbon Company grade AUC high purity graphite.
4. R. B. AUST and H. G. DRICKAMER, Science 140, 817 (1963).
5. E. R. DUFRESNE and E. ANDERS, in *The Moon, Meteorites and Comets*, B. M. MIDDLEHURST and G. P. KUIPER, Eds. (Univ. of Chicago Press, Chicago, 1963), vol. 4, pp. 496-526.
6. We thank PROF. PAUL RAMDOHR for pointing out this possibility.
7. This study was carried out, in part, while M. E. LIPSCHUTZ was in the U.S. Army at the Goddard Space Flight Center. We thank PEDRO C. SARMIENTO of the Goddard Space Flight Center for his valuable assistance, and P. S. DECARLI (Stanford Research Institute) and E. P. HENDERSON (U.S. National Museum) for providing samples. One of us (M.E.L.) thanks the National Science Foundation for a postdoctoral fellowship at Bern, and the Schweizerischen Nationalsfonds.

See N64-29205

A.C. CONDUCTIVITY OF A PLASMA*

C. S. SHEN† AND R. L. W. CHEN‡

This paper computes the electrical conductivity of a fully ionized, spatially homogeneous plasma under the influence of a uniform, periodically alternating electric field. The velocity distribution of the electrons is determined by solving the linearized Fokker-Planck equations. All the terms in the collision integral are retained, including those representing electron-electron interactions. The resultant values of conductivity is expected to be valid in the range of frequencies from zero to below the plasma frequency.

I. INTRODUCTION

The purpose of this paper is to calculate the a.c. conductivity of a spatially homogeneous plasma using the Fokker-Planck equation. The d.c. conductivity of a plasma has been calculated in the well-known works of COHEN *et al.* (1950) and SPITZER Jr. and HÄRM (1953). Their results are in good agreement with the later experimental works of LIN *et al.* (1955). BERNSTEIN and TREHAN compute the a.c. conductivity assuming a Lorentz gas model (1960). The a.c. conductivity of a real gas should approach that of a Lorentz gas at high frequencies (see detailed discussions in Section 4). Toward lower frequencies their departure is expected to increase so that their ratio becomes nearly 2 in the d.c. limit, in accordance with COHEN *et al.* and SPITZER and HÄRM. The recent works on a.c. conductivity by DAWSON and OBERMAN (1962) consider the time variation of the two-particle distribution, which is necessary when dealing with a.c. currents of ultra-high frequencies. However, the domain of applicability of their work is limited to frequencies much higher than the collision frequency. Thus, a more precise calculation for the low and intermediate range of ω appears desirable and we proceed to do this in accordance with methods to be described in the next section. After completion of most of the numerical work, a paper by ROBINSON and BERNSTEIN (1962) came to our attention. They computed the a.c. conductivity using a variational technique. Our results obtained by direct integration of the Fokker-Planck equation will be compared to theirs in Section 4.

We begin with the Boltzman equation:

$$\frac{\partial f_i}{\partial t}+\mathbf{v}\cdot\frac{\partial f_i}{\partial \mathbf{r}}=\frac{F}{m}\cdot\frac{\partial f_i}{\partial \mathbf{v}}+\left(\frac{\delta f_i}{\delta t}\right)_c \qquad (1)$$

where f_i is the distribution function of particles of type i, $\left(\frac{\delta f_i}{\delta t}\right)_c$ is the change of f_i produced by collisions.

Equation (1) is deduced from Liouville theorem to describe a many-particle system under two assumptions:

(i) That the characteristic dimensions of the inhomogeneities are much larger than the average impact parameter for the particles participating in the collision.

(ii) That the characteristic time variation of the process is much longer than the duration of an average collision, or in other words, a collision is completed and the correlation function is 'relaxed' before the distribution function itself makes any appreciable change.

It should be noted here that the term 'duration of collision' is different from the so-called 'collision time'; collision time is the time between two collisions. For particles interacting through long-range forces, this time may be regarded as the time in which deflexions gradually deflected the considered particle by 90°. Duration of collision is the time during which an interaction takes place. In a plasma it is of the order ω_p^{-1}. In Fig. 1 a

*Published in *Plasma Physics* (Journal of Nuclear Energy Part C), 6:389–399, 1964.

†National Academy of Sciences—National Research Council Research Associate with NASA and Research Fellow of Institute of Physics, Academia Sinica, The Republic of China.

‡National Academy of Sciences—National Research Council Research Associate with NASA.

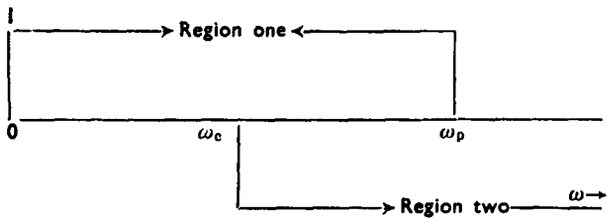


FIGURE 1.—The ranges of validity computed a.c. conductivities. The values of a.c. conductivity obtained in this paper is valid in region one. When omega exceeds ω_c—region two—the values calculated by Dawson and Oberman begin to be valid. Here ω_c is the collision frequency, ω_p is the plasma frequency.

time-scale diagram is drawn, and the validity of our calculation and those of DAWSON and OBERMAN are indicated.

The explicit expression of $\left(\frac{\delta f_i}{\delta t}\right)_c$ depends on the nature of the interaction force. In a fully ionized plasma, the particles interact through the long-range Coulomb forces. The cumulative effect of 'weak' deflexions resulting from the relatively distant collisions outweighs the effect of occasional large deflexions due to relatively close collisions, so one may neglect the contribution by those very close encounters (COHEN *et al.*)—encounters which result in deflexions of 90° or larger.

Also, the effect of distant particles lying outside the Debye length λ_D may be neglected because of the shielding of inner particles. Thus, in the computations of $\left(\frac{\delta f_i}{\delta t}\right)_c$, it is only necessary to consider the collisions with impact distance intermediate between λ_D and b_0, where $b_0=\frac{e^2}{KT}$ is the impact parameter yielding a 90° deflexion. The effects of these collisions are cumulative, and the total deflexion produced in an interval of time is similar to that of the Brownian motion; hence, one may expand $\left(\frac{\delta f_i}{\delta t}\right)_c^2$ in powers of $[\Delta \mathbf{v}]$, where $[\Delta \mathbf{v}]$ is the average velocity change due to collisions (COHEN *et al.*, CHANDRASEKHAR, 1943). This procedure leads to the following Fokker–Planck collision integral (ROSENBLUTH *et al.*, 1957):

$$\left(\frac{\delta f_i}{\delta t}\right)=\Gamma_i\left\{-\frac{\partial}{\partial \mathbf{v}}\left(f_i\frac{\partial h_i}{\partial \mathbf{v}}\right)+\frac{1}{2}\frac{\partial^2}{\partial \mathbf{v}\partial \mathbf{v}}\left(f_i\frac{\partial^2 g}{\partial \mathbf{v}\partial \mathbf{v}}\right)\right\} \quad (2)$$

where

$$h_i=\sum_j\frac{m_i+m_j}{m_j}\int d\mathbf{v}' f_j(\mathbf{v}')\,|\mathbf{v}-\mathbf{v}'|^{-1} \quad (3)$$

$$g=\sum_j\int d\mathbf{v}' f_j(\mathbf{v}')\,|\mathbf{v}-\mathbf{v}'| \quad (4)$$

and

$$\Gamma_i=\frac{4\pi z^2 e^4}{m_i^2}\ln\frac{m_i m_j v_{th}^2 \lambda D}{2(m_i+m_j)e^2}. \quad (5)$$

The summation in h_i and g sums over all species, m_i is the mass of the i-th species, e is the electronic charge, $\lambda_D=\left(\frac{KT}{4\pi ne^2}\right)^{\frac{1}{2}}$ is the Debye length, and v_{th} is the relative thermal velocity.

In this paper we consider only plasma with singly-charged ions. The extension of the present method to those with multiply-charged ions is straightforward.

2. DERIVATION OF EQUATIONS AND FORMULAE

If the distribution function f has an azimuthal symmetry about a certain axis, then, following ROSENBLUTH *et al.*, the collision term may be written down explicitly in spherical polar co-ordinates in velocity space:

$$\begin{aligned}\left(\frac{\delta f_i}{\delta f}\right)_c=\Gamma_i\Bigg\{&-v^{-2}\frac{\partial}{\partial v}\left[f_i v^2\frac{\partial h_i}{\partial v}\right]-v^{-2}\frac{\partial}{\partial \mu}\left[f_i(1-\mu^2)\frac{\partial h_i}{\partial \mu}\right]\\&+(2v^2)^{-1}\frac{\partial^2}{\partial v^2}\left[f_i v^2\frac{\partial^2 g}{\partial v^2}\right]\\&+2(v^2)^{-1}\frac{\partial^2}{\partial \mu^2}\Bigg[f_i\Bigg\{v^{-2}(1-\mu^2)^2\frac{\partial g^2}{\partial \mu^2}\\&+v^{-1}(1-\mu^2)\frac{\partial g}{\partial v}-v^{-2}\mu(1-v^2)\frac{\partial g}{\partial \mu}\Bigg\}\Bigg]\\&+v^{-2}\frac{\partial^2}{\partial \mu\partial v}\left[f_i(1-\mu^2)\left\{\frac{\partial^2 g}{\partial \mu\partial v}-v^{-1}\frac{\partial g}{\partial \mu}\right\}\right]\\&+(2v^2)^{-1}\frac{\partial}{\partial v}\Bigg[f_i\Bigg\{-v^{-1}(1-\mu^2)\frac{\partial^2 g}{\partial \mu^2}-2\frac{\partial g}{\partial v}\\&+2\mu v^{-1}\frac{\partial g}{\partial \mu}\Bigg\}\Bigg]+(2v^2)^{-1}\frac{\partial}{\partial \mu}\Bigg[f_i\Bigg\{v^{-2}\mu(1-\mu^2)\frac{\partial^2 g}{\partial \mu^2}\\&+2\mu v^{-1}\frac{\partial g}{\partial v}+2v^{-1}(1-\mu^2)\frac{\partial^2 g}{\partial \mu\partial v}-2v^{-2}\frac{\partial g}{\partial \mu}\Bigg\}\Bigg]\Bigg\}\end{aligned} \quad (6)$$

where $\mu=\cos\theta$ is the direction cosine between $\mathbf{v}$ and $\mathbf{E}$. Equation (6) is an exact expression of the Fokker–Planck equation in spherical co-ordinates for a distribution function with azimuthal symmetry.

We assume that the system is subject to a weak electric field $\mathbf{E}_0 e^{i\omega t}$ whose direction lies along z-axis. Then following CHAPMAN and COWLING (1939) and SPITZER Jr. we expand f_i in a power series of E:

$$f_i(\mathbf{v},t)=f_i^{(0)}(v)+E_0 f_i^{(1)}(\mathbf{v},t)+E_0^2 f_i^{(2)}(\mathbf{v},t)+\cdots \quad (7)$$

where $f_i^{(0)}(v)$ is a time-independent Maxwellian distribution and $f^{(1)}(\mathbf{v},t)$, $f^{(2)}(\mathbf{v},t),\cdots\cdots$ are the perturbed part due to applied electric field. When a steady state has been reached and no transient current exists, the time-dependent part of $f_i^{(j)}(\mathbf{v},t)$ must be proportional to $e^{i\omega t}$. Since the average energy imparted to the electrons between encounters is small compared with their kinetic energy, the velocity-dependent part of $f_i^{(j)}(\mathbf{v},t)$ can be written as $e^{-(m_i v^2/2kt)}D_i^{(j)}(v)\mu$. Therefore, we have

$$f_i^{(j)}(\mathbf{v},t)=\frac{m^{3/2}}{(2\pi KT)^{3/2}}e^{-m_i v^2/2KT}D_i^{(j)}(v)\mu e^{i\omega t}. \quad (8)$$

Combining equations (1), (6) and (7) keeping only terms linear in E_0, we obtain

$$\frac{\partial f_i^{(1)}}{\partial t}+\frac{e}{m_i}e^{i\omega t}\mathbf{E}_0\frac{\partial}{\partial \mathbf{v}}f_i^{(0)}=\left(\frac{\delta f_i^{(1)}}{\delta t}\right)_c \quad (9)$$

where $(\delta f_i/\delta t)_c{}^{(1)}$ is the linearized Fokker-Planck collision integral.

Since the ions' contribution to electric current is negligible compared to electrons, we will consider only electron distributions and drop the subscript i in the distribution function hereafter.

Substituting equation (8) into equation (9) we find, after some algebraic manipulations, the following second-order linear integral-differential equation:

$$D''(\chi)+P(\chi)D'(\chi)+Q(\chi)D(\chi)=R(\chi)+S(\chi) \quad (10)$$

where

$$P(\chi)=-2\chi-\frac{1}{\chi}+\frac{2\chi^2\Phi'(\chi)}{H(\chi)} \quad (11)$$

$$Q(\chi)=\frac{-iB\chi^3+2(1+\Phi=2\chi^3\Phi')}{H(\chi)}+\frac{1}{\chi^2} \quad (12)$$

$$R(\chi)=-\frac{2d\chi^4}{H(\chi)}-\frac{8(2\cdot 4\chi^6-2\chi^4)}{3\pi^{1/2}H(\chi)}I_0(\infty) \quad (1\upsilon)$$

$$S(\chi)=\frac{16}{3\pi^{1/2}H(\chi)}\{\chi I_3(\chi)-1\cdot 2\chi I_5(\chi) -\chi^4 I_0(\chi)(1-1\cdot 2\chi^2)\} \quad (14)$$

$$\Phi(\chi)=\int_0^\chi e^{-y^2}dy \quad (15)$$

$$H(\chi)=\Phi(\chi)-\chi\Phi'(\chi) \quad (16)$$

$$I_n(\chi)=\int_0^\chi y^n D(y)e^{-y^2}dy \quad (17)$$

$$I_0(\infty)=\int_0^\infty D(y)e^{-y^2}dy \quad (18)$$

with

$$\alpha=-E_0KT/\pi e^3 n\ln\lambda \qquad \lambda=\lambda_D/b_0$$

$$\chi=v\Big/\left(\frac{2KT}{m}\right)^{1/2} \qquad B=\frac{4\sqrt{2}\lambda\omega}{\omega_p\ln\lambda}=\frac{\omega}{\omega_c},$$

where ω_c is approximately the 90° deflexion time of a particle with thermal velocity. When $\omega=0$, equation (10)* reduces to equation (8) of SPITZER Jr. which considers d.c. electric conductivity.

In a d.c. electric field, the electrons are not accelerated in a steady state. Hence, the inertia force term is zero and $I_0(\infty)=\frac{3\pi^{\frac{1}{2}}}{8}\alpha$.

3. SOLUTION OF EQUATION

Equation (10) is a linear integral-differential equation whose unknown $D(\chi)$ is a complex function of a real variable. The present section will discuss the method of its solution. As will be evident in what follows, the procedure for numerical integration is far from straightforward.

On the one hand, we encounter the problem of the instability of the solution at small and at large χ. Because of the existence of singularities in equation (10) at $\chi=0$ and at $\chi=\infty$, a slight deviation of $D(\chi)$ at either small or large χ, tends to be built up quite rapidly. In order to obtain a physically acceptable solution, it is required that $D(\chi)$ does not approach infinity too fast, leading to infinite conductivities. The starting value of D at small χ can be obtained by means of a series solution. Because of the instability, we cannot proceed to integrate in a step-wise manner. To overcome this difficulty, we adopted a scheme

*Note that $I_0(\infty)$ is essentially the total change of momentum of electrons arising from electronion interactions. Since the mutual electronic interaction cannot change the total momentum of the electrons, $I_0(\infty)$, by Newton's second law, must equal the total force exerted on the electrons by the applied field minus the inertia force of electrons. This relation gives us:

$$I_0(\infty)=\frac{3\pi^{\frac{1}{2}}}{8}\alpha-\frac{iB}{2}I_3(\infty).$$

used by COHEN *et al.* We shall refer to their paper for full details.

On the other hand, we note that $I_0(\infty)$ is no longer a known quantity as it is in the case of d.c. conductivity; it depends on the solution $D(\chi)$ itself. We proceed as follows: Since we want conductivities at different frequencies, it is necessary to obtain solutions for different values of the parameter B. We begin with a small value $B=0{\cdot}05$. Using an $I_0(\infty)$ taken from the d.c. case, i.e. $I_0(\infty)=0{\cdot}655$, we obtain a solution to equation (10) from which we get a new $I_0(\infty)$. Next, we pass on to $B=0{\cdot}1$ using the $I_0(\infty)$ obtained for the previous B. In this way, we proceed to ever-increasing values of B, until the initial adopted $I_0(\infty)$ and the final calculated $I_0(\infty)$ differ by no more than 2 percent. This occurs at $B=1{\cdot}37$. From this point on, we resort to a method of systematic trials. The initial and final $I_0(\infty)$ for all values of B agree to within 2 percent, which is considered sufficiently accurate for the present purposes.

4. RESULTS AND DISCUSSION

The current is given by

$$\mathbf{J}=-e\int \mathrm{d}\mathbf{v}\mathbf{v}f_1[(\mathbf{v},t)=AE_0e^{i\omega t}I_3(\infty) \qquad (19)$$

where

$$I_3(\infty)=\int_0^\infty \chi^3 D(\chi)\mathrm{e}^{-x^2}\mathrm{d}\chi$$

$$A=\frac{2}{3}\frac{(2KT)^{3/2}}{\pi^{3/2}m^{1/2}e^2\ln\lambda}. \qquad (20)$$

Since $J=\sigma E$, we have the complex conductivity

$$\sigma=AI_3(\infty), \qquad (21)$$

the impedance

$$Z=\frac{1}{AI_3(\infty)}=\frac{c}{I_3(\infty)}, \qquad (22)$$

the resistance

$$R=\frac{1}{A}\frac{ReI_3(\infty)}{|I_3(\infty)|^2}=\frac{cReI_3(\infty)}{|I_3(\infty)|^2}, \qquad (23)$$

and the reactance

$$X=\frac{1}{A}\frac{ImI_3(\infty)}{|I_3(\infty)|^2}=\frac{cImI_3(\infty)}{|I_3(\infty)|^2}, \qquad (24)$$

with

$$c=\frac{1}{A}=\frac{3}{2}\frac{1}{(2KT)^{3/2}}\pi^{3/2}m^{1/2}e^2\ln\lambda.$$

It may be remarked here that the a.c. conductivity depends on three factors:

(i) The inertia of the conducting electrons.
(ii) The mutual interaction among electrons and ions.
(iii) The mutual interaction among electrons themselves.

The mutual electronic interactions have no direct effect on conductivity since the total change of momentum due to such interactions is zero. Nevertheless, they alter the distribution of electrons and thereby modify the effect which electron-ion collisions and electron inertia have in impeding the current. When ω is small, the conductivity is primarily determined by collisions. The inclusion of electron-electron interactions reduces the conductivity by a factor of approximately two. As ω increases and becomes of order ω_c, this effect becomes less and less important because there is then insufficient time in each a.c. cycle to allow an effective modification of the distribution by electron-electron interactions. When ω well exceeds ω_c, we may neglect this effect and $D(\chi)$ reduces to

$$D_i(\chi)=\frac{\chi^4}{1+\frac{iB}{2}\chi^3} \qquad (25)$$

and the corresponding conductivity becomes

$$\sigma_i(\chi)=A\int_0^\infty \frac{\chi^7\mathrm{e}^{-x^2}}{1+\frac{iB}{2}\chi^3}\mathrm{d}\chi \qquad (26)$$

which is just the a.c. conductivity of a Lorentz gas (BERNSTEIN and TREHAN).

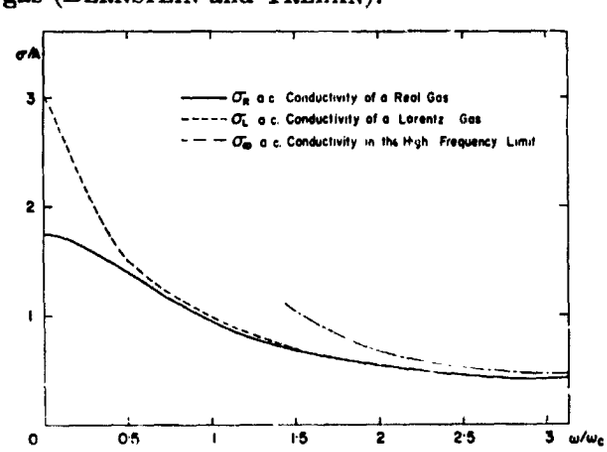


FIGURE 2.—A. C. conductivities.

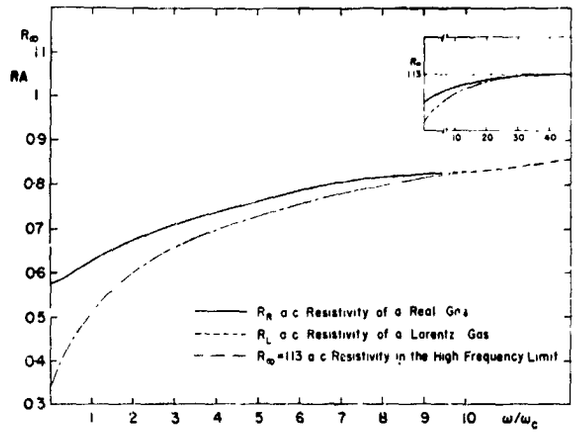


FIGURE 3.—A. C. resistivities

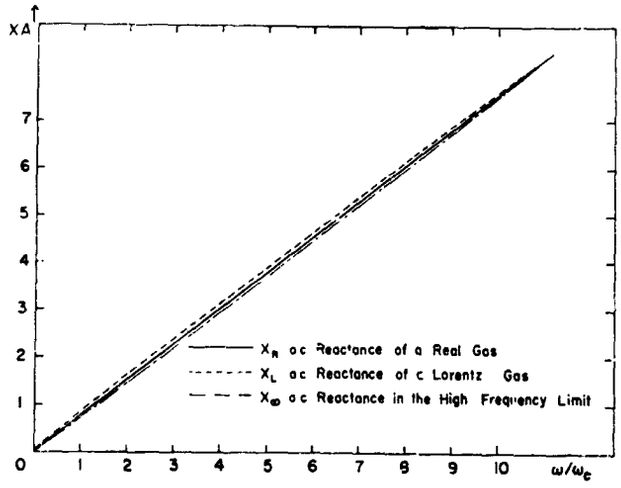


FIGURE 4.—A. C. reactances

If we further increase ω, the inertia of electrons become dominant. Then we may treat collision effect as a perturbation and obtain

$$\sigma_\infty = A\left(-\frac{3\pi^{1/2}}{4}i+\frac{2}{B^2}\right)\frac{1}{B} \tag{27}$$

$$Z_\infty = \frac{1}{A}\,\frac{4B^2}{64+9\pi B^2}(8+3\pi^{1/2}iB) \tag{28}$$

$$R_\infty = \frac{1}{A}\,\frac{32}{9\pi}\left(1-\frac{64}{9\pi B^2}\right) \tag{29}$$

$$X_\infty = \frac{1}{A}12\pi^{1/2}B\left(1-\frac{64}{9\pi B^2}\right). \tag{30}$$

In Table 1 the values of $D(\chi)$ for $B=1$ are given and compared with the corresponding values for the d.c. case obtained by SPITZER Jr. and HÄRM.

In Table 2 the resistance, the reactance and the absolute value of conductivity are given for various B from 0 to 10. For $B>10$, one may use

TABLE 1.—*Values of* $D(\chi)$ *for* $\omega=\omega_c$ *and* $\omega=0$ $\omega_c=\dfrac{\omega_p \ln\lambda}{4\sqrt{2}\lambda}$

x	$\omega=\omega_c$		$\omega=0$
	$Re[D(x)]$	$Im[D(x)]$	$D(x)$
0.10	0.0005887	−0.0002029	0.0008093
0.11	0.0009252	−0.0003438	0.001300
0.12	0.001376	−0.000542	0.001970
0.13	0.001956	−0.0008068	0.002847
0.14	0.00268	−0.001149	0.003955
0.15	0.00356	−0.00158	0.005317
0.16	0.00461	−0.00210	0.006955
0.17	0.00583	−0.00272	0.008886
0.18	0.00724	−0.00346	0.01113
0.19	0.00884	−0.00431	0.01370
0.20	0.01063	−0.00528	0.01660
0.22	0.01483	−0.00761	0.02347
0.24	0.01985	−0.01048	0.03180
0.26	0.0257	−0.0139	0.04165
0.28	0.0324	−0.0180	0.05304
0.30	0.0400	−0.0226	0.06601
0.32	0.0483	−0.0279	0.08057
0.34	0.0575	−0.0339	0.09672
0.36	0.0675	−0.0405	0.1145
0.38	0.0783	−0.0478	0.1338
0.40	0.0899	−0.0557	0.1548
0.44	0.1153	−0.07366	0.2015
0.48	0.1435	−0.09436	0.2545
0.52	0.1744	−0.1179	0.3137
0.56	0.2080	−0.1442	0.3792
0.60	0.2439	−0.1734	0.4508
0.64	0.2882	−0.2055	0.5285
0.68	0.3227	−0.2405	0.6123
0.72	0.3652	−0.2785	0.7023
0.76	0.4096	−0.3196	0.7983
0.80	0.4559	−0.3637	0.9005
0.88	0.5535	−0.4813	1.123
0.96	0.6570	−0.5718	1.371
1.04	0.7656	−0.6957	1.645
1.12	0.8782	−0.8335	1.945
1.20	0.9937	−0.9858	2.273
1.28	1.111	−1.1531	2.630
1.36	1.2290	−1.3359	3.017
1.44	1.3457	−1.5347	3.435
1.52	1.4598	−1.7500	3.887
1.60	1.5693	−1.9820	4.375
1.76	1.7657	−2.4962	5.465
1.92	1.1915	−3.0739	6.728
2.08	1.9973	−3.7049	8.190
2.24	1.9913	−4.3694	9.880
2.40	1.8852	−5.0382	11.83
2.72	1.4266	−6.2497	16.62
2.88	1.2198	−6.6809	19.53
3.04	1.3105	−6.8066	22.74
3.20	2.1113	−5.6758	26.00

TABLE 2.—*The conductivity, the resistance and the reactance of a.c. current*

ω/ω_c	σ/A		R/c		X/c	
	Real gas	Lorentz gas	Real gas	Lornetz gas	Real gas	Lorentz gas
0.0	1.734	3.0	0.577	0.333	0.0	0.0
0.05	1.729	2.880	0.577	0.340	0.045	0.70
0.1	1.713	2.653	0.577	0.354	0.089	0.130
0.15	1.687	2.430	0.578	0.367	0.134	0.185
0.2	1.651	2.233	0.579	0.380	0.178	0.236
0.25	1.608	2.061	0.58	0.393	0.223	0.285
0.3	1.561	1.913	0.582	0.404	0.267	0.332
0.35	1.510	1.784	0.584	0.415	0.311	0.377
0.4	1.458	1.671	0.587	0.425	0.354	0.421
0.45	1.406	1.572	0.590	0.434	0.397	0.465
0.5	1.354	1.483	0.593	0.443	0.440	0.508
0.55	1.303	1.405	0.597	0.452	0.482	0.550
0.6	1.255	1.334	0.600	0.460	0.525	0.592
0.65	1.208	1.270	0.603	0.467	0.567	0.634
0.7	1.164	1.212	0.607	0.475	0.608	0.675
0.75	1.122	1.159	0.610	0.482	0.650	0.716
0.8	1.082	1.111	0.614	0.488	0.691	0.756
0.85	1.045	1.066	0.617	0.495	0.732	0.797
0.9	1.009	1.025	0.620	0.501	0.773	0.807
0.95	0.976	0.987	0.624	0.507	0.813	0.877
1.0	0.944	0.952	0.627	0.513	0.853	0.916
1.1	0.886	0.889	0.634	0.524	0.934	0.996
1.2	0.834	0.835	0.640	0.534	1.013	1.074
1.3	0.786	0.785	0.645	0.544	1.094	1.159
1.4	0.743	0.741	0.750	0.553	1.172	1.230
1.5	0.721	0.702	0.657	0.561	1.222	1.308
2.0	0.560	0.557	0.672	0.599	1.588	1.693
3.0	0.408	0.398	0.711	0.655	2.335	2.453
4.0	0.313	0.305	0.734	0.696	3.086	3.207
5.0	0.255	0.248	0.766	0.728	3.842	3.960
6.0	0.214	0.211	0.784	0.754	4.601	4.710
7.0	0.183	0.182	0.790	0.776	5.380	5.460
8.0	0.162	0.161	0.796	0.794	6.08	6.21
9.0	0.144	0.143	0.814	0.810	6.69	6.81
10.0	0.130	0.129	0.830	0.824	7.51	7.6

equation (26) to compute them. The error will be within 2 percent. For $B>50$ the collisions become unimportant and equations (27)–(30) will give the correct values to within 2 percent. However, there the validity of the Fokker-Planck equation already becomes questionable and one should use DAWSON-OBERMAN'S values instead of ours.

In Table 3, the complex conductivity calculated in this paper are compared with those obtained by ROBINSON and BERNSTEIN. They showed that transport coefficients obtained from the Fokker-Planck equation should possess an extremal nature, and proceed to calculate conductivities using the variational technique. Their Table 6 gives conductivities for various values of the logarithm of ω/ω_c including very large values of ω_c. In our Table 3 only those values are included for comparison which fall within the range of validity of the Fokker-Planck equation. The discrepancy between the two results are generally within 5 percent.

TABLE 3.—*Comparison of the a.c. conductivities obtained in this paper (direct integration) with those obtained by Berstein and Robinson (variational calculation)*

ω/ω_c	*Log X* $(X=4\sqrt{2\omega/\omega_c})$	$Re\sigma/A$		$-Im\sigma/A$	
		BERNSTEIN ROBINSON	SHEN–CHEN	BERNSTEIN–ROBINSON	SHEN–CHEN
0.0	$-\infty$	1.734	1.734	0.000	0.000
0.0057	−3.0	1.734	1.734	0.004	0.002
0.0179	−2.5	1.730	1.732	0.649	0.408
0.0565	−2.0	1.705	1.722	0.154	0.151
0.179	−1.5	1.605	1.608	0.447	0.443
0.565	−1.0	0.992	0.987	0.816	0.813
0.901	−0.8	0.637	0.632	0.785	0.787
1.425	−0.6	0.357	0.353	0.645	0.649
2.261	−0.4	0.183	0.179	0.477	0.482
3.559	−0.2	0.087	0.084	0.329	0.336
5.65	0.0	0.040	0.038	0.201	0.204

We should add that our results can be readily applied to the case of conductivity in the presence of a uniform magnetic field. The addition of the magnetic field leads to equations which are entirely similar to (10). If the electric field is parallel to the magnetic field, the conductivity is not affected. If it is perpendicular to the magnetic field, the conductivity becomes $\sigma_H(\omega)\equiv\sigma(\omega+\omega_H)$, where σ is the function obtained in this paper and $\omega_H=eH/mc$.

REFERENCES

BERNSTEIN, I. B. and TREHAN, S. K. (1960) *Nucl. Fusion* **1**, 3.

CHANDRASEKHAR, S. (1943) *Rev. Mod. Phys.* **15**, 1.

CHAPMAN, S. and COWLING, T. G. (1939) *The Mathematical Theory of Non-Uniform Gases*, Cambridge University Press, London.

COHEN, R., SPITZER, L. JR., and ROUTLY, P. (1950) *Phys. Rev.* **80**, 230.

COWLING, T. G. *Proc. Roy. Soc.* (1945) **A183**, 453.

DAWSON, J. and OBERMAN, C. *Phys. Fluids* **5**, 517.

LIN, S., RESLER, E. L., and KANTROWITZ, A. (1955) *J. Appl. Phys.* **26**, 95.

OBERMAN, C., RON, A., and Dawson, J. (1962) *Phys. Fluids* **5**, 1514.

ROBINSON, B. and BERNSTEIN, I. B. (1962) *Ann. Phys.* **18**, 110.

ROSENBLUTH, M., MACDONALD, W., and JUDD, D. (1957) *Phys. Rev.* **107**, 1.

SPITZER, L. JR. and HÄRM, R. (1953) *Phys. Rev.* **89**, 977.

PARTIAL WAVE THEORY OF DIATOMIC MOLECULES, II*

N64-20589

A. TEMKIN AND A. K. BHATIA†

The method of expanding the wave function of diatomic molecules in a series of orbital angular momentum eigenfunctions (partial waves) is extended to two-electron homonuclear molecules. The angular momentum is a function of the Euler angles only; a symmetric choice of these angles is used which greatly facilitates the description of the exchange character of the wave function. As a result, explicit equations for the 3 dimensional "radial" functions can be derived for all the different magnetic parity, and exchange states. For Σ states the partial wave sums only go over alternate values of the angular momentum giving rise to the expectation of an even more rapidly convergent series in these cases.

I. INTRODUCTION

In a previous paper[1] as well as others,[2] the idea of expanding the wave function of a (homonuclear) diatomic molecule in a series of angular momentum eigenfunctions was introduced for the purpose of affording a very natural method of successive approximations in these problems.

Previous calculations[1,2] using this idea have almost exclusively been confined to the one-electron molecule H_2^+. In that case, the formalism is rather easily constructed, and the results are in the nature of a check as the problem is amenable to exact calculation.[3] The comparison with the exact results is nevertheless most encouraging.[1]

The main immediate goal of this approach is two-electron molecules for which there is no analytically exact solution. As a prerequisite for effecting this approach, however, it is necessary that two electron atoms be completely understood. Although the general approach to the two-electron atomic problem is quite old[4], only results for specific angular momentum states have until recently been derived. The difficulty lay in the description of the character of the wave function under the exchange of electron coordinates which becomes very complicated[4] with the original choice of Euler angles[5]. The explicitly complete solution has required a symmetrical choice of Euler angles, which Holmberg[6] has provided, and has therefore only just been carried out[7].

In section II, we very briefly review the main results of that investigation and present some additional transformation properties of the Euler angles and vector spherical harmonics which bear on the present application. In section III we describe the construction of the molecular wave functions, and in section IV we derive the radial equations. These equations couple states of different angular momentum through the potential energy, but the kinetic energy acts in the same way as in the atomic case. The new feature is, therefore, only the coupling terms of the molecular potential and this in turn depends on integrals over three vector spherical harmonics. The heart of section III is the resulting formulae for these integrals and the selection rules for its non-vanishing. This latter consideration leads to the decoupling of the Σ states according to even $(+)$ or odd $(-)l$. This decoupling is not present for other m states corresponding to the absence of the $\pm$ quantum numbers in those cases; thus the Σ states provide particularly favorable cases for the rapid convergence of the partial wave expansion.

Section V presents the zeroth-order equation for the Σ_g^+ states and discusses a specific calculation which can be related to the accuracy of this approximation and which by inference bears upon the convergence properties of the entire sequence of approximations which this paper projects.

*Published as *Goddard Space Flight Center Document X-640-64-66*, March 1964.

†Goddard Space Flight Center; National Science Foundation, National Research Council resident research associate.

II. THE EULER ANGLES AND TRANSFORMATION PROPERTIES

The Schrödinger equation of a two-electron homonuclear molecule whose nuclei are considered fixed (Born-Oppenheimer approximation) will depend on the two vectors of the electrons $\mathbf{r}_1$ and $\mathbf{r}_2$ which can be measured relative to an origin defined as the midpoint between the nuclei and parametrically on the internuclear separation R_{AB}.

In place of the six coordinates $\mathbf{r}_1$, $\mathbf{r}_2$ one can introduce the three Euler angles and three residual variables; the Euler angles describe the orientation of the plane containing the two unit vectors $\hat{\mathbf{r}}_1$ and $\hat{\mathbf{r}}_2$ relative to a space-fixed coordinate system (whose z-axis is by definition the internuclear axis).

The particular Euler angles we use where introduced by Holmberg[6] and have the advantage of being symmetrically disposed with respect to the

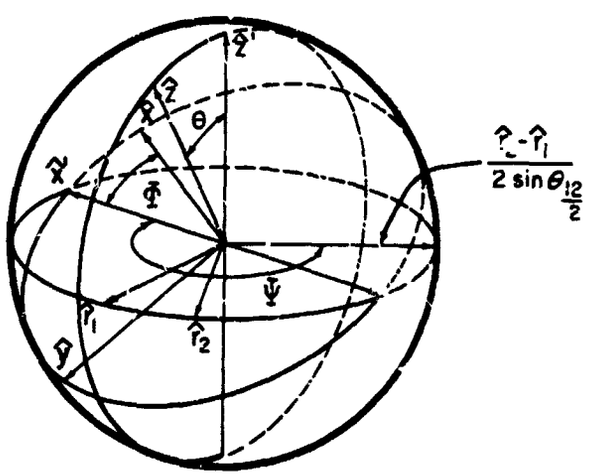


FIGURE 1.—Perspective drawing of (Holmberg's) Euler angles and the unit vectors of the problem.

two-electron. Figure 1 contains a diagrammatic representation of these angles and the unit vectors with which we shall be concerned.

The axes are defined by

$$\hat{\mathbf{z}}' = (\hat{\mathbf{r}}_1 \times \hat{\mathbf{r}}_2)/\sin\theta_{12} \tag{2.1}$$

$$\hat{\mathbf{x}}' = (\hat{\mathbf{z}} \times \hat{\mathbf{z}}')/\sin\theta \tag{2.2}$$

$$\hat{\mathbf{y}}' = \hat{\mathbf{z}}' \times \hat{\mathbf{x}}' \tag{2.3}$$

The Euler angles are obviously: θ = angle between $\hat{\mathbf{z}}$ and $\hat{\mathbf{z}}'$; Φ = angle between $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$; Ψ = angle between $\hat{\mathbf{x}}'$ and $(\hat{\mathbf{r}}_1 - \hat{\mathbf{r}}_2)$. Trigonometric relations between these angles and the particles spherical angles have been worked out in reference 7; the reader is encouraged to familiarize himself with that discussion, although we shall attempt to make this treatment reasonably self-contained. Specifically, we shall repeat here the relation of the x, y and z components of $\hat{\mathbf{r}}_1$ and $\hat{\mathbf{r}}_2$ to the Euler angles in order that the transformation properties given below can readily be checked.

$$\frac{x_1}{r_1} = \sin\vartheta_1\cos\varphi_1 = \cos\Phi\sin(\psi - \tfrac{1}{2}\theta_{12}) + \cos\theta\sin\Phi\cos(\psi - \tfrac{1}{2}\theta_{12}) \tag{2.5}$$

$$\frac{y_1}{r_2} = \sin\vartheta_1\sin\varphi_1 = \sin\Phi\sin(\psi - \tfrac{1}{2}\theta_{12}) - \cos\theta\cos\Phi\cos(\psi - \tfrac{1}{2}\theta_{12}) \tag{2.6}$$

$$\frac{z_1}{r_1} = \cos\vartheta_1 = -\sin\theta\cos(\psi - \tfrac{1}{2}\theta_{12} \tag{2.7}$$

The formulae for analogous components of $\hat{\mathbf{r}}_2$ may be obtained from the above by letting $\theta_{12} \rightarrow -\theta_{12}$. The angle θ_{12} is the angle between $\hat{\mathbf{r}}_1$ and $\hat{\mathbf{r}}_2$. It is not one of the Euler angles, but rather one choice of residual coordinate, the other two usually being chosen as the two distances r_1 and r_2. The residual coordinates are often called generalized radial coordinates; an alternate choice is r_1, r_2, r_{12}, the latter being the interelectron distance.

The distinguishing feature of the Euler angles as opposed to residual coordinates is the fact that the angular momentum depends only on the Euler angles. In this connection, it is well known that square of the total orbital angular momentum, $\mathbf{M}^2$, can be conserved in atoms but not in diatomic molecules, whereas the z-component of the angular momentum, M_z, is a possible constant of the motion in both cases. A central role in the atomic problem is therefore played by the eigenfunctions of the total angular momentum $D_l^{m,k}(\theta,\Phi,\Psi)$:

$$D_l^{m,k}(\theta,\Phi,\psi) = \frac{\sqrt{2(2l+1)}}{4\pi} e^{i(m\Phi+k\psi)} d_l^{m,k}(\theta) \tag{2.8}$$

where

$$d_l^{m,k}(\theta) = (-1)^{\frac{1}{2}(|k-m|+k-m)} \frac{4\pi}{\sqrt{2(2l+1)}} N_{lmk} \sin^{|k-m|}\left(\frac{1}{2}\theta\right)\cos^{|k+m|}\left(\frac{1}{2}\theta\right) F\left(-l+\frac{\beta}{2}-1,\ l+\frac{\beta}{2},1+|k-m|,\sin^2\frac{\theta}{2}\right) \tag{2.9}$$

$$\beta = |k+m| + |k-m| + 2 \tag{2.10}$$

and

$$N_{\ell mk} = \frac{1}{(|k-m|)!}\left[\frac{(2\ell+1)}{8\pi^2}\,\frac{(\ell+\frac{1}{2}|k+m|+\frac{1}{2}|k-m|)!}{(\ell-\frac{1}{2}|k+m|-\frac{1}{2}|k-m|)!}\times\frac{(\ell-\frac{1}{2}|k+m|+\frac{1}{2}|k-m|)!}{(\ell+\frac{1}{2}|k+m|-\frac{1}{2}|k-m|)!}\right]^{\frac{1}{2}} \tag{2.11}$$

The eigenfunction character of these functions is then given by

$$\mathbf{M}^2 D_\ell^{m,k} = \hbar^2 \ell(\ell+1) D_\ell^{m,k} \tag{2.12}$$

$$\mathbf{M}_z D_\ell^{m,k} = \hbar m D_\ell^{m,k} \tag{2.13}$$

where the explicit form of $\mathbf{M}^2$ and $\mathbf{M}_z$ has been derived in reference 7 and in any case is known[8]. In the atomic problem if we are describing a state of definite orbital angular momentum ℓ with z-component m, then the wave function is at most a finite sum over the $(2\ell+1)$ function for the given ℓ and m:

$$\Psi_{\ell m}(\mathbf{r}_1,\mathbf{r}_2) = \sum_{k=-\ell}^{\ell} g_\ell^{m,k}(r_1,r_2,\theta_{12}) D_\ell^{m,k}(\theta,\Phi,\psi) \tag{2.14}$$

Actually the above sum is more general than what is needed. Specifically one has an additional operation which commutes with the Hamiltonian and therefore gives rise to an additional constant of the motion. This is the operation of space inversion (called parity, p, in the atomic application[7]), usually labelled iE in molecular applications[9] corresponding to the operation $\mathbf{r}_1 \rightarrow -\mathbf{r}_1$, $\mathbf{r}_2 \rightarrow -\mathbf{r}_2$ in all cases. It turns out that only the Euler angles are involved in this transformation, and if we denote them by a three component vector, (θ, Φ, Ψ), then[7]

$$iE\begin{pmatrix}\theta\\ \Phi\\ \psi\end{pmatrix} = \begin{pmatrix}\theta\\ \Phi\\ \pi+\psi\end{pmatrix} \tag{2.15}$$

and

$$iE D_\ell^{m,k}(\theta,\Phi,\Psi) = (-1)^k D_\ell^{m,k}(\theta,\Phi,\Psi) \tag{2.16}$$

Therefore, the wave function (2.14), if it is to be an eigenfunction of iE, must be restricted to even or odd k. Since the action of the Schrödinger equation is to couple all possible $g_\ell^{m,k}$ functions for the given ℓ and m, it can be seen that the invocation of parity halves the number of functions, and thus among other things is of considerable practical utility.

In the case of diatomic molecules the invariance against rotations, except those around the z-axis, is lost. Thus according to the above arguments only the rotation quantum number m survives. The idea of the partial wave theory, then, is to expand the molecular wave function in terms of the orbital angular momentum:

$$\Psi_m = \sum_{\ell=|m|}^{\infty} \Psi_{\ell m}$$

The above sum too may be delimited according to the symmetry properties of the states we intend to describe. These symmetry operations are well known for the diatomic molecule,[9], and except for exchange, ϵ_{12}, they all involve only the Euler angles. The transformation involved in the residual variables for ϵ_{12}, is $r_1 \rightleftarrows r_2$ independent of whether the third variable is taken as θ_{12} or r_{12}.

In Table I we have summarized the transformation properties of the Euler angles and the vector spherical harmonics under some of the major symmetry operations of the two-electron homonuclear diatomic molecule.

In the next section we shall utilize these properties to construct wave functions with the desired symmetry properties. We emphasize that *not* all these properties are independent. Thus for example a full space inversion is the product of the inversion about each of the axes separately:

$$iE = \sigma_h \sigma_{vx} \sigma_{vy}$$

where the factors on the right can be written in any order. Therefore, the operation σ_h is redundant.

The effect of additional symmetry operations such as σ_v on the angular momentum eigenfunctions and variables can easily be worked out from Table I by noting, in this case, that

$$i\sigma_v = iE \cdot \sigma_v$$

Not all symmetry operations commute with M_z. Specifically σ_v does not commute with M_z. This may be easily ascertained by envisaging a vector

TABLE 1.—*Symmetry properties of the Euler angles and the angular momentum eigenfunctions.*

Transformation Property	θ	Φ	Ψ	$D_l^{m,k}$
iE space inversion (parity)	θ	Φ	$\Psi+\pi$	$(-1)^k D_l^{m,k}$
σ_h reflection through x=y plane (inversion of z-coordinates	θ	$\Phi+\pi$	$\Psi+\pi$	$(-1)^{k+m} D_l^{m,k}$
σ_{vz} reflection through z−x plane (inversion of y-coordinates)	$\pi-\theta$	$2\pi-\Phi$	Ψ	$(-1)^{l-k} D_l^{-m,k}$
σ_v, reflection through z−y plane (inversion of x-coordinates)	$\pi-\theta$	$\pi-\Phi$	Ψ	$(-1)^{l+m-k} D_l^{-m,k}$
ϵ_{12} exchange of particles*	$\pi-\theta$	$\Phi+\pi$	$2\pi-\Psi$	$(-1)^l D^{m,-k}$

*This operation also affects the residual coordinates. See text.

whose projection in the x-y plane is in the first quadrant. If first you reflect it through the y-z plane (σ_{vy}), thus putting in the second quadrant and then rotate around the z-axis by 90°, this vector will end up with a projection in the third quadrant. If first you rotate by 90°, putting it in the second quadrant, and then reflect through the y-z plane, you put it back into the first quadrant. Thus

$$[\sigma_v, M_z] \neq 0 \tag{2.19a}$$

This means that in general one can't construct simultaneous eigenfunctions of M_z and σ_v.

There is one exception to that situation, and that is Σ states.

For in that special case

$$M_z D_l^{0,k} = 0, \tag{2.17}$$

so that

$$[\sigma_v, M_z] D_l^{0,k} = 0, \tag{2.18}$$

or

$$[\sigma_v, M_z]_{m=0} = 0 \tag{2.19b}$$

We shall see in the next section that because of this fact the structure of the Σ wave function and the resulting Σ radial equations is quite different from all other states.

III. WAVE FUNCTIONS AND THE POTENTIAL ENERGY

In order to facilitate the derivation of the radial equations, it is convenient to construct an alternate set of orthonormal vector spherical harmonics which for $m=0$ are real[7]. We define

$$D_l^{(m,\varkappa)+} \equiv (\sqrt{2}+\delta_{0\varkappa}(2-\sqrt{2}))^{-1} (D_l^{m,\varkappa}+D_l^{m,-\varkappa}) \tag{3.1a}$$

$$D_l^{(m,\varkappa)-} \equiv (\sqrt{2}i)^{-1}(D_l^{m,\varkappa}-D_l^{m,-\varkappa}) \tag{3.1b}$$

The important point about these alternate sets of vector spherical harmonics is that they are eigenfunctions of exchange:

$$\epsilon_{12} D_l^{(m,\varkappa)\pm} = \pm(-1)^l D_l^{(m,\varkappa)\pm} \tag{3.2}$$

while still retaining their eigenfunction character with respect to parity, $\mathbf{M}^2$, and M_z (with eigenvalues $(-1)^\varkappa$, $\hbar^2\ell(\ell+1)$, $\hbar\, m$ respectively).

For $m \neq 0$ we have already indicated in the previous section that the correct expansion of the complete wave function is

$$\Psi_m = \sum_{l=|m|}^{\infty} {\sum_{\varkappa}}'' (f_l^{(m,\varkappa)+}(r) D_l^{(m,\varkappa)+} + f_l^{(m,\varkappa)-}(r) D_l^{(m,\varkappa)-}) \tag{3.3a}$$

The double prime on the summation indicates that every second value of the summation index (in this case $\varkappa$) is to be taken. This insures as stated, that the wave function is an eigenfunction of iE. It is precisely this property which distinguishes the homonuclear from the heteronuclear molecule. Thus there is established in the present homonuclear case two classes of solutions for a given m, one for even and one for odd $\varkappa$. These are of course the gerade and ungerade states respectively.

The radial functions $f_l^{(m,\varkappa)\pm}(r)$ are written as functions of r which is short for the generalized radial coordinates, either r_1, r_2, θ_{12} or r_1, r_2, r_{12}. The superscript indicates that the radial functions depend on the magnetic quantum number m. However the well-known degeneracy of the energy with respect to $\pm m$ indicates that the radial functions for $\pm m$ states must be simply related to each other. The relation is undefined up to a $\pm$ sign, but a consistent choice is (see below)

$$f_l^{(-m,\varkappa)\pm}(r_1,r_2,\theta_{12})=(-1)^{\beta}f_l^{(m,\varkappa)\pm}(r_1,r_2,\theta_{12})$$

The derivation of radial equations divides itself into two parts, the kinetic energy and the potential energy. The kinetic energy is diagonal with respect to ℓ and m but by no means trivial with respect to how mixes the allowable $\varkappa$ components. However, this aspect of its behaviour has been derived in detail in reference 7 and will not change in this application.

The potential energy in the atomic problem depends only on the generalized radial variables r_1, r_2, r_{12} and therefore acts as additional, completely diagonal terms in the atomic radial equations. On the other hand the molecular potential energy is precisely what destroys the spherical symmetry of the Hamiltonian. Specifically for nuclei of charge z (energy in units of rydbergs) to complete molecular potential is

$$v_{m0\ell}+\frac{2}{r_{12}}=-\frac{2z}{R_{A1}}-\frac{2z}{R_{A2}}-\frac{2z}{R_{B1}}-\frac{2z}{R_{B2}}+\frac{2}{r_{12}} \tag{3.4}$$

The last term is the interelectron repulsion and since it commutes with the angular momentum as discussed above need not be considered further. For the rest we can expand

$$V_{m0\ell}=-4z\sum_{\lambda\ \text{even}}{}''\, g_\lambda(\tfrac{1}{2}R_{AB},r_1)P_\lambda(\vartheta_1) -4z\sum_{\lambda\ \text{even}}{}''\, g_\lambda(\tfrac{1}{2}R_{AB},r_2)P_\lambda(\vartheta_2) \tag{3.5}$$

where

$$g_\lambda(x,y)\equiv\begin{cases}\dfrac{x^\lambda}{y^{\lambda+1}} & x<y\\[2ex] \dfrac{y^\lambda}{x^{\lambda+1}} & y<x\end{cases} \tag{3.6}$$

R_{AB} is the internuclear separation. The Legendre polynomials $P_\lambda(\vartheta)$ (note what we mean is eg. $P_1(\vartheta)=\cos\vartheta$ *not* $P_1(\vartheta)=\vartheta$) can further be expanded in terms of the Euler angles and θ_{12}[7]:

$$P_\lambda(\vartheta_1)=\frac{4\pi}{(2\lambda+1)^{\frac{1}{2}}}\sum_{\mu\ \text{even}}{}''\,[\alpha_\lambda^{\mu+}(\theta_{12})D_\ell^{(0,\mu)+} -\alpha_\lambda^{\mu-}(\theta_{12})D_\ell^{(0,\mu)-}] \tag{3.7a}$$

Applying ϵ_{12}, we get from the above

$$P_\lambda(\vartheta_2)=\frac{4\pi}{(2\lambda+1)^{\frac{1}{2}}}\sum_{\mu\ \text{even}}{}''\,[\alpha_\lambda^{\mu+}(\theta_{12})D_\lambda^{(0,\mu)+} -\alpha_\lambda^{\mu-}(\theta_{12})D_\lambda^{(0,\mu)-}] \tag{3.7b}$$

where

$$\alpha_\lambda^{\mu+}(\theta_{12})=(-1)^\mu P_\lambda^\mu(\pi/2)\left[\frac{(\lambda-\mu)!}{(\lambda+\mu)!}\right]^{\frac{1}{2}}\left\{\frac{\delta_{0\mu}}{\sqrt{2}} +(1-\delta_{0\mu})\cos\left(\frac{\mu}{2}\theta_{12}\right)\right\} \tag{3.8a}$$

$$\alpha_\lambda^{\mu-}(\theta_{12})=(-1)^\mu P_\lambda^\mu(\pi/2)\left[\frac{(\lambda-\mu)!}{(\lambda+\mu)!}\right]^{\frac{1}{2}}\sin\left(\frac{\mu}{2}\theta_{12}\right) \tag{3.8b}$$

The associated Legendre polynomials are understood to have the phase of Magnus and Oberhettinger[10], and their argument is understood to be inserted in the transcendental form of these functions as parenthesized above.

Since $P_\lambda(\vartheta)$ has parity $(-1)^\lambda$ and from (3.5) λ is even, it follows the sums in (3.7) go over only even μ as indicated.

The point of these expansions is that when the wave function (3.3) is substituted into the Schrödinger equation

$$H\Psi_m=E\Psi_m \tag{3.9}$$

where

$$H=T+V_{m0\ell}+2/r_{12} \tag{3.10}$$

and T is the kinetic energy, only $V_{m0\ell}$ will couple terms of different ℓ. The decompositions, Eqs. (3.7), then tell us that the coupling from a state (radial function) with quantum members ℓ, $\varkappa$, m to one with quantum numbers ℓ_f, $\varkappa_f$, m (the molecular interaction is still diagonal in m) will be governed by integrals over *three* vector spherical harmonics. Two of these angular functions come from the coupled states and one from the interaction; the integral in question is therefore a linear

combination of integrals of the form

$$\int_0^{\pi}\int_0^{2\pi}\int_0^{2\pi}(D_{\ell_i}^{m,\varkappa_i})^* D_{\lambda}^{0,\mu} D_{\ell_f}^{m,\varkappa_f}\sin\theta d\theta d\Phi d\psi = \sqrt{\frac{(2\ell_i+1)(2\lambda+1)}{8\pi^2(2\ell_f+1)}}(\ell_i\lambda-m0\,|\,\ell_f-m)(\ell_i\lambda-\varkappa_i\mu\,|\,\ell_f-\varkappa_f) \quad (3.11)$$

The formula for this integral given on the *rhs* of the above equation is well known[11] in terms of the Clebsch-Gordan (C-G) coefficients. The selections for the vanishing of this integral can therefore be deduced from those of the C-G coefficients.

Consider first Σ states ($m=0$). In this case the first C-G coefficient becomes $(\ell\ \lambda\ 00/\ell'0)$. This coefficient vanishes unless $\ell+\lambda+\ell'=$even (integer). But λ is an even integer, Eq. (3.5, therefore:

$$\ell+\ell'=\text{even}$$

This says that for Σ states only radial functions of even ℓ will be coupled together and only odd ℓ states will be coupled but there will be no intermixing. Thus for Σ states the expansion of the wave function, Eq. (3.3a), separates into two classes defined by the evenness or oddness of ℓ:

$$\Psi_0=\sum_{\ell}^{''}\sum_{\varkappa}^{''}(f_\ell^{(0,\varkappa)+}(r)D_\ell^{(0,\varkappa)+} + f_\ell^{(0,\varkappa)-}(r)D_\ell^{(0,\varkappa)-}) \quad (3.12)$$

The molecular quantum numbers $\pm$ are determined by the parity of ℓ which is the eigenvalue of $i\,\sigma_{xy}$. This then constitutes the exception to the noncommutivity of σ_{xy} with M_z which was noted in Eq. (2.19).

The quantum lables of the Σ states are compiled in the entries of the following simple array:

$\varkappa$ \ ℓ	even	odd
even	gerade, +	gerade, −
odd	ungerade, +	ungerade, −

For $m\neq 0$, the relevant C-G coefficient is not necessarily zero when $\ell+\lambda+\ell'=$odd. Thus the decoupling according to even or odd ℓ ($\pm$symmetry) does not occur, corresponding to the noncommutivity of σ_y with M_z, Eq. (2.19b). Here, however, there exists a well known degeneracy in energy with respect to $\pm m$ eigenfunctions. If the eigenfunction for $+m$ is given by (3.3a) and that for $-m$ is given by

$$\Psi_{-m}=\sum_{\ell=|m|}\sum_{\varkappa}^{''}(f_\ell^{(-m)\varkappa)+}D_\ell^{(-m,\varkappa)+} + f_\ell^{(-m,\varkappa)-}D_\ell^{(-m,\varkappa)-}), \quad (3.3b)$$

the obvious question is what is the relation of the $f_\ell^{(m,\varkappa)\pm}$ to $f_\ell^{(-m,\varkappa)\pm}$. The answer to this question still resides in the same C-G coefficient, in this case in the symmetry property[11]:

$$(\ell\lambda m0\,|\,\ell'm)=(-1)^{\ell+\lambda+\ell'}(\ell\lambda\text{--}m0\,|\,\ell'-m) \quad (3.13)$$

Since $\lambda=$ even, we see the sign of the coupling of an ℓ to an ℓ' state is reversed for Ψ_{-m} relative to what it is for Ψ_m if $\ell+\ell'=$odd (i.e., ℓ and ℓ' have opposite parity). If the radial functions change sign in the same way, then the equations (and hence the eigenvalues) will in fact not change. i.e., we must have

$$f_\ell^{(-m,\varkappa)\pm}=(-1)^l f_\ell^{(m,\varkappa)\pm} \quad (3.14a)$$

This relation could be altered by a minus sign; the relation

$$f_\ell^{(-m,\varkappa)\pm}=(-1)^{\ell+m} f_\ell^{(m,\varkappa)\pm} \quad (3.14b)$$

is equally acceptable and has the advantage that the first radial function that appears in the expansion of Ψ_m will be the one which doesn't change sign when used in Ψ_{-m}.

In all cases we have by virtue of the very simple exchange property of $D_\ell^{(m,\varkappa)\pm}$, Eq. (3.2), that the complete spatial wave function Ψ_m will be symmetric (upper sign) or antisymmetric (lower sign) if the radial functions have the exchange property

$$f_\ell^{(m,\varkappa)+}\left(r_2,\ r_1,\begin{Bmatrix}\theta_{12}\\ r_{12}\end{Bmatrix}\right) = \pm(-1)^\ell f_\ell^{(m,\varkappa)+}\left(r_1,\ r_2,\begin{Bmatrix}\theta_{12}\\ r_{12}\end{Bmatrix}\right) \quad (3.15a)$$

and

$$f_\ell^{(m,\varkappa)-}\left(r_2,\ r_1,\begin{Bmatrix}\theta_{12}\\ r_{12}\end{Bmatrix}\right) = \pm(-1)^{\ell+1} f_\ell^{(m,\varkappa)-}\left(r_1,\ r_2,\begin{Bmatrix}\theta_{12}\\ r_{12}\end{Bmatrix}\right) \tag{3.15b}$$

These properties essentially halve the independent variable space over which the equations need to be integrated[7].

IV. RADIAL EQUATIONS

The foregoing discussion together with the knowledge of the kinetic energy operator[7] essentially allows us to write down the "radial" equations for arbitrary m. Specifically substituting (3.12) or (3.3a) into the Schrödinger equation, multiplying on the left by $D_\ell^{(m,\varkappa)*}$ and integrating over the Euler angles, one obtains

$$\left[L_{\theta 12}+\frac{2m}{\hbar^2}\left(E-\frac{2}{r_{12}}\right)\right]f_\ell^{(m,\varkappa)+}(r)-\left(\frac{1}{r_1^2}+\frac{1}{r_2^2}\right)\left[\left\{\frac{\ell(\ell+1)-\varkappa^2}{2\sin^2\theta_{12}}+\frac{\varkappa^2}{4}-\frac{\cot\theta_{12}}{2\sin\theta_{12}}\delta_{1\varkappa}(1-2A_\varkappa^\ell)\right\}f_\ell^{(m,\varkappa)+}(r)\right.$$
$$+(\varkappa+1)(\varkappa+2)\frac{\cot\theta_{12}}{\sin\theta_{12}}B_{\ell,\varkappa+2}f_\ell^{(m,\varkappa+2)+}(r)+\frac{\cot\theta_{12}}{4\sin\theta_{12}}\frac{(1+3\delta_{0,\varkappa-2})}{(1+\delta_{0,\varkappa-2}(\sqrt{2}-1))}\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell}{\beta_{\ell\varkappa}}f_\ell^{(m,\varkappa-2)+}(r)\Big]$$
$$+\left(\frac{1}{r_1^2}-\frac{1}{r_2^2}\right)\left[\varkappa\left(\frac{\cos\theta_{12}}{2\sin(\frac{1}{2}\theta_{12})}-\cot\theta_{12}-\frac{1}{2\sin\theta_{12}}\right)f_\ell^{(m,\varkappa)-}(r)-\varkappa\frac{\partial f_\ell^{(m,\varkappa)-}(r)}{\partial\theta_{12}}\right.$$
$$-\frac{\delta_{1\varkappa}(1-2A_\varkappa^\ell)}{2\sin\theta_{12}}f_\ell^{(m,\varkappa)-}+\frac{(\varkappa+1)(\varkappa+2)}{\sin\theta_{12}}\beta_{\ell,\varkappa+2}f_\ell^{(m,\varkappa+2)-}(r)$$
$$\left.-\frac{1}{4\sin\theta_{12}}\frac{(1-\delta_{0,\varkappa-2})}{(1+\delta_{0,\varkappa-2}(\sqrt{2}-1))}\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell}{\beta_{\ell\varkappa}}f_\ell^{(m,\varkappa-2)-}(r)\right]$$
$$+\frac{2m}{\hbar^2}4\pi\sum_{\gamma_f}\sum_{j=1}^{2}\sum_{\gamma_\lambda(\text{even})}^{\prime\prime}g_\lambda\left(\frac{1}{2}R_{AB},\ r_j\right)(1-2\delta_{2j}\delta_{\epsilon_\lambda -})\alpha_\lambda^{\mu\epsilon_\lambda}(\theta_{12})I_m(\ell\varkappa+,\ \gamma_\lambda,\ \gamma_f)f_\ell^{(m,\varkappa_f)\epsilon_f}(r)=0 \tag{4.1a}$$

$$\left[L_{\theta 12}+\frac{2m}{\hbar^2}\left(E-\frac{2}{r_{12}}\right)\right]f_\ell^{(m,\varkappa)-}(r)-\left(\frac{1}{r_1^2}+\frac{1}{r_2^2}\right)\left[\left\{\frac{\ell(\ell+1)-\varkappa^2}{2\sin^2\theta_{12}}+\frac{\varkappa^2}{4}+\frac{\cot\theta_{12}}{2\sin\theta_{12}}\delta_{1\varkappa}(1-2A_\varkappa^\ell)\right\}f_\ell^{(m,\varkappa)-}(r)\right.$$
$$+\frac{\cot\theta_{12}}{\sin\theta_{12}}(1-\delta_{2,\varkappa+2})(\varkappa+1)(\varkappa+2)B_{\ell,\varkappa+2}f_\ell^{(m,\varkappa+2)-}(r)+\frac{\cot\theta_{12}}{4\sin\theta_{12}}(1-\delta_{0,\varkappa-2})\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell}{\beta_{\ell\varkappa}}f_\ell^{(m,\varkappa-2)-}(r)\Big]$$
$$+\left(\frac{1}{r_1^2}-\frac{1}{r_2^2}\right)\left[-\varkappa\left(\frac{\cos\theta_{12}}{2\sin\left(\frac{\theta_{12}}{2}\right)}-\cot\theta_{12}-\frac{1}{2\sin\theta_{12}}\right)f_\ell^{(m,\varkappa)+}(r)+\varkappa\frac{\partial f_\ell^{(m,\varkappa)+}(r)}{\partial\theta_{12}}-\frac{\delta_{1\varkappa}(1-2A_\varkappa^\ell)}{2\sin\theta_{12}}f_\ell^{(m,\varkappa)+}(r)\right.$$
$$\left.-\frac{(1-\delta_{2,\varkappa+2})(\varkappa+1)(\varkappa+2)}{\sin\theta_{12}}B_{\ell,\varkappa+2}f_\ell^{(m,\varkappa+2)+}(r)+\frac{1}{4\sin\theta_{12}}\frac{(1+3\delta_{0,\varkappa-2})}{(1+\delta_{0,\varkappa-2}(\sqrt{2}-1))}\frac{A_{\varkappa-2}^\ell A_{\varkappa-1}^\ell}{\beta_{\ell\varkappa}}f_\ell^{(m,\varkappa-2)+}(r)\right]$$
$$+\frac{2m}{\hbar^2}4\pi\sum_{\gamma_f}\sum_{j=1}^{2}\sum_{\lambda_\gamma(\text{even})}^{\prime\prime}g_\lambda\left(\frac{1}{2}R_{AB},\ r_j\right)(1-2\delta_{2j}\delta_{\epsilon_\lambda -})\alpha_\lambda^{\mu\epsilon_\lambda}(\theta_{12})I_m(\ell\varkappa-,\ \gamma_\lambda,\ \gamma_f)f_\ell^{(m,\varkappa_f)\epsilon_f}(r)=0 \tag{4.1b}$$

The index γ in the above formulae represents the triple indices ℓ, $\varkappa$, ϵ. For example

$$\gamma_f=\ell_f,\ \varkappa_f,\ \epsilon_f \tag{4.2}$$

A summation with respect to this index means:

$$\sum_{\gamma_f}\equiv\sum_{\ell_f}\sum_{\varkappa_f}\sum_{\epsilon_f} \tag{4.3a}$$

and in particular

$$\sum_{\gamma_\lambda(\text{even})}^{\prime\prime}=\sum_{\lambda\,(\text{even})}^{\prime\prime}\sum_{\mu\,(\text{even})}^{\prime\prime}\sum_{\epsilon_\lambda}^{\prime\prime} \tag{4.3b}$$

ϵ is a symbol which can be + or −, so that a sum over ϵ consists of only these two terms.

Aside from these abbreviations, the radial equations have been written above in a fairly explicit

form (with a corresponding sacrifice of elegance). For an arbitrary ℓ the coupling is reasonably complicated although as we shall see in the next section for ℓ small the equations simplify greatly. The nature of the coupling is nevertheless contained in the equations themselves via the selection rules for the various coefficients. Regarding the effect of the kinetic energy coupling, a verbal description of the nature of that coupling is given in reference 7.

The equations in the form given here are appropriate to the radial coordinates r_1, r_2, θ_{12} with

$$L_{\theta_{12}}=\frac{1}{r_1}\frac{\partial^2}{\partial r_1{}^2}r_1+\frac{1}{r_2}\frac{\partial^2}{\partial r_2{}^2}r_2 +\left(\frac{1}{r_1{}^2}+\frac{1}{r_2{}^2}\right)\frac{1}{\sin\theta_{12}}\frac{\partial}{\partial\theta_{12}}\left(\sin\theta_{12}\frac{\partial}{\partial\theta_{12}}\right) \tag{4.4}$$

Also the fact that one has two functions (and hence two equations) for a given ℓ and $\varkappa$ indicates that the functions are in symmetric form. Hence the solutions may be confined to the region, say, $r_1>r_2$ with the boundary condition

$$\left[\frac{\partial f_\ell^{(m,\varkappa)\epsilon_1}}{\partial\eta}\right]_{r_1=r_2}=0 \tag{4.5a}$$

and

$$[f_\ell^{(m,\varkappa)\epsilon_2}]_{r_1=r_2}=0 \tag{4.5b}$$

where $\partial/\partial n$ is the normal derivative and ϵ_1 and ϵ_2 are the two choices of $+$ and $-$, being so selected according to (3.15) to give either the space symmetric (singlet) or the space antisymmetric (triplet) solution of the Schrödinger equation.

The coupling due to the potential energy is given in the last term of each of the radial equations. It can be seen that it alone is m dependent, and coupling coefficients I_m are:

$$I_m(\gamma,\gamma_\lambda,\gamma_{\ell'})=\frac{4\pi}{(2\lambda+1)^{\frac{1}{2}}}\int_0^{\pi}\int_0^{2\pi}\int_0^{2\pi}(D_\ell^{(m,\varkappa)\epsilon})^*D_\lambda^{(0,\mu)\epsilon_\lambda}D_{\ell'}^{(m,\varkappa')\epsilon'}\sin\theta d\theta d\Phi d\psi \tag{4.6a}$$

This quadrature, which is a linear combination of integrals of the form (3.11), yields:

$$I_m(\gamma,\gamma_\lambda,\gamma_{\ell'})=\frac{[(1+\delta_{0m}\delta_{0\varkappa}(\sqrt{2}-1))(1+\delta_{0\mu}(\sqrt{2}-1))(1+\delta_{0m}\delta_{0\varkappa'}(\sqrt{2}-1))]^{-1}}{\{\epsilon\}^*\{\epsilon_\lambda\}\{\epsilon_{\ell'}\}}$$
$$\sqrt{\frac{2(2\ell+1)}{(2\ell'+1)}}(1+\epsilon\epsilon_\lambda\epsilon_{\ell'}(-1)^{\ell+\ell'})(\ell\lambda-m0|\ell'\varkappa')[(\ell\lambda-\varkappa\mu|\ell'-\varkappa')+\epsilon_{\ell'}(\ell\lambda-\varkappa\mu|\ell'\varkappa')$$
$$+\epsilon_\lambda(\ell\lambda-\varkappa-\mu|\ell'-\varkappa')+\epsilon(\ell\lambda\varkappa\mu|\ell'-\varkappa')] \tag{4.6b}$$

Here the ϵ's are to be thought as the numbers ±1, and

$$\{\epsilon_\sigma\}=\{(1+\epsilon_\sigma)^{\frac{1}{2}}+i(1-\epsilon_\sigma)^{\frac{1}{2}}\} \tag{4.7}$$

These coefficients are subject to many restrictions if they are not to vanish. Also only one of the C-G coefficients in the final square bracket will not vanish when any of these indices are non-zero. (When $\varkappa=\mu=0$ all four terms contribute.)

V. TRUNCATION AND THE Σ_g^+ STATE OF H_2

The coupling in ℓ' and λ contained in the potential terms in principle includes a doubly infinite sum. If one truncates the sum in ℓ' at $\ell'=L$, then only a finite number of terms in the λ sum survive. The inclusion of all λ terms for a given L, then constitutes a given order of approximation in the context of the partial wave theory.

It is now our assertion that the solutions of successive approximation converges usefully. By this we mean two things; first that for a given L the $f_\ell'r)^{(m,\varkappa)\epsilon}$ from $\ell=\ell_i$ to $\ell=L$ get smaller in some sense. Secondly in going from one L to the next the $f_\ell(r)^{(m,\varkappa)\epsilon}$ for a given set of indices $(\ell,\varkappa,\epsilon)$ does not change much.

In view of the complicated nature of these equations this would be a very difficult thing to prove rigorously. The heuristic argument, however, follows that for the one-electron diatomic molecule[1]. Namely, in the centrifugal terms, which are those multiplied by $(r_1^{-2}+r_2^{-2})$ in the above

equations, there is a repulsive term proportional to $\ell(\ell+1)$ which tends to diminish the amplitude of each succeeding $f_\ell(r)$ component in the region of interaction.

Since, however, there are considerably more terms than this $\ell(\ell+1)$ term, the cogency of this argument is considerably more obscure than its one-electron counterpart[1] and in any event it is known from at least one atomic example that there can be resonance regions in which an additional ℓ component can in fact dominate its predecessors.[12]

We shall therefore examine one case in which there exist numerical results. Consider the Σ_g^+ states of H_2. In zeroth order ($L=0$) these states which include the ground state, are governed by the equation ($z=1$):

$$\left[\frac{1}{r_1}\frac{\partial^2}{\partial r_1^2}r_1+\frac{1}{r_2}\frac{\partial^2}{\partial r_2^2}r_2+\frac{1}{\sin\theta_{12}}\frac{\partial}{\partial\theta_{12}}\left(\sin\theta_{12}\frac{\partial}{\partial\theta_{12}}\right)+\frac{2m}{\hbar^2}\left(E-2/r_{12}+4g_0\left(\frac{1}{2}R_{AB},r_1\right)+4g_0\left(\frac{1}{2}R_{AB},r_2\right)\right)\right]f(r_1,\,r_2,\,\theta_{12})=0 \quad (5.1)$$

This is a three-dimensional partial differential equation for a single function ($f\equiv f_0^{(0,0)+}$). The expression in square brackets is manifestly invariant with respect to exchange, $r_1\rightleftarrows r_2$, and the singlet and triplet solutions are distinguished by the boundary conditions:

$$\left[\frac{\partial}{\partial n}f(r_1,\,r_2,\,\theta_{12})\right]_{r_1=r_2}=0 \qquad \text{singlet} \quad (5.2a)$$

$$[f(r_1,\,r_2,\,\theta_{12}]_{r_1=r_2}=0 \qquad \text{triplet} \quad (5.2b)$$

Below we shall outline a systematic approximation procedure for solving this equation. For completeness, however, we shall give this equation with r_1, r_2, and r_{12} as the independent variables:

$$\left[\frac{1}{r_1}\frac{\partial^2}{\partial r_1^2}r_1+\frac{1}{r_2}\frac{\partial^2}{\partial r_2^2}r_2+\frac{2}{r_{12}}\frac{\partial^2}{\partial r_{12}^2}r_{12}+\frac{r_1^2+r_{12}^2-r_2^2}{r_1r_2}\frac{\partial^2}{\partial r_1\partial r_{12}}+\frac{r_2^2+r_{12}^2-r_1^2}{r_1r_2}\frac{\partial^2}{\partial r_2\partial r_{12}}+\frac{2m}{\hbar^2}\left(E-2|r_{12}+2g_0\left(\frac{1}{2}R_{AB},r_1\right)+2g_0\left(\frac{1}{2}R_{AB},r_2\right)\right)\right]f(r_1,\,r_2,\,\theta_{12})=0 \quad (5.3)$$

This form of the zeroth order equation is appropriate for solution in terms of Hylleraas-type expansions. In reference 7 other ways of writing these equations are given, involving in particular asymmetric functions $F\left(r_1,\,r_2,\begin{Bmatrix}\theta_{12}\\ r_{12}\end{Bmatrix}\right)$.

A methodical way of solving the θ_{12} equation is to make a *relative* partial wave expansion of f:

$$f(r_1,\,r_2,\,\theta_{12})=\sum_{n=0}^{\infty}\frac{1}{r_1r_2}\Phi_n(r_1r_2)P_n(\theta_{12}) \quad (5.4)$$

The Legendre polynomials $P_n(\theta_{12})$ are selected because they are the eigenfunctions of the θ_{12} derivative term in (5.1). Also the interelectron repulsion $1/r_{12}$ has a well known expansion in terms of $P_n(\theta_{12})$, so that one can easily derive an infinite set of two dimensional partial differential equation, which are essentially those that already have been derived in the case of the scattering of electrons from atomic hydrogen[13]. One can then truncate the sum at $n=N$ and expect convergence in N in a completely analogously way as one expects convergence in L in the partial wave theory as a whole.

Alternatively one can expand the total wave function in the form (considering here only the singlet case):

$$\Psi(1,2)=\sum_{n_1n_2}\sum_{\nu_1\nu_2}\sum_{\mu}C_{n_1n_2\nu_1\nu_2\mu}[R_{n_1\nu_1}(r_1)r_2^{\nu_2}\cdots)Y_{\nu_1\mu}^{n}(\Omega_1)Y_{\nu_2\mu}(\Omega_2)+(1\rightleftarrows 2)] \quad (5.5)$$

If one restricts this sum to $\nu_1=\nu_2$ and all values of μ for each ν_1, then this expansion is equivalent

to (5.4). A detailed investigation of the ground Σ_g^+ state of H_2 using (5.5) has been carried out by Hagstrum and Shull[14]. They have in fact relaxed the condition $\nu_1 = \nu_2$, so that what they approximate in those cases is a hybrid of approximations within the partial wave theory. They have, however, examined the $\nu_1 = \nu_2 = 0$ (spherically averaged) approximation. This is equivalent to including one term in (5.4) within the $L = 0$ (zeroth order) approximation. They obtain $E = -2.0894$ ryd. (1 ryd. = 13.6 eV) at $R_{AB} = 1.4$ Bohr radius. It is quite clear that the inclusion of all $\nu_1 = \nu_2$ and with their associated μ terms will lower this energy to below -2.1 ryd. Such a value is to be compared to $E = -2.34\ 884$ ryd., the very accurate nonrelativistic value (including the nuclear repulsion) of Kolos and Roothan[15]. The differences between these numbers, less than 0.2 ryd., is very close to the difference between the zeroth order and exact H_2^+ energies near their minimum ($R_{AB} = 2$)[1]. This is just what our heuristic argument would lead us to expect in view of the fact, this being a Σ state, that the next angular momentum, $\ell = 2$, is two units greater than the included $\ell = 0$ component of the zeroth order problem, in complete analogy with the case for the H_2^+ molecular ion[1]. In addition for the exact solutions of the partial wave approximations the analogy carries through regarding the upper boundedness of the energies of the ground and excited state solutions and of their mutual orthogonality[1].

REFERENCES

1. A. Temkin, J. Chem. Phys. *39*, 161 (1963).
2. M. Cohen and C. A. Coulson, Proc. Cambridge Phil Soc. *57*, 96 (1961). For additional papers involving this approach, cf. references 7 and 8 of our paper (reference 1).
3. D. R. Bates, K. Ledsham, and A. L. Stewart, Phil Trans. Roy. Soc. (London) *A 246*, 215 (1953).
4. G. Breit, Phys, Rev. *35*, 569 (1930).
5. E. A. Hylleraas, Z. Physik *48*, 469 (1928).
6. B. Holmberg, Kungl. Fysiografiska Sälskapets I Lund Förhandlinger *26*, 135 (1956).
7. A. K. Bhatia and A. Temkin, Rev. Mod. Phys. (submitted for publication).
8. P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York. 1953), p. 1719.
9. H. Evring, J. Walter, and G. E. Kimball, *Quantum Chemistry* (John Willy and Sons, Inc., London, 1944).
10. W. Magnus and I. Oberhettinger, *Formulas and Theorems for the Functions of Mathematical Physics*, (Chelsea Publishing Co., New York, 1949).
11. A. R. Edmonds, *Angular Momentum in Quantum Mechanics*, (Princeton University Press, Princeton, 1957), p. 62.
12. A. Temkin and R. Pohle, Phys. Rev. Letters *10*, 22, 268 (E) (1963). A. Temkin, NASA Technical Note D-1720.
13. A. Temkin, Phys. Rev. *126*, 130 (1962).
14. S. Hagstrum and H. Shull, J. Chem. Phys. *30*, 1314 (1959). Cf. references 3 and 7–12 of their paper for further references to papers applying this idea to H_2.
15. W. Kolos and C. C. J. Roothan, Rev. Mod. Phys. *32*, 178 (1960).

N66 37973

HYPERFINE STRUCTURE IN THE MICROWAVE SPECTRUM OF NH_2D*

P. THADDEUS,† AND L. C. KRISHER‡ AND P. CAHILL‖

Hyperfine structure in the $3_{13} \rightarrow 3_{03}$ rotation-inversion transition of deuterated ammonia NH_2D at 18 807.7 Mc/sec has been studied with a beam maser microwave spectrometer giving a linewidth of 7 kc/sec. All details of the observed hyperfine structure have been accounted for in terms of the various quadrupole, spin-rotation, and spin-spin interactions of the four coupling nuclei. The hyperfine coupling constants which have been determined with the aid of a digital electronic computer are, for the 3_{13} state: $(eq_zQ)_D = -62.5 \pm 1.0$ kc/sec, $C_D = -2.7 \pm 0.3$ kc/sec, $C_H = -13.6 \pm 0.5$ kc/sec; and for the 3_{03} state: $(eq_zQ)_D = -73.9 \pm 1.0$ kc/sec, $C_D = -2.7 \pm 0.3$ kc/sec, $C_H = -12.3 \pm 0.5$ kc/sec. On the assumption that the electric field at the deuteron is cylindrically symmetric about the N–D bond direction ξ, we find that $(eV_{\xi\xi}Q)_D = 282 \pm 12$ kc/sec, or $V_{\xi\xi} = 1.38 \pm 0.06 \times 10^{15}$ statV/cm², where $V_{\xi\xi}$ is the second derivative of the electrostatic potential along ξ.

I. INTRODUCTION

In a previous paper,[1] which we refer to as I, the results were reported of a study with a beam-maser microwave spectrometer of hyperfine structure (hfs) due to two nuclei in the rotational spectrum of the asymmetric-top molecules HDO, HDS, CH_2O, and CHDO. Linewidths in the centimeter region of about 5 kc/sec were obtained, allowing an investigation of deuteron quadrupole, spin-rotation, and spin-spin hyperfine interactions. The K-type doublet transitions of such light asymmetric-top rotors are especially favorable for beam-maser study, since a given state is well populated, and electrostatic state selection is not usually inhibited by other nearby rotational states.

In the present paper we report the observation with the same beam-maser spectrometer [2] of hfs in the $3_{13} \rightarrow 3_{03}$ rotation-inversion transition of the ammonia molecule NH_2D. Hfs in this molecule is due to four nuclei, and is so complex that it was not found possible to interpret the spectrum by means of calculations done on a desk calculator alone. With an electronic computer and automatic curve plotter, however, we have succeeded in determining the hyperfine coupling constants to the same precision obtained for HDO, HDS, CH_2O, and CHDO.

II. THEORY

The rotation-inversion spectrum of the partially deuterated ammonias has been extensively studied by Weiss and Strandberg[3] with a Stark-modulated microwave spectrometer. Successive deuteration progressively lowers the ammonia inversion frequency from about 24 000 Mc/sec for NH_3 to about 12 000, 5000, and 1500 Mc/sec for NH_2D, NHD_2, and ND_3, respectively. *Partial* deuteration in addition destroys the molecular symmetry (Fig. 1), and splits the degenerate symmetric-top levels into the well-known K-type doublets. The doublet splitting is comparable to the inversion frequency, and varies considerably from doublet to doublet. Therefore, while the strong lines of the NH_3 inversion spectrum are clustered around 24 000 Mc/sec, the rotation-inversion spectra of the partially deuterated ammonias are scattered over the entire microwave region.

The $(3_{13}, 3_{03})$ K-type doublet of NH_2D, for example, is shown in Fig. 2. The allowed transitions at 18 807 and 43 042 Mc/sec correspond to

*Published in *The Journal of Chemical Physics*, 41(6): 1542–1547, September 15, 1964. This work was supported in part by the Joint Services (The U.S. Army, the Office of Naval Research, and the U.S. Air Force Office of Scientific Research) and in part by the National Science Foundation.

†Goddard Space Flight Center and Columbia University, New York.

‡University of Maryland, College Park, Maryland and Columbia University, New York.

‖Columbia University, New York.

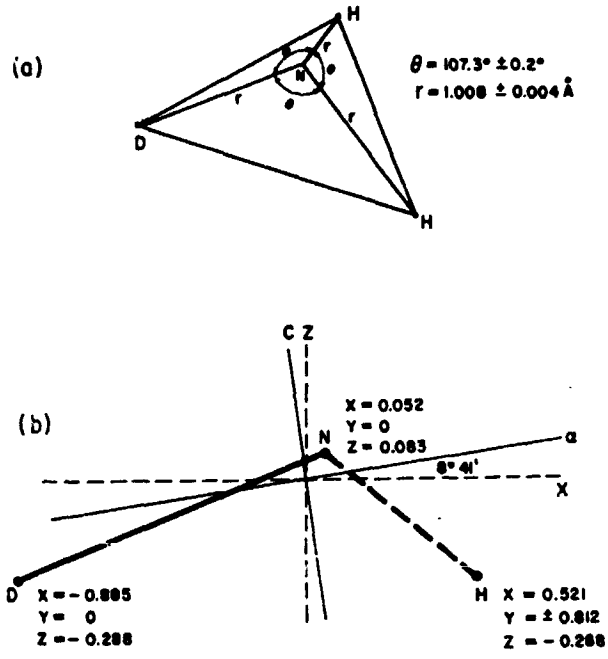


FIGURE 1.—Geometry of the NH_2D molecule. (a) shows the bond angle and length taken from Ref. 3. (b) shows how deuteration rotates the principal axes of inertia away from the symmetry axis. The xyz frame is centered on the center of mass, with Z parallel to the axis of charge symmetry.

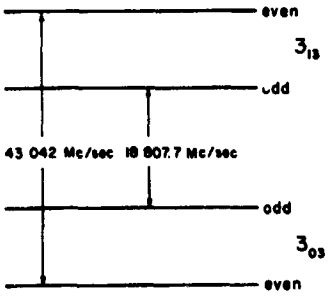


FIGURE 2.—The $(3_{13}, 3_{03})$ K-Type doublet of NH_2D split by the inversion frequency showing the allowed microwave transitions. The symmetry of the wave function on exchange of the protons is indicated as even (e) or odd (o).

the NH_3, $J=3$, $K=3$ inversion transition at 23 870 Mc/sec—the strongest absorption line in the centimeter region of any molecule, and the one with which Gordon, Zeiger, and Townes[4] first obtained maser oscillation in 1954. The 18 807-Mc/sec line in absorption is the strongest NH_2D transition at X or K band, and was selected for study since it was accessible to the microwave apparatus used in I.

All four nuclei in NH_2D contribute to the hfs. An estimate of the hierachy of the hyperfine interactions can be made from the many previous studies of NH_3,[4-6] and the measurement of the deuteron quadrupole coupling constant by Herrman,[7] who studied with a Stark-modulated spectrometer the inversion spectrum of ND_3 near 1500 Mc/sec. The quadrupole interaction of the nitrogen in NH_2D is expected to be over an order of magnitude larger than the quadrupole interaction of the deuteron, or the spin-rotation $(\mathbf{I}\cdot\mathbf{J})$ interactions of the protons. The following coupling scheme was therefore used to theoretically interpret the hfs:

$$\begin{aligned}\mathbf{J}+\mathbf{I}_N&=\mathbf{F}_1,\\ \mathbf{F}_1+\mathbf{I}_D&=\mathbf{F}_2,\\ \mathbf{F}_2+\mathbf{I}_H&=\mathbf{F},\end{aligned}\qquad(1)$$

where $\mathbf{I}_H$ is the sum of the proton spins,

$$\mathbf{I}_H^{(1)}+\mathbf{I}_H^{(2)}=\mathbf{I}_H.\qquad(2)$$

Since the deuteron quadrupole and proton $\mathbf{I}\cdot\mathbf{J}$ interactions are of nearly the same strength, the order of coupling following F_1 is in fact unimportant.

The coupling scheme (2) yields an $I_H=1$ symmetric, and an $I_H=0$ antisymmetric spin state. In Ref. 3 it is shown that both of the rotation-inversion states of the 18 807-Mc/sec transition are antisymmetric on exchange of the protons, and therefore for these states only the $I_H=1$ spin state exists. The two protons thus couple to the rest of the molecule like a single spin 1 nucleus, and the hyperfine Hamiltonian in the notation of I can be written in the symmetrical form

$$\mathcal{H}=\frac{(eq_JQ)_N}{2I_N(2I_N-1)\,J(2J-1)}\left[3(\mathbf{I}_N\cdot\mathbf{J})^2+\frac{3}{2}(\mathbf{I}_N\cdot\mathbf{J})-\mathbf{I}_N^2\mathbf{J}^2\right]+C_N(\mathbf{I}_N\cdot\mathbf{J})\qquad(3a)$$

$$+\frac{(eq_JQ)_D}{2I_D(2I_D-1)\,J(2J-1)}\left[3(\mathbf{I}_D\cdot\mathbf{J})^2+\frac{3}{2}(\mathbf{I}_D\cdot\mathbf{J})-\mathbf{I}_D^2\mathbf{J}^2\right]+C_D(\mathbf{I}_D\cdot\mathbf{J})\qquad(3b)$$

$$+\frac{g_H^2\mu_N^2\,(d_J)_{HH}}{2I_H(2I_H-1)J(2J-1)}\left[3(\mathbf{I}_H\cdot\mathbf{J})^2+\frac{3}{2}(\mathbf{I}_H\cdot\mathbf{J})-\mathbf{I}_H^2\mathbf{J}^2\right]+C_H(\mathbf{I}_H\cdot\mathbf{J})\qquad(3c)$$

$$+\frac{g_H g_D \mu_N^2 (d_J)_{HD}}{J(2J-1)}\left\{\frac{3}{2}[(\mathbf{I}_H\cdot\mathbf{J})(\mathbf{I}_D\cdot\mathbf{J})+(\mathbf{I}_D\cdot\mathbf{J})(\mathbf{I}_H\cdot\mathbf{J})]-(\mathbf{I}_H\cdot\mathbf{I}_D)\mathbf{J}^2\right\} \quad (3d)$$

$$+\frac{g_H g_N \mu_N^2 (d_J)_{HN}}{J(2J-1)}\left\{\frac{3}{2}[(\mathbf{I}_H\cdot\mathbf{J})(\mathbf{I}_N\cdot\mathbf{J})+(\mathbf{I}_N\cdot\mathbf{J})(\mathbf{I}_H\cdot\mathbf{J})]-(\mathbf{I}_H\cdot\mathbf{I}_N)\mathbf{J}^2\right\} \quad (3e)$$

$$+\frac{g_D g_N \mu_N^2 (d_J)_{DN}}{J(2J-1)}\left\{\frac{3}{2}[(\mathbf{I}_D\cdot\mathbf{J})(\mathbf{I}_N\cdot\mathbf{J})+(\mathbf{I}_N\cdot\mathbf{J})(\mathbf{I}_D\cdot\mathbf{J})]-(\mathbf{I}_D\cdot\mathbf{I}_N)\mathbf{J}^2\right\}. \quad (3f)$$

Terms (3a) are the quadrupole and spin-rotation interactions of the nitrogen, and (3b) are the same interactions for the deuteron. The coupling constants for these interactions, together with C_H of (3c), will be determined from experiment for both the 3_{13} and 3_{03} states. The coupling constants of the four spin-spin interactions of (3c), (3d), (3e), (3f), however, can be calculated to higher precisions from the geometry of Fig. 1 than they can be experimentally determined with the present resolution. The spin-spin constants calculated from Eq. (9) of I are listed in Table I. It should be noticed that the mutual spin-spin interaction of the protons has the same dependence on the coupling momenta as a quadrupole interaction.

III. EXPERIMENT

a. Sample Preparation

Since hydrogen is readily exchanged between ammonia molecules, a sample of 100 percent of the partially deuterated ammonias cannot be prepared. A maximum of $\frac{4}{9}NH_2D$ can be produced in a sample in which the ratio of protons to deuterons is 2:1. The sample used in this experiment was prepared by mixing commercial ammonia in a vacuum system with the appropriate mixture of ordinary water and heavy water, condensing the mixture out in a liquid-nitrogen trap and warming to room temperature several times to assure equilibrium. The ammonia was subsequently separated from the water by vacuum distillation.

TABLE I.—*The spin-spin constants calculated from the molecular geometry of Fig. 1.*

Nuclei	$g_A g_B \mu_N^2 (d_J)_{AB}$, kc/sec	
$A-B$	3_{13} state	3_{03} state
H–H	8.59	2.99
H–D	1.82	1.94
H–N	1.09	0.29
D–N	0.58	0.71

b. Microwave Spectrometer

The microwave spectrometer used in this work has been described in detail elsewhere.[2] Sufficient microwave power (about 1 mW) to operate the superheterodyne receiver was at first furnished by the second harmonic of a very stable Varian X–13 klystron using a Varactor doubler. Subsequently a Varian V40B klystron was used directly at 18 807 Mc/sec.

Frequency markers less than 1 kc/sec wide were obtained by mixing the output of the klystron with the appropriate harmonic of a Gertsch FM–4 frequency generator locked to the Laboratory frequency standard. The difference frequency was coaxially mixed with a tunable low-frequency oscillator, a Rohde-Schwarz XUA frequency synthesizer, and then detected with a Collins receiver. The receiver was kept at a constant frequency, and the marker frequency changed by varying the frequency of the Rohde-Schwarz synthesizer.

c. Results

The observed hfs, except for the $\Delta F_1=\pm 1$ nitrogen quadrupole satellites which are over 1 Mc/sec to either side, is shown in Fig .3. In all three traces the most intense $\Delta F_1=\Delta F_2=0$ transitions have been deliberately saturated to bring out the finer detail, and therefore appear weaker than calculated. Equation (3a) shows the three $\Delta F_1=0$ transitions, and the theoretical structure calculated only on the basis of the nitrogen[6] and deuteron[7] quadrupole coupling. Deuteration, as expected, by destroying the molecular symmetry spreads these transitions into a wide triplet, and the much smaller deuteron quadrupole interaction adds even further structure. It is apparent from the additional structure seen in the more magnified traces of (3b) and (3c), however, that the magnetic

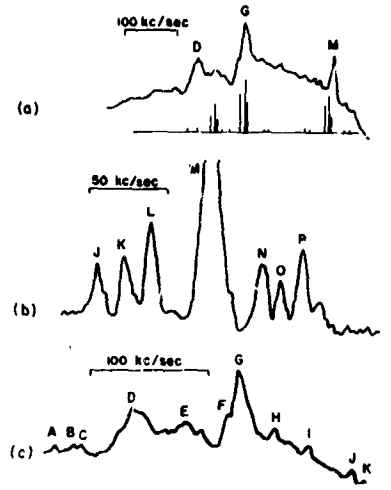


FIGURE 3.—Oscilloscope traces of the observed hfs. (a) shows the three $\Delta F_1=0$ transitions against the background of the maser cavity response, with, for comparison, the theoretical structure due to the nitrogen and deuteron quadrupole interactions. (b) shows the $F_1=3\rightarrow3$ transition in detail. (c) shows the $F_1=4\rightarrow4$ and $2\rightarrow2$ transitions. The letters indicates the features whose frequency was accurately measured. The vertical scale varies from trace to trace.

interactions of the protons are important for a detailed understanding of the hfs.

The frequencies of the prominant features of the hfs lettered in Fig. 3 are listed in Table II. All frequencies are given relative to the strong central line M of the $F_1=3\rightarrow3$ transition.

IV. HYPERFINE CONSTANTS

All the matrix elements of the Hamiltonian of Eq. (3) and the intensity matrix were calculated from Eqs. (22), (23), (24), and (31) of I, using tabulated values of the $6j$ coefficients.[8] For the spin-spin interactions, after the various 27×27 (highly reducible) operators $\mathbf{I}_i\cdot\mathbf{I}_j$ and $\mathbf{I}_i\cdot\mathbf{J}$ were calculated, the indicated matrix products and summations of Egs. (3d), (3e), (3f) were performed with an IBM 7094 digital computer.

Since the nitrogen quadrupole interaction is much larger than the deuteron quadrupole interaction, terms in the Hamiltonian off-diagonal in F_1, which contribute diagonal frequencies (energies) somewhat smaller than $(eq_JQ)_D{}^2/(eq_JQ)_N \approx 2$ kc/sec, can be ignored. This leads to the considerable simplification that any one of the three $\Delta F_1=0$ transitions can be analyzed separately.

TABLE II.—*Measured frequencies relative to M of the features of the hfs identified in Figs. 3 and 5. The uncertainties are estimated to be 1 kc/sec. The intervals calculated on the basis of the best-fit constants of Table 3 are also listed.*

Feature	Observed frequency (kc/sec)	Calculated frequency (kc/sec)
A	−344.6	−345.7
B	−318.0	−318.9
C	−295.7	−294.3
D	−268.1	−2[illegible]1
E	−238.3	−236.1
F	−205.9	−204.9
G	−195.1	−195.1
H	−162.5	−160.1
I	−125.7	−122.9
J	−70.0	−69.4
K	−49.6	−49.4
L	−34.2	−35.0
M	0.0	0.0
N	35.3	36.0
O	48.7	50.2
P	61.2	62.4

Accordingly, a machine program was written for the 7094 computer that for a given set of coupling constants calculated the hfs for a given one of the $\Delta F_1=0$ transitions. A Gaussian lineshape was assumed. The results were plotted out directly with a General Dynamics SC 4020 plotter to allow direct comparison with the observed hfs. A typical plot of the $F_1=3\rightarrow3$ transition, which required less than 1 min of machine time, is shown in Fig. 4. Figure 3 shows that the structure of this transition is more highly resolved than that of the overlapping $F_1=4\rightarrow4$ and $2\rightarrow2$ transitions. The coupling constants were therefore determined from this transition alone, the other two being used only as a check.

Since in practice the six coupling constants which determine the hfs within one of the $\Delta F_1=0$ transitions could not all be varied, it was important to use the symmetry of the molecule and the results of previous work to reduce the number of independent constants. It was therefore at first assumed that the electric field at the deuteron is cylindrically symmetric along the D–N bond direction ξ, and the coupling constant along the

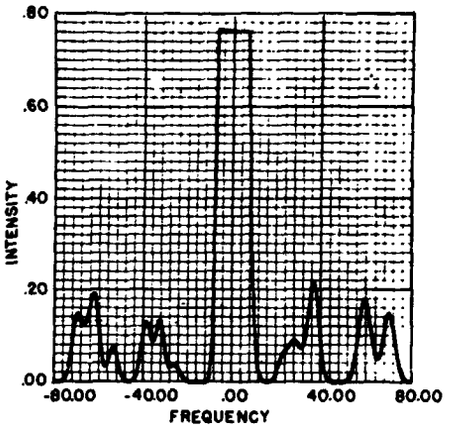


FIGURE 4.—Direct plotter output for the initial theoretical estimate of the $F_1=3\rightarrow3$ hfs.

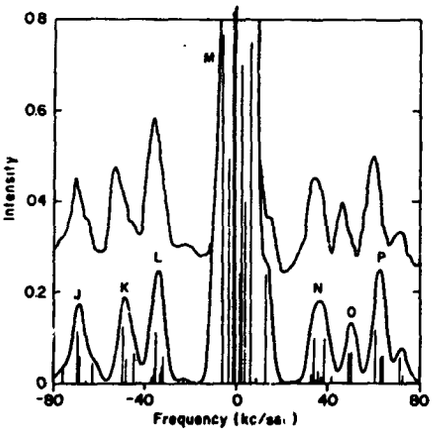


FIGURE 5.—Theoretical spectrum of the $F_1=3\rightarrow3$ transition, with Figure 3(b) traced in to show the close agreement with observation. The hump on the right side of the main line is weaker than calculated due to power saturation. (see text).

bond, $(eV_{\xi\xi}Q)_D$ was taken from the work of Herrman on ND_3 to be 200 kc/sec. On this basis $(eq_JQ)_D$ is calculated to be -42.6 kc/sec for the 3_{13} state and -52.9 kc/sec for the 3_{03} state.

Moreover, since the c axis is only rotated some 8° from the electronic symmetry axis, the magnetic field at the protons is expected to be nearly equal to that at the deuteron, so it was assumed that

$$C_D=(g_D/g_H)C_H=0.18\ C_H.$$

It was further assumed that

$$C_H(3_{13})=C_H(3_{03})=-18.8\ \text{kc/sec},$$

the value found for the (3, 3) state of NH_3 by Gordon.[6] The hfs calculated on these assumptions is shown in Fig. 4. It is seen to fit the observed $F_1=3\rightarrow3$ transition rather poorly.

By increasing $(eV_{\xi\xi}Q)_D$ and decreasing C_H a much better fit was soon found. Close agreement with experiment, however, was arrived at only after varying the hyperfine constants independently by slight amounts in either state. The theoretical best-fit $F_1=3\rightarrow3$ transition is shown in Fig. 5—a tracing of the plotter output with the 40 odd significant lines drawn in. All observed features were fit to within 3 kc/sec, and the well-resolved details of the $F_1=3\rightarrow3$ transition were fit to within 1.5 kc/sec. The best-fit full line-width at half maximum was found to be 7 kc/sec.

Taking the six constants found from the 3→3 transition, the hfs of the $F_1=4\rightarrow4$ and $2\rightarrow2$ transitions was calculated, and was found to reproduce all the features of the observed spectrum. Finally, from the intervals D and G (Table II) between the main peaks of the three $\Delta F_1=0$ transitions, it was possible to determine the difference between the nitrogen coupling constants for the 3_{13} and 3_{03} states. No attempt was made to analyze the nitrogen quadrupole satellites, which would have given the individual nitrogen constants, since the signal-to-noise ratio for these transitions was low. All the hyperfine constants found are listed in Table III, and the three calculated $\Delta F_1=0$ transitions are shown in Fig. 6.

V. FIELD GRADIENT AT THE DEUTERON

From the measured value of $(eq_JQ)_D$ in both the 3_{13} and 3_{03} states, the coupling constant $(eV_{\xi\xi}Q)_D$ along the D–N bond direction ξ, and the asymmetry parameter η may be calculated. It is found that

$$(eV_{\xi\xi}Q)_D=282\pm12\ \text{kc/sec},$$

TABLE III.—*The best-fit hyperfine coupling constants for the deuteron and the protons. Uncertainties are estimated. In addition, the nitrogen coupling constants in the 3_{13} minus those in the 3_{03} states are found to be: $\Delta(eq_JQ)_N=-197.3\pm0.6$ kc/sec, and $\Delta C_N=0.3\pm0.2$ kc/sec.*

State	$(eq_JQ)_D$ (kc/sec)	C_D (kc/sec)	C_H (kc/sec)
3_{13}	-62.5 ± 1.0	-2.7 ± 0.3	-13.6 ± 0.5
3_{03}	-73.9 ± 1.0	-2.7 ± 0.3	-12.3 ± 0.5

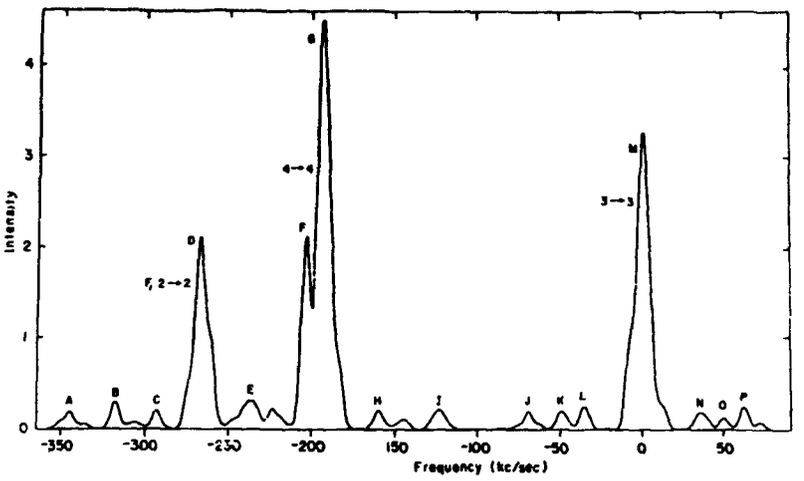


FIGURE 6.—Theoretical spectrum of all the $\Delta F_1 = 0$ transitions.

somewhat higher than the value 200 ± 20 kc/sec found by Herrman[7] from his study of ND_3. Taking the quadrupole moment of the deuteron to be[9] 2.82×10^{-27} cm^2, this gives for the second derivative of the electrostatic potential along the bond at the deuteron

$$V_{\xi\xi} = 1.38 \pm 0.06 \times 10^{15} \text{ statV/cm}^2,$$

comparable to the values found for the deuteron in HDO, HDS, and CHDO.[1]

If the asymmetry parameter is defined as

$$\eta = (V_{\xi\xi} - V_{\chi\chi})/V_{\xi\xi},$$

where ξ is parallel to the b axis, and χ lies in that plane of molecular symmetry, it is calculated that

$$\eta \approx 0.05 \pm 0.06.$$

It may therefore be concluded that to within experimental uncertainty the electric field at the deuteron is cylindrically symmetric about the D–N bond.

VI. DISCUSSION

A word should be said concerning the most effective plan of attack on the hfs of molecules as complex as NH_2D. The presence of equivalent nuclei implies that the structure of certain states will be much simpler than that of others. Both rotation-inversion states, for example, of the high-frequency $3_{13} \rightarrow 3_{03}$ transition of NH_2D at 43 042 Mc/sec are even on exchange of the protons (Fig. 2), which therefore have a total spin $I_H = 0$, and do not contribute to the hfs. Qualitatively, the structure is then expected to resemble the simple theoretical spectrum shown in Fig. 3(a), from which the deuteron quadrupole and spin-rotation constants can be easily and precisely determined. In the present case, it is clear that prior precise knowledge of the deuteron constants would have considerably simplified the analysis of the 18 807-Mc/sec transition, and that in general in a study of NH_2D and similar molecules, it is most reasonable to first study states which are even on exchange of the protons (for equivalent deuterons, the converse is true).

The frequency range of the available microwave apparatus did not permit this procedure in the present case. However, the necessity to interpret a very complex hyperfine spectrum has led to the development of electronic data-processing programs that should be adequate for the majority of the transitions in the deuterated ammonias, and indicates that a number of other molecules with three and four spins can now be profitably studied.

Further experimental investigations of the deuterated ammonias and other molecules are now in progress in this Laboratory, with emphasis on attaining higher frequencies. In particular, we are interested in finding for NH_2D all three of the diagonal elements M_{gg} of the tensor that determines the spin–rotation constants in the rigidrotor approximation,[1] or in observing any departures from the theory based on this approximation, such as have been observed in HDO.[1,10]

ACKNOWLEDGMENTS

We are indebted to the staff of the Radiation Laboratory for support and assistance given during the course of this work. H. Lecar and E. Ballik aided in the taking of data during the initial phase of the experiment, and E. Schreier was of invaluable assistance in checking the many numerical calculations. We wish especially to thank Ohseun Koh, who wrote the program for the calculation of the hyperfine structure.

REFERENCES

1. P. THADDEUS, L. C. KRISHER, and J. H. N. LOUBSER, J. Chem. Phys. **40**, 257 (1964).
2. P. THADDEUS and L. C. KRISHER, Rev. Sci. Instr. **32**, 1083 (1961).
3. M. T. WEISS and M. W. P. STRANDBERG, Phys. Rev. **83**, 567 (1951).
4. J. P. GORDON, H. J. ZEIGER, and C. H. TOWNES, Phys. Rev. **99**, 1264 (1955).
5. C. H. TOWNES and A. L. SCHAWLOW, *Microwave Spectroscopy* (McGraw-Hill Book Company, Inc., New York, 1955), Chap. 12.
6. J. P. GORDON, Phys. Rev. **99**, 1253 (1955).
7. G. HERRMAN, J. Chem. Phys. **29**,, 875 (1958).
8. M. ROTENBERG, R. BIVENS, N. METROPOLIS, and J. K. WOOTEN, *The 3j and 6j Symbols* (The Technology Press, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1959).
9. J. P. AUFFRAY, Phys. Rev. Letters **6**, 120 (1961).
10. E. B. TREACY and Y. BEERS, J. Chem. Phys. **36**, 1473 (1962).

AUTHOR INDEX

Aikin, A. C., 325, 335
Alexander, J. K., 841
Allison, L. J., 846, 1183
Arking, A., 856

Balasurbrahmanyan, V. K., 419
Bandeen, W. R., 861, 1027
Bartko, F., 969
Bauer, S. J., 343, 348
Beard, D. B., 423
Bhatia, A. K., 1198, 1336
Blum, R., 441
Blumle, L. J., 348
Boggess, A. III, 1, 4
Borgman, J., 4
Bourdeau, R. E., 355, 363, 376, 879
Boyd, R. L. F., 384, 408
Bozis, G., 58
Brace, L. H., 884, 902, 1148
Bracken, P. A., 498
Brancazio, P. J., 907
Brunstein, K. A., 445
Bryant, D. A., 449, 466

Cahill, P., 1346
Cain, J. C., 471
Cameron, A. G. W., 7, 474, 770, 907, 1216
Chandra, S., 387, 879
Chaney, L., 946
Chen, R. L. W., 1329
Chiu, H.-Y., 13, 27, 55
Cline, T. L., 449, 466, 476
Contopoulos, G., 58

Dalgarno, A., 884, 912
Davis, T. N., 480
Desai, U. D., 449, 466
Donn, W. L., 819
Dunkleman, L., 987

Felsentreger, T. L., 191
Fichtel, C. E., 482, 493, 496
Fowler, W. B., 987
Frost, K. J., 291
Fussell, W. B., 294

Giacaglia, G. E. O., 196
Goldberg, R. A., 387, 307
Gross, S. H., 936
Guss, D. E., 299, 482, 493, 496

Halev, M., 861
Hanel, R. A., 946, 969
Henderson, C. L., 412
Hendricks, S., 471
Hennes, J. P., 987
Hess, W. N., 431
Hicks, G. T., 480
Hoffman, R. A., 498
Horowitz, R., 1024, 1148
Huang, S.-S., 67, 108, 113, 120, 129, 137, 206
Hultqvist, B., 513, 588, 604

Inman, C. L., 142

Jaggi, R. K., 612
Jastrow, R., 1216

Kalra, G. L., 155
Kane, J. A., 335
Kastner, S. O., 993
Katchen, L., 1106
Kaufmann, R. L., 618
Kniffen, D. A., 493, 496
Kohler, R., 1231
Konradi, A., 618
Krisher, L. C., 1346
Kunde, V., 1027

Laflin, S., 408
Lipschutz, M. E., 1239, 1326
Ludwig, G.H., 476, 628

Maeda, K., 997
McDonald, F. B., 419, 449, 466, 476, 628, 802
Michels, T. E., 1244
Montgomery, D., 632
Musen, P., 215, 233

Neelakantan, K. A., 482, 493, 496
Ness, N. F., 642, 675, 689, 697
Neupert, W. M., 301, 306, 879
Newton, G. P., 1024, 1148
Nordberg, W., 1027, 1172

Omidvar, K., 1261
Osterbrock, D. E., 169

Panagakos, N. W., 907
Prabhakara, C., 1300
Priester, W., 1024

Raitt, W. J., 384, 408
Rasool, S. I., 936, 1049, 1055, 1059
Reber, C. A., 1060, 1148
Reeves, H., 160
Roper, R. G., 1073
Ruderman, M. A., 142

SACHER, P., 1106
SALPETER, E. E., 55, 165
SCEARCE, C. S., 642
SCHM'D, L. A., 701
SCHUTT, J. B., 294
SCIACCA, T. P., 1326
SCOLNIK, R., 1066
SEEK, J. B., 642
SERBU, G. P., 750
SHEN, C. S., 1333
SINGH, M. P., 757, 768
SMITH, L. L., 926
SMITH, W., 1106
SPENCER, N. W., 884, 902, 1148
STECHER, T. P., 1158
STOCKHAUSEN, R. E., 169, 315
STOKES, A. P., 612
STOLOV, H. L., 770
STONE, R. G., 841
STOTHERS, R., 174, 184
STRANGE, I., 861
SUGIURA, M., 777
SWARTZ, P., 1106

TALWAR, S. P., 155, 796
TEMKIN, A., 1198, 1336
THADDEUS, P., 1231, 1346
THEON, J., 1106
TIDMAN, D. A., 632
TROIM, J., 335

WAGNER, C. A., 243
WALKER, J. C. G., 912, 1162
WARNECKE, G., 1027, 1172, 1183
WATANBE, T., 997
WEBBER, W. R., 802
WHITE. W. A., 319
WICKLESS, W. J., 191, 282
WILCOX, J. M., 697
WILLMORE, A. P., 412

☆ U.S. GOVERNMENT PRINTING OFFICE. 1966—0-786-496