NASA SP-361

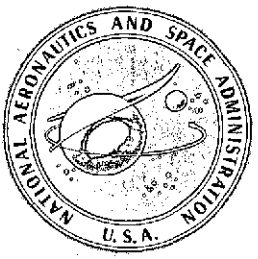
SIGNIFICANT ACCOMPLISHMENTS IN SCIENCE AND TECHNOLOGY

Goddard Space Flight Center, 1973

A symposium held at

GODDARD SPACE FLIGHT CENTER

December 18-19, 1973



NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

NASA SP-361

SIGNIFICANT ACCOMPLISHMENTS IN SCIENCE AND TECHNOLOGY

GODDARD SPACE FLIGHT CENTER, 1973

The proceedings of a symposium held at the NASA
Goddard Space Flight Center, December 18-19, 1973

Prepared by Goddard Space Flight Center

Scientific and Technical Information Office 1975
NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Washington, D.C.

For sale by the National Technical Information Service
Springfield, Virginia 22151 -- Price $8.00

FOREWORD

This is an almost verbatim transcript of a symposium held at Goddard Space Flight Center, Greenbelt, Maryland, on December 18 and 19, 1973. No attempt has been made to introduce editorial or stylistic uniformity in the talks; on the contrary, an effort has been made to retain the informality of the proceedings.

The sole major change results from NASA policy, which now requires in all formal publications the use of international metric units according to the Systeme International d'Unites (SI). However, in certain cases utility requires the retention of other systems of units in addition to the SI units. The conventional units stated in parentheses following the computed SI equivalents are the basis of the measurements and calculations reported here.

CONTENTS

CONTENTS (continued)

CONTENTS (continued)

CONTENTS (continued)

CONTENTS (continued)

OPENING REMARKS

Dr. John F. Clark
Director, Goddard Space Flight Center

Each year, more and more, we are building on the foundation of the years that have come before. So each year should be more exciting and more fruitful, and this one is certainly no exception to this rule.

I would like to acknowledge the presence in the audience of several people who have been very important to the space program, particularly the space science program. In front of me is Dr. Homer Newell, NASA Associate Administrator and my friend of many years, who is probably more responsible than any other individual for the past successes of our space science programs. Another man whose presence I would like to acknowledge is Dr. John Nangle, NASA Associate Administrator for Space Science and another valued colleague of mine over the past decade.

With these brief comments, I should like to turn this podium and microphone over to our General Chairman, Dr. Meredith.

REMARKS

Dr. Leslie H. Meredith, General Chairman

As you will notice from the program, we are in tune with the energy problems in the United States. We are starting off with the icy comets and we are ending up the morning session with freezing your blood. In the middle, of course, we will warm things up appreciably and we do have coffee planned in about an hour and a half. So without further ado, the first speaker will be Jack Brandt from the Cometary Program.

OVERVIEW

John C. Brandt

Naturally, a certain amount of work goes into preparing one of these sessions, but I cannot recall one that has been more fun. It is a great pleasure to chair this session, because comets are so fascinating, and now we have the blessing of blessings, a comet in the sky. Part of this program was labeled "Round Table Discussion" ("Scientific Results Obtained and/or Expected from Kohoutek"), simply because of our inability to predict results and create titles for talks while the research is being carried out.

Historically, many great astronomers and philosophers were involved in the study of comets, and issued weighty opinions on what they were or should have been. The views of Aristotle, who described comets as being atmospheric phenomena, dominated the scene for a very long time. When comets were labeled as atmospheric, several things were accomplished, for at that time, comets were not a part of astronomy. If you check in old astronomy books of the sixteenth century or earlier, it will be found that comets are generally not discussed in them because they were not part of astronomy. Comets were, however, thought to herald disasterous events such as wars, deaths of kings, and hot winds. This all made a certain amount of sense if the comets were atmospheric in nature and could influence events on earth. The situation was fundamentally changed by Tycho Brahe, while studying the comet of 1577. In effect, he tested the atmospheric hypothesis and was able to establish that the comet had no parallax as viewed from two different places on the earth, or as viewed from the same place at two different times. In other words, the comet's position did not change with respect to the background stars. Tycho Brahe was an excellent observer, and he estimated that the comet had to be at least four times as far away from the earth as the moon. Clearly, the comet of 1577 was an astronomical object.

Much of the research after this time deals with the Newton and Halley era that is familiar to most people. The movement of planetary bodies in conic sections was established. By extending these ideas to comets, Halley was able to make his famous prediction of the return of the comet which now bears his name. It was verified on Christmas Day in 1758, and the comet passed perihelion in 1759. The fact that the comets move in highly elliptical orbits was at that time considered to have great cosmological

significance. The planets were supposed to be embedded in solid crystalline spheres. Comets in their orbits were passing through many of these so-called solid crystalline spheres and this cosmology had to be abandoned.

The physical study of comets began in the early nineteenth century. The great impetus in this work was Bessel (who observed Halley's comet in 1835) and when you read in textbooks about cometary forms you will find the so-called Bessel-Bredikhin theory mentioned quite prominently. At the same time, we should be aware of the characterization of comets by Sir William Herschel, and the following is his view as summarized by Agnes Clerke. It dates from 1800 and is remarkably prophetic. Comets are described as "fragments of nebula–the scattered debris of embryo worlds." Recently, detection of OH and methylcyanide CH_3CN in comets has been reported. Now, CH_3CN is known in the interstellar medium, and the view of the comets as the debris of the formation of the solar system seems to be gaining strength.

In the 1950's, the cometary nuclear body was held to be a single (or small number of) solid ice body on Whipple's "dirty snowball" model. The ices were thought to be complex or relatively complex parent molecules. When the comet approaches the sun, the solar radiation evaporates or sublimates the icy material, which ultimately forms the other parts of the comet that we will talk about later. In the 1960's, the picture was somewhat completed even though I hesitate to say anything is complete in cometary physics. There was the realization that we not only needed photons from the sun to make comets look like themselves, but we also needed the solar particle emission, the solar wind. Specifically, comets are embedded in the solar wind and this fact must be considered if we are to achieve a physical understanding of comets.

Figure 1 is a photograph of Comet Bennett taken in 1970. There are two kinds of comet tails: The broad, homogeneous tail is composed of dust. This is one reason to call the comet a dirty snowball. As the cometary ice sublimates, it releases the dust embedded in it. This dust is blown by solar radiation pressure away from the sun. It is not too efficient and as a result we often get a large curvature. If we had a striking color photograph, the dust tail would appear to be yellow, and it is essentially reflected sunlight. The tail with the striations, little knots, kinks, and other irregularities is the ion tail. It is generally straight, and points close to the radial direction from the sun. The principle constituent in the visible wavelength range is CO^+. There are other molecules, such as OH^+, but the CO^+ molecules dominate most pictures. The principle emission bands of CO^+ are in the blue range, and ion tails show up blue in color pictures. The coma, or the head of the comet, is a basically round area of 10^5 to 10^6 kilometers in

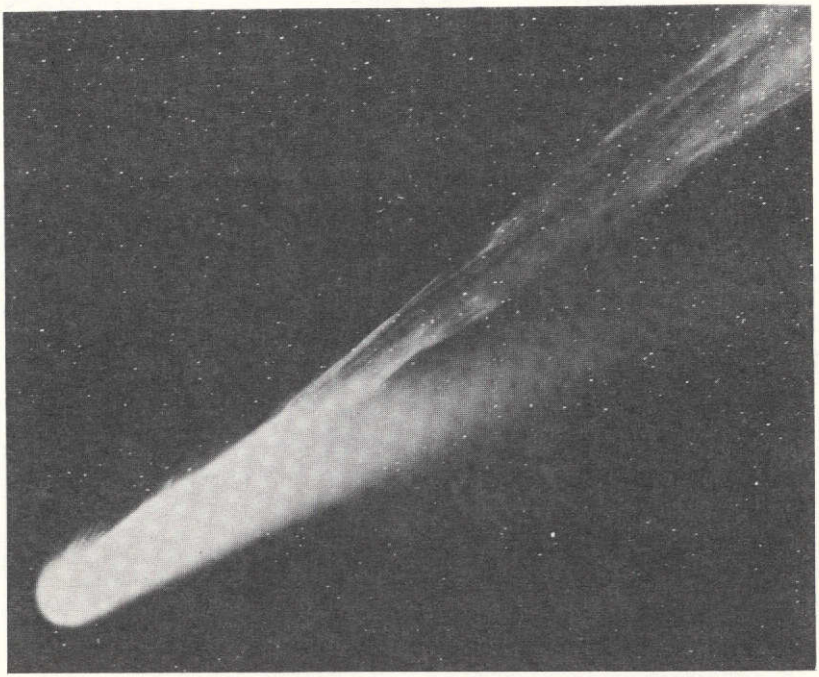


Figure 1. Comet Bennett, 1970.

diameter that surrounds the central part of the comet. It does not show on this picture of Comet Bennett. It is composed of dust and neutral molecules. Finally, we have every reason to expect, but have no direct proof, that there is a nucleus of 1 to 10 kilometers in diameter at the center of the coma.

Figure 2 is a summary, schematic diagram showing the principal parts and dimensions of comets. I am not too keen on the larger sizes listed for the nucleus, but they are certainly within the uncertainties. We do not know the masses of comets, but there are many ways of calculating upper limits.

I would like to point out another cometary feature, not visible on figure 1. This is the neutral hydrogen cloud which is visible in Lyman-α at 121.6 nm (1216 Å) or in the OH lines. The fact that the comet is surrounded by clouds of water-ice dissociation products—namely, H and OH—and that densities are approximately equal, argues strongly for the idea that the predominant ice in the nucleus is simply water-ice. The coma has the sizes listed in figure 2. Ion tails can be 10^8 kilometers long, and there are some on record that are even larger. In other words, the ion tail lengths can be

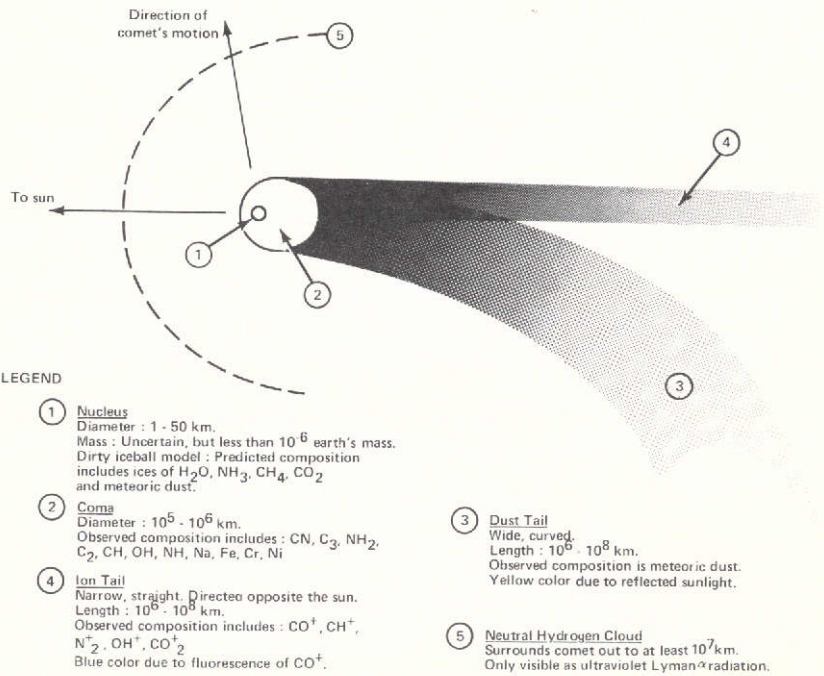


Figure 2. Schematic drawing of a comet (not to scale).

larger than the distance from the earth to the sun. This gives you some idea of our current views. I emphasize that this sketch gives what we think a comet is at the present time. With the number of observations that are going on now with unprecedented techniques, and with the Skylab astronauts observing above the atmosphere, it is clear that we may have to change some of our ideas about comets.

In the round table discussion, we will hear about new observatories designed completely from beginning to end to observe comets (Dr. Hobbs). We will hear about radio observations designed to search for water (Dr. Clark). The structure in the ion tails clearly implies filaments. The filamentary structure observed in the ion tails needs a physical explanation; perhaps it is a magnetic field. A possible way of measuring this will be described (Dr. Jackson). We know what the dissociation products of the parent molecules are in the coma, but we are not sure of the identity of the parent molecules themselves. Detection may be possible by observations in the infrared (Dr. Pearl). Some new techniques for comet studies, such as heterodyne receivers and infrared up converters may yield new results (Dr. Kostiuk).

We are always interested in the origin of comets, and perhaps the ratio of deuterium to hydrogen could be a test of the region of comet formation (Dr. Donn). Finally, even if our most optimistic hopes for the observations planned for Comet Kohoutek are all successful, there is still a complete class of comet observations that will not have been touched: direct sampling or in situ observations. Such observations should be a part of any balanced cometary research program (Dr. Mumma).

I cannot overstress the concept that we are beginning a new era in cometary physics and many facts and ideas could change. One of my favorite philosophers, if you will let me call him a philosopher, is Ambrose Bierce. In his Devil's Dictionary, he defined an observatory: "Observatory, n. A place where astronomers conjecture away the guesses of their predecessors." This is no reflection either on the predecessors or on the astronomers. The problem in cometary physics is that the observatory must be staffed, and as far as cometary physics is concerned, this has not been the case. One can quibble about the exact numbers, but in the United States there have been, to my knowledge, only three fulltime comet workers. If we consider the man-years of effort that are going into the observations of Comet Kohoutek, it is certainly proper to expect that we will be making substantial progress. Finally, with a comet of this nature and the publicity involved, there is a rather large communications and operations effort that involves scientists. There is also a very substantial communications effort that involves public information and the general status of the public's view of comets.

Figure 3 is a cartoon taken from a French publication in 1857. When I was originally preparing this talk I thought I would include it so we could all have a good laugh. That was before I saw the December 17, 1973, article on comets in *Time* magazine, which was about half filled with material that is at least as bad as figure 3. Also, much of the television coverage of this comet has been concentrated on the frankly bizzare response of some people. In essence, some of the public response is a return to the Aristotelian view of comets. I regard that as discouraging. Hence, it is important that we tackle some of the educational aspects of the comet, and Dr. Robert Chapman has prepared a booklet entitled "Comet Kohoutek 1973-1974" for high school teachers. Various people have collaborated in producing other leaflets relating to the comet. It is important to bear the educational aspects of this comet in mind. After all, the comet is in the sky, and we are the public agency most identified with investigating such phenomena. It is important that we explain the situation to the public.

I would like to proceed to Dr. Steve Maran, who will present "Operation Kohoutek—An Update."

Figure 3. French cartoon of a comet, 1857.

OPERATION KOHOUTEK-AN UPDATE

Stephen P. Maran

A photograph from the Hale Observatories, figure 1, shows Comet Kohoutek (1973f) as it was photographed on November 24, 1973, with the Palomar Schmidt telescope. An appropriate filter-emulsion combination was used to isolate the dust tail. Note the prominent notch in the tail. Figure 2, which is from the Lowell Observatory in Flagstaff, shows the comet 2 weeks later, on December 7. So some very striking changes have already been recorded.

We had, apparently, a successful Kohoutek observation on Skylab yesterday. The coma of Comet Kohoutek occulted a bright B-star in Scorpius, and the event was recorded from the scientific airlock with the S019 experiment (ultraviolet objective prism spectrograph) of Dr. Karl Henize. The measurements with the ultraviolet airglow grating spectrometer on Mariner-10 are now scheduled to begin on January 10, 1974. Acquisition of relevant solar wind data with Pioneer-6 is now underway, and the Pioneer-8 measurements will occur shortly after the comet reaches perihelion.

Measurements with the Princeton Experiment Package on the OAO-3 (Copernicus) satellite will begin on January 29, 1974, and will be repeated at intervals of 4 days. OSO-7 may be used on December 20 and 21, when the comet is near Antares and it will, in any case, make observations for a few days around perihelion, when both pointed experiments will be employed.

The first sounding rocket launch, for the Johns Hopkins University and Caltech/Bendix experiments, is presently listed for January 4, 1974. The first successful infrared observations of the comet with the NASA/Ames Research Center Lear Jet were accomplished on December 10; the Erickson 1- to 4-μm photometer was used.

Our new Joint Observatory for Cometary Research is now in operation and Dr. Hobbs will present a report on it. Among the several special NASA grants for Comet Kohoutek observations is one to Dr. F. Scherb at the University of Wisconsin. His work at the solar telescope on Kitt Peak under this grant has already resulted in the first reported detection of a cometary hydrogen cloud from the ground. (Dr. B. Donn of GSFC is a collaborator in that project.) Ground-based observations have also yielded

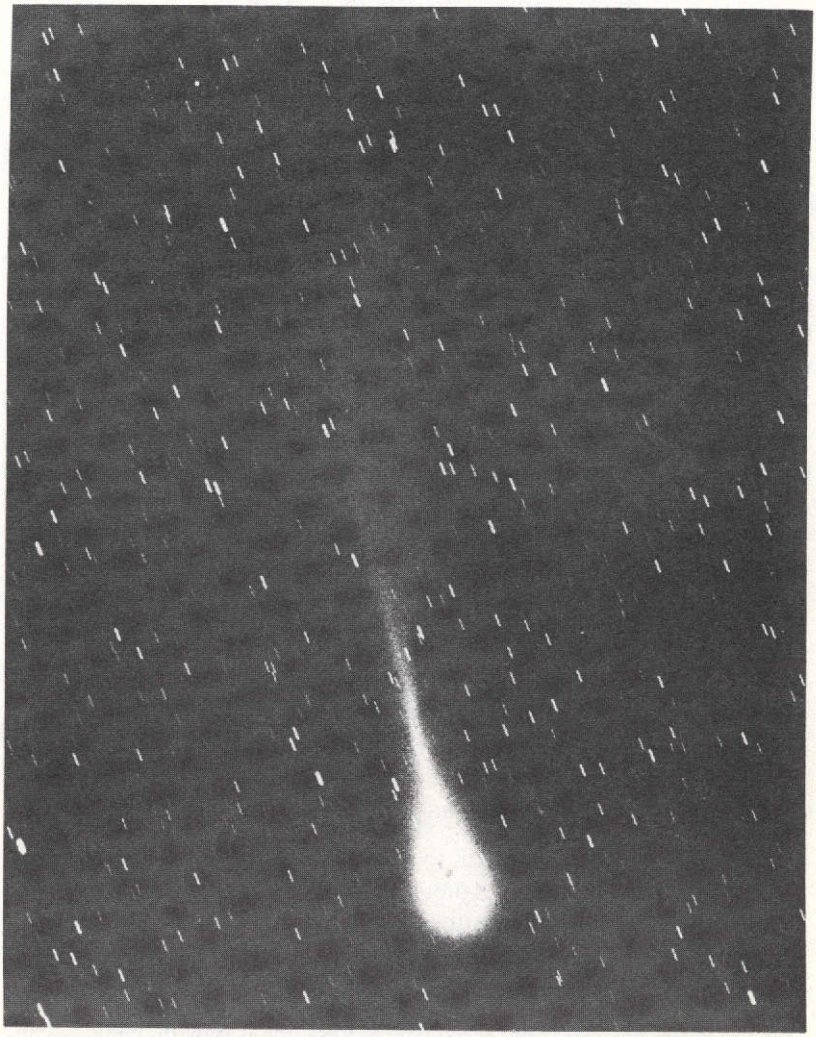


Figure 1. Comet Kohoutek, photographed at the Hale Observatories by A. Sandage, November 24, 1973, 12:58 UT. Emulsion-filter combination favored yellow light, to emphasize the dust tail.

the first strong evidence for radio signals from a comet (methyl cyanide emission), as well as evidence for radio absorption (by OH molecules) against the galactic background.

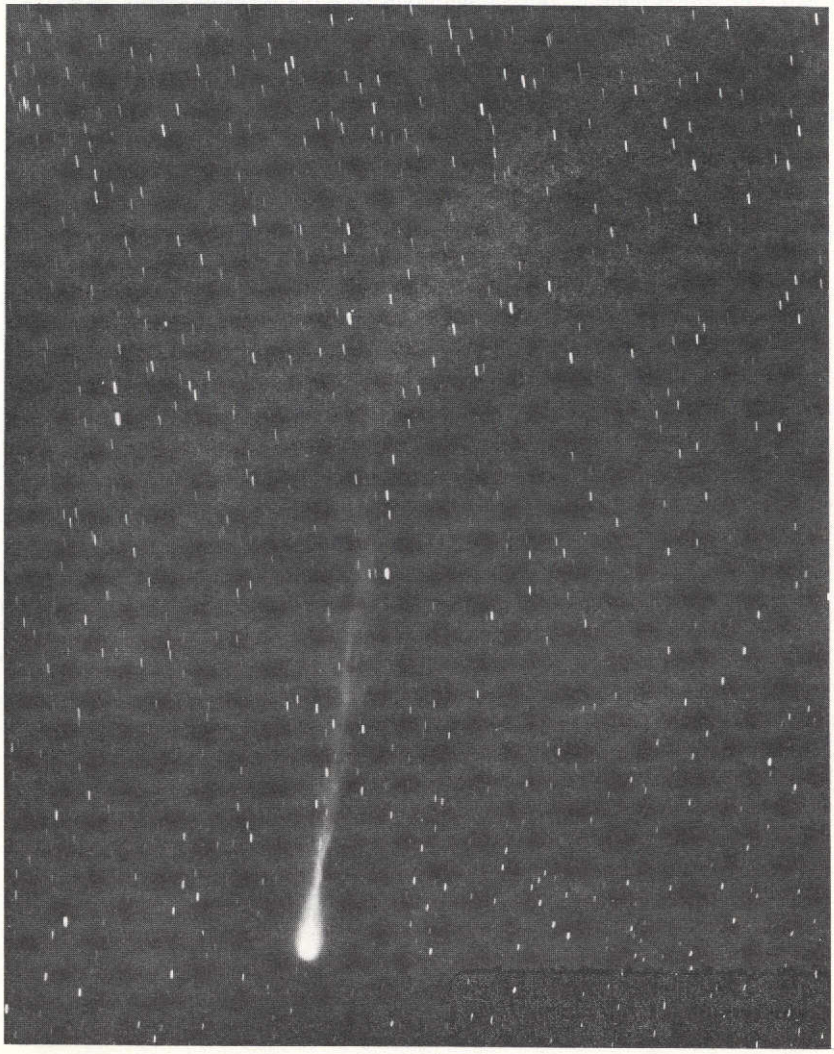


Figure 2. Comet Kohoutek, photographed on December 7, 1973, 12:35 UT, courtesy of Lowell Observatory. The structure in the ion tail should be noted.

A new element in our program is the Convair 990 (Galileo II), which will fly a variety of experiments out of NASA/Ames and also over Alaska. We are also cooperating with the AEC's Los Alamos Scientific Laboratories; their NC-135A aircraft had a successful observation flight on December 14.

An information center on Comet Kohoutek has been established here at GSFC and is engaged in relaying reports on observation results and schedules. In addition, detailed ephemerides have been produced for individual radio observatories by Dr. T. Clark.

THE JOINT OBSERVATORY FOR COMETARY RESEARCH

Robert Hobbs

The Joint Observatory for Cometary Research is the only observatory in the world that was conceived for and dedicated exclusively to cometary research. It is operated under a cooperative agreement with the New Mexico Institute of Mining and Technology and the Goddard Space Flight Center. The observatory is on South Baldy Mountain at 3180 meters (10,600 feet) elevation near Socorro, New Mexico. The site is very dark, located in a National Forest, with almost no nearby lights. Fortunately, it is situated near the Langmuir Laboratory so that new roads and other facilities did not have to be constructed.

Figure 1 shows the twin-dome structure of the observatory when it was in the final stages of completion. One of the twin domes houses an f/35 telescope capable of taking large-scale photographs of the coma of a comet. The other dome houses a 14-inch, f/2 Schmidt telescope which is shown in figure 2.

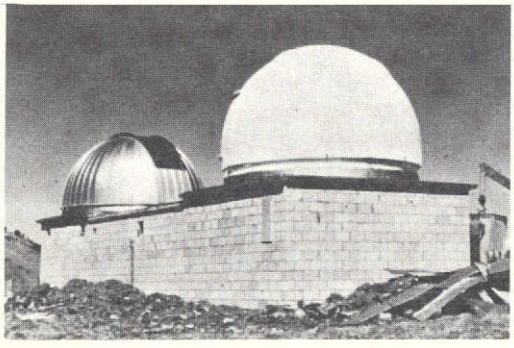


Figure 1. Joint Observatory for Cometary Research.

The Schmidt telescope was developed especially for cometary research. It is called a 14-inch telescope because this is the size of the corrector plate. The primary mirror, however, is 28 inches in diameter. The whole telescope is of a very special design having a temperature-compensated fixed focus, so one does not have to take a number of focus plates at sunset or sunrise

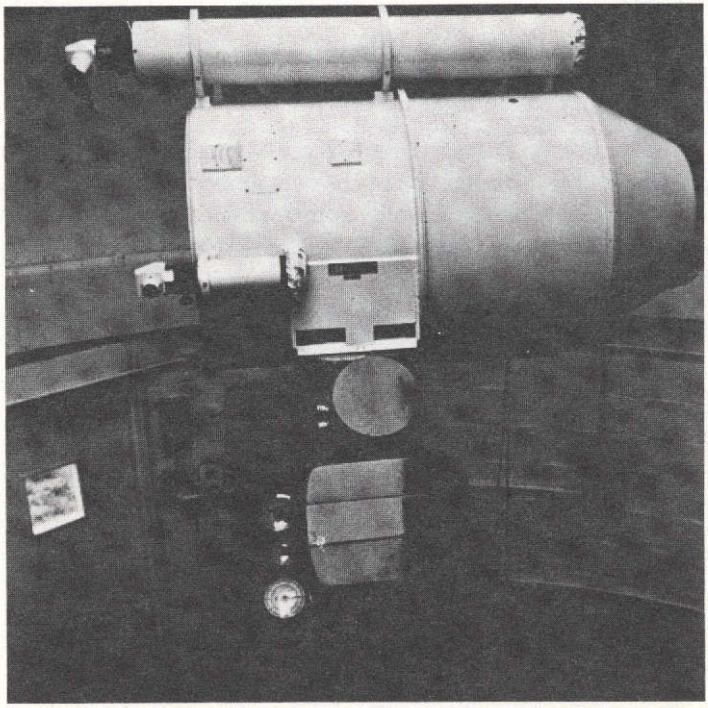


Figure 2. Schmidt 14-inch, f/2 telescope.

when the temperature is changing most and when one is most likely to observe a comet. It has an additional feature that has been developed specifically for this telescope—roll film transport. Instead of changing plates for each exposure, a procedure which might take several minutes, the roll film transport is a completely automatic film handling device combined with a shutter, making it possible to take a new exposure within a second or two after the old exposure has been completed. This also is

especially valuable for cometary research where one does not have a lot of time in which to observe. At present, both of the telescopes, the f/2 Schmidt and the f/35 instrument are in operation obtaining data on Comet Kohoutek.

Figure 3 is a photograph of Comet Kohoutek that was taken on December 11, 1973. One can see two tails here quite clearly. The tail length is about 6°. The scientific objectives of this observatory are to study the interaction of comet tails with the solar wind and, hence, to further our knowledge of solar physics. A special effort was made to get the observatory ready for Comet Kohoutek because a comet that comes as close to the sun as Kohoutek is rather rare. The close approach to the sun provides a unique opportunity for solar wind research.

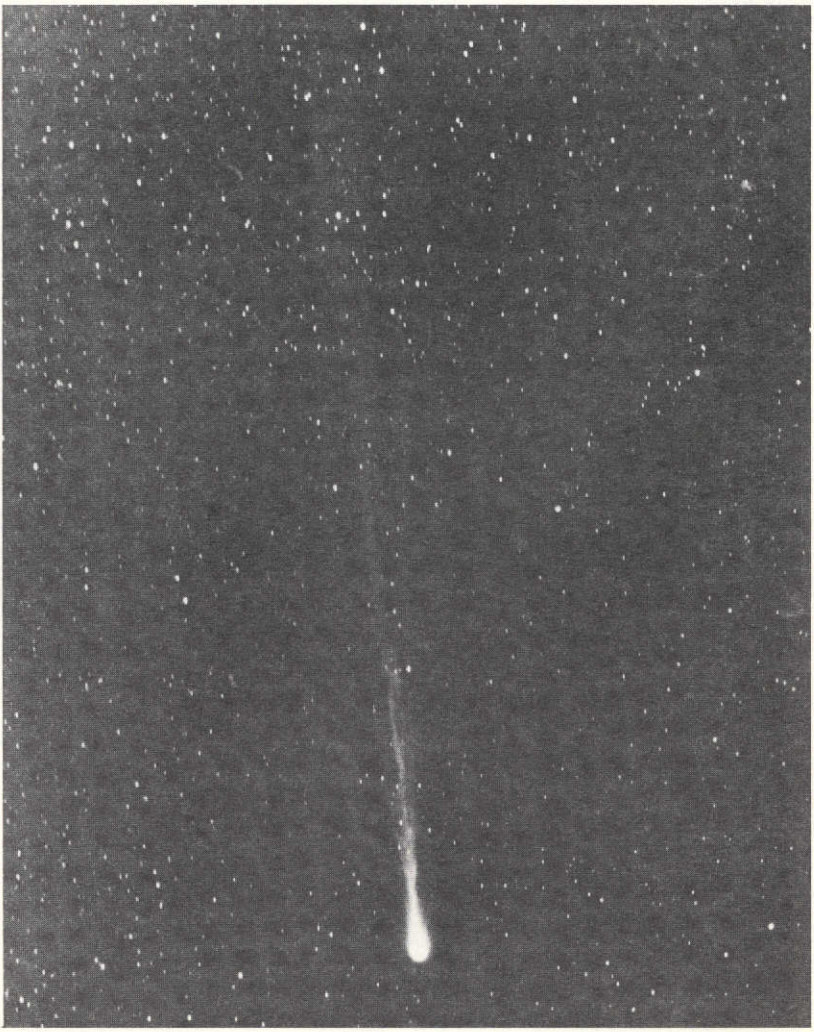


Figure 3. Comet Kohoutek, with two tails visible, photographed December 11, 1973.

WHAT REMAINS TO BE MEASURED AFTER COMET KOHOUTEK

Michael Mumma

It may be somewhat pretentious to talk about what remains to be measured before we have results from the Kohoutek experiments, but there are certain questions that will not be addressed by the experiments currently underway and which, in fact, could not be addressed by instruments either on the earth or in earth orbit. The talks immediately preceding this one have described a range of experiments that will potentially yield new data of a kind never before obtained on comets. The most conspicuous gap in our knowledge of comets is that up until the Kohoutek passage, earth men had not been able to identify any of the parent molecules directly. We now have a radio detection of methyl cyanide (CH_3CN), a possible detection of water, and direct detection of OH and a number of other fragment radicals that were seen in earlier comets, but the detection of parent molecules is probably the single most important result to come out of this set of experiments.

Some of the important questions which will not be answered are as follows:

- Existence of the nucleus; physical parameters
- Identity and abundance of parent molecules; mechanisms of destruction
- Composition of dust; size distribution
- Nature of the solar wind-comet interaction
- Mechanisms responsible for ion production

The first, and perhaps most important question that will not be addressed directly is whether in fact there is a single solid cohesive nucleus. Most scientists at the present time believe in the existence of the nucleus, but there is at least one cometary scientist, Raymond Littleton, who suggests that rather than being a solid cohesive body, the nucleus could be a swarm of particles. It is impossible to resolve this question through near-earth photography. It is interesting to note that even a 3-m telescope, such as the LST in earth orbit, operating at its maximum resolution (the diffraction limit), would only be able to resolve objects with diameters greater than 60 km, 0.1 AU away from the earth. Most comets pass much farther away

from the earth than that and the nuclear diameters which are in vogue these days run from 1 to 30 km. Thus, even with this superb optical telescope, there is no chance of seeing whether there is a solid nucleus, let alone determining its physical parameters such as size, shape, rate of rotation, and so forth.

The second point is that while we may identify certain parent molecules which are relatively abundant, such as water or perhaps methyl cyanide (some of the infrared measurements may result in detection of ammonia, methane, and so forth), there undoubtedly will be a problem in determining the abundance ratios. In order to determine the absolute abundances, one needs to know rotational distributions, which will certainly be nonthermal. Furthermore, the scarcer parent molecules will probably not be picked up at all, and there are probably some molecular species present that we do not now suspect. For these reasons, it is best to go through a comet and actually measure these molecules in situ with an ionization mass spectrometer. In addition to determining their identities and abundance ratios, one also could measure their spatial distributions which could then be related to theoretical models. We know very little about the specific mechanisms of molecular destruction. These can best be studied by direct in situ measurements.

There were several remarks made earlier today that silicate absorption has been seen in the spectrum of Comet Kohoutek, although I was not aware of that. I knew that the emission peak at 10 μm due to the thermal brightness of the dust had been seen, but of crucial importance to determining whether the dust is of silicate nature is the determination of the elemental abundances. This can only be done by in situ measurements with a device such as an impact ionization dust analyzer. Furthermore, the size distribution of the dust particles is a very important piece of information for cometary physics and this, again, can best be determined by direct measurements.

The nature of the solar wind-comet interaction is quite important, not only for studying the solar wind in regions inaccessible to spacecraft (that is, at high inclinations to the ecliptic), but also for studying the physics of comets and for use as a tool in studying plasma physics. The cometary region is an ideal plasma physics laboratory and provides a collision-free domain which we cannot achieve on earth. Finally, we really have no clear-cut knowledge of the mechanisms responsible for the production of ions which we see in fluorescence in the tail. We know, of course, that once neutral molecules become ionized, they will interact with the magnetic field carried by the solar wind. They will be locked onto the field lines which act as guiding centers, but we do not know how they are produced.

Figure 1 shows Goddard's concept of an intercept mission to Comet Encke. We had worked long and hard on the Grigg-Skjellerup Giacobini-Zinner intercept, which we thought was the best possible early comet mission, but that has fallen by the wayside due to funding problems. The next possible mission is a ballistic flyby of Comet Encke in 1980. The sun-comet direction is illustrated in figure 2. To first approximation, the coma and the comet itself can be thought of as being cylindrically symmetric about the sun-comet line. Under those circumstances, it seemed to make most sense to cross the cylinder at a steep angle to the sun-comet line so that by taking measurements at different positions along the trajectory, one could later generate a surface of revolution, characterizing first the interaction at the bow shock, then coming into the outer coma, through the inner coma, measuring parent molecules and photographing the nucleus at high resolution (<100 m), and then proceeding out through the bow shock again on the other side. Because the physics and the environment are different in the tail, we feel that a tail probe would also be highly desirable with a complimentary set of instruments. Both probes would be launched with a single rocket and would carry essentially the same instruments where possible; for example, the ion mass spectrometers would be identical.

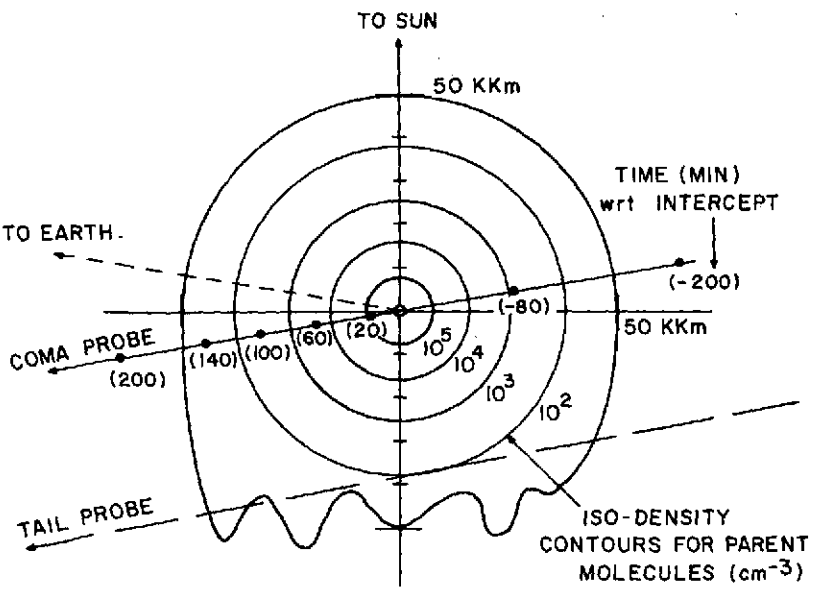


Figure 1. Probe trajectories at parent Encke encounter (0.34 AU).

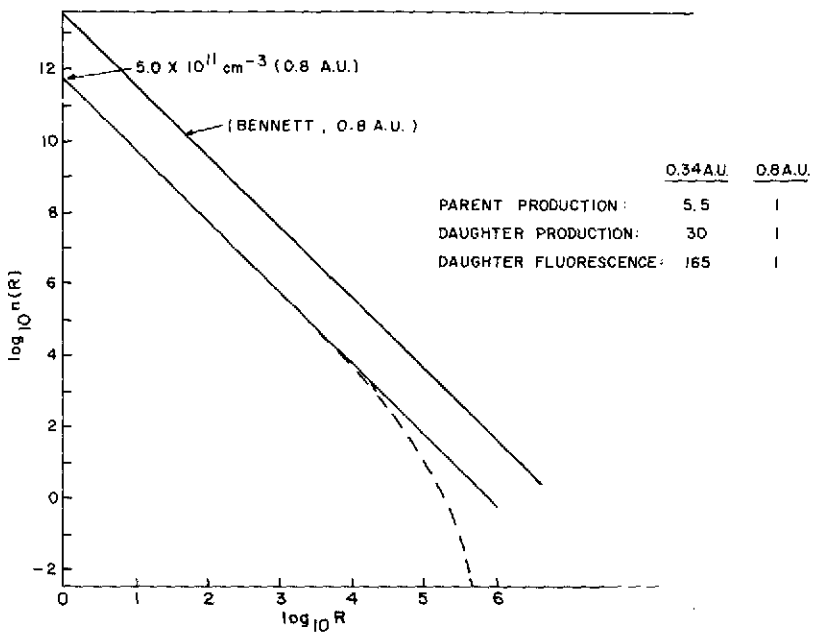


Figure 2. Parent Encke: parent density versus distance from the nucleus.

To give you some idea of the critical nature of the distance of closest approach, figure 2 shows how the parent molecule density varies for Comet Encke at an intercept occurring at a heliocentric distance of 0.8 AU. These data are normalized to the results of Bertaux, Blamont, and Festou, and we see that the parent water density falls rapidly from a value of about 5 $\times 10^{11}$ cm^{-3} at 1 km from the center of the nucleus. In fact, however, any neutral mass spectrometer that one could fly would have a threshold sensitivity of $\sim 10^3$ cm^{-3}. Presuming that the instrument spends 5 percent of the available time on a particular mass, then the H_2O density would have to be at least 2×10^4 cm^{-3} in order to be measurable. This density occurs at ~5000 km from the nucleus at a heliocentric distance of 0.8 AU. One obviously has to go in much closer than this range in order to make spatial measurements at decent integration time. Thus, a distance of closest approach of 500 km is almost mandatory to get a decent measurement of water for Comet Encke. Of course Comet Bennett, which was a much brighter comet, had correspondingly higher water densities, and in fact, Kohoutek could be somewhat brighter than that. This kind of argument illustrates the importance of the minimum flyby distance.

I hope I have given you an understanding of some of the questions that need to be answered, and I think it is quite clear that the best chance to answer these questions is to fly there directly and intercept the comet with a spacecraft. It is especially important because comets, no doubt, represent our best chance to study unmodified matter from the presolar nebula.

POSSIBLE OBSERVATIONS OF MAGNETIC FIELDS IN COMETS

William Jackson

I would like to discuss an experiment that has been designed to detect magnetic fields in Comet Kohoutek. Observations of comet tails, as Dr. Brandt has told you, led several previous investigators to suggest the presence of magnetic fields in comets. Figure 1 illustrates our present thoughts about the origin of this cometary magnetic field.

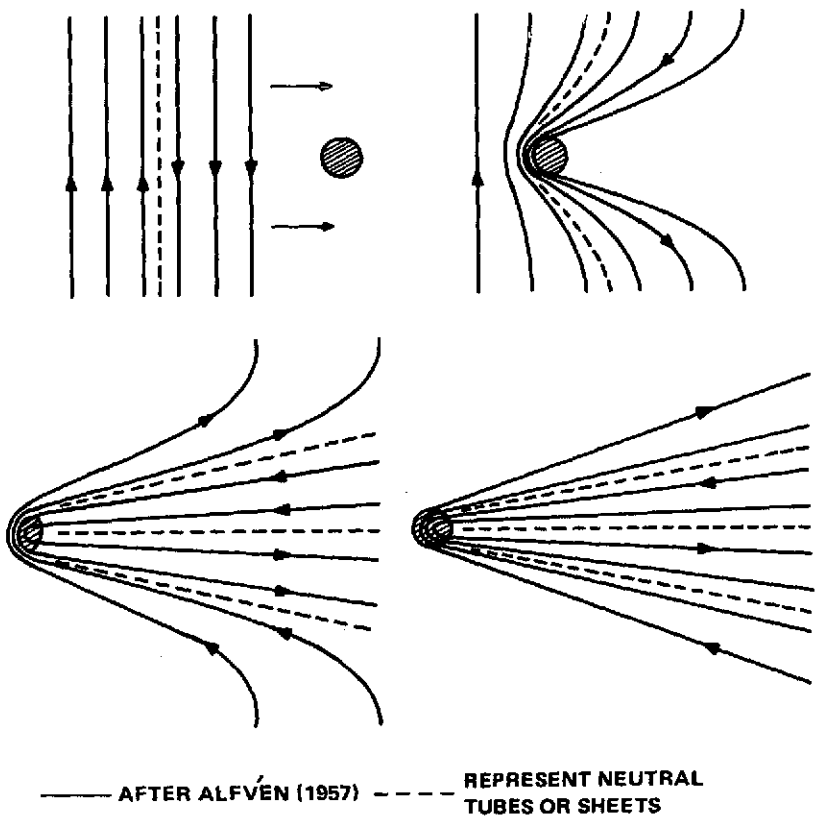


Figure 1. Cometary capture of interplanetary magnetic field.

The solar wind with its magnetic field comes along and hits the coma of the comet. Some of the magnetic field lines of the solar wind are trapped and compressed here by the gas in the head of the comet. The magnetic field lines are dragged back in the tail where they form rays that somehow trap the cometary ions. This is a very general explanation of the origin of ion tails in comets. Since there are no measurements of the strength of this magnetic field, the following experiment was designed to measure fields as high as a few milligauss.

As we all know, all of the molecular emission lines in comets occur because of residence fluorescence. Suppose we take OH—a photon from the sun then excites the OH which will then reemit another photon. It is this photon that we see and detect and it is this photon that gives a comet the characteristic molecular emission lines. In the absence of collisions and magnetic fields, we would expect this light to be polarized. But, as I said, we expect magnetic fields in comets, so that we expect some depolarization of the normally expected polarized light. This fact was appreciated by earlier observers and they tried to use a sodium resonance radiation to measure the magnetic field in comets. These observations led to an upper limit of the order of 2 gauss.

The polarization (P) in the presence of the magnetic field divided by the polarization (P_o) in the absence of a magnetic field, for mutually perpendicular geometry, is given by equation 1. In this equation, H is the magnetic field strength, m the mass of the electron, c the velocity of light, g the λg factor, and τ, the radiative lifetime of the transition. The main point here is that this ratio is inversely proportional to the square of the magnetic field strength, the g factor, and τ, the lifetime. The g factors for OH are about the same. So, unless there is a big difference in τ, we do not expect any increase in sensitivity. But of course there is a big difference in τ.

$$\frac{P}{P_o} = \frac{1}{1 + \left(\frac{eH}{mc}\, g\tau\right)^2}$$

The relative ratio of the OH lifetime to the Na lifetime is about 30, so that the limiting magnetic field for OH lines compared to sodium lines would be something like 2 gauss times 1 over the square of 30, which is on the order of 1 milligauss. This is something like 20 times the magnetic field in the

solar winds at the earth. If no depolarization is measured then we would know that the compression ratio of the solar wind in comets is smaller than a factor of 20.

INFRARED SPECTRAL OBSERVATIONS IN THE 10-μm ATMOSPHERIC WINDOW

John C. Pearl and Virgil G. Kunde

Observations of Comet Kohoutek in the near and middle infrared provide an opportunity to study the molecular and particulate components of the cometary coma. At the present time, we have begun observations at the Coude focus of the 107-inch reflector at the McDonald Observatory. These observations will continue until December 21, 1973, and a second observing period will begin shortly after perihelion passage. Our spectrometer is a Michelson interferometer which has previously been used at McDonald for planetary observations. Cometary spectra are being recorded in the 10-μm region of the spectrum with a spectral resolution of about 1 part in 1000 and a spatial resolution of 20 to 30 arc-s. Possible cometary constituents which may be visible in the 10-μm regions of the spectrum include various molecular species, dust, and water ice. The line positions and spectral signatures of such materials are shown in figure 1.

The abscissa here is in wave numbers, which is the reciprocal of the wavelength, so that on a wavelength scale this corresponds to about 8 μm, 10 μm, and 12 μm. The portions of this spectral region which are available for ground-based observations correspond to the gaps between these heavy black lines. The black lines themselves indicate where atmospheric absorption exceeds 50 percent.

To date, none of the so-called parent molecules postulated in various cometary models have been detected experimentally. However, spectral features of many such molecules lie in this region of the spectrum. At the top of figure 1 are shown line positions for ammonia and corresponding to these lines, methane, and band centers for formaldehyde and NH_2. Potentially, the easiest parent molecule to observe is ammonia, since it has very strong lines and is in a very transparent region in the spectrum. The two NH_3 Q-branches at 830 and 870 cm^{-1}, which I will show you subsequently, are exceedingly strong. The ammonia molecule is a possible parent for the NH and NH_2 radicals which have already been observed in comets. Our spectral measurements should detect and allow identification of strong molecular emissions or allow strict upper limits to be set on some of the molecular abundances.

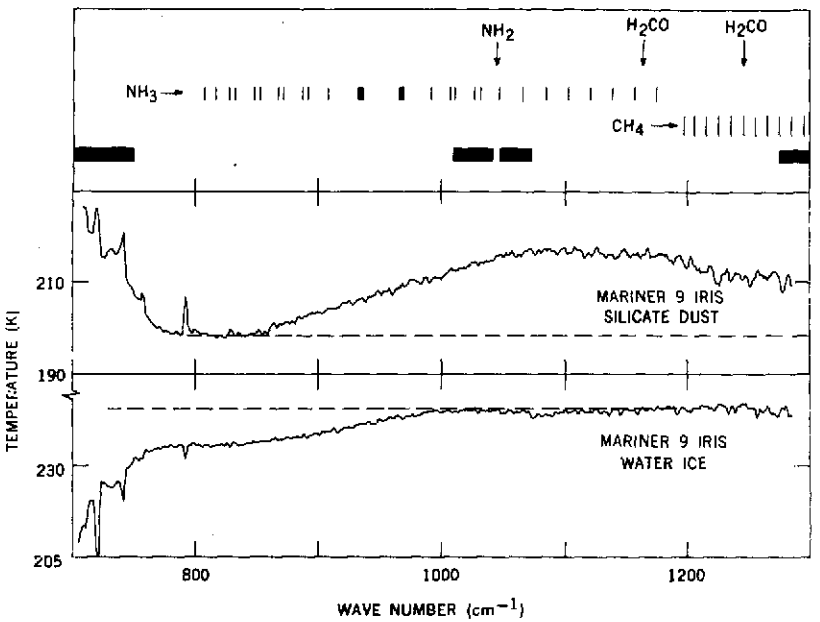


Figure 1. Cometary composition schematic, 10-μm region.

Silicate dust shows a strong spectral feature in this region as is illustrated by the feature in one of our Mariner-9 IRIS spectra shown in the center of figure 1. Based on wideband photometric measurements, an emission feature at 10 μm in cometary spectra has already been identified with silicates. However, since such wideband measurements can be contaminated with molecular emissions, such as NH_3, high spectral resolution measurements are important to establish the relative contributions of the dust and gas contributions of this feature.

Water ice crystals also show a strong spectral feature in this range. This is evident on a Mariner-9 IRIS spectrum shown at the bottom of the figure. Since water ice is a key component in most models of the cometary nucleus, its identification would be of great significance.

Returning now to the molecular constitution of the comet, we have computed a synthetic ammonia spectrum for a simple model of the coma as is shown in figure 2.

The model used, which is shown schematically in the insert in figure 2, consists of a warm inner atmosphere at 400 K surrounded by a cooler ammonia atmosphere. The spectrum itself, as you can see, is very rich,

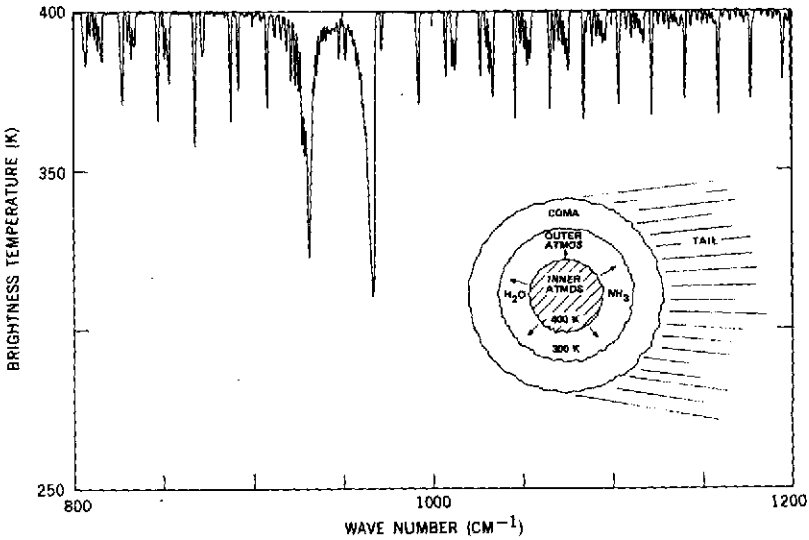


Figure 2. Synthetic cometary NH_3 absorption spectrum.

containing well over a dozen lines which should be detectable through the atmospheric window. I might add here that the next to last line on the right is going to be investigated by Dr. Mumma and his group. They will be actually looking at the line profile. Dr. Kostiuk's talk will deal with the heterodyne experiment which will be used for that measurement.

From a study of the relative intensities of these various lines, information on the populations of various energy levels can be derived. The level populations, in turn, yield information on some of the other physical parameters and processes in the coma. To conclude, we anticipate that cometary spectra in the 10-μm region will yield substantial information on gaseous components of the comet, such as ammonia and methane, and on dust and water ice particulates. From these data, a better understanding of the structure and processes within cometary bodies will emerge.

NEW METHOD FOR SPECTRAL MEASUREMENTS ON COMETS

Theodor Kostiuk

In this paper I will describe two highly sensitive methods for measuring infrared emissions from comets. We will attempt to apply these methods to measurements of molecular emissions from Comet Kohoutek. The devices employed are based on the mixing of two electromagnetic frequencies to obtain a third, which can then be easily enhanced and interpreted.

The first device, illustrated in figure 1, is the infrared heterodyne spectrometer. The scheme here is to combine two infrared signals, one (which is generally fairly weak) from the telescope and the other from an infrared laser local oscillator, to obtain a difference frequency. This difference is in the radio frequency range and can then be electronically processed and analyzed. The results provide us with the desired spectrum or line profile of the cometary emission. The spectral tuning range of this device, as well as its signal-to-noise ratio, or its sensitivity, is controlled by the local oscillator laser. By tuning it one can tune the spectrometer and by increasing its power one can optimize its sensitivity.

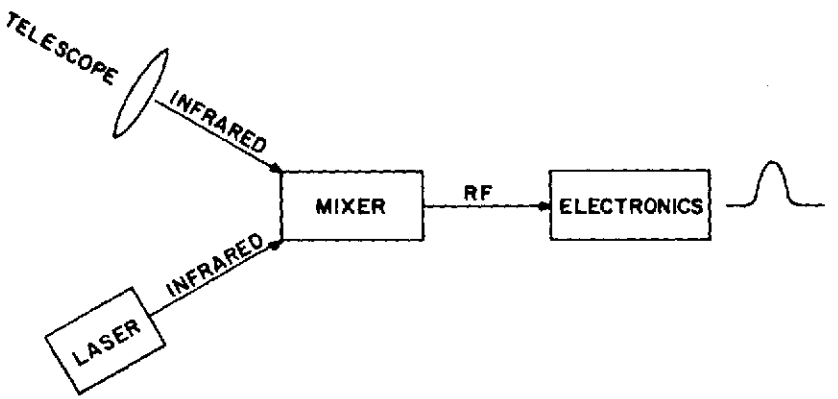


Figure 1. Infrared heterodyne spectrometer.

The second device, shown in figure 2, is an up-conversion spectrometer. It employs a very similar idea where you mix two signals—the infrared frequency from the comet and a visible frequency from a strong visible laser. The usable output is the sum frequency which is also in the visible. Visible photons can be fairly easily counted by photon counting techniques; in fact, this is one of the advantages. Counting several visible photons per second does not present much of a problem, whereas it would take many orders of magnitude of infrared photons in order to be able to count them. Table 1 will show some of the relative advantages and disadvantages of the two new devices and some typical numbers for conventional devices now available. The heterodyne spectrometer that we have operates at about 8.5 μm. Its tuning range is not very great, only about 1 part in 100, yet its spectral resolution is the highest obtainable at these wavelengths. It is able to resolve 1 part in 10^8. This fact, along with its high sensitivity, would enable us to pick out, for instance, one of the NH_3 lines at 8.5 μm that Dr. Pearl discussed. With our high resolution, we would see that it is really composed of several lines, and that would be enough to identify a molecule of ammonia. Some of the other molecules we will be after are given in the last column of the table. Those within our tuning range at 8.5 μm are formalydehyde, N_2O, and deuterated formic acid.

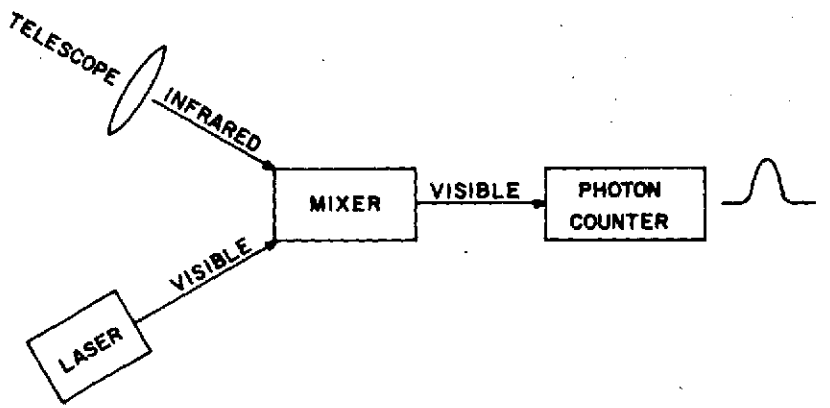


Figure 2. Infrared up-conversion spectrometer.

The up-conversion spectrometer has a much wider tuning range. Its spectral resolution, however. is much less, but its wide spectral range and its relatively high sensitivity also enable us to detect molecules such as ammonia and methane, which are abundant in comets, as well as many other possible molecules.

Table 1
State-of-the-Art Infrared Spectrometers

Device	Wavelength (μm)	Tuning Range	Spectral Resolution	Sensitivity	Target Molecules
Heterodyne Spectrometer	8.5	1/100	$1/10^8$	$1/10^{19}$	NH_3 H_2CO N_2O HCOOD
Up-conversion Spectrometer	2.2-4.3	1/2	$1/10^3$	$1/10^{16}$	NH_3 CH_4 HCN } Many
Conventional Devices	Various	Various	$1/10^5$	$1/10^{14}$	Many

The sensitivity of conventional devices cannot compare with these new methods. Their sensitivity, essentially their ideal sensitivity, is not greater than about 1 part in 10^{14}, which is at least a factor of 100 less than what we can achieve. The up-conversion device does not require cryogenic temperatures, a major hindrance in many applications, whereas infrared detectors, in general, do require such temperatures in order to be useful in detecting infrared radiation.

With the help of personnel in the Laser Data Systems Branch we are now completing the interfacing of our heterodyne system with the 30-inch telescope at the Goddard Optical Site, and weather permitting, we hope to begin measurements soon. It can be added that the usefulness of these methods is not limited to cometary investigation and they can be applied to the study of a great number of astronomical objects.

PLACE OF ORIGIN AND COMPOSITION OF COMETS

Bertram Donn

It would appear that I am going to play the role of Ambrose Bierce's speculating astronomer that Dr. Brandt referred to. What I want to do here is follow up on the cosmological role by which comets tell us something about the early stages of formation of stars or planets.

It has been recognized for a long time, and generally accepted, that comets might form in the outer regions of the solar system where the giant planets originated. For a long time, there has existed the idea that comets may have formed in the interstellar medium. The fact that interstellar clouds fragment to give rise to clusters is now well established. The sun initially may very likely have been a member of a cluster. What I want to consider here are some of the implications of these two modes of formation, solar nebula and interstellar cloud, on the composition of the comet, and whether we can tell something about where comets may have formed from their composition.

The so-called new comets, such as Kohoutek, which are making their first or at least very early approaches to the sun, are presumably relatively unchanged in composition because they have always been very cold and not affected appreciably by radiation. These comets, however, have two kinds of spectra: In one class, the dust is very strong and compares in intensity with the molecular emissions or dominates the spectrum. In the second class, the molecular emissions dominate the spectrum and the dust continuum is very weak, in some cases almost missing. One question is, how do we explain these two different characteristics in comets that presumably accumulated in the same place? Why does one type have lots of gas and the other dust? One way of looking at this, which also leads into a more definitive role of deuterim, is shown in table 1.

The first four radicals, or ions, in column one represent the four visible species in comets that give rise to the characteristic spectrum, and the last line is the dust continuum which I have taken as a sort of standard. In the second column are some of the suggested parent molecules for the observed species. In the solar nebula, where the conditions of high density would tend to give rise to the thermodynamic equilibrium, the relative abundance of these species should be low compared, for example, to

hydrogen and water, which do not affect the visible appearance of the spectrum. In the case of the interstellar medium, however, a fact that we were struck by has been the discovery by radio astronomers that these complex molecules are, in fact, very abundant. This suggests that comets formed in this solar nebula would be relatively weak in the visible emissions, compared to the comets formed in the interstellar cloud. One suggestion then is that comets having a strong dust continuum would tend to represent solar system comets, and comets with very strong molecular emissions would tend to be interstellar comets. A method of further research is to pursue this in terms of the orbits and the dynamics of the comets.

Table 1
Relationship Between the Strength of Visible Features in Cometary Spectra and the Region of Comet Formation

Visible Radical	Molecule	Strength	
		Solar Nebula	Interstellar
NH_2	NH_3	medium	strong
CN	HCN, (CH_3CN)	weak	strong
C_2, C_3	Organics	weak	strong
CO^+	CO	medium	strong
Cont.	Dust	strong	strong

Another feature of interstellar molecules, particularly HCN, is that the deuterated species, DCN, is anomalously strong. In fact, the ratio of DCN to HCN is the order of 1:100, unlike the case of molecules in the solar nebula in Jupiter, for example, where the ratio of deuterated to normal hydrogen molecules is about $1:10^5$. There is also evidence in the interstellar medium that HD forms very readily compared to H_2. We may expect that comets formed in interstellar clouds would have a high deuterium content.

One investigation of Comet Kohoutek which is currently being carried out in several ways is to try to detect deuterium. The same scheme that has studied H-α from the ground can be used to look for deuterium-α line. There is a narrow window for deuterium detection, December 17, 18, and 19. On December 17, the equipment failed just as the observers were getting ready to look for deuterium. I have not received any report this morning, as they finished at Tucson at about 9 o'clock our time. The

idea is that the presently observed, thousand rayleigh surface brightness for atomic hydrogen in the H-α line may yield 1 to 10 rayleighs for the equivalent deuterium line. The sensitivity of the Fabry-Perot spectrometer used in the observations is the order of a few rayleighs and could detect deuterium. This becomes a very critical measurement, and we should have some results on this, hopefully in the next few days.

The other way of looking for deuterium is to look for the deuterated species, OD and CD, in comparison to the normal OH and CH which are well known in comets. High resolution spectra are being taken at several places. I have not yet heard any results nor seen any analysis of the spectra.

These are two methods, starting with Comet Kohoutek and continuing with new, bright comets, by which we can seek ways of relating the composition of comets with proposed places of origin.

OVERVIEW

Frank McDonald

I would like to say that I think Goddard has a right to be particularly proud of its high energy astrophysics program. As we pointed out previously, in the 1960's we were exploring our local environment in space. To us that was sort of the earth's magnetosphere, the interplanetary media, the upper atmosphere. We are pursuing those now with more detailed programs such as IME and AE, and now the new frontiers have moved on to areas like X-rays and gamma-ray astronomy. Today we would like to give you some of the high spots from some of these programs. We have not been able to include everything from what has been a most unusual and, for us, productive year, but we will try to give sort of a broad-brush impressionistic view of this.

SAS-2 was launched a little over a year ago. Today you will hear from Carl Fichtel and others about the most exciting results that have come out of that. Another follow-on would be the opportunity to use some of the more insensitive telescopes that we know have been developed and have them ready on such missions as HEAO-C. In X-ray astronomy overall there continues to be a wealth of new developments as we start to explore the new physics of black holes and other exotic objects. We have managed to keep pace with our rocket program while preparing the HEAO-A, OSO-1, and UK-5 programs.

The third part of all of this is cosmic rays. After all, the gamma rays and the X-rays really represent the electromagnetic of nuclear interactions of charged particles, albeit at a great distance. For the past several years, we have been telling you about the ways that the composition of cosmic rays change as we go to the very high energies, the energies above 10 billion electron volts. Now, with Pioneer-10 and the IMP-7, we have been able to sample what might be considered the local galactic population of cosmic rays as we have gone down to very low energies, and we will find that the charge distribution differs very remarkably from what one observes at high energies. There has been an unusual amount of attention devoted to the very exciting Pioneer-10 encounter with Jupiter—What people tend to overlook is that often in many things, half the fun is getting there, and for us the interplanetary region has been a very exciting part of the whole mission. One of the nice things that sometimes happens, but not in the Federal Government, is that one gets

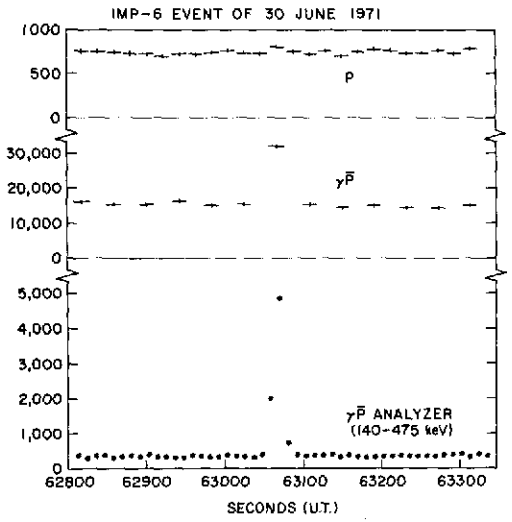
unexpected bonuses. For us, this year, this has come in the form of a gamma-ray burst that Tom Cline and Upendra Desai will tell you about. As any Madison Avenue type knows, when you have got a good product, it is much better to let it speak for itself and so I will call for the first talk.

COSMIC GAMMA-RAY BURSTS: NASA OBSERVATIONS AND THEIR IMPLICATIONS

Thomas L. Cline

A new celestial phenomenon was recently discovered with the Vela bomb-test monitoring program, when several satellites occasionally recorded brief but intense simultaneous increases in the interplanetary 100-keV to 1-MeV background. The energy region monitored by the Los Alamos group is appropriate to thermonuclear reaction products, and the rise times of the observed bursts are at least as short as the detector resolution permits—1/64 s. However, the source directions, determined by triangulation of the individual onset times of the bursts at the satellite locations, are found to be generally far away from the earth or sun, and off both the ecliptic and galactic planes. The rate of appearance of the bursts averages four times per year at presently resolved intensity levels, but so far neither the occurrence times nor the approximate celestial directions of the bursts have been correlated with any known astronomical phenomenon.

NASA observations of several of the same bursts, from Goddard detectors on IMPs and OGOs, immediately provided the first independent confirmation of the existence of these bursts. More to the point, our results established the hard X-ray or soft gamma-ray nature of the incident flux, which could only be inferred from the Vela observations. Figure 1 shows the response of our semi-omnidirectional IMP-6 CsI crystal to a burst observed in 1971. The detector element used consisted of a 25-cm^2 by 4-cm thick crystal, entirely surrounded by a thin plastic anticoincidence scintillator for particle rejection, and viewed by one PM tube. During a 5-s interval, the counting rate in the plastic scintillator, P, increased by roughly 10 percent. Simultaneously, the neutral counts in the crystal, labeled as $\gamma\,\overline{P}$, increased to about double. The gamma-ray counts in the 140- to 475-keV interval, as determined by an onboard differential analyzer, surged from about 400 to over 5000, representing more than an order of magnitude increase from the total omnidirectional and secondary background. This response shows that the behavior of the incident flux is entirely consistent with that of hard X-rays or soft gamma-rays. Second, it illustrates the remarkable intensity of the bursts: a response of this magnitude is



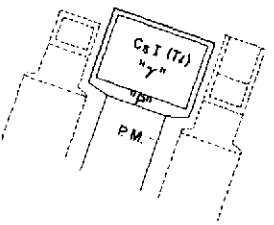


Figure 1. Response of semi-omnidirectional IMP-6 CsI crystal to a June 30, 1971, event.

comparable to that for a typical large solar flare. Yet, unless something very strange is going on in the solar system, the source of this radiation is most likely at either typical stellar or galactic distances, that is, at over 10^6 or 10^{12} times as far away as the sun, thereby yielding an instantaneous power, radiated in this energy band, of 10^{12} or 10^{24} times as much as from a flare. Another feature of these bursts is illustrated in figure 2, which compares our data with Vela results for one event. In two cases our detector, turned off for half the time, luckily caught each of the pulses in these double-pulsed events. This also shows the brevity of the bursts: Each pulse has a width of barely 2 s. The temporal pulse complexity, actually seen in all six IMP-6 events, is a curious phenomenon that, of course, should come out of any useful model. The energy spectra we obtained will be discussed by Dr. Desai.

Our observations also confirmed the directional properties of the bursts observed with the Vela satellites. That is, in spite of our relatively rough time resolution, the burst profiles could be intercompared in order to obtain self-consistent burst onset times; these IMP onset times, in turn, fitted the wave front propagation pattern across the Vela and IMP orbits with consistent relative delays.

The frequency of occurrence of these events, together with the flux and energy information obtained in the IMP results, also make possible an

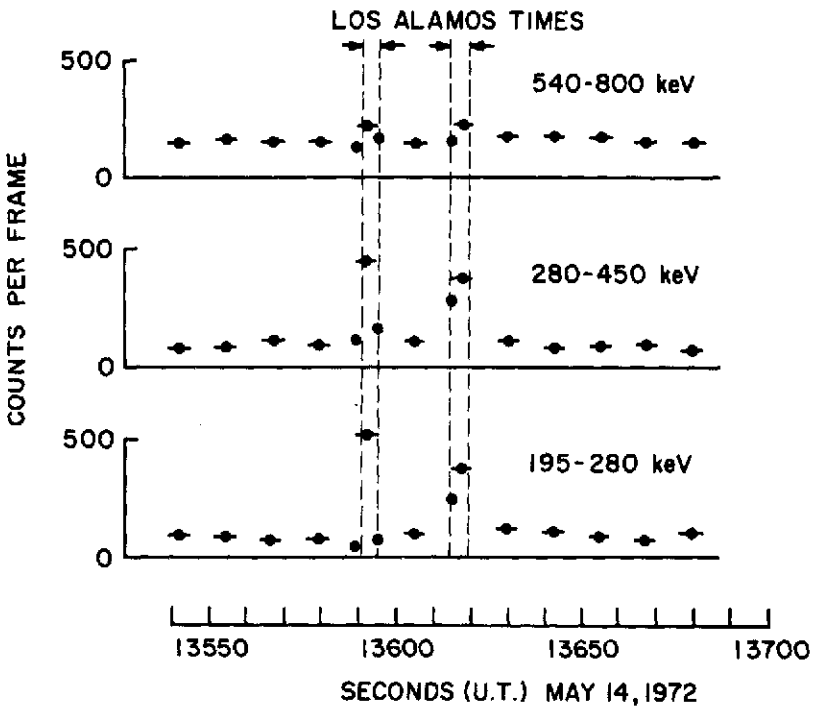


Figure 2. Comparison of NASA satellite solar flare data with that obtained from Vela.

estimate of the observed size spectrum of gamma-ray bursts. Figure 3 shows the integral size histogram of the IMP-6 events observed in 1971 and 1972. If we arbitrarily normalize this to a distance scale appropriate to the nearby galactic clusters, the frequency of occurrence of the bursts is found to be essentially the same as that expected for metagalactic supernovae. Also, the theoretically predicted emission of $10^{40\text{-}41}$ J ($10^{47\text{-}48}$ ergs) would give rise to the same energy flux as that which is observed, namely 10^{-12} to 10^{-11} J/cm^2 (10^{-5} to 10^{-4} ergs/cm^2) above 100 keV. This may be coincidental, since there are no correlations of the burst times with known supernovae. The bursts may therefore represent quite different metagalactic phenomena, or instead $\approx 10^{31}$ J (10^{38} erg) processes in nearby stellar regions, or even $\approx 10^{47}$ J (10^{54} erg) processes at cosmological distances. Dr. Ramaty will outline one possible source model.

Studies of the size spectrum may be able to provide clues as to the spatial distribution of sources, if not as to their distances. Since the IMP sensitivity is about an order of magnitude greater than that of Vela, we should be able

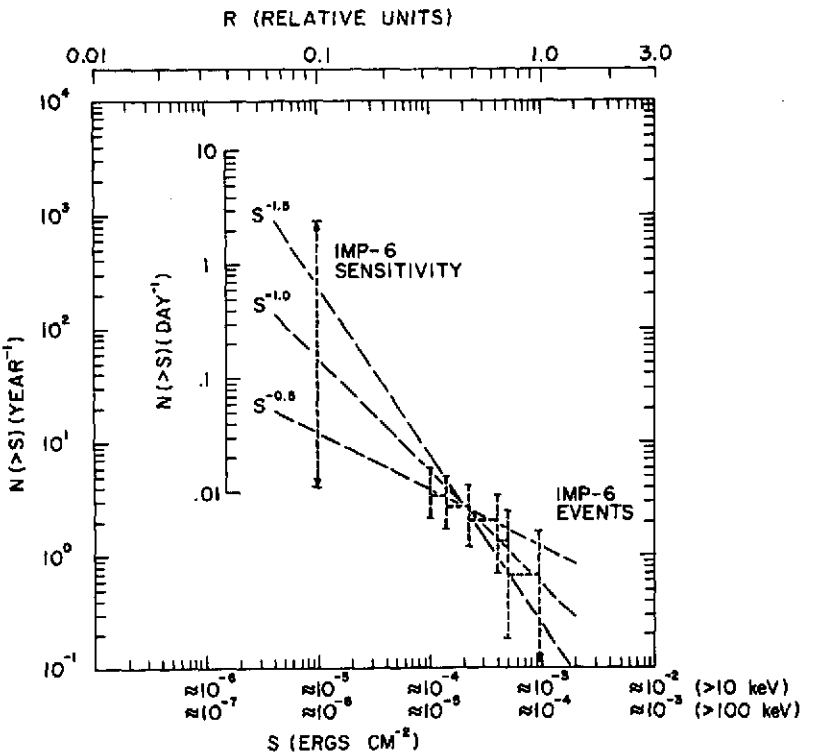


Figure 3. Integral size histogram of the IMP-6 events observed in 1971 and 1972.

to extend this distribution to the smaller size region, and we have, in fact, compiled a list of several hundred smaller increases that can be considered to be candidate events. If other similarly sensitive detectors in orbit can provide coincidences with our candidate events, it may be possible to infer a relatively well determined slope to the size spectrum. However, second-generation satellite experiments will ultimately be needed to describe the detailed variations in the size spectrum that reflect the shape of the emitting regions, the features of the instantaneous, differential energy spectra, and the very fast time variations. Finally, very accurate directional resolution will be required to identify the sources of these celestial gamma-ray bursts, and this may be possible only by using deep space probes in a long baseline array.

THE UNIQUENESS OF THE ENERGY SPECTRUM OF COSMIC GAMMA-RAY BURSTS, AS OBSERVED BY IMP-6

Upendra D. Desai

The observations of cosmic gamma-ray bursts with IMP-6 have contributed significantly to the study of this new and exciting phenomenon in various ways. In addition to unambiguously identifying these bursts as due to photons, the IMP-6 results confirmed some of the interesting features, such as double pulsation, and for the first time, provided differential energy spectra in the 100- to 1200-keV range.

The energy spectra of the incoming photons were accumulated in a 14-channel differential pulse height analyzer. The spectral data were collected over half the spin of the satellite, from sunrise to sunset of the detector, yielding 50-percent live-time coverage. This storage time of about 6.3 s was further subdivided by the telemetry clock every 5.1 s into two or three spectral observation periods. The gain of the system was cycled for in-flight calibration, and so some of the events were analyzed over the 50- to 880-keV range while some were in the 70- to 1150-keV range.

IMP-6 detected six of the eight events reported by Vela satellites during the overlapping period of observations. Each of the six events has at least two distinct pulses of up to a few seconds duration. Figure 1 shows the time history of the May 14, 1972, event in which this feature is clearly revealed. Since IMP-6 accumulated data for several seconds at a time, and since the peaks of the bursts last over a second or so, the spectral information was accumulated over the entire time duration of each burst, providing average energy spectra. Spectral information during one or both pulses was available during some events, while, in some cases, only the decay phase was monitored. Figure 2 shows the photon number spectra for several events when the sources are (a) well within the field of view, and (b) near the horizon. The third group illustrates a spectral measurement during the decay phase, rather than during pulse activity of the event. The E_o values for pulse maximum intensity cluster at 150 keV, for (a), and for (b), around 250 keV. During (c), the decay phase, E_o is about 70 keV. The directions of origin of the six events are known with varying accuracy, but in the case of the June 30 event, it is known that the first spectrum was recorded when the source was below the detector horizon. Thus, the

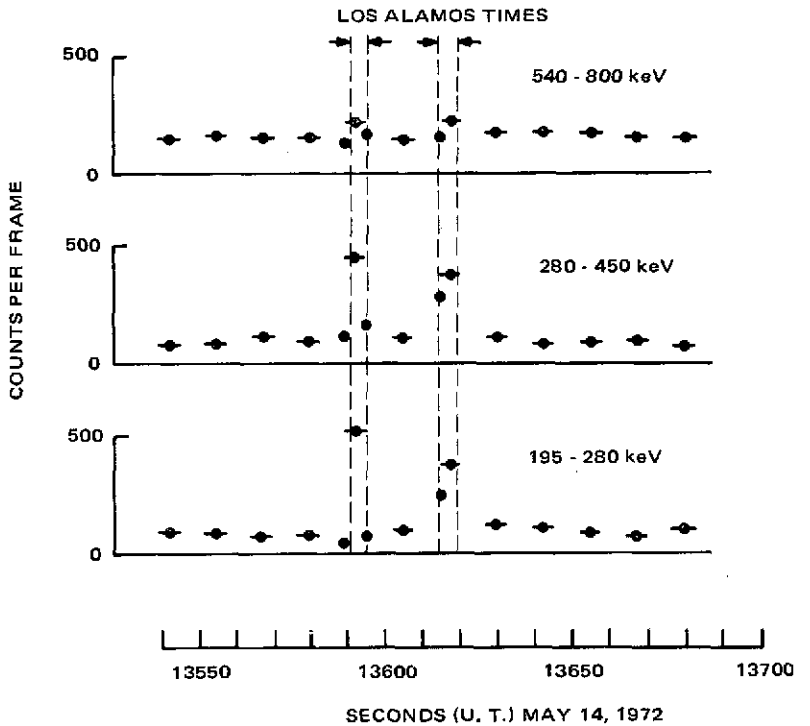


Figure 1. Time history of the May 14, 1972, solar flare event.

harder spectrum, E_o = 250 keV, may be accounted for by attenuation of low energy photons in the satellite skin. If that is also the case for the March 29, 1972, event, then all eight or nine spectra observed during the pulse maxima are consistent with 150-keV spectra. Two of the six events, March 15 and May 14, 1972, have unambiguously known source directions which are not far from the center of the field of view, and these are 150-keV spectra. In addition, a portion of the March 18, 1971, event and the decays of other events are consistent with softer spectra, suggesting the presence of a soft component with a slow time constant.

Figure 3 shows E(dN/dE) against E_γ for two events where each source is in the field of view. These are energy spectra in which E is multiplied by dN/dE, the number spectra shown previously. We compare low-energy results from OSO-7 with high-energy results of IMP-6 for the May 14, 1972, event, and OGO-5 with IMP-6 for the March 15 event. In both cases the energy at maximum power is emitted around 150 keV. This may indicate

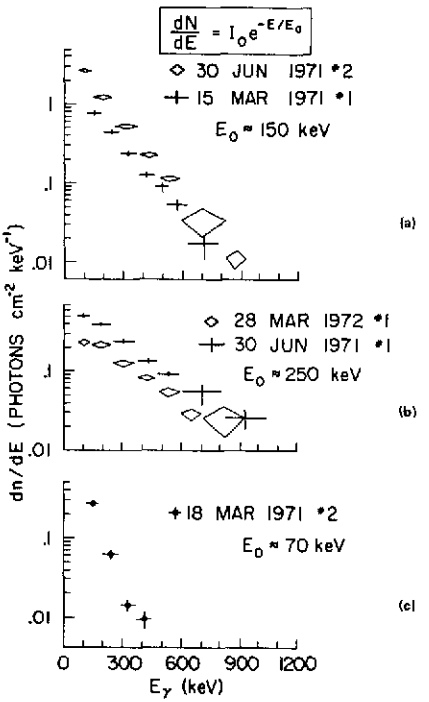


Figure 2. Photon number spectra for several events. (a) Source in field of view; (b) Source on horizon; (c) Decay phase.

that the photons released from the source are essentially gamma ray in nature and not X-ray distributions with spectral tails in the gamma-ray region.

It is clear that the observed gamma-ray bursts represent an entirely novel mode of energy release. The duration of the individual pulses during a burst are typically only a few seconds with separation between them of 6 s or 24 s. Because of the similarity of such temporal structures, as well as the constancy of the spectral characteristic, a uniform source model, such as a collapsing object, would be suggested as responsible for the bursts. The 6-s pulse separation could also be a characteristic vibrational period of such objects. On the other hand, the temporal structure could instead be compared to that of solar flares, but with time scales one or two orders of magnitude shorter, suggesting a conceivable gamma-ray

flare star origin. However, one would expect diversity in spectral characteristics with such a model, which is not the observational fact. The good fit to an exponential number spectrum and the constancy of E_o are the unique features of this phenomenon.

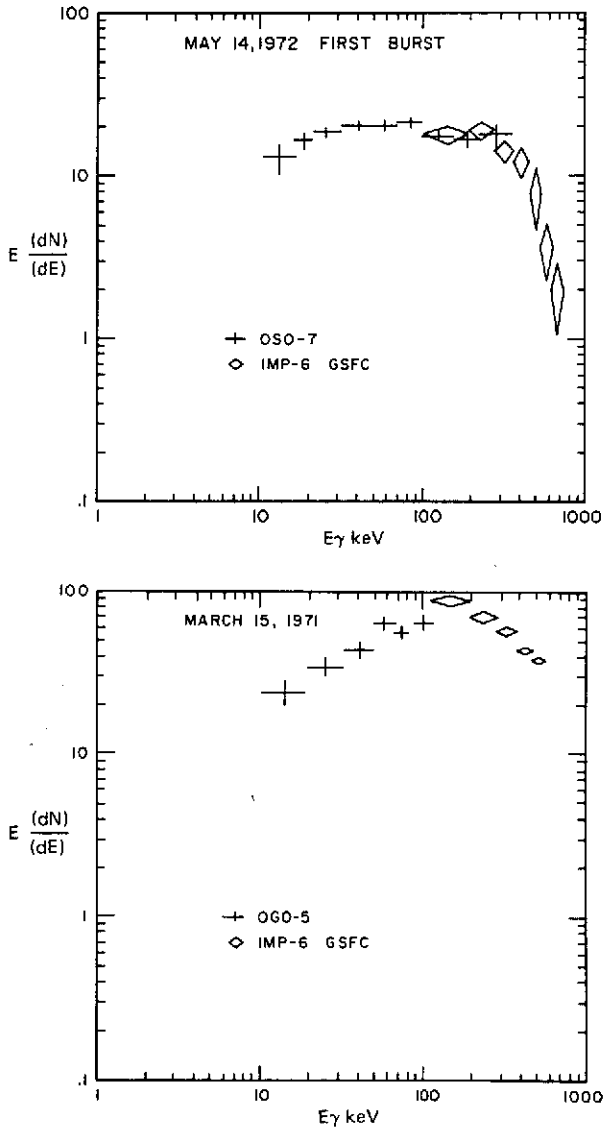


Figure 3. Energy spectra for events where each source is in the field of view.

COOLING BLACKBODY, A MECHANISM FOR COSMIC GAMMA-RAY BURSTS

Reuven Ramaty

I would like first to make one point which occurred to me during Dr. Brandt's talk on comets. In a sense, the study of gamma-ray bursts resembles cometary studies before Halley's time, when a multitude of theories existed to account for the phenomenon of comets. At present, there are a multitude of theories for the gamma-ray bursts as well. They range all the way from the very close in to the very far out, and the crucial measurement and interpretation have yet to come. In addition to the theory that I am going to describe here, there are other theories for gamma-ray bursts. These include cometary collisions with neutron stars, relativistic grains impinging upon the solar system, and stellar super flares. Another model, which was proposed before the discovery of the gamma-ray bursts, is Colgate's supernova model. His predictions, however, are generally not consistent with the present observations.

The model that I am going to discuss has to do with blackbody radiation. The reason that we have considered this mechanism is that for objects of a given size and temperature, blackbody radiation is the most efficient noncoherent mechanism, a property of great importance for the understanding of the large intensities of the bursts.

Equation 1 gives the photon spectrum of a spherical object of radius r, at the distance d, radiating as a blackbody at constant temperature T.

$$\psi(E) = \frac{2\pi}{c^2 h^3}\left(\frac{r}{d}\right)^2 \frac{E^2}{\exp\,[E/kT-1]} \quad \frac{\text{photons}}{\text{cm}^2\ \text{s keV}} \qquad (1)$$

This, of course, is the well known Planck function, given in units of photons per keV, rather than energy per units frequency. The spectrum shown in the first equation, cannot, however, be directly compared with observed spectra of the bursts, because the spectral measurements are made over a time interval that is of the same order as the duration of the bursts, and the temperature is expected to vary over this interval. Therefore, it is necessary to integrate the instantaneous emission given in the first equation over time.

This is done in equation 2, where a time derivative of T is the rate of change of the temperature with time. We have no detailed knowledge of this variation.

$$\Phi(E) = \int_{T_0}^{T_1} \psi(E) \frac{dt}{dT} dT \quad \frac{\text{photons}}{\text{cm}^2 \text{ keV}} \tag{2}$$

We can, however, assume that the object cools mainly by blackbody emission. Then the above derivative varies as the fourth power of T, and the integral in the second equation can be evaluated in detail. The results are shown in figure 1, together with observed spectrum of the May 14, 1972, gamma-ray bursts which has already been mentioned in the previous two talks. As can be seen, there is reasonable agreement between theory and observation. We estimate that the initial temperature of the burst, kT_0, is about 150 keV and that it cools during the duration of the burst to temperature kT_1 of a few keV.

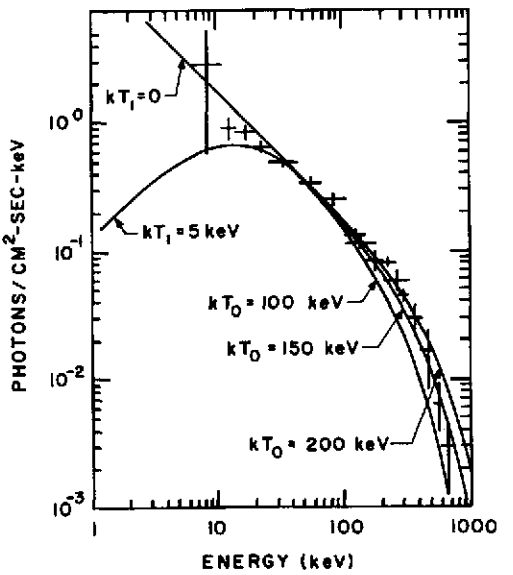


Figure 1. Rate of change of temperature with time.

If the gamma-ray bursts are indeed blackbody emission, we can estimate their distance and luminosity. Some representative numbers are given below.

Distance and Luminosity Estimate

$$L = 4\pi r^2\ \sigma T^4 = 4\pi d^2 \psi$$

for $kT = 150$ keV and $\psi = 5 \times 10^{-5}$ ergs $cm^{-2}\ s^{-1}$

$$d\ (\text{Mpc}) \cong r(10^6\ \text{cm})$$

$$L\ (\text{ergs}\ s^{-1}) \cong 6 \times 10^{45}\ r^2\ (10^6\ \text{cm})$$

The luminosity of the source is L, r is its radius, σ is the Stephan-Boltzman constant, T is the temperature, d is the distance to the source, and ψ is observed energy flux of the burst. By substituting typical values, it is shown that the distance to the source of the burst, measured in megaparsecs, approximately equals the radius of the source in units of 10 km. Because the smallest size of known stellar objects is about 10 km (this could be neutron stars or black holes), the distance has to be about 1 Mpc or larger. This means that the bursts are extragalactic because a distance of 1 Mpc exceeds typical galactic dimensions of kiloparsecs. In fact, we believe that the gamma-ray bursts are due to the formation of neutron stars. These formations, however, should not be accompanied by optical supernovae because no supernovae event was found to correlate with any other gamma-ray burst. We propose that the neutron stars are formed in binary systems where a white dwarf accretes matter from a companion star. When the mass of the white dwarf exceeds the Chandrashekhar limit, at which an instability sets in, it collapses. The collapse releases the binding energy of the newly formed star which is about 10^{53} ergs. Most of this energy is carried away by neutrinos, but a small fraction could be emitted as an electromagnetic blackbody emission at high temperature. By comparing 10^{53} ergs with the required gamma-ray luminosities, we see that less than about a tenth of one percent of the total energy has to be in the form of prompt gamma rays. This does not appear to place a significant strain on the model.

The proposed model can also account for some of the observed time of variations, and this was mentioned in the previous talk by Dr. Desai. The rebound of the neutron star could lead to the observed secondary peaks of the bursts which are separated by intervals of a few seconds. Variations on shorter time scales, on the order of tenths of seconds, could be due to vibrations of the star during rebound. Because the radius increases with time during the rebound, the vibration period should also increase with time. This seems to be consistent with the temporal structure of at least part of the bursts.

In conclusion, the present model seems to be capable of accounting for most of the observed features of gamma-ray bursts. The main observational tests of the model will be the observation of a blackbody spectrum by spectral measurements with high temporal resolution. You recall that the reason we do not see a blackbody spectrum with the present measurement is that these are made over time intervals too far apart, and the temperature of the source varies very rapidly. More detailed observations of time variations are also required. Finally, a neutrino burst should accompany the gamma rays. Its detection, however, is beyond the capabilities of present experiments.

HIGH ENERGY DIFFUSE GAMMA RAYS: EXPERIMENTAL RESULTS AND THEORETICAL INTERPRETATION

David Thompson

As Dr. McDonald just said, the second small astronomy satellite, SAS-2, was launched just over 1 year ago. During its lifetime it produced a large quantity of data on cosmic gamma rays with energies above 30 MeV. In this and the next two talks, we will be presenting some preliminary results from SAS-2. I emphasize the word preliminary for two reasons: First of all, we have by no means analyzed all the data from the satellite; and second, we have not yet completed analysis of the experimental detector calibration. The we in this case includes Drs. Fichtel, Kniffen, Hartman, Bignami, and myself, all of whom have been involved in various aspects of the work on SAS-2.

Figure 1 shows a schematic diagram of the detector system. Basically, it is a 32-deck, magnetic core, digitized spark chamber which is triggered by a gamma ray entering through its opening angle and undergoing pair production. From the paths of the created positron and negatron, we can determine the arrival direction of the gamma ray within a few degrees and estimate the gamma-ray energy. Combined with the detector efficiency and the exposure time, this data gives the gamma-ray flux measurement. For the portion of the sky which we have examined thus far, there is a diffuse flux of high energy gamma rays which seems to be uniform for regions away from the galactic plane. Three such regions have been analyzed. Within the experimental uncertainties, the gamma-ray intensity is the same for all three. Data from these areas has been combined into the diffuse energy spectrum shown in figure 2.

The diffuse spectrum measured by SAS-2, as seen in figure 2, is quite steep. By way of contrast, this figure also shows the atmospheric secondary gamma-ray spectrum measured by one of our balloon-borne detectors, near the top of the atmosphere. I think this figure makes it abundantly clear why a satellite such as SAS-2 was necessary in order to measure the diffuse flux. In terms of both flux and energy spectrum, the atmospheric secondaries would completely dominate any attempt to measure the diffuse flux from balloons. As I said a minute ago, the diffuse energy spectrum seen by SAS-2 is steep. In fact, it is the steepest spectrum observed by the detector and it is steeper than any spectra observed by the balloon work in our group.

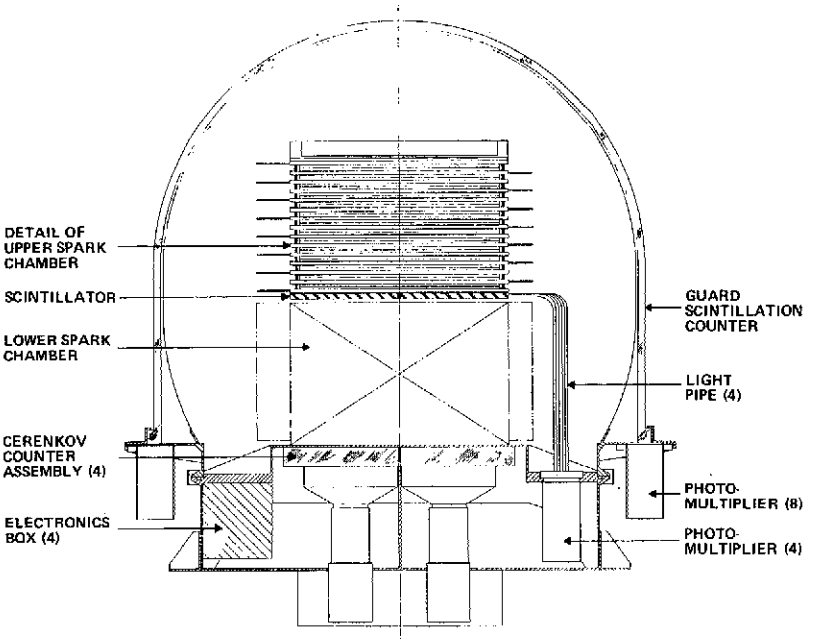


Figure 1. SAS-2 gamma-ray experiment.

Figure 3 shows the SAS-2 data again, along with experimental data of lower energies. The uncertainty shown by the shading for the SAS-2 data is due primarily to the incomplete calibration, and much of this uncertainty will be resolved in the near future. Taken together, these measurements show a complex spectrum for the diffuse gamma rays, with an apparent enhancement in the 1- to 10-MeV energy range, relative to a power law extension from the X-ray region, and then a rapid decrease in intensity in the energy range above 30 MeV.

Now that we have these measurements, the next question is where do these gamma rays come from? I would like to discuss briefly several possibilities, some of which have been stimulated by the SAS-2 results. One of these is that gamma rays come from diffuse electrons interacting with matter, photons, or magnetic fields. There seem to be some problems with this approach in the energy range above 1 MeV, mainly in trying to reproduce the shape of the gamma-ray spectrum and the total intensity without requiring substantial changes in many estimates of the intergalactic parameters.

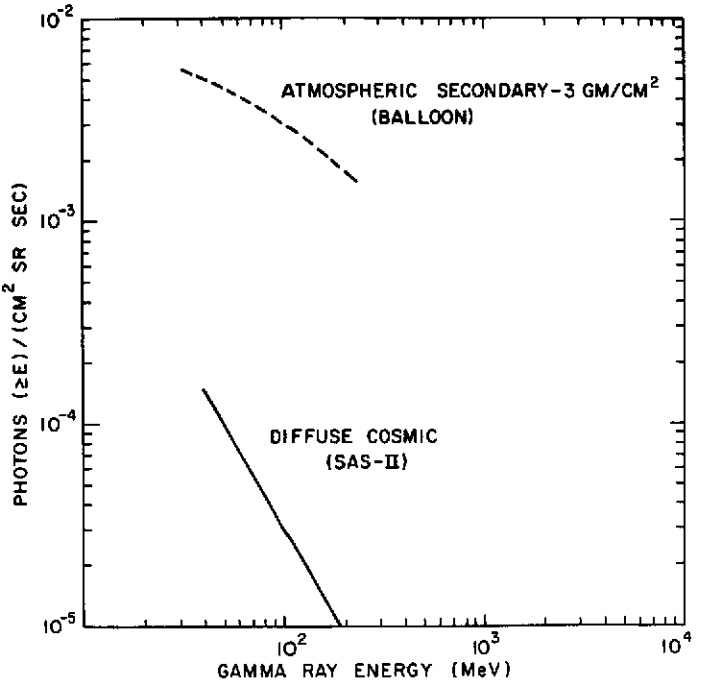


Figure 2. Diffuse energy spectrum measured by SAS-2.

Two other possible origins of the diffuse gamma rays are cosmological in nature. Dr. Stecker from Goddard has been involved with both models. For both of these, at an early point in cosmological time, large quantities of neutral pi mesons were produced, either by collisions or by matter-anti-matter annihilation. The decay of these mesons produced high energy gamma rays. By the time they reach the earth, the gamma ray energies have been shifted downward by the cosmological redshift caused by the expansion of the universe. Both models predict approximately the same energy spectrum for gamma rays. This is shown by the dashed line in figure 3 and is seen to be a reasonably good fit to the experimental data. The agreement of these models with the data makes them particularly attractive. In turn, having the possibility of a direct measurement of a cosmological effect emphasizes again the importance of these diffuse gamma-ray measurements.

Another model has recently been proposed by a group including Dr. Ramaty. In this model, neutral pions are produced by matter falling onto the surface of neutron stars. The energies of the resulting gamma rays are gravitationally redshifted to lower energies. This model also predicts a spectrum consistent with the preliminary SAS-2 observations.

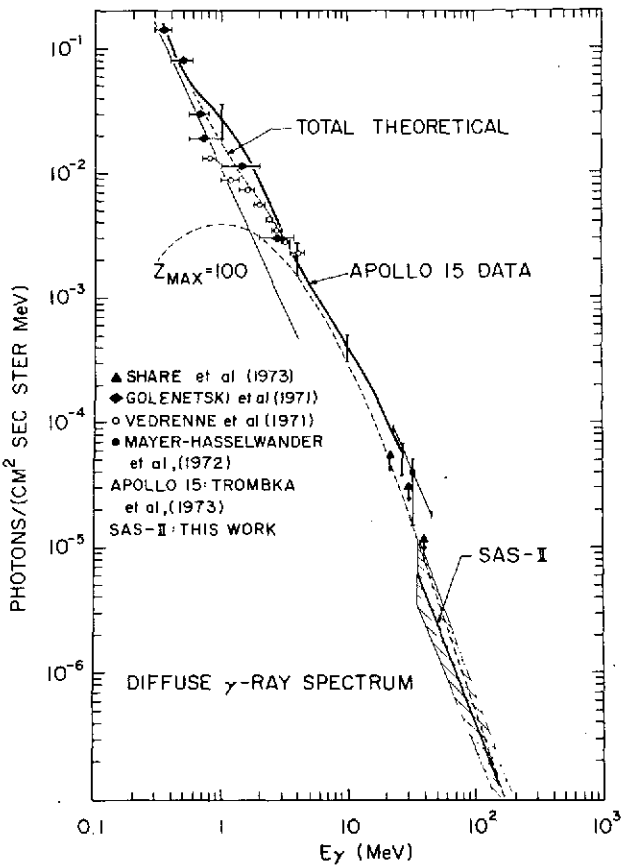


Figure 3. Diffuse energy spectrum from SAS-2 and experimental data of lower energies.

In conclusion then, several candidate models exist to explain the diffuse cosmic gamma-ray spectrum and its uniquely steep spectrum above 30 MeV. As more of the SAS-2 data is analyzed and refined, it should be possible to tell just how well these models fit the data and perhaps to distinguish between the various models.

SAS-2 HIGH ENERGY GAMMA-RAY OBSERVATIONS OF GALACTIC STRUCTURE AND THE CRAB NEBULA

Donald Kniffen

It has long been realized that high energy gamma rays would be expected from the galactic plane if cosmic rays exist throughout the galaxy at the intensity they are observed to have in the near vicinity of the earth. Kraushaar, Clark, Garmire, and their colleagues, with a gamma-ray experiment flown on OSO-3, observed an enhanced radiation of gamma rays above 50 MeV coming from the galactic plane, whose angular width in latitude was less than its detector angular resolution, about 35° full-width-half-max. One of the studies of most immediate interest with SAS-2 was the radiation from the galactic plane with the larger sensitivity and the improved angular and energy resolution possible with SAS. Although the attitude information for the orientation of the spacecraft is still uncertain, I will present some preliminary results that have been obtained for two regions of the galactic plane.

One is a 90° interval in galactic longitude in the region of the galactic center, and the other, from 160° to 230° in longitude, directly away from the galactic center. This provides an example of the galactic plane emission in the direction of the center and a region well away from the center. An enhanced flux of high energy (>30 MeV) gamma rays has been observed from both these regions, relative to the diffuse background that was presented by Dr. Thompson. The galactic radiation in the region within 40° of longitude from the galactic center region is particularly intense.

Figure 1 shows the intensity of gamma rays above 100 MeV. The observed intensity is summed from galactic latitudes from $-10^\circ \leq b_{II} \leq 10^\circ$ and plotted as a function of longitude. The enhanced radiation extends from about 320° through the galactic center region to 40° of longitude. The important features are that the intensity remains uniformly high through most of this region, and particularly, that there is no further enhancement at the galactic center. This will be important in some of the models that will be discussed in the next two papers.

Summing the galactic center radiation above 100 MeV into bins of 2.5° width in galactic latitude, the distribution in figure 2 is obtained. The one σ half-width of this distribution obtained from SAS-2 is 4.5° in this

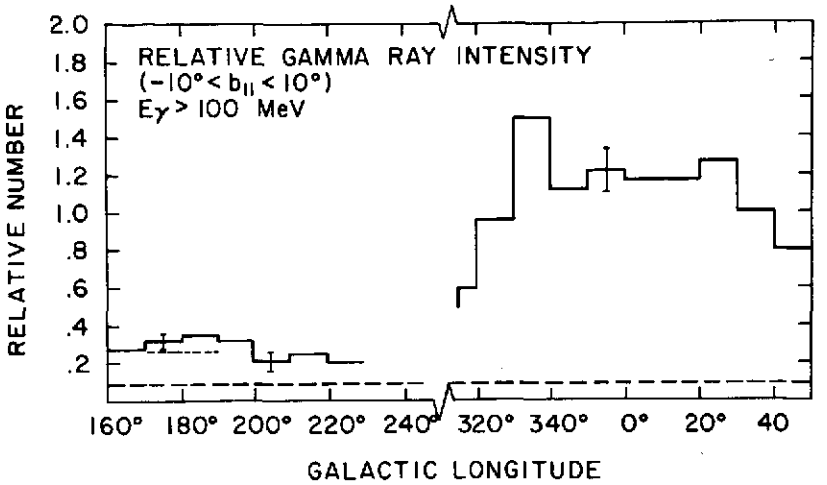


Figure 1. Intensity of gamma rays above 100 MeV.

particular graph. In more recent data the resolution of the experiment has been improved and this width is correspondingly reduced. However, the remaining uncertainties in the knowledge of the pointing direction and the known accuracy for determining the arrival direction of individual gammas are still significant factors, even in the new distributions. When all of these effects have been removed there is still a finite width in galactic latitude to the emission that is observed.

The energy spectrum for the galactic center gamma radiation is shown in figure 3. The spectrum is quite flat, especially as compared to the very steep spectrum obtained in the diffuse radiation as shown by Dr. Thompson. Here the spectrum, including the dashed line, is the total emission that we have observed. The solid curve is obtained if the diffuse background is subtracted from the total observed galactic center radiation under the assumption that this diffuse background is universal. The energy flux above 100 MeV is 1.15×10^{-4} photons/cm^2 rad s, where the error includes uncertainties due to the fact that the analysis of the calibration data is not yet complete.

Figure 1 shows that, in the direction away from the galactic center, the intensity of the emission is much lower, specifically, in the 160° to 220° longitude region. The contribution of the diffuse radiation, if it is universal, is denoted by the lower dashed line, and the upper dashed line refers to the total flux, if the contribution that we have observed from the region of the Crab Nebula is subtracted. The net intensity in this region, if the

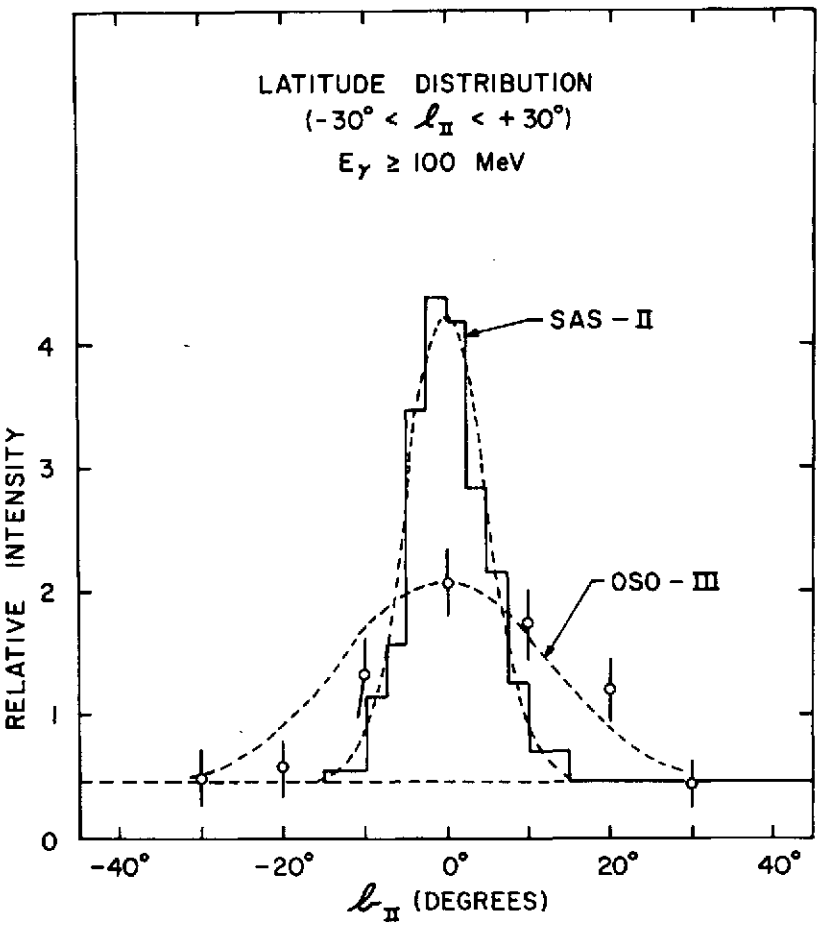


Figure 2. Summation of the galactic center radiation above 100 MeV.

diffuse radiation is subtracted, is about 0.2×10^{-4} photons/cm^2 rad s above 100 MeV. Particularly significant from this figure is the very high ratio between the galactic center region and the anticenter region. Under the assumption that the cosmic rays pervade the entire galaxy at the intensity level they are observed to have locally, there will be high energy gamma rays with a distinctive energy spectrum produced by the interaction of the cosmic rays with the interstellar matter. The gamma rays arise principally from the decay of π^0-mesons and as a result, the differential spectrum has a maximum at 70 MeV and extends to very high energies. Calculations made of the expected galactic intensity indicates that it should be between 2

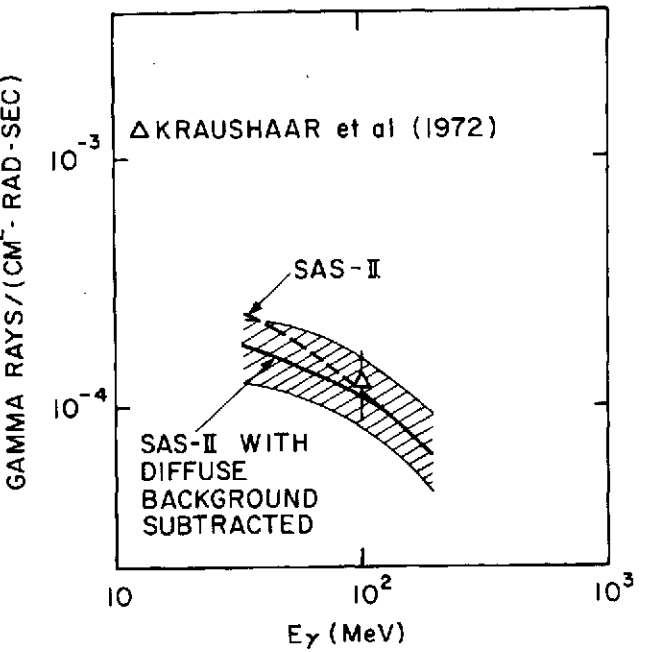


Figure 3. Energy spectrum for the galactic center gamma ray radiation.

and 4×10^{-5} photons/cm^2 rad s above 100 MeV along the galactic plane. The SAS-2 data that we present here indicate an average experimental value just under 2×10^{-5} for the region from 160° to 230° in galactic longitude. This is not surprising since this is the region in the anticenter direction where the intensity might be expected to be the least. Since the intensity in the galactic center region is about three to five times the expected value, there must be another component which contributes to this radiation. It must have the following characteristics: a reasonably hard energy spectrum, a very flat longitude distribution from −30° to +30° in galactic longitude, and a very narrow latitude distribution within ±3° of the galactic plane.

The next two papers will focus on possible origins of this enhanced radiation.

THEORETICAL INTERPRETATION OF THE SAS-2 HIGH ENERGY GAMMA-RAY OBSERVATIONS OF THE GALACTIC PLANE IN TERMS OF THE GALACTIC COSMIC-RAY AND MATTER DISTRIBUTIONS

Carl Fichtel

As Don Kniffen has just discussed, the galactic gamma-ray emission from the central region of our galaxy has turned out not only to be three or four times the expected level, but to have a distribution which was surprising in that it was quite flat over a 60° to 80° interval in galactic longitude. In pursuing this problem theoretically, it is important, first of all, to realize that the angular resolution of the SAS-2 detector is of the order of a few degrees, and therefore, the gamma radiation observed by this detector from any feature with a thickness comparable to the disk of the galaxy, will fall off inversely with distance once it is more than about 2 kiloparsecs away, and faster if it is also small in extent within the plane. Hence, more distant regions of the galaxy would have to be substantially more intense than local ones to explain an observed intensity of gamma-rays in any given direction. This consideration, together with the remarkable geometrical distribution of the gamma-rays, (that is, the very flat distribution over about 60° centered on the galactic center) has suggested to those of us involved in the reduction of the gamma-ray data (namely, G. Bignami, R. Hartman, D. Kniffen, D. Thompson, and myself), that the source of the enhancement is possibly predominantly diffuse radiation from the spiral arm segments closest to the sun in the direction of the galactic center.

In pursuing this further, I wish first to give the reasons for proposing an enhanced gamma radiation in the arm segments due to the interaction of cosmic rays with the matter in the galactic arms. Second, I want to discuss the specific arm structure. Finally, I will show the distribution of gamma radiation predicted by this model.

With regard to the enhanced cosmic radiation in the arm segments, the assumption is made here that the cosmic rays and magnetic fields are galactic and not universal, or at least that they are not confined to the galaxy by extragalactic pressures. Then, as shown by Bierman and Davis (1960) and Parker (1966), the magnetic fields must be contained by the weight of the gas through which the magnetic fields penetrate, and hence, the

magnetic fields are tied to the matter. The general picture that emerges is then one in which the magnetic field lines have their greatest density in the regions of the highest matter density, and tend to diverge in less dense regions. The galactic cosmic rays are inevitably tied to the magnetic fields and, indeed, their energy density cannot substantially exceed that of the magnetic fields, or the cosmic-ray pressure will cause a bulge in the magnetic fields which will allow the cosmic rays to escape. The local energy density of the cosmic rays is about equal to that estimated for the energy density of the average magnetic fields. This feature together with other considerations strongly suggests that the magnetic fields are nearly saturated with cosmic rays and that the cosmic-ray density may generally approach the limit the magnetic fields can contain. As a working hypothesis, it will be assumed that the energy density of the cosmic rays is essentially at its saturation value and, therefore, a reasonable specific trial hypothesis is that the cosmic-ray density is proportional to the matter density. If this picture is correct, the fluctuations in matter density, N_H, become quite important in determining the expected gamma-ray intensity in any direction since the gamma radiation becomes proportional to $N^2{}_H$. This hypothesis is applied, and indeed is most relevant on the scale of galactic arms. As gamma-ray astronomy improves in angular resolution, it can also be tested on the scale of clouds.

Let us now turn to the picture of the galaxy that has emerged thus far from radio data, and particularly the 21-cm data. We shall assume (and there is some support for this assumption) that the ionized and molecular hydrogen are similarly distributed or at least are not dense enough elsewhere to alter this general picture.

The matter is tied preferentially to the spiral arm segments and figure 1 shows one of the pictures currently available from the radio data giving the distribution of the principal arm segments (Simonson, C., 1973, "A Density-Wave Map of the Galactic Spiral Structure," University of Maryland preprint, 1973).

There are some difficulties associated with the direct interpretation of the 21-cm data, and here we rely on radio astronomers such as Simonson and Kerr at Maryland. First, there is clearly significant absorption of the 21-cm line over a band in galactic longitude of the order of 30° or more about the galactic center, and there are also indications of high optical depth along spiral arm segments. Second, the interpretation of the observed intensity in the 21-cm line in terms of density depends on the assumed galactic velocity field, and there is increasing reason to believe the velocity pattern is not as simple as assumed in the earliest models. It is actually this

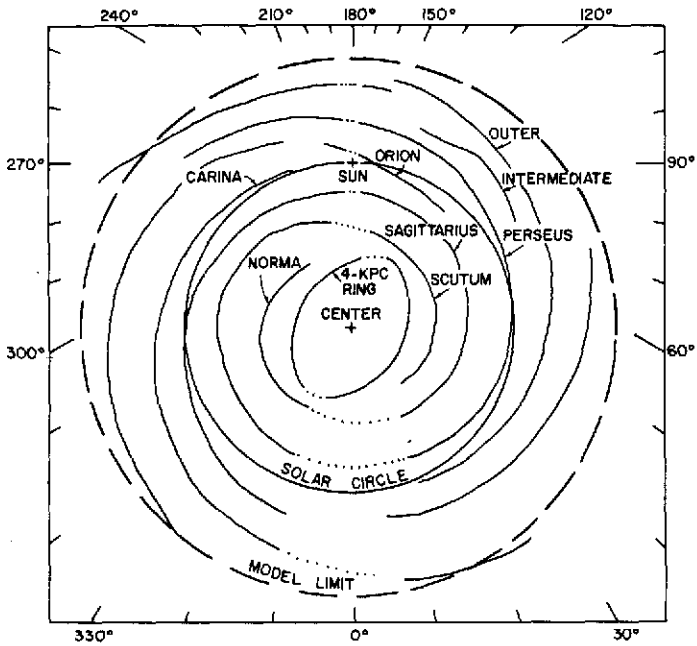


Figure 1. A smoothed spatial diagram of the locations of the maxima of the matter density deduced from 21-cm HI line measurements, and the density-wave theory by Simonson [21].

latter problem which is of greater concern here because it affects the peak-to-valley ratio of the matter density distribution. It now seems plausible, based partially on observations of other galaxies, to assume this ratio might be 4 or 5 to 1.

Notice particularly that there is a strong arm segment around 30° or just inside 30°; also, at 330°, one views along an arm segment. The Sagittarius arm also will make a strong contribution, and there is an enhanced region in the center. There are also arm segments as one views toward 80° and toward 260° to 270°.

If we assume the model discussed above, then one gets the results shown in figure 2 for the gamma-ray distribution. The Sagittarius arm contribution is seen to be quite large. It is more flat because it is very close and therefore, produces a relatively high intensity in the central position. The other arms produce greater peaks as one looks along the arm segment. Lastly, one has the inner two arms, and the shaded region

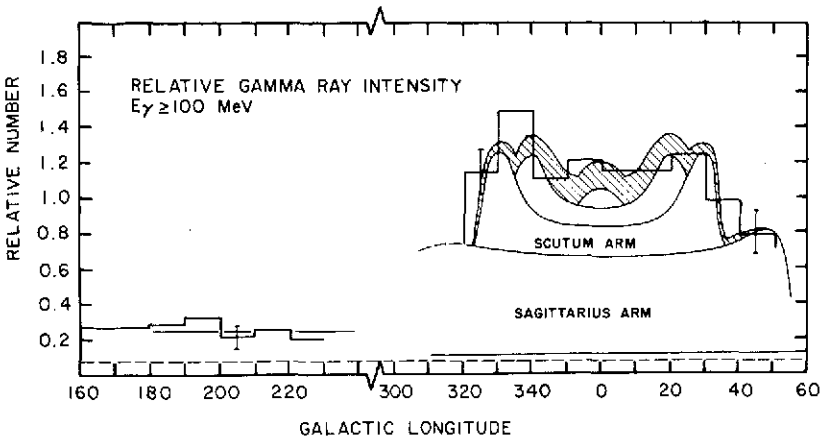


Figure 2. Observed gamma-ray distribution compared to theoretical prediction of this paper.

is an estimate of the arms from the back side of the galaxy. Generally a reasonable fit to the data is obtained, particularly the broad central distribution of 60°, with the additional extension out to the order of 50° on each side. One specific prediction that can be made by this theory is that there should be enhancements in the directions of about 260° to 270°, and also 70° to 80° where there are strong arm segments.

I should also indicate that, in the future, one would also like very much to examine individual clouds as well as arm segments to look for the expansion of cosmic rays. This is unfortunately going to have to wait for the next generation high energy gamma-ray satellite experiment with somewhat more sensitivity and angular resolution than SAS-2.

REFERENCES

Bierman, L. and L. Davis, 1960, *Zs. f. Ap. (Zeitshrift für Astrophysik)*, **51**, p. 19.

Parker, E. N., 1966, *Astrophys. J.*, **145**, p. 811.

GAMMA-RAY EVIDENCE FOR GALACTIC FERMI ACCELERATION OF COSMIC RAYS

Floyd Stecker

The SAS-2 galactic gamma-ray data give significant new information on the distribution of cosmic rays in the galaxy. The observed gamma-ray energy spectrum reveals that about 70 percent of the radiation in the central galactic region arises from the decay of neutral pi mesons produced by cosmic-ray nucleon interactions with interstellar gas.

Figure 1 shows the galactic longitude distribution of the gamma radiation observed by SAS-2 as compared with the distribution produced by a model which we have constructed for the gamma-ray production rate as a function of galactocentric distance. What we are assuming here is that the gamma-ray production rate, $Q(\tilde{\omega})$, is only a function of distance from the galactic center, $\tilde{\omega}$, and so we get a symmetrized curve shown by the solid line for the intensity-longitude distribution. A perfect fit to the longitude data can be obtained by a mathematical unfolding of the data. This unfolding reveals the galactic low energy cosmic-ray intensity as a function of galactocentric distance, as shown in figure 2.

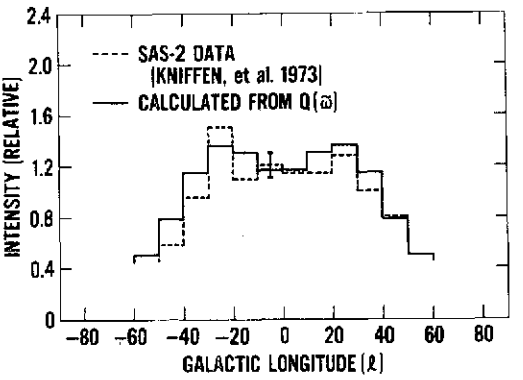


Figure 1. Comparison of gamma-ray distribution observed by SAS-2 with that produced using $Q(\tilde{\omega})$.

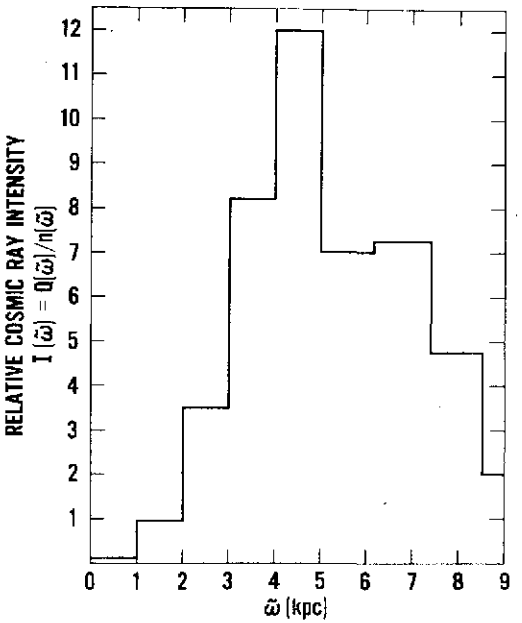


Figure 2. Galactic cosmic-ray intensity plotted as a function of galactocentric distance.

In the second figure, the distance from the galactic center in kiloparsecs is denoted by $\tilde{\omega}$. The relative cosmic-ray intensity, denoted by I, is the intensity we have calculated from unfolding the longitude distribution and dividing by the mean galactic gas density $n(\tilde{\omega})$. You can see a very strong peak in the region between 4 and 5 kpc from the galactic center. The enhancement is about an order of magnitude, so it is a strong effect; it is presumably located in a toroidal ring in the vicinity of 5 kpc from the galactic center, and implies a trapping of low energy (1 to 10 GeV) cosmic rays in this region.

Figure 3 shows the galaxy in terms of our model. Extraordinary features in the region of the inner galaxy are revealed by 21-cm radio data. The region inside 3 kpc is surprisingly free of gas which appears to have been swept out from the galactic center. In the 3- to 4-kpc region an expanding arm of gas is observed to be moving outward from the galactic center with velocities between 50 and 200 km/s. Just outside of this region is a ring of ionized hydrogen marking the position where the outward moving gases are stopped, roughly around 5 kpc out. The region of peak gamma-ray production is shown by the dotted area.

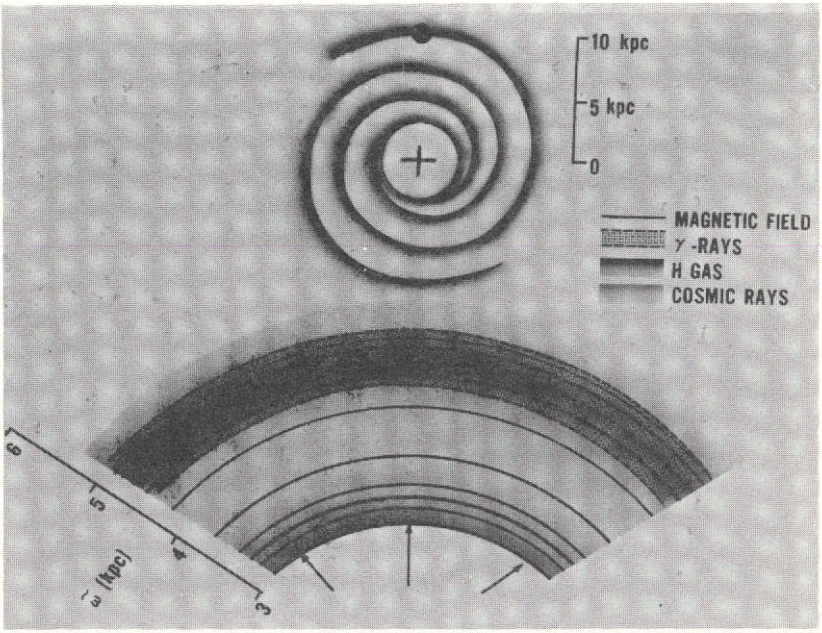


Figure 3. Model of galactic structure showing uniform distribution of gamma radiation.

Similar gas structure is revealed in the studies of our twin galaxy M-31 in Andromeda. We propose that in the 4- to 5-kpc region, cosmic rays are accelerated by a first order Fermi process acting over a time scale which we have calculated to about 7 million years. In this picture, momentum is transferred from the gas in the expanding feature to the cosmic rays, and the cosmic rays are trapped in the contracting region between 4 and 5 kpc both by the gas moving outward in the plane and by gas falling back in from higher galactic latitudes. The acceleration and contraction enhances the cosmic-ray intensity in this region enough to explain the increased gamma-ray production. Enhancement of subrelativistic cosmic rays in this region may also explain the ring of ionized hydrogen observed. Our picture ties in with the density wave theory of spiral arm formation. But our theory predicts a more uniform distribution of gamma radiation at intermediate galactic longitudes than does the model Dr. Fichtel just discussed. Further observations should provide a test between these hypotheses with profound consequences to our understanding of the overall dynamics of galactic structure.

CYGNUS X-1 TEMPORAL MICROSTRUCTURE: EVIDENCE FOR A BLACK HOLE

Richard Rothschild

On October 3, 1973, a rocket-borne multilayer, multianode proportional counter experiment was launched from White Sands Missile Range as a part of our continuing program of X-ray astronomy investigations. The purpose was to study the spectral and temporal character of three X-ray sources whose present configurations are thought to be the result of massive gravitational collapse. Even though they are all considered collapsed binary objects they exhibit quite different X-ray profiles. The three sources are Hercules X-1, Cygnus X-3, and Cygnus X-1, and their respective X-ray profiles are shown in figure 1 which comprises 15 s of data from each of the sources.

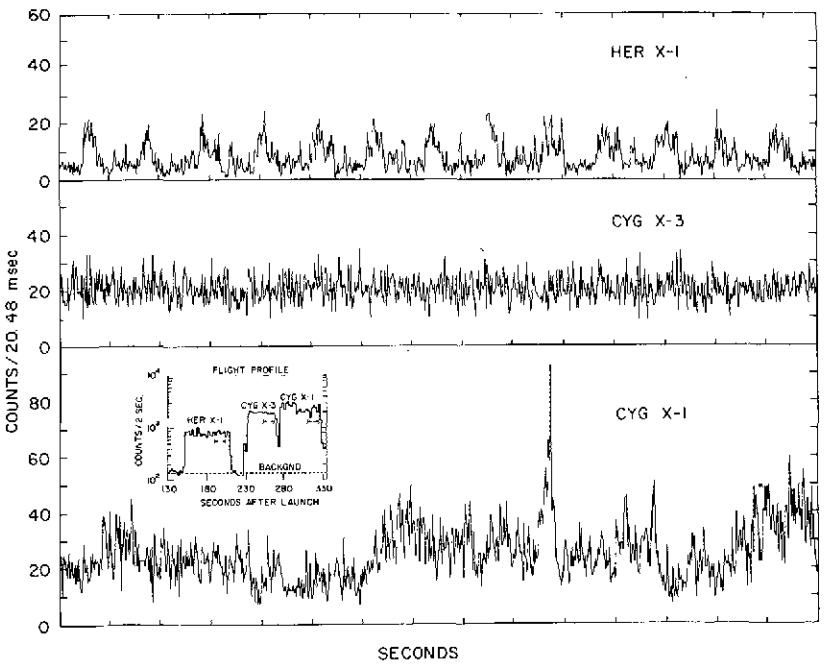


Figure 1. X-ray profiles of Hercules X-1, Cygnus X-3, and Cygnus X-1; the inset shows the overall flight profile.

The inset in figure 1 shows the flight profile, and the arrows show where each 15-s set of data came from. The top trace is that of Hercules X-1, an X-ray pulsar. The prominent feature of its profile is the regularity in the pulses. This is generally believed to be caused by a rotating neutron star in orbit about a companion star. Dr. Holt will elaborate further on this source in the next paper. The middle trace is that of Cygnus X-3. Note that its X-ray profile was that which would be expected from a source with purely random fluctuations on the time scale shown. The lower trace is that of Cygnus X-1, the leading candidate of all X-ray sources for identification with a black hole. Its X-ray profile displays considerable temporal structure which can provide evidence toward establishing the possible black hole nature of the source. During this exposure to Cygnus X-1, the intensity experienced several periods of enhanced activity, especially from 318 to 319 s after launch, as shown in figure 2.

Figure 2 shows some 4 s of data, binned every 20.48 ms, encompassing this enhanced activity when the mean luminosity over 400 ms doubled.

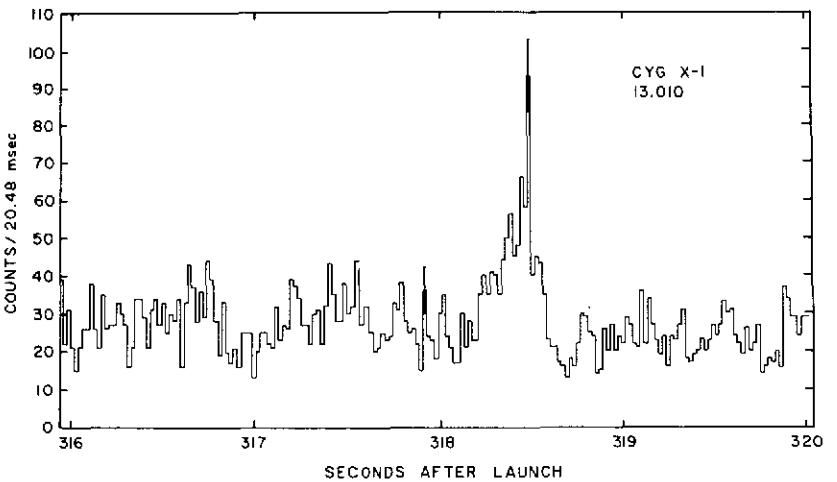


Figure 2. Four seconds of Cygnus X-1 exposure including 0.33-second enhancement that contains some of the bursts.

Figure 3 shows 80 ms of data, binned every 640 μs about the peak channel of the enhancement at 318 s. On three occasions during a 50-ms interval at the peak of the enhancement, the instantaneous count rate more than tripled during 1.28-ms samples. These bursts are shaded in the figure. Poisson statistics would predict 0.01 such events in 409.6 ms. In order to confirm the experimental validity of the bursts, the data for the entire rocket flight was examined for bursts outside of the Cygnus X-1 exposure, and none

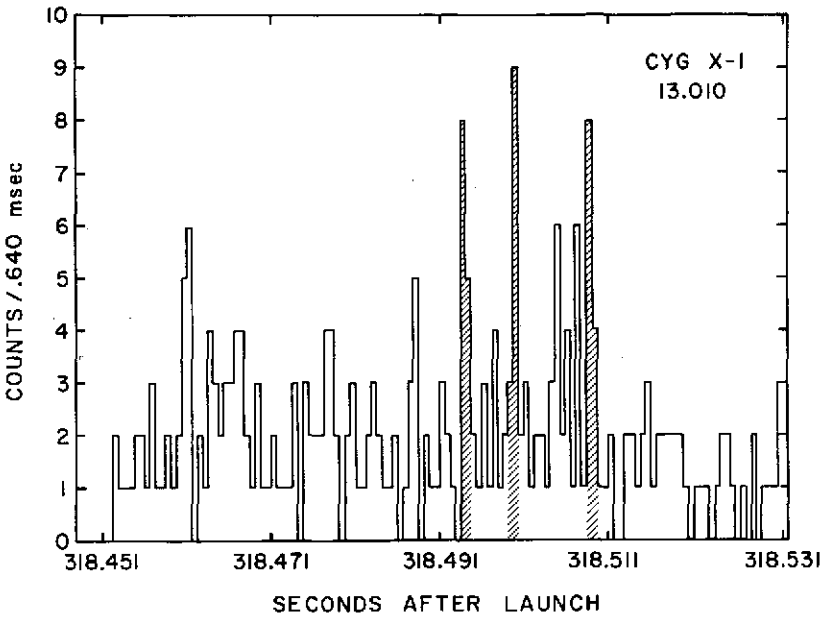


Figure 3. Eighty milliseconds at peak of enhancement showing three of the bursts; the 1-millisecond bursts are shaded.

were found. The rate of rejected (that is, non-X-ray) events was then examined during the time of the bursts and no significant variations were seen. This then minimizes the chance that the bursts were due to pick up in the electronics from such things as attitude control valves firing or camera relays changing state. Finally, the combined rate from the inner layers of the argon detector was compared with the first layer from both detectors, and all three sets of data were similar within the statistical accuracy available. This eliminates the possibility that precipitating energetic electrons entering through the windows were responsible for the observed bursts. If they had been, the inner argon layers, which benefit from having an active layer above as well as below for anticoincidence, would not have shown the enhancement, while the first layer would have shown it since it lacks an anticoincidence layer above it.

Current theories of a black hole orbiting a companion star predict that gas spiraling into the collapsed object from the other star will form an accretion disk about the black hole, from whose inner orbits X-ray emission by bremsstrahlung is expected. When there is disk accretion into a black hole and associated turbulence, chaotic fluctuations of the observed luminosity,

this emission may be expected. The typical time of these fluctuations in the observed luminosity are expected to be on the order of milliseconds for a disturbance at the innermost stable orbit of the disk about a 10-solar-mass black hole. We have positive evidence of the variability of Cygnus X-1 on time scales down to milliseconds, consistent with turbulence and disk accretion. The period of enhanced activity, containing three of these bursts, lasts for about one-third second. The changes in luminosity associated with these bursts are greater than six times the mean luminosity of the entire exposure. The spectrum of the bursts, which is consistent with bremsstrahlung by energetic subrelativistic electrons and nucleons, and which could correspond to thermal emission with KT greater than 10 KeV, is consistent with conditions expected at the inner orbits of accretion disks about collapsed objects. Recent mass determinations of Cygnus X-1 placed its minimum mass greater than the 3.2 solar mass upper limit for neutron stars of any type. If one assumes that variations in emission come from turbulence in accretion, then millisecond bursts imply that Cygnus X-1 must have a radius less than 100 km, which is less than the radius of a white dwarf. This evidence, combined with the mass determinations, points toward a black hole nature for Cygnus X-1.

MULTIPLE HERCULES X-1 PERIODICITIES: A SIMPLE KINEMATIC EXPLANATION

Stephen Holt

In the same rocket flight in which we looked at Cygnus X-1, we also spent about a minute looking at Hercules X-1 (Her X-1). We wanted to study its 1.24-second periodicity, which just about everybody believes is a manifestation of the rapid rotation of a magnetized neutron star accreting mass and a close binary system. The binary nature of the system is apparent from many experimental facts, not the least of which is a regular eclipsing of the X-ray source each 1.7 days. There are other facets of the X-ray emission which are not so well understood, however. For example, there is a third periodicity, one of about 35 days, and several rather disquieting, let me call them anomalies for the moment, which have no real explanation at present. It was very surprising, therefore, that in trying to understand our own 1-minute measurement of the 1.24-second variation, we were led to a model which appears to almost naturally account for virtually all of these anomalies on all time scales up to about 35 days.

Pulsars, you remember, are rotating magnetized neutron stars which beam their X-ray emission along the magnetic dipole axis M. Accreting systems, on the other hand, are expected to have their emission maximized in a sort of donut shape around the magnetic equator, which we see in cross section in figure 1. In the most general motion of the neutron star, the dipole axis is inclined with respect to the fast rotation axis ω, by an angle α, and the fast rotation axis may precess with angle P about a fixed axis in space, Ω. For the remaining discussion, I have chosen the simplest set of parameters which can synthesize the observations, and I will hold everything constant except the angle β between you and the fast rotation axis, which is the geometric sum of a fixed angle β_0 and the precession angle P.

Measurements of the 1.24-second variation before ours have always yielded double-peaked profiles, with the two peaks either 90° or 180° out of phase. We measured a triple-peaked profile, which cannot be synthesized by a donut alone, and requires something like beaming along the pole as well. Considering the better sensitivity of our experiment when compared to Uhuru, the model calculations of figure 2 for various values of β are consistent with all measured profiles. Uhuru can only see the interpulse peak

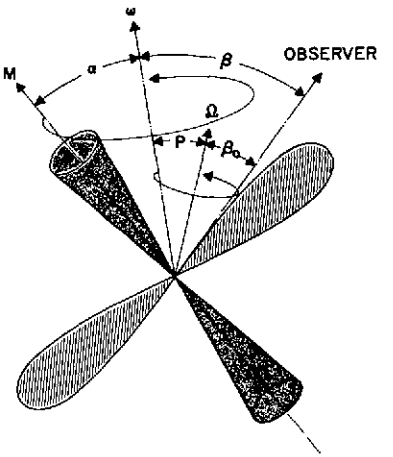


Figure 1. General motion of a rotating dipole.

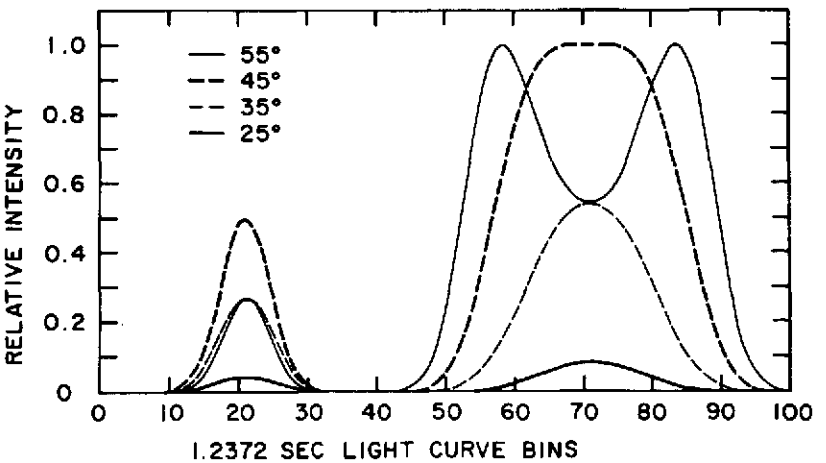


Figure 2. Model calculations for various values of β.

when β is close to 45°, because it is not sensitive enough to see it otherwise. In fact, when β is less than about 30°, Uhuru cannot see any profile at all. Uhuru has measured a pulse profile corresponding to a change in β in the model of about 10°, on a time scale of hours. Therefore, if we assume that the 35-day periodicity is the precession period of the fast rotation axis, we just cannot get β to change rapidly enough. It will change like the slowly varying trace in figure 3, where 10 percent means 10 percent of the maximum intensity which Uhuru measures (effectively, their observation threshold). But we need these wiggles to explain the anomalies I keep promising to tell you about, and I will do that now.

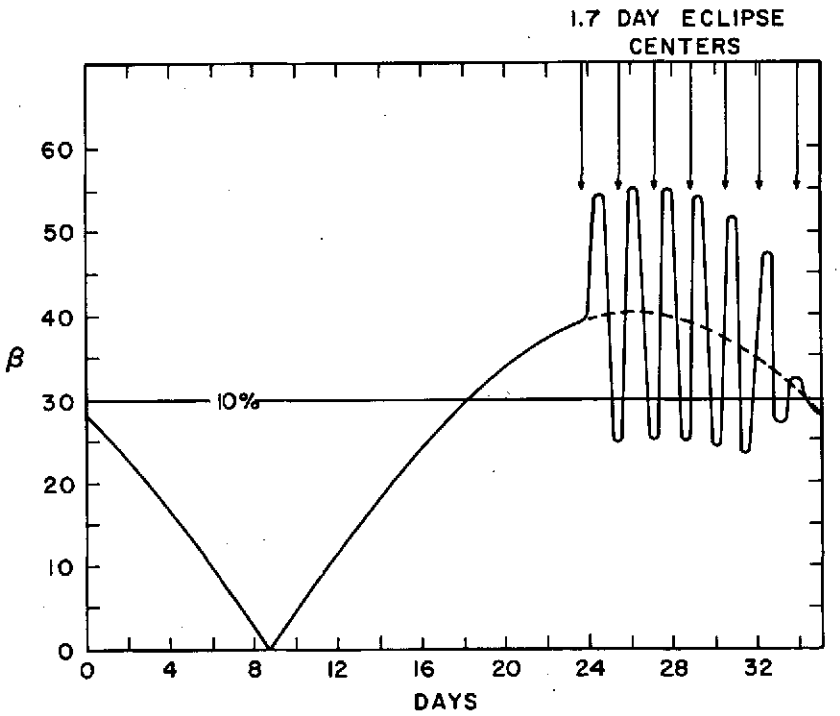


Figure 3. The possible variation of β as a function of precession phase.

Interestingly enough, for the parameters I have chosen, the place where the wiggles begin is the only point in the whole 35-day period when the poles just skim the accretion disk. In other words, this is the point at which the poles spend a maximum amount of time in the accretion disk. Furthermore, there are only two phases in the 1.7-day eclipse period that correspond to either one pole or the other moving directly toward the central star of the binary system. One of the anomalies that I keep alluding to is the fact that in

each 35-day cycle there is a rapid increase in the X-ray emission at one or the other of these two phases in the 1.7-day period. Our conjecture is that this ideal geometry turns the accretion process on hard, which tries to keep the pole pulled into the accretion disk, and which effectively means an increase in β. Geometrically, this looks like a forced nutation of the precession. The pulse profile can then change on a time scale of hours, as observed, and should even disappear when it drops below the 10° level. The frequency of the pseudo-nutation should be slightly higher than the eclipse frequency, because in each succeeding eclipse period, the accretion turns on a little earlier because the precession brings the rotation axis a little bit closer to the accretion disk. Such preeclipse dips, as they are called, which occur a little earlier each eclipse period, are yet another of the experimental anomalies with which the model can be reconciled.

Now believe me, I appreciate your difficulty in actually visualizing what I am describing, because I have been agonizing over it myself for a month now. But I am convinced that everything that has been observed from Her X-1 can be reconciled with the kinematic model I have outlined, in a fairly straightforward way (although it is going to take some time before the dynamics are worked out properly). I suggest that the complete consistency with a rather complicated set of anomalies, as I have called them, indicates that we have probably stumbled on a picture for Her X-1 which no longer has any substantial loose ends.

X-RAY ASTRONOMY EVIDENCE FOR SUBRELATIVISTIC COSMIC RAYS FROM SUPERNOVAE

Elihu Boldt

X-ray astronomy appears to be a useful tool in measuring otherwise elusive subrelativistic cosmic rays in regions remote to the solar system. First, unlike ultrarelativistic particles of high magnetic rigidity, subrelativistic cosmic rays are subject to exclusion from the inner solar system by the solar wind. Second, while relativistic cosmic rays lose a small fraction of their energy during their galactic confinement lifetime of more than a million years, subrelativistic cosmic rays lose all their energy to heating and ionizing interstellar gas and only propagate a small distance away from their origins. This effect of proximity of such cosmic ray particles to their origins is illustrated in figure 1.

The x-axis gives the particle energy at injection in MeV per nucleon. The subrelativistic cosmic rays considered here are protons, alphas, and silicon and iron nuclei. The y-axis gives the rectilinear propagation distance in kiloparsecs away from the origin when the particle has lost all of its initial energy. This distance is based upon a random walk of step length 100 pc, the half-thickness of the galactic hydrogen disk. As such, this estimate is likely to be an upper limit to the propagation away from the origin. Notice that for protons and alphas, this propagation exceeds 10 kpc (that is, the distance to the galactic center), only for energies exceeding about 80 MeV per nucleon, where the range exceeds about 3 g/cm^2. For lower energies, the distance is small compared to the radius of the galactic disk. This situation is even more severe for silicon and iron nuclei. For example, at about 10 MeV per nucleon the rectilinear excursion is only a few hundred parsecs, comparable to the stellar half-thickness of the galactic disk.

Of necessity, we are left with the problem of examining subrelativistic cosmic rays near their origins. This means that we need to detect them by virtue of the photons they generate. However, relativistic emission mechanisms such as synchrotron, inverse Compton, and transition radiation are just not available in this regime. For continuum radiation we look to the bremsstrahlung X-ray emission that occurs during collisions of subrelativistic cosmic rays with interstellar electrons. For example, X-rays up to 10 keV may be generated by energetic nuclei of 20 MeV per nucleon. In addition to continuum

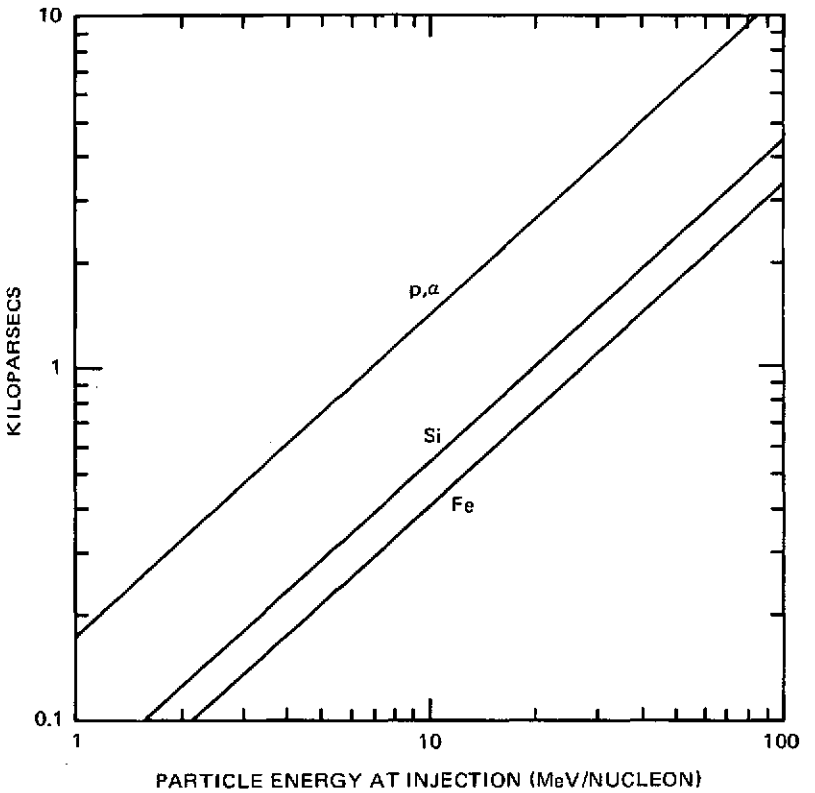


Figure 1. Propagation of subrelativistic cosmic rays away from origin.

radiation, though, subrelativistic cosmic rays may also give characteristic line radiation in the X-ray region. This is related to the phenomenon that a subrelativistic nucleus traversing atomic matter picks up and loses K shell electrons, giving an effective partially shielded charge that fluctuates about a value somewhat less than the bare nuclear charge. This effect was noticed by early workers in nuclear physics who examined tracks of stopped particles in photographic emulsions, and is illustrated in figure 2.

Here again, in figure 2, the x-axis is particle energy in MeV per nucleon. The y-axis gives the expectation value for the number of atomic electrons. Two nuclei of particular interest, silicon and iron, are considered here. First, for silicon we note that for energies below a few MeV per nucleon the K shell is likely to be filled with its two electrons. On the other hand, at much above 10 MeV per nucleon, the nucleus is likely to be bare. The intermediate region is one where K electron charge exchange with ambient

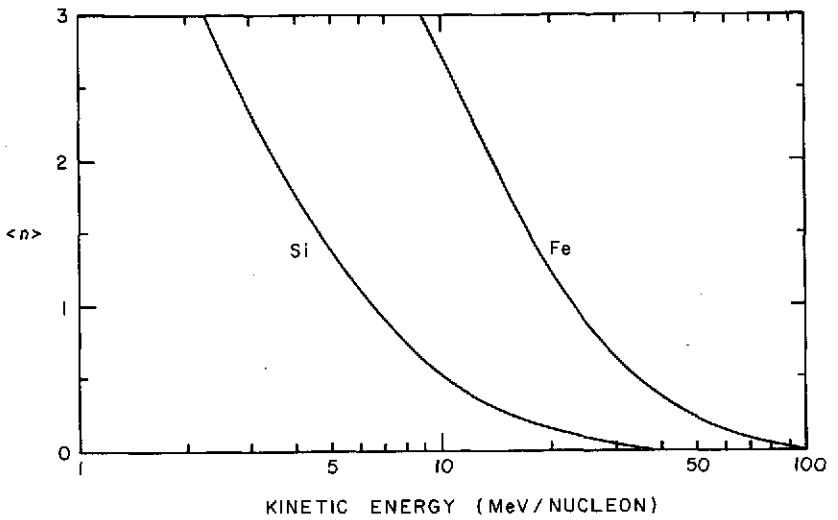


Figure 2. Effective number (<n>) of atomic electrons versus energy.

atoms becomes a prominent effect. As an electron is captured to the K shell, characteristic line radiation is emitted at about 2 keV. For iron nuclei, the characteristic line radiation occurs at about 7 keV. Here the particle energy region of particular interest is about 10 to 30 MeV per nucleon. For iron nuclei of 20 MeV per nucleon the line so emitted is Doppler broadened by about 20 percent.

X-ray emission from the vicinity of the two youngest supernova remnants in our galaxy, Cassiopeia A and Tycho, appears to indicate the presence of subrelativistic cosmic-ray nuclei such as iron and possibly silicon.

Figure 3 shows the X-ray spectrum for Cassiopeia A obtained by our laboratory in a rocket-borne experiment. Serlemitsos, Brisken, Holt, Ramaty, and I considered the bulge at about 7 keV as arising from K X-rays emitted by monoenergetic iron nuclei. The corresponding particle energy is about 13 MeV per nucleon. Considering a power-law particle spectrum with a low energy cutoff, the dominant energy turns out to be 17 MeV per nucleon. In either case, the energy is well situated within the characteristic region of K electron charge exchange, shown in figure 2. Since the measured spectrum only begins at about 2 keV we cannot say much about silicon emission. However, the higher energy emission can be described entirely in terms of line emission from higher Z nuclei, mainly iron, and continuum bremsstrahlung emission from all particles in the associated subrelativistic cosmic-ray population.

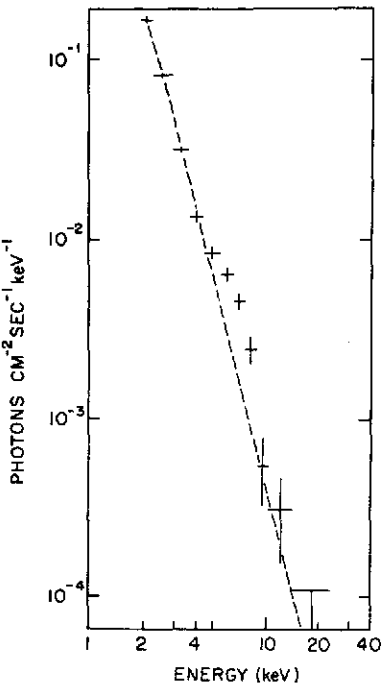


Figure 3. Incident X-ray spectrum for Cassiopeia A.

With somewhat poorer statistics the same situation appears to hold for Tycho. However, in this instance the characteristic particle energy seems to be a few MeV per nucleon lower. By X-ray observations then, we have obtained the evidence to be expected for young subrelativistic cosmic rays still in the immediate vicinity of their supernova origins.

THE JOVIAN ENCOUNTER EXPERIENCE

James H. Trainor

This paper is a preliminary account of the Jovian encounter as viewed by the Goddard/University of New Hampshire particle detector systems on Pioneer-10. These systems were designed to measure the charge, energy, and angular distribution of galactic and solar cosmic rays as well as the energetic particles in the Jovian magnetosphere. In the outer Jovian magnetosphere, the two cosmic-ray detector systems provided excellent data on high energy protons and electrons as well as measuring energetic helium nuclei. A third detector system, especially designed for high intensity fluxes, measured protons from 0.5 to 21 MeV and electrons from 0.1 to 1 MeV.

The first detection of energetic particles near Jupiter came very early, when Pioneer-10 was at a distance of ~200 R_J from the planet. These detections were in the form of temporary electron increases with energies extending above 1 MeV. While such upstream effects are not uncommon in front of the earth's magnetosphere, their characteristic energies are much smaller, generally in the 10- to 100-KeV range.

Pioneer-10 entered Jupiter's magnetosphere some 6 days later at a distance of approximately 109 R_J. Time histories for 0.4- to 2.0-MeV proton fluxes and 0.4- to 1-MeV electron fluxes are shown in figure 1. Unlike our own magnetosphere, the boundary for energetic particles is not well defined. The low energy protons are seen well before the bow shock and extend farther out than the electrons. Ratios of the electron-to-proton flux for the 0.4- to 1-MeV energy range are of the order 10:1 in the outer region and are generally decreasing as periapsis is approached.

Since Jupiter's magnetic axis is tilted with respect to its axis of rotation, the magnetic latitude of the spacecraft should vary with the rotation period of the planet (~10 hours). The points at which the expected magnetic latitude was a minimum (that is, when the spacecraft was closest to or on the magnetic equator) are shown in figures 1 to 3 as solid vertical bars. Generally, on the inbound pass, the predicted magnetic latitude varied between ~+2° and -20° (depending on the assumed value for this dipole tilt). There are no obvious 10-hour periodicities in the particle fluxes in the distant outer region (~80 to 100 R_J). Inside 80 R_J, first the 1-MeV

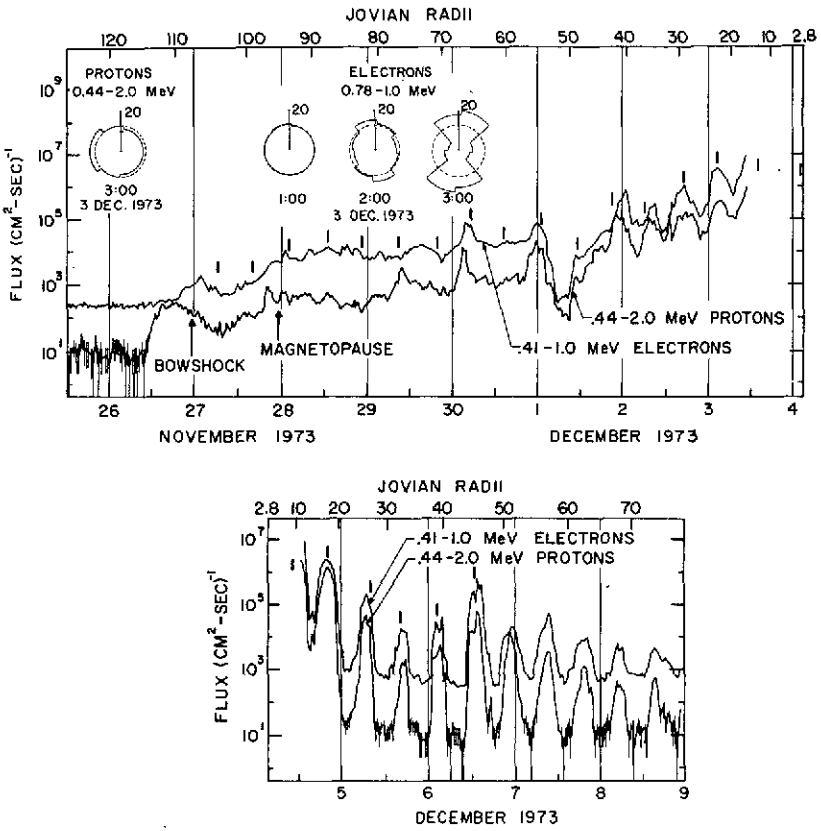


Figure 1. Time histories for 0.4- to 2.0-MeV proton fluxes and 0.4- to 1-MeV electron fluxes.

protons, and then the electrons, show increased fluxes near the magnetic equator. As we move inward these particles are apparently becoming more and more confined near the magnetic equator, consistent with a thin disk model for Jupiter's magnetosphere as defined by energetic protons and electrons. Inside ~40 R_J this latitude confinement is even more pronounced. These low energy counting rates were in saturation for ~12 to 15 R_J either side of periapsis. On the outbound pass, the 10-hour periodicity is much more striking with peak-to-valley ratios as much as 4 to 5 decades. Outbound, the predicted magnetic latitude varied generally between 0 and 22°. The inbound trajectory enters the magnetosphere at an angle of ~36° to the Jupiter-sun line, whereas the outbound trajectory leaves at close to 90°. This implies that, in the region of the magnetosphere traversed

on the outbound pass, protons and electrons are much more concentrated in low latitude regions. This behavior persists out to at least 80 R_J and is strikingly different from the inbound pass. The strong differences in behavior on the inbound and outbound legs suggest that there are significant azimuthal variations in the structure of the Jovian magnetosphere.

The experiment accumulates angular information for 12 different particle/energy combinations every 96 s. The figure 1 insert shows four polar plots of particle angular distributions. From 90 to 60 R_J, irregular and changing proton distributions were often seen. Inside 60 R_J, one typically saw a substantial anisotropy (such as that shown in figure 1), which is probably produced by corotation and/or particle gradients. Over the majority of the magnetosphere it was surprising how often the electron fluxes were found to be isotropic. There are transient periods of anisotropy, however, and a good example is outlined in the three polar plots for electrons in figure 1. This dumbbell distribution is observed in the earth's tail region for lower energies and is associated with particles flowing along the field line and mirroring very far away. Two hours later, the distribution was once again isotropic. This event occurred during a period of rising fluxes. Either these electrons had a very short lifetime or they were restricted to a very narrow region of field lines.

Figure 2 shows proton and electron time histories at higher energies ($>$ 6 MeV). These do not extend as far out in the magnetosphere as the lower energy particles in figure 1. However, a 10-hour periodicity is seen much earlier in the high energy electrons, indicating that the outer disk structure is more pronounced for these particles. There is a consistent phase difference between the latitude minima and high energy electrons in the 70- to 100-R_J region. This may indicate distortion of the outer disk structure. The electron-proton ratio at these energies is substantially higher ($\sim 10^3$ in the outer regions), indicating that the proton spectrum is much steeper than the electron spectrum. Inside 80 R_J these protons sometimes show large increases uncorrelated with the electrons, and so this ratio decreases at times to less than a factor of 10. The higher energy particles show even larger peak-to-valley ratios than the low energy particles in figure 1. They are, therefore, even more strongly confined to the equatorial regions.

Three proton energy intervals from the high intensity monitor are shown in figure 3. There may be some saturation effects within 5 R_J so this portion should be treated with caution. Note that inside 17 R_J the proton flux $>$16 MeV increases very strongly and the energy spectra appear to become increasingly harder. The times at which the shells of the several Galilean satellites were crossed are also indicated in figure 3. Local

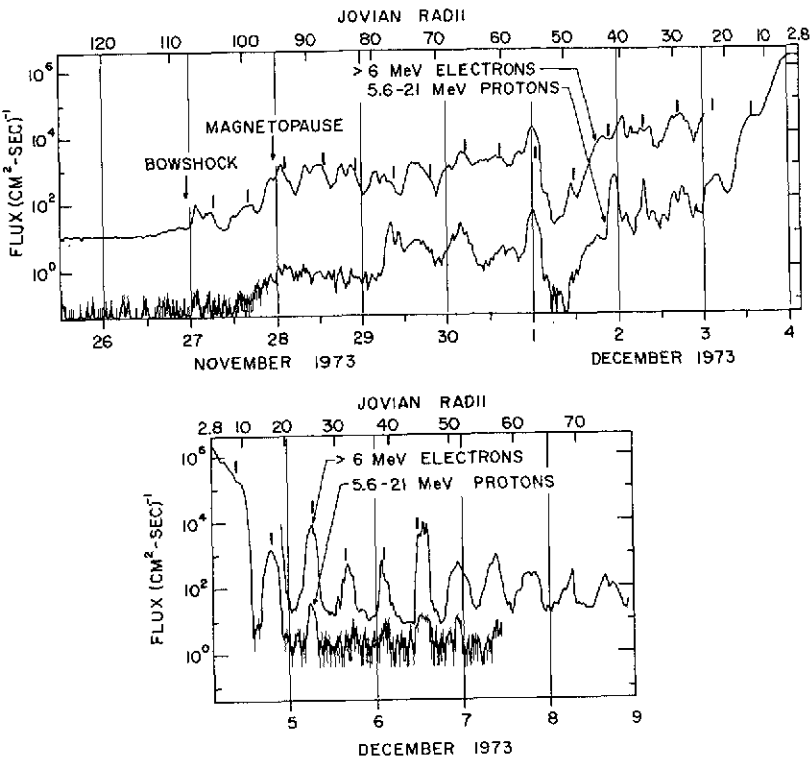


Figure 2. Proton and electron time histories at energies >6 MeV.

decreases are seen at both crossings of Io's (JI) shell and on the outbound crossing of Europa's (JII) shell. These times do not coincide with the 10-hour minima and are therefore probably caused by the sweeping effect of the satellites. Helium nuclei in the 3.3- to 21-MeV per nucleon range are clearly identified by multiparameter analysis between 25 and 80 R_J. These show marked increases during proton peaks such as the 40-R_J peak. The helium/proton ratio at this time is $\sim 3 \times 10^{-3}$.

This experiment is mounted outside the spacecraft and several electronic malfunctions in logic and switching circuits were induced during the several hour period of closest approach. While annealing effects have restored most of these, they provide yet another indication of the severity of the inner Jovian radiation region.

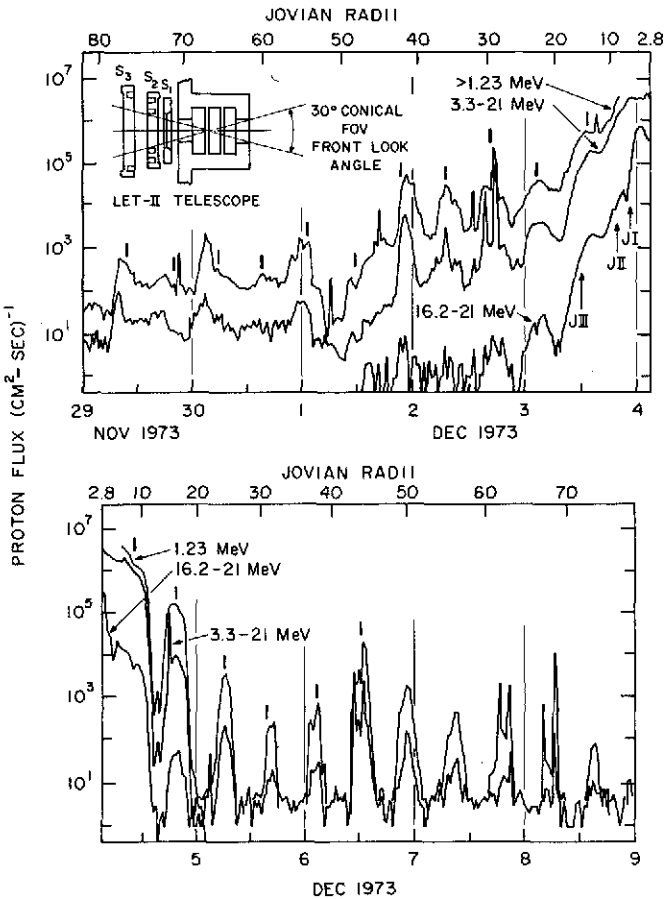


Figure 3. Three proton energy intervals from the high intensity monitor.

PIONEER-10 MEASUREMENTS OF THE COMPOSITION OF LOW ENERGY GALACTIC COSMIC RAYS

Bonnard Teegarden

This paper is concerned with interplanetary results from the Goddard Space Flight Center/University of New Hampshire cosmic-ray experiment on Pioneer-10, which show dramatic changes in the composition of cosmic rays at low energies. We believe this is the first positive evidence for the existence of a new type of source for cosmic rays. In particular, I will be discussing measurements of the cosmic-ray energy and charge spectra in the charge range of 2 to 14; in other words, helium through silicon nuclei in the energy range of 5 to 30 MeV per nucleon.

Figure 1 illustrates the standard dE/dX versus the E technique that we use to separate the various constituents of the particles that we measure. The individual elements, some of which are shown here, fall on separate, easily separable tracks. It is evident from figure 1 that we have no difficulty resolving the elements in question. Note, in particular, the relative behavior of the carbon and oxygen lines. Near the high energy ends of the tracks, the densities of events are roughly similar. As we proceed toward lower energies, you can see that the carbon 12 track becomes depleted, whereas the oxygen track becomes strongly enhanced, leading to a very large oxygen-to-carbon ratio and a very strong energy dependence in this ratio. This is the effect referred to earlier, and it is quite apparent even in the raw data shown here in figure 1.

In figure 2, we show energy spectra from our experiment plus spectra from two other experiments. The solid line indicates results that have been measured in the past for the carbon and oxygen spectra. A single line is shown since the spectra are essentially identical at high energies ($\gtrsim$30 MeV per nucleon). The Pioneer-10 results span the region approximately between 30 and 5 MeV per nucleon. We can see that at above 30 MeV per nucleon the oxygen spectrum turns up sharply. Carbon, however, continues downward and does not turn up until at least 10 MeV per nucleon. The data is sparse in this region, so that we can only place upper limits for carbon below 10 MeV per nucleon. The carbon spectrum may, in fact, continue to decrease toward lower energies. The oxygen-to-carbon ratio in this energy interval is in excess of 20:1. At lower energies, we have shown data from the IMP-7 experiment of Hovestadt and Gloeckler. In their data the oxygen

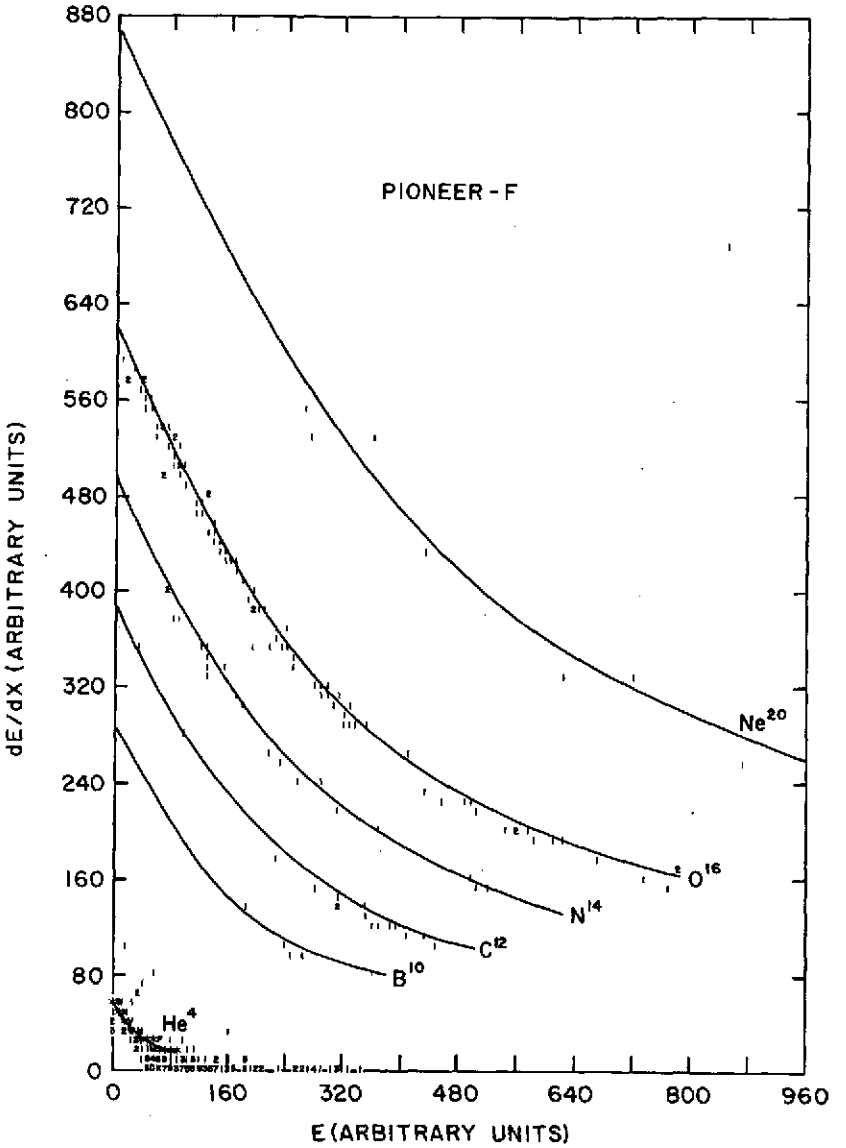


Figure 1. The standard dE/dX technique versus the E technique for separating constituents of cosmic rays.

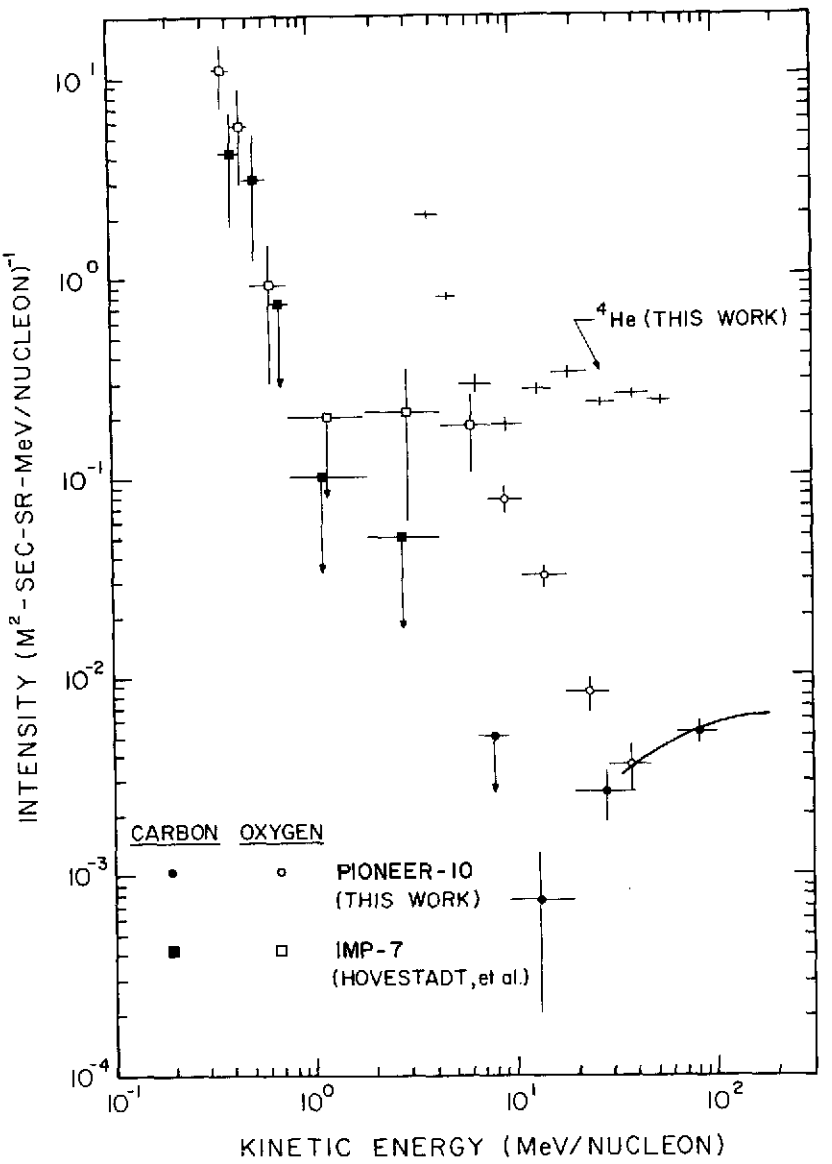


Figure 2. Energy spectra observed from Pioneer-10 and IMP-7 experiments.

and carbon apparently reconverge. The tentative interpretation is that, as we go toward lower energies, we enter a region where solar particles dominate. An oxygen-to-carbon ratio of 2 is characteristic of solar abundances and consistent with the low energy data. We should address the question of whether the particles that we measure on Pioneer-10, where the anomalous composition is seen, could in fact be of solar origin. The huge oxygen-to-carbon ratio that we observe argues against this by itself. In all solar flare measurements of these abundance ratios, values consistent with spectroscopic ratios (~2:1) for oxygen to carbon have been seen. In addition, we have been very careful to eliminate any active periods from the data. Finally, we have preliminary evidence for a positive gradient in these oxygen nuclei. We think these three points argue very strongly for an extrasolar origin for these particles.

In figure 3 we summarize the results for the various elements that we have measured. The figure shows the enhancement over so-called normal cosmic-ray abundances, where normal is defined as the abundances at energies greater than 1 GeV. Both the normal and anomalous are abundances normalized to helium. It is evident that for all elements except nitrogen, oxygen, and possibly neon, the value for the enhancement is less than 1. This could be explained in terms of propagation effects in the interstellar medium. The enhancements, however, that we see for nitrogen and oxygen are very difficult to explain in terms of any propagation models. We are forced to postulate the existence of a new type of cosmic-ray source, rich in oxygen and nitrogen, which is supplying us with low energy particles. The anomalous chemical composition of these particles is a positive signature which may well lead to an eventual positive identification of the source.

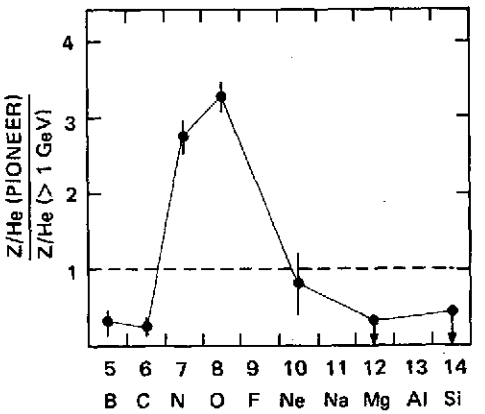


Figure 3. Enhancement over normal cosmic-ray abundances (normalized to helium).

WHAT PIONEER-10 MEANS TO COSMIC-RAY MODULATION THEORY

Leonard Fisk

The observations which Bonnard Teegarten just described present an interesting challenge to current modulation theory–the theory of how cosmic-ray particles behave in the interplanetary medium. The challenge comes not from being able to explain the unusual composition, since the composition is presumably preserved during the modulation process. Rather, it comes from being able to explain how these low energy particles, which we will assume here are galactic in origin, got into the inner solar system in the first place.

Shown in figure 1 is the observed differential intensity spectrum of oxygen in the energy range from a few to 100 MeV per nucleon. The spectrum turns up dramatically at low energies as Bonnard reported. It is this turn-up which is very difficult to understand in terms of the current theory, the reason for the difficulty being the dominant role that adiabatic energy loss plays in the current theory. In the current theory, we assume that cosmic rays diffuse into the solar cavity among the irregularities that are present in the interplanetary magnetic field. As the particles undergo this diffusion, they find themselves in a medium which is expanding–the solar wind–and the particles are adiabatically cooled. The longer the particles spend in the solar cavity, the more cooling they experience. In the current theory this predicted cooling can actually be quite large.

One way to see why this adiabatic cooling makes it difficult to account for these oxygen observations is to consider the observed proton spectrum in a somewhat higher energy range, from about 20 to 100 MeV as is shown in figure 1. This proton spectrum is measured at the same time as the oxygen spectrum is measured. Of course, we do not believe that this measured proton spectrum is the interstellar proton spectrum, but rather that these protons have been modulated. A reasonable guess for a minimum interstellar proton spectrum might be the one that is shown in figure 1.

If we want to modulate this interstellar proton spectrum into the observed proton spectrum, which is falling off very steeply, the current theory requires that these particles lose about 100 MeV of energy. That is, all particles that are observed below 100 MeV in the inner solar system came

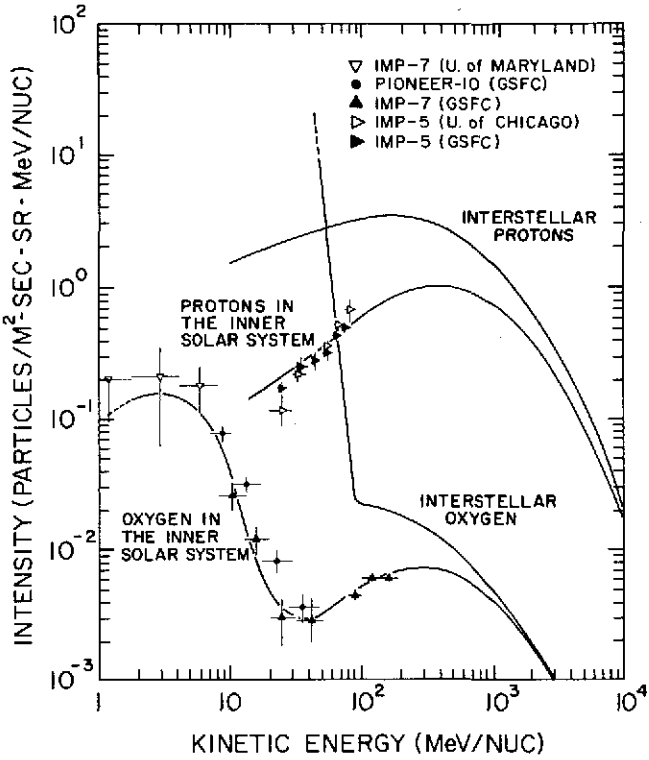


Figure 1. The observed differential intensity spectrum of oxygen.

into the solar cavity at higher energies and were cooled down. Modulation, however, like any magnetic process, is rigidity dependent. Particles with the same momentum per unit charge will experience the same modulation conditions. Protons in this energy range of 20 to 100 MeV, however, have the same rigidity as the oxygen in the turned-up portion of the spectrum (< 30 MeV/nucleon). So superficially then, if these protons lose 100 MeV of energy, we might expect in the current theory an energy loss of about 100 MeV per nucleon or so for the oxygen particles, and we would never be able to sample the low energy interstellar oxygen.

However, it is not quite that simple because this 100 MeV per nucleon is the mean energy loss of a smooth distribution of cosmic rays, not one with a sharp turn-up like we have here. Nevertheless, if the protons lose ~100 MeV, the energy loss for the oxygen particles should be large compared to the energy at which we observed them, which is around a few MeV per

C-2

nucleon. The only way that we could then produce such a turn-up in the current theory, in the presence of such energy loss, is to have a very steeply increasing interstellar spectrum like the one shown in figure 1. With such a steep turn-up in the interstellar spectrum, however, we would soon exceed an acceptable boundary on the energy density of interstellar oxygen, and we would blow up the galaxy. Such a large modulation is untenable. I must construct, therefore, a modulation theory in which the adiabatic cooling does not play such a significant role. We have, in fact, done this by taking into account variations in the modulation with heliocentrical latitude, which is something that is not considered in the current theory.

Consider, for example, a model in which the modulation above and below the ecliptic plane is larger than the modulation in the ecliptic plane, as is depicted in figure 2. The cosmic-ray number density above and below the ecliptic plane will then be inherently smaller than it is in the ecliptic plane.

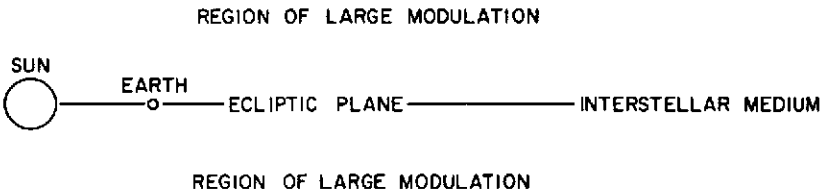


Figure 2. Cosmic-ray model with the modulation above and below the ecliptic plane larger than that in the plane.

Suppose also that particles which attempt to penetrate from the interstellar medium in the ecliptic plane can do so relatively easily, which is consistent with the small Pioneer gradient measurements. The particles, in that case, will not spend much time in this penetration, and as a result will not lose much energy. But suppose also that there is a high probability that, as the particles penetrate inward in the ecliptic plane, they will be lost into this region of low density or large modulation above and below the ecliptic plane (and this probability can be made quite high). So then we have a model in which only a small fraction of the particles will penetrate into the inner solar system, but those that do will lose very little energy. We have constructed a theory in which the modulation can be arbitrarily large but with only a minimum of energy loss. That this theory can adequately account for the observations can be seen in figure 3.

The curves shown in figure 3 are the result of numerical solutions to the modulation equation. As you can see, we have the prescribed modulation for the protons, but now the oxygen modulation is much more reasonable and it yields only a minor and acceptable energy density for interstellar oxygen.

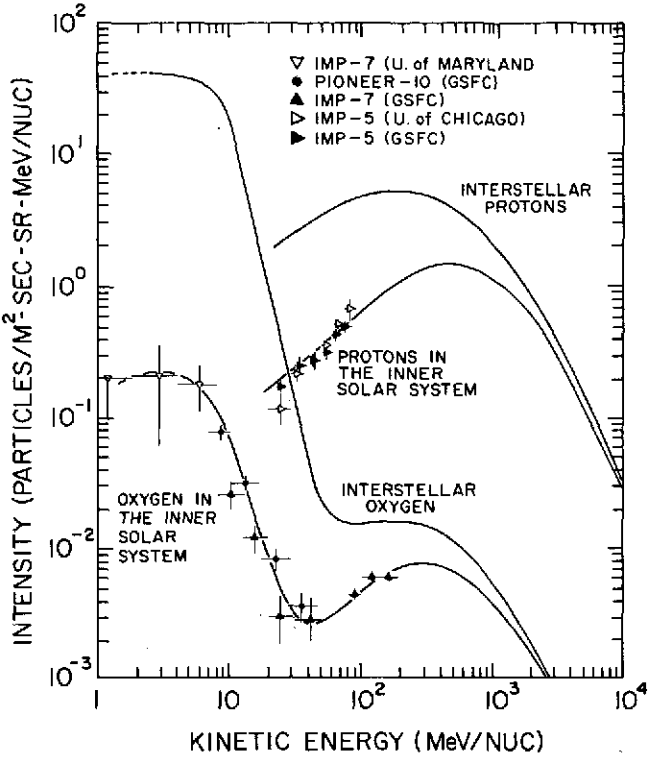


Figure 3. Results of numerical solutions to the modulation equation.

In conclusion then, we would argue that these oxygen observations strongly suggest that adiabatic cooling, at least during this portion of the solar cycle, is not as important as was previously thought, and these observations suggest also that our local modulation is strongly influenced, indeed controlled, by conditions out of the ecliptic plane. It follows, of course, that if we ever should hope to really understand the modulation, we will someday have to make direct spacecraft measurements of these off-ecliptic conditions.

BALLOON OBSERVATIONS OF COSMIC-RAY NEGATRONS AND POSITRONS

Robert Hartman

I would like to report briefly on some balloon investigations we are conducting on the electron component of the cosmic radiation. These experiments differ from most of the earlier electron measurements in having the capability for separating positive and negative electrons. This provides some new approaches to the study of solar modulation. In particular, a measurement of the interplanetary positron spectrum in the 20 to 500 MeV range should provide a measure of the absolute modulation of the positrons in that energy range, since it is possible to calculate rather well the interstellar positron spectrum expected from cosmic-ray collisions with the interstellar gas. These are the same collisions that produce the gamma rays which were discussed earlier by Carl Fichtel. In addition, it should be possible to observe directly the energy loss, or cooling, which Len Fisk just mentioned, predicted by modulation theory, by studying the time variations in the positron-negatron ratio.

Figure 1 shows the equipment we are using to make these measurements. The experiment used digital spark chambers as shown, 16 layers in 4 groups, to determine the deflections of individual charge particles passing through the gap of a permanent magnet. This deflection determines the charge sign and also the energy of those electrons. The spark chambers are triggered by a counter telescope including the scintillators and also a gas Cerenkov counter at the top which provides for accepting only downward-moving particles and also has a velocity threshold sufficient to exclude most of the cosmic-ray nuclear component. A shallow sandwich of lead plates and plastic simulators at the bottom enhances our capability for separating the positrons from the large background of protons. With such a system, a unique identification of individual electrons is possible.

Figure 2 shows our measured total electrons spectrum along with those from some other experiments in the same energy range. The solid curves at the top are estimates of the interstellar electron spectrum outside the solar system derived from measurements of the nonthermal radio noise. Now, using curve A and our measurements, we obtain an estimate of the modulation function of these electrons in this energy range. It turns out that we can describe this with a rather simple form for the solar modulation function.

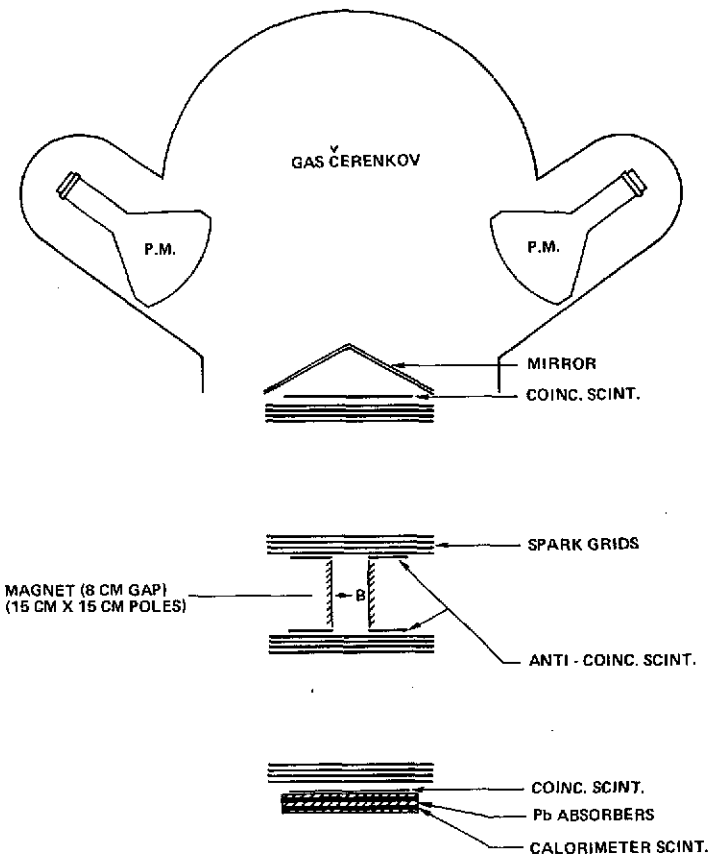


Figure 1. Electron-positron spectrometer (low energy).

Our measured positron spectrum is shown in figure 3, along with a theoretical curve. This is the spectrum calculated by Dr. Ramaty for positrons from the collisions of nuclear cosmic rays with the interstellar gas. It is quite important to note here that this theoretical spectrum is completely independent of the theoretical spectrum shown in figure 2. The dashed curve shown here is derived by using the modulation function obtained from figure 2 to modulate this predicted interstellar positron spectrum. As you can see, the agreement with our experimental results is reasonably good, although of course we do have rather large errors on these measurements. Nevertheless, we feel that this consistency forms a good basis for judging the modulation picture that we have, and we are really quite encouraged by it.

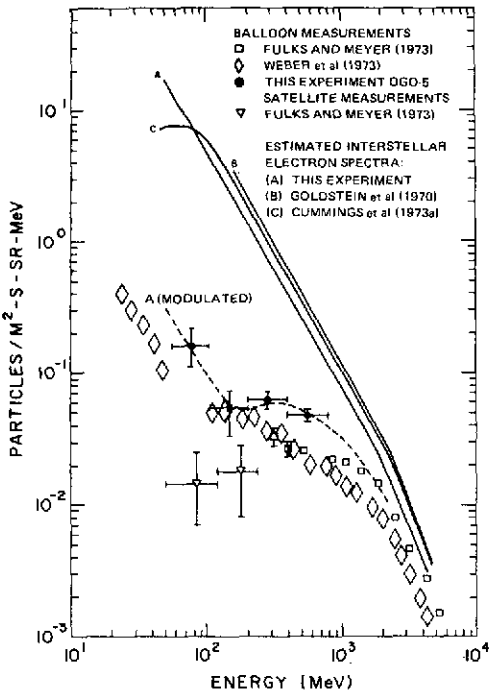


Figure 2. Total electron spectrum, 1972 measurements.

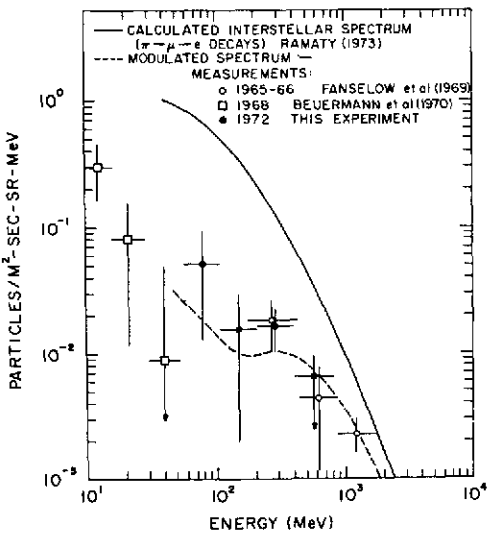


Figure 3. Positron spectrum (1973) and theoretical curve.

We feel that we have shown that these balloon measurements can obtain information on the modulation of these positrons. It is an extremely difficult measurement but we feel that with certain improvements in our hardware and in our experimental technique, we can make rather dramatic improvements in the accuracy of these measurements, and therefore, pin down this modulated spectrum even better. We are planning additional balloon flights during next summer to attempt this improvement.

OVERVIEW

Siegfried Bauer

This afternoon we are going to retreat to the local corner of the galaxy. Compared with the exciting discoveries about the universe at large about which we have heard this morning, the study of the solar system and even more so, our own planet, appear rather mundane. It seems to be a basic facet of human nature that the things which are far away are mysterious and thus are very fascinating, while things which are close by are, of course, more familiar. Now I do not want to go to the extent to say that familiarity breeds contempt, however, familiarity also does not necessarily imply understanding. As a result of over a decade of spacecraft observations, we now have a reasonably good picture of the topology and morphology of the near space and biomental earth and its solar control, although some of the physical processes, especially relating to the energetics of this control, are still elusive.

Figure 1, in a very schematic way, illustrates this control. Of course, most of the energy from the sun comes directly to the planetary surface and controls our weather. Very small, but highly variable fractions of the sun (those in the extreme UV and in the corpuscular component) and the solar wind have very profound effects on the near space and biomental earth, the upper atmosphere, and the magnetosphere.

More recently, there has been an old notion revived which was the topic of a recent conference at Goddard on "Possible Effects of Solar Control on the Weather." Here again, one has the hope that some of the highly variable components of the sun's energy might effect the weather; although this kind of link is rather tenuous it definitely is a rather interesting concept which definitely bears some investigation. On the other hand, it is rather quite well known that the large-scale pressure system on the troposphere caused by internal gravity wave propagations affect the upper atmosphere. In fact, gravity waves have been observed in the upper atmosphere and this kind of propagation path from the lower level up is now generally well accepted. I cannot help but remember that 15 years ago I showed that the ionosphere responds locally to the passage of hurricanes.

The upper atmosphere, the ionosphere, which of course is primarily controlled by the EUV radiation from the sun, has been studied over the years

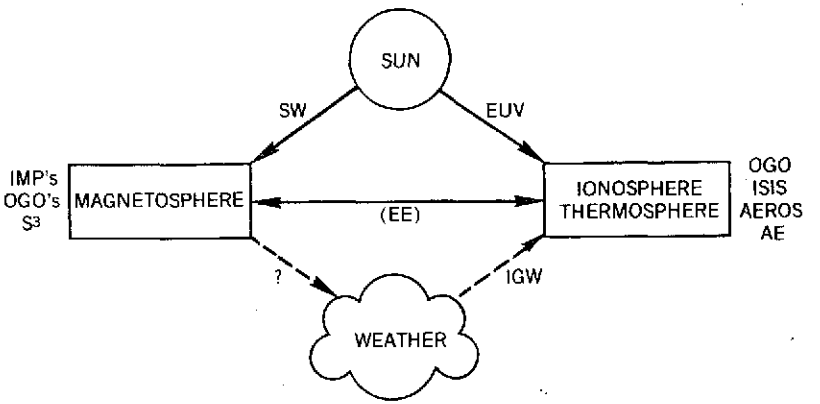


Figure 1. Schematic showing solar control on the near space and biomental earth, the upper atmosphere, and the magnetosphere.

by satellites of the type of OGO, ISIS, a more recently aeronomic satellite like the German AEROS, and to a much higher degree of sophistication with the AE satellite system, the first one of which was just recently launched. On the other hand, the magnetosphere, the corpuscular input, and the control of the topology of the magnetosphere by the solar wind have been studied with the IMPs and OGOs while the inner magnetosphere with some of the processes occurring there, like magnetic storms and drink currents, have been particularly studied with the S^3 satellite.

The interrelationship of the magnetosphere and the ionosphere has up to date still been neglected and is still not fully understood; for that reason, the so-called electron dynamics explorer has been suggested by a number of scientists as a possible project to further the understanding of the interrelationship between the magnetosphere and the ionosphere. One of the parameters which up until recent times has been very difficult to measure, yet has a profound influence on the dynamics of the magnetosphere and the ionosphere is electric fields. Reliable measurements of this type have been made only recently and they are definitely leading to a great advance in our understanding of the dynamics of this part of the medium. The following speakers will concentrate on virtually all aspects of the solar-terrestrial relationships, from the sun as an energy source both in the EUV and the corpuscular component to phenomena occurring within the magnetosphere and the upper atmosphere and ionosphere, to suggested relationships occurring as far down as the mesosphere which brings us a little bit closer to this elusive problem of the weather.

SOLAR CYCLE VARIATION OF EUV RADIATION

Robert D. Chapman

I want to talk today about some work that has been done with data from the OSO-1 and OSO-3 satellites. We know that the radiation in the wavelength region from 15 to 40 nm (150 to 400 Å) contributes something on the order of 30 to 50 percent of the ionizations in the ionosphere at times when the sun is quiet. We know also that it contributes roughly the same fraction of the energy input to the ionosphere, and this energy is input at a level in the earth's atmosphere on the order of 180 km. It would be very useful if we could have a day-to-day picture of the magnitude of the flux in this wavelength region that falls on the earth's atmosphere, so that we would have day-to-day knowledge of the energy input into the atmosphere.

It has been known for some time that the EUV radiation in this wavelength region is correlated with the radio emission from the sun, and so we have asked ourselves the following question: Given a measure of the radio radiation, can we predict the total flux of the EUV radiation on the earth's upper atmosphere? In order to answer this question we have carried out correlations between the level of the EUV flux in various lines in the 15- to 40-nm (150- to 400-Å) region and the solar radio flux. It turns out that this region of the spectrum is almost entirely line radiation; there is very little, if any, continuum present. We took the data from OSO-1 and OSO-3, determined correlation coefficients between the EUV flux and the radio flux, and found that for the 2,000-MHz radio flux and the 2800-MHz radio flux, the correlation coefficients are significant at the 99-percent confidence level. This means that we have a very good correlation and have some hope of using the measured radio fluxes to give us some idea of the EUV flux. Just to test this a little bit further, we took data from OSO-1 which covers a 3-month period in 1962 (approximately three solar rotations when the level of solar activity was low), and calculated linear regression lines for the EUV intensities of a few of the lines in this wavelength region versus the radio flux. We used these regression lines to try to predict the general trend of the OSO-3 data that was taken in 1967 at a higher level of solar activity and which covered something on the order of 300 days, about 10 solar rotations. The absolute calibration of the two instruments is not known at present, and so we adjust the level of the curves to the same relative level.

Figure 1 shows the data on the 30.4-nm (304-Å) line of He II compared with predictions from the 2800-MHz radio flux. We picked the 2800-MHz flux rather than the 2000-MHz flux, strictly because it was more readily available on a day-to-day basis. The abscissa is days into 1967, the ordinate is the OSO-3 data in counts. The solid curve is the prediction from the OSO-1 regression lines of the line intensity, and the dots are the observed intensities. You can see that the agreement is quite good.

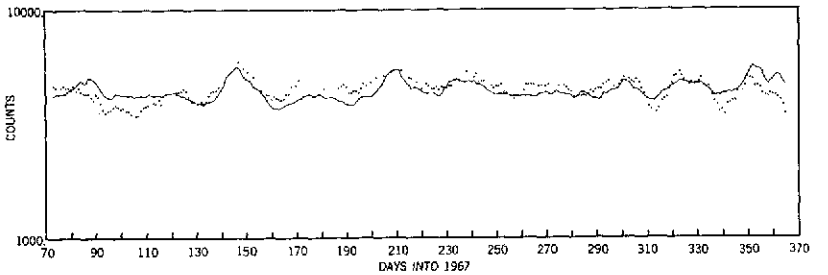


Figure 1. Data on 30.4-nm (304-Å) line compared with prediction from 2800-MHz radio flux.

Figure 2 shows Fe XV, 28.4-nm (284-Å) results. The 30.4-nm (304-Å) line is formed at a fairly low level in the solar atmosphere with temperatures of about 50,000 K. The 28.4-nm (284-Å) line is formed at a level where the temperature is about 2×10^6 K. Again, you can see that the general agreement is really quite good; the observations may be above or below the predictions on any given day, but the agreement over a solar rotation is very good.

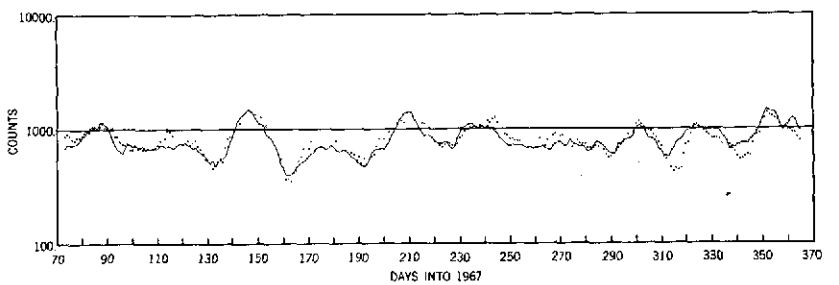


Figure 2. Data on 28.4-nm (284-Å) line compared with prediction from 2800-MHz radio flux.

We took the OSO-3 data and tried to apply an absolute calibration to it. That was done by first using the laboratory calibration of the flight instrument, and second, the absolute calibration of Hinteregger's data which were

taken at the same time on the same satellite, but in a slightly different wavelength region. There are a few strong lines in the region where the two wavelength regions overlapped.

Figure 3 shows a calibrated OSO-3 spectrum. The lower contour is the level of the continuum, while the two other contours indicate the 0.01 and 0.1 erg/cm^2/s flux levels. The continuum is purely scattered radiation. The total flux of all the lines in the 15- to 40-nm (150- to 400-Å) region is given as a slope times F, the 2800-MHz flux, plus an additive constant. At quiet sun conditions, F is on the order of 100 so that the two numbers are roughly of the same magnitude; at that time the EUV energy input to the atmosphere is about 1.4 ergs/cm^2/s. That number is about twice what has been indicated by past work in this wavelength region and more in line with what the upper atmospheric physicists would like to see.

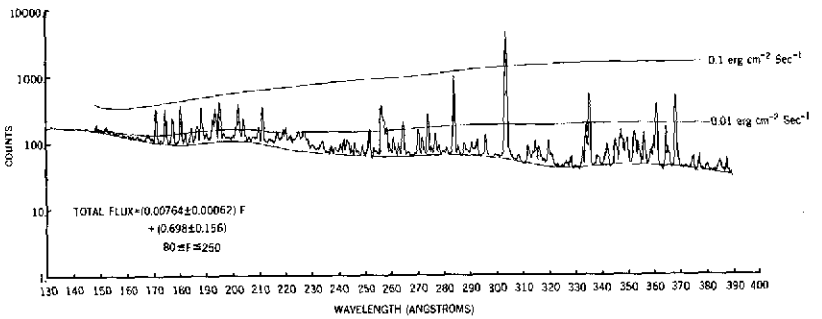


Figure 3. Calibrated OSO-3 spectrum.

SOLAR CYCLE VARIATION OF SOLAR WIND CHARACTERISTICS

Keith W. Ogilvie

This talk is really an interim report on a study of the solar cycle variation of helium abundance in the solar wind. The relative abundance of helium in the sun is not well known and all the presently proposed methods of measuring it are more or less in disrepute.

One such method is measuring the relative abundance of helium in the solar wind which is, of course, an extension of the solar corona. But a difficulty associated with this is that although the photospheric abundance may be regarded as fixed and characteristic of the sun, the abundance of helium in the corona is expected to vary from place to place and from time to time because of the effects of gravitational settling, thermal diffusion across the transition layer, and convective mixing.

Indeed, the measured relative abundance of helium in the solar wind at 1 AU, where most of the measurements have been made, shows wide variation on all time scales. Normally, these variations lie in the range of 0.01 to 10 percent. There are also large impulsive increases when $N\alpha/Np$, the ratio of the number densities of helium to hydrogen, goes up to 20 to 25 percent for periods of an hour or more; these have been ascribed to material ejected by solar flares from the lower corona where one might expect on general grounds that there be a reservoir of high helium concentration.

One quantitative regularity which has been noted is that the average abundance of helium in the solar wind varies with solar wind speed; in figure 1 we see some regression lines for various different experiments and periods of time. In general, this increase of relative abundance of helium with increasing speed appears as a general characteristic of all the available data.

Dr. Hirschberg, of the High Altitude Observatory in Boulder, and I have been examining the available evidence to see if it indicates a real variation of $N\alpha/Np$ over the last solar activity cycle, in addition to what one would expect from any variation of average bulk speed and variation of the rate of occurrence of major flares accompanied by impulsive enhancements. In figure 2a we see the smoothed sunspot number for the last solar cycle. In figure 2b we have the solar rotation averages of $N\alpha/Np$ for various experiments; and figure 2c shows the number of solar flares and sudden commencements per year. We note that the major flow index, not plotted here, shows

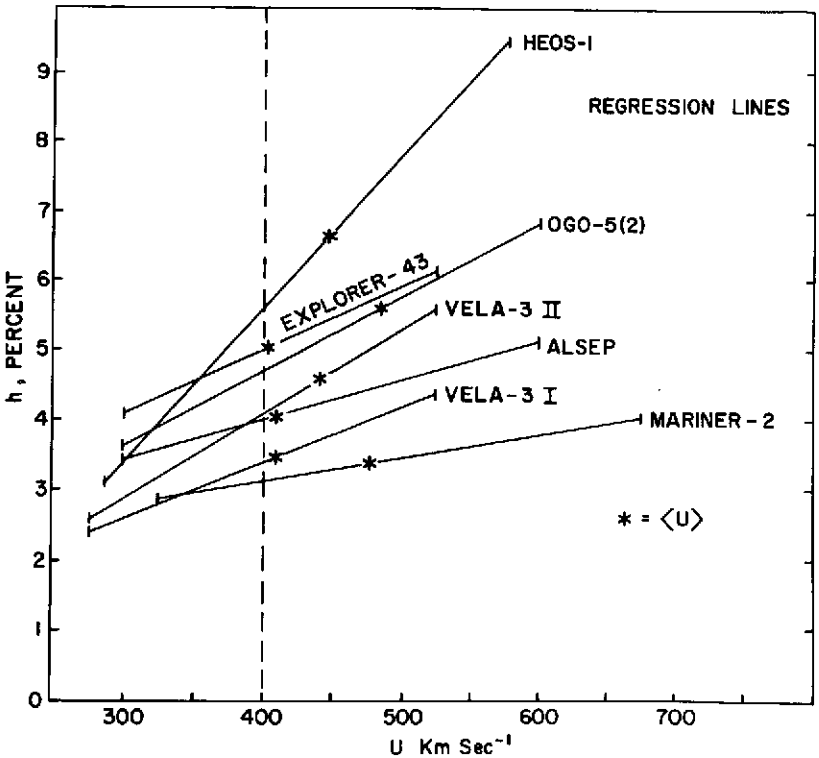


Figure 1. Regression lines Nα/Np versus U, the solar wind bulk speed.

a pronounced peak in 1969, just after the sunspot maximum. The data points in figure 2a are averages over a single solar rotation which removes the short period variations, but as you can see, major fluctuations remain. Let us examine this data, bearing in mind that the time required for the solar wind to remove material equal to that in the corona is about 11 days and for the chromosphere about 8 days.

Consideration of both systematic and random errors associated with each of these measurements, which is what this work has mainly consisted of, leads us to conclude that there was a real variation during the last solar cycle in the relative abundance of helium between solar minimum and solar maximum from about 3 percent to about 4.5 percent. The question is then, "Can we explain this variation as a combined result of the variation in solar wind speed and the rate of impulsive events?" If we use the average observed solar wind speed from each rotation to correct Nα/Np to a fixed speed, the observed variation is only slightly reduced. In other words, it is not an effect due to the variation in bulk speed, at least not to any great extent.

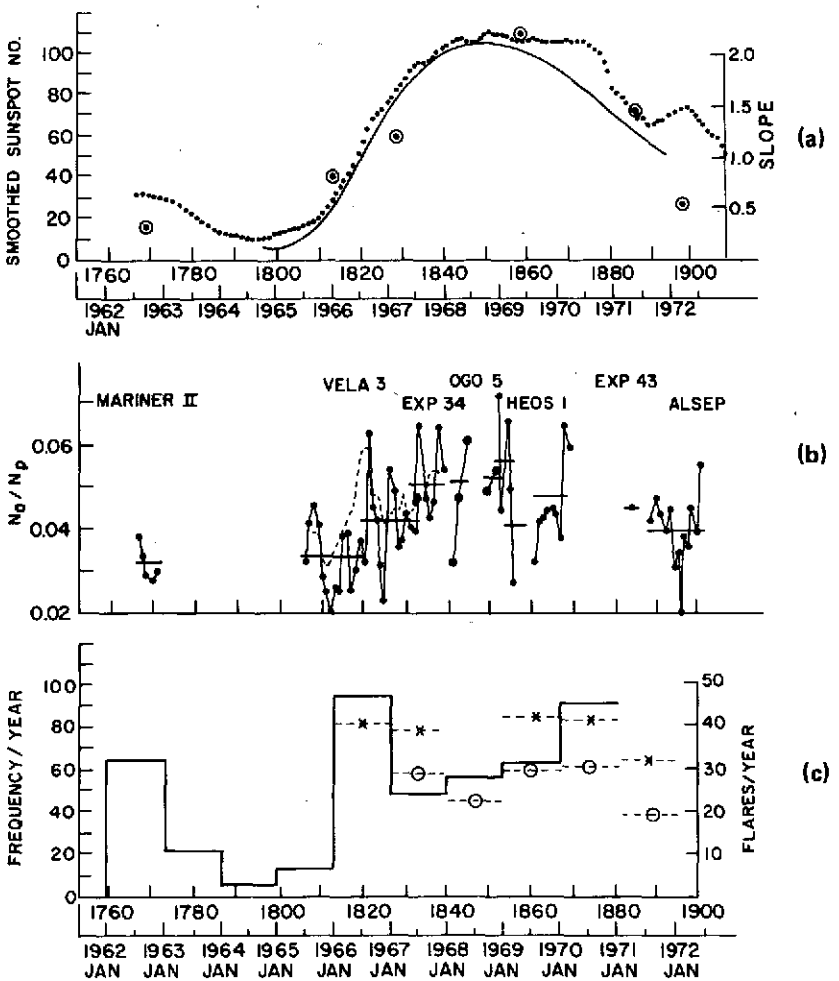


Figure 2. (a) Smoothed sunspot number; (b) Solar rotation averages of Nα/Np from various experiments; (c) Solar flares per year.

Using representative values for the size and duration of major solar-flare-associated increases, we can show these to be insufficiently frequent at solar maximum to increase the average value of Nα/Np by as much as 50 percent. So our present hypothesis is that another process must produce increases which are not recognizable as distinct impulsive events at 1 AU, though they may be associated with the kind of impulsive events which are seen by Skylab. The rate of impulsive events that are seen there are probably sufficient to explain this data.

SHOCK WAVES FROM THE SUN

Leonard F. Burlaga

Shock waves are a key link between solar processes and terrestrial processes. One of the oldest and perhaps most important questions that you can ask about such shocks is "How do they propagate from the sun to the earth?" There is a large literature on this subject, but essentially all of it considers the motion of a shock into an unstructured medium. Work by members of our group showed that even the most sophisticated of these theories does not qualitatively describe most of the observations. A recent study by Ogilivie and myself showed that if you want to understand post-shock flows, then you have to consider the interaction of shock waves with interplanetary streams, which are always present.

This talk concerns a model built by Jack Scudder and myself to describe how a shock moves through an interplanetary stream. The elements of the model are illustrated in figure 1.

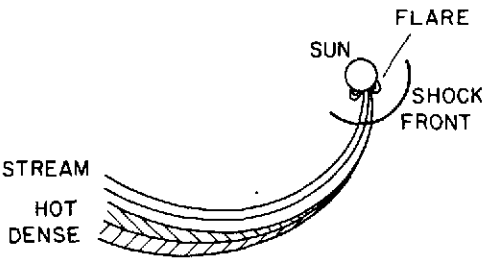


Figure 1. Model describing how a shock moves through an interplanetary stream.

A shock wave starts out near the sun, presumably the consequence of a flare at the sun. This moves through a stream, a region of fast plasma which is corotating with the sun. Ahead of the fast stream, material is heated and compressed as a result of the stream interaction; generally, the dense material precedes the hot region. We have shown in an earlier study that this is a natural consequence of generating a stream by addition of heat in the solar envelope. So the question at hand is, "How does the

shock wave move through a configuration like this." The answer is given in figure 2.

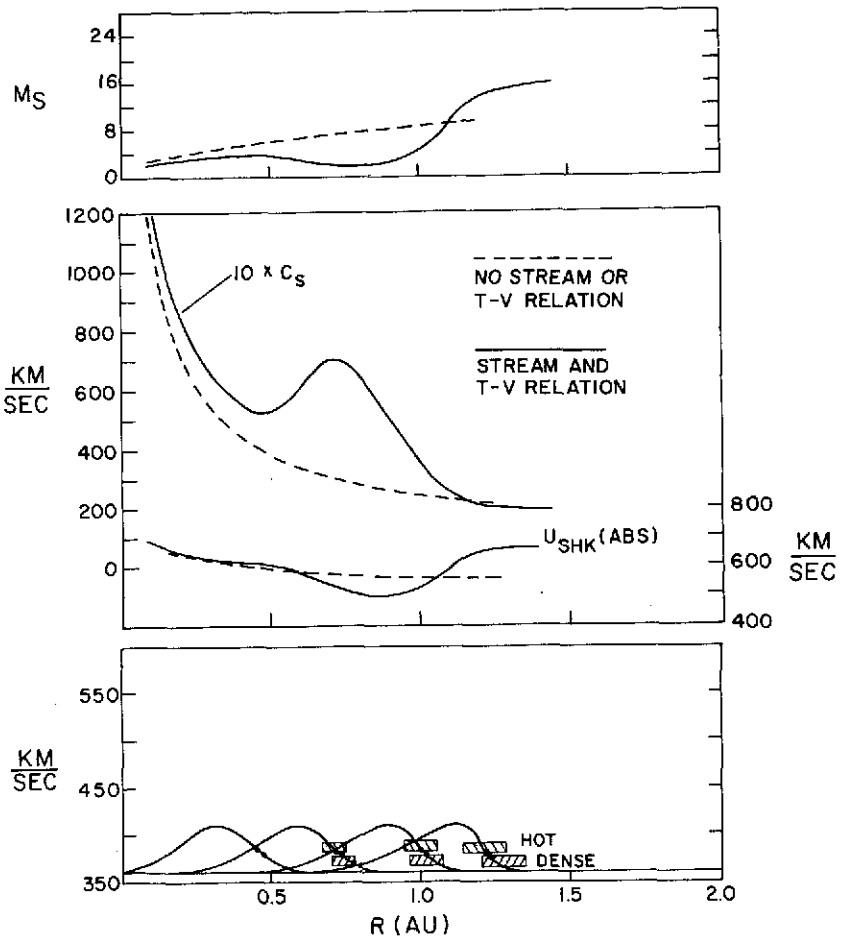


Figure 2. Evolution of a stream and the variation of shock speed and strength as it moves through the stream.

I will digress for a moment to describe how a shock moves in the absence of a stream, for reference. The abscissa gives the distance from the sun, the ordinate gives the shock speed in kilometers per second. The results in the absence of a stream are shown by the dashed lines. The shock simply decelerates slowly as it moves away from the sun in the absence of a stream. The temperature profile as seen by the shock is shown by the acoustic speed. It simply decreases monotonically with distance from the sun, as in

the model shown in figure 2. The shock strength measured by the Mach number simply increases monotonically as the shock moves away from the sun, in the absence of a stream.

Now let us ask what happens when we have a shock moving through a stream. We model the stream as shown in figure 2. Near the sun we introduce a Gaussian profile and this, as it moves outward, simply steepens kinematically as the faster part overtakes the slower part. As a result of this steepening, material is compressed and heated as I have shown schematically in figure 2, leading to a pressure pulse in the interaction region. Now, if we use our model to follow the motion of a shock through this configuration, we see the following results. Near the sun, the stream starts out ahead of the shock, but the shock is moving faster so it overtakes the stream. Eventually the shock reaches the stream maximum where it actually moves a little faster than it would if there were no stream, just because the stream is carrying it along. Now, as the shock moves on into the interaction region, it slows down as a consequence of the pressure pulse. Then it moves through the interaction region and accelerates to a faster speed. The Mach number, in this case, shows a very interesting feature, namely, an abrupt decrease as the shock moves through the interaction region.

There are three results: As the shock moves through the stream, it accelerates near the stream maximum; it decelerates in the interaction region; and the Mach number can decrease to a very small value. This has a number of very interesting consequences. The effect on the shock geometry is shown in the next figure.

The acceleration that I mentioned above causes an outward bulge, as seen in figure 3. The deceleration in the interaction region causes a retardation and therefore an inward bulge. The result is a large distortion from spherical symmetry. Thus, all the previous analyses, which assumed spherical symmetry, are not valid. This also has significance for comet studies, for example, when attempting to relate shock observations to disturbances in cometary tails.

The decrease in Mach number that I mentioned has at least two interesting consequences. One is that the shock wave can degenerate essentially into a nonlinear wave as it moves through the interaction region, and then as it moves on ahead the wave can steepen once again into a shock. An example of such a process has been demonstrated by Dr. Chao. Another possibility is that the Mach number may actually go to unity, in which case the shock may disintegrate. In this case, we might not see a correlation between flares and interplanetary shock waves.

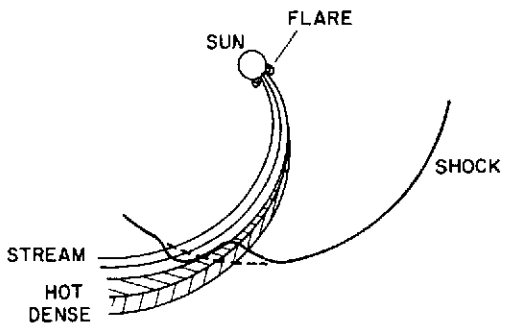


Figure 3. Distortion of shock surface caused by interaction with a stream.

To summarize, the work that I have described provides some new insights and new answers to the old question, "How does a shock move from the sun to the earth?"

RADIO TRACKING OF PARTICLES FROM THE SUN

Joseph Fainberg

I would like to show how observations of solar radio bursts can be used to track the motion of energetic particles from regions close to the sun out to the earth's orbit. These solar radio bursts are thought to be caused by groups of energetic electrons with energies above 5 keV, which are abruptly accelerated in the low corona. As the electrons move out along the solar magnetic field lines swept out by the solar wind, they excite plasma oscillations in the solar wind. A small part of this energy is converted into radio waves that propagate freely and which can be used as a remote means to get information on the many physical processes involved.

Goddard's radio experiment on the IMP-6 spacecraft has permitted these radio observations to be made over a wide frequency range, corresponding to emission occurring from close to the sun out to 1 AU. Figure 1 is an example of a type 3 solar radio burst observed with this experiment. Plotted here are a series of envelopes of the burst intensity profiles for frequencies from 950 kHz at the top of the figure down to 30 kHz at the lower part of the figure. Since the lower frequencies are omitted in less dense regions of the corona, they are produced further out from the sun and occur later in time. Because of the satellite rotation, the dipole reception pattern of the antenna is swept past the direction of the source many times during the course of the measurement. This results in a spin-modulated pattern which is shown in the upper right-hand part of figure 1. We have developed a technique using this spin modulation to determine the direction of arrival of the radio radiation. With several minutes of data, we can determine this direction of arrival to within several degrees of accuracy, and we can make this determination many times during the course of the burst. From a study of limb events, we can also determine an emission level scale which tells how far out from the sun the radiation at a given frequency originates. For example, the 250-kHz radiation originates at a level of about 70 solar radii from the sun.

Using the emission level scale and the measured direction of arrival, we can then locate the radio emission regions. Figure 2 illustrates this procedure. The plane of this figure is the ecliptic plane with the sun at the center and the IMP-6 spacecraft and the earth located at the bottom. Also shown are

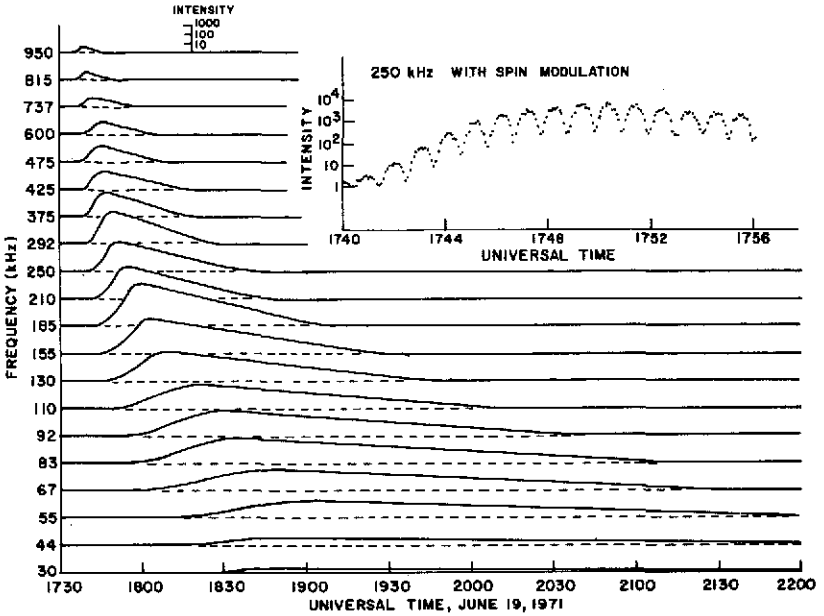


Figure 1. Type 3 solar radio burst observed with GSFC experiment on IMP-6.

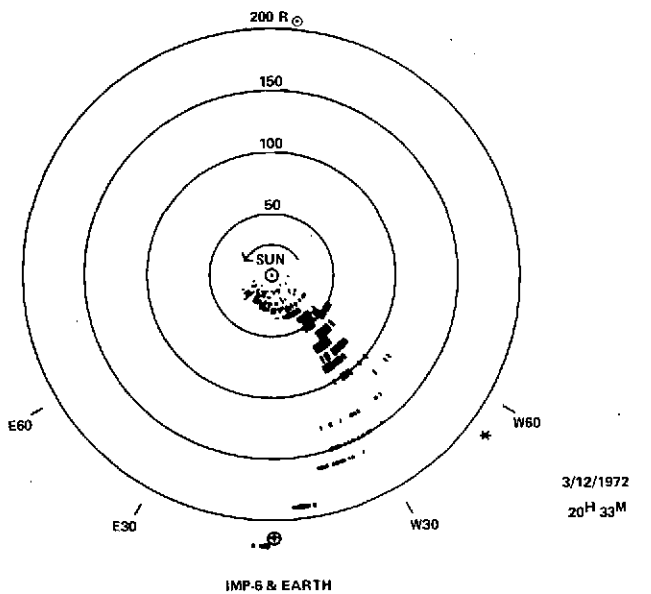


Figure 2. Location of radio emission regions.

concentric circles, separated by 50 solar radii out to 1 AU. The direction of the solar rotation is indicated by the arrow. For each minute of spin-modulated data we determine the direction of arrival of the radiation, and with the emission level scale we determine the intersection point. At this intersection point, which is the location of the emission region, we plot a radial line segment whose length is proportional to the log of the intensity of the radio signal. Figure 2 is a collection of these intersections for 32 frequencies measured in a 2-minute interval. This 2-minute snapshot, then, shows a solar burst in progress with the main part of the burst located between 40 and 80 solar radii; the faster particles of the exciter have already moved out to 1 AU. The plot of the radio burst trajectory then outlines the spiral interplanetary magnetic field. In addition, there is a long-term radio storm present within the first 50 solar radii which lasts for several days.

A sequence of figures like this can be used to study the dynamic processes involved, and we have assembled such a collection in which I would like to show this burst and several others. Figure 3 illustrates the progress of a radio burst occurring on March 12, 1972, from regions close to the sun to

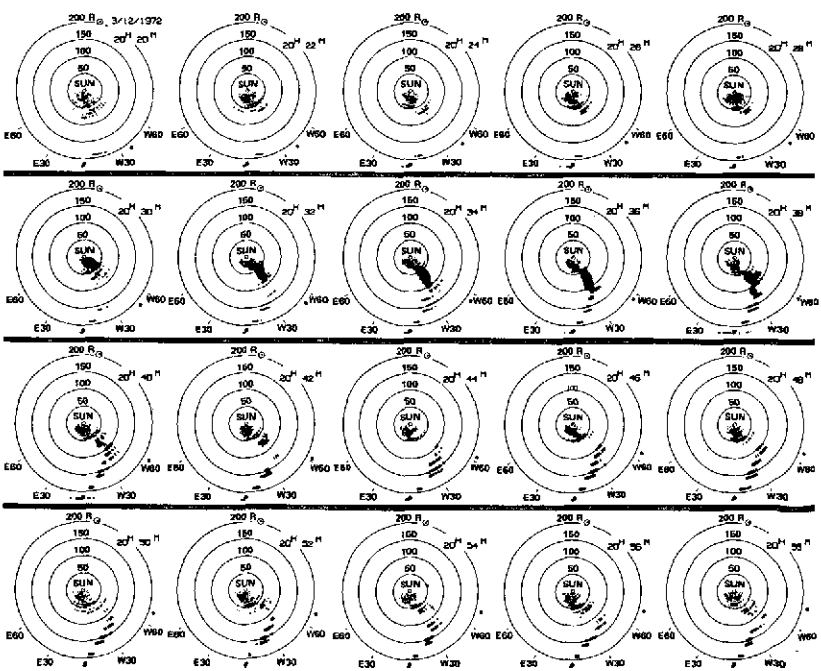


Figure 3. Progress of radio burst that occurred March 12, 1972.

past 1 AU. A solar storm which lasts for days is visible during this entire sequence within the first 50 solar radii from the sun. At about $20^h\ 26^m$ a large individual solar radio burst first became visible close to the sun. Each successive frame of this selection from the movie is separated by 2 minutes and shows the radio emission generated by the energetic particles moving out from the sun. Note that this is very analogous to the way the path of a charged particle is made visible by its ionization loss in a cloud or bubble chamber. Near $20^h\ 36^m$, the spiral structure of the interplanetary magnetic field caused by the solar rotation is quite apparent. At still later times the region of intense excitation moves progressively outward and finally passes 1 AU. From a study of other bursts such as this in conjunction with data from the University of California experiment on IMP-6, the burst exciter particles have been identified as electrons in the 5- to 200-keV energy range. It is clear that low frequency radio astronomy studies of this type provide a new source of information about remote regions of interplanetary space.

MODELS OF THE MAGNETOSPHERIC ELECTRIC FIELD

David P. Stern

When the radiation belt was first discovered, people tried to explain its properties using only the earth's magnetic field. Some of these explanations worked well–like the theory of neutron albedo origin for the inner belt–but they did not explain the ease with which flows of plasma often seemed to violate the rules of magnetic confinement. The reason, we know now, is that there also exists an electric field, which strongly influences the motion of low-energy particles. The trouble is, of course, with our instruments– they easily observe magnetic fields down to very low levels, while the electric field is so weak that only recently did we get good quantitative measurements of it. This is a report about model representations of this electric field and of application of such models to explain observations. However, let me precede this with a few words about this field, for those who might not be familiar with the details.

First of all, how does this field originate? To get an electric field in space, one usually needs a bulk flow of one part of the conducting medium relative to other parts. Viewed from the frame of the solar system, two such flows can contribute electric fields, and apparently do: namely, the flow of the solar wind and the rotation of the earth and of its ionosphere.

In the ionosphere, around the magnetic poles, there exist two roughly circular regions with a radius of some 1120 to 1440 km (700 to 900 miles)– they are usually called the polar caps– from which magnetic field lines seem to connect directly to interplanetary space, as shown by solar particle precipitations and other effects. Figure 1 shows one model of the magnetic field which was used and which contains such a connection. You can see the northern polar cap and also–as a broken line–one of the interplanetary magnetic field lines leading to it. Such field lines can carry the interplanetary electric field to the polar caps and indeed, the electric field there is stable and on the average points from dawn to dusk with a total drop of 40,000 V, although there also exist considerable fluctuations. The open field line drawn in the figure is considerably distorted. A study was made of the topology of such distortions and their effect on the structure of the electric field, and the results explain two polar effects correlated with interplanetary sector structure. One is due to Dr. Svalgaard's work in Denmark and one was found with OGO-6 by Dr. Heppner of Goddard. The details appear in the November issue of the *Journal of Geophysical Research.*

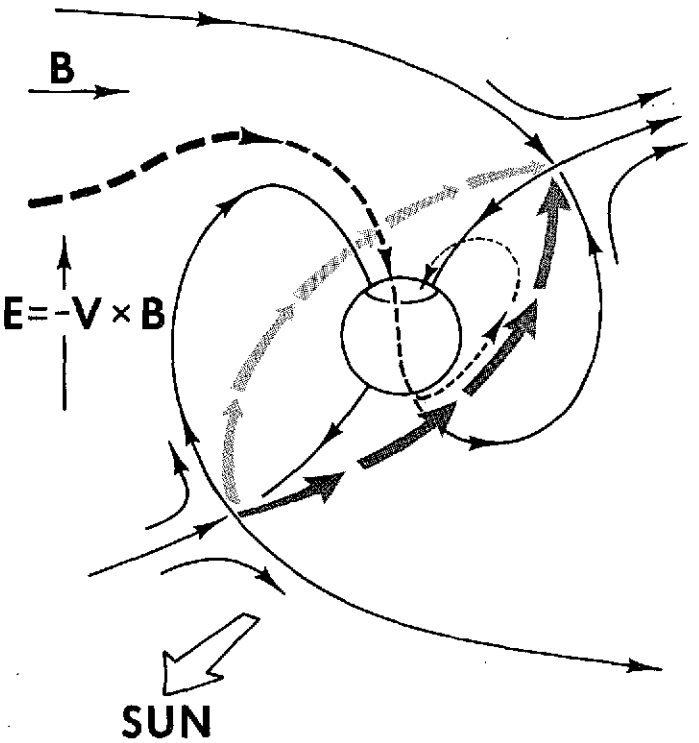


Figure 1. Model of the magnetic field in which magnetic field lines connect directly to interplanetary space.

Because the polar caps are covered by a conducting ionosphere, their electric fields will spill over the edges, so to speak, and extend to closed field lines. In these models the electric field can be described by a potential function, and then, if the polar electric field is assumed to be constant and directed from dawn to dusk, the polar equipotential lines run parallel to the noon-midnight direction, as shown on the left in figure 2. If the ionosphere were an ohmic conducting sheet, then the spill-over gives the pattern on the right in figure 2. There is some distortion due to anisotropic conductivity, which we neglected, but we did introduce an adjustable parameter describing the rate at which the field falls off near the polar cap boundary, choosing for it a value that fits the OGO-6 data.

Magnetic field lines from such fringe areas lead to opposite hemispheres and, by the nature of electric conductivity in space, tend to be equipotentials. This lets one map the ionospheric potentials all over space, and there exist mathematical tools for this—the Euler potentials α and β. The condition

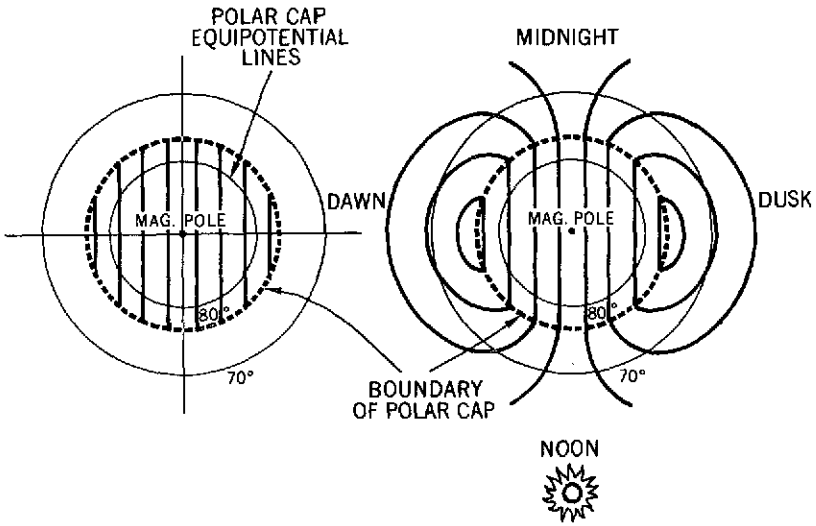


Figure 2. Electrical equipotential lines in and near the polar cap.

that field lines be equipotentials reduces to the demand that the electric potential is a function of α and β.

We now want to see how charged particles move. At low energies, such particles stick to surfaces of constant potential and accordingly may or may not be trapped, depending on where they start out. The boundary between the two types of motion is generally identified with the plasmapause, and Dr. Chen has calculated that this model represents the plasmapause particularly well.

With higher energies, magnetic drifts enter and complicate the analysis so much that one usually confines attention to equatorial particles. An interesting application is shown in figure 3, for which I am grateful to Dr. Paul Smith. Figure 3 shows the equatorial proton energy spectrum observed aboard Explorer-45 at 5 earth radii. There is a peak around 200 keV and a dip around 20 keV, both explained quite well by Paul Smith and Bob Hoffman by invoking charge exchange. But at lower energies, where charge exchange should have wiped everything out, you can see a big peak, too.

When one analyzes the motion of such protons, one finds that some orbits are trapped and others are open to the tail. For protons in particular—not for electrons—some of these open orbits penetrate quite far, because for them the electric corotation drift and the magnetic drift oppose each other. If these cancel out, this leaves only the drift due to the convection field spilling out from the polar caps, and this tends to bring the protons inward.

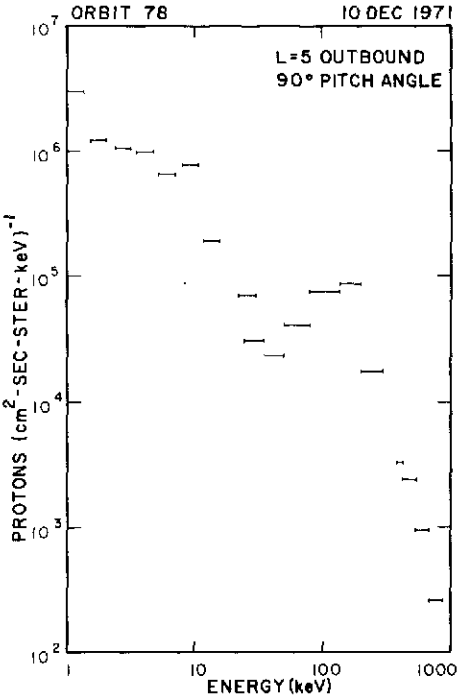


Figure 3. Equatorial proton energy spectrum observed aboard Explorer-45 at 5 earth radii.

We have developed a rather simple graphic method for identifying which energies correspond to such orbits for any observation point, and this seems to indicate that the low-energy peak shown in figure 3 indeed contains open orbits.

Let me end with a rather tantalizing conjecture, less than 2 weeks old. As said before, the mechanism shown here can drive protons deep into the magnetosphere, where they can contribute to the ring current which, indeed, is mainly due to protons. Now for the last 2 weeks we have known that Jupiter also seems to have a ring current and a rather powerful one, capable of distending its magnetic field to an extent observed near earth only in the tail region. Furthermore, it appears as if the major component of Jupiter's trapped radiation consists of electrons. It is interesting to note that the dipole polarity of Jupiter has an opposite orientation (upside down) relative to its rotation axis than is the case with the earth, so the same mechanism of partial cancellation of drifts, which brings protons earthward and energizes them, would in Jupiter's case act on electrons. Maybe that is the reason why Jupiter's magnetosphere seems to differ so much from that of the earth.

ELECTRIC FIELDS IN THE MIDNIGHT SECTOR OF THE MAGNETOSPHERE

Nelson C. Maynard

It has been recognized for some time that during magnetic substorm activity a definite boundary or fault line exists in the midnight region near local magnetic midnight. Across this region many phenomena change from a characteristic late evening behavior to a totally different early morning behavior. In table 1, I have summarized a few of these characteristics. The list is easily expanded. Ionospheric electric currents, primarily Hall currents, are known to reverse in the midnight sector from a west-to-east to an east-to-west pattern. A net field-aligned current is believed to flow out of the ionosphere in this region. The magnetic disturbances resulting from the currents change polarity near magnetic midnight, and the auroral behavior changes from stable quiet arcs to more active rayed arcs, later changing to diffuse patchy aurora. Note that the most intense magnetic activity and aurora occur just on the morning side of the boundary. The convective motion of auroral forms and the electric field are also reversed in the early morning region compared with the late evening region.

In the above we have mentioned a few phenomena that are occurring at the foot of the magnetic field lines. Out in the tail of the magnetosphere, this boundary projects into the area where substorm phenomena such as tail collapse, neutral sheet thinning, and particle injection and energization are occurring. Thus, this boundary serves as the focal region of much of the substorm activity both in the ionosphere and the magnetosphere. The average character of the boundary as seen in ground magnetic disturbances was investigated by Harang in the 1940's using statistical studies of magnetograms. Recently this boundary has been dubbed the "Harang discontinuity." The purpose here is to study the characteristics of the Harang discontinuity as revealed by electric fields.

The electric field can be used to define the bulk motion of magnetic field lines and accompanying plasma from convection. Figure 1 shows two schematic representations of the polar regions; 1a shows the direction of convective flow and 1b shows the electric field direction associated with that flow. The boundaries in each view represent averages of those seen by the electric field experiment on OGO-6 during moderately-disturbed to disturbed times. The lower latitude boundary is defined as the place where

Table 1
Magnetic Substorm Characteristics

Substorm Phenomena	Late Evening	"Midnight" Boundary	Early Morning
Ionospheric:			
Electric Currents	W E	Current Reversal	E W
Magnetic Disturbance	Weak to moderate (+ΔH)		Intense (-ΔH) Decreasing
Aurora	Homogeneous arcs		Brilliant rayed forms Diffuse and patchy forms
Auroral (Convective) Motions	E W		W E
Electric Field	Northward		Southward
Magnetospheric:			
Magnetic Disturbance and Particles		Tail collapse Neutral sheet thinning Particle injection and energization	

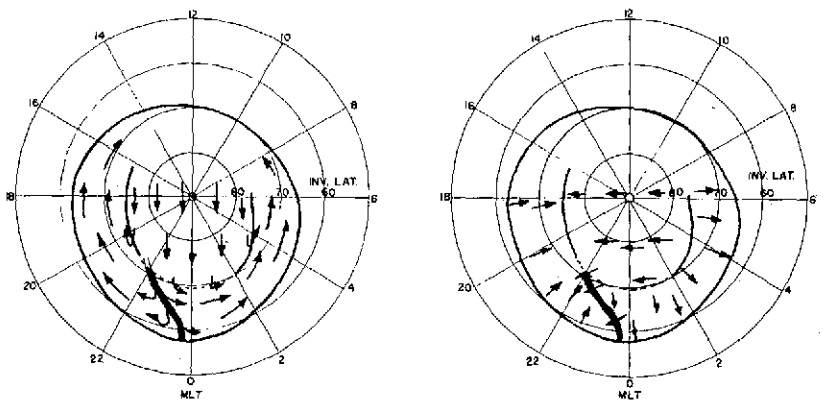


Figure 1. Schematic representations of the polar regions. (a) Direction of convective flow; (b) Electric field direction associated with (a).

significant electric fields are first seen, and the upper boundary defines the reversal of the electric field as we move to the region of antisunward convection on the open field lines across the polar cap.

Note the boundary between the evening and morning regions appearing near midnight. As one crosses this Harang discontinuity heading poleward,

the electric field rotates through the west from a northward to a southward direction, appearing as a reversal in a single axis measurement. A second smaller change of direction occurs as one crosses into the polar cap. Thirty-one Harang reversals have been identified from the OGO-6 electric field data during November and December 1969. The data set is limited by the facts that only one component was measured and that attitude information in satellite eclipse was often inadequate. It has been shown from this data that:

- The boundary is a dynamic one and can shift over several hours in local time.
- In active periods the boundary moves south and steepens its latitude profile.
- As activity decreases the boundary region spreads out as the effects tend toward corotation and relax poleward.

The most significant point from this study is that the Harang discontinuity exists in the electric field data in extreme quiet conditions as well as during substorms. Figure 2 shows the electric field measured during a crossing of the boundary at a time when no significant substorm activity had occurred over the previous 9 hours and generally quiet conditions had prevailed for almost 2 days. The electric field strengths are strong on both sides of the boundary, which is denoted by the black bar near 73°. The positive directions of the electric field component being measured are shown by the

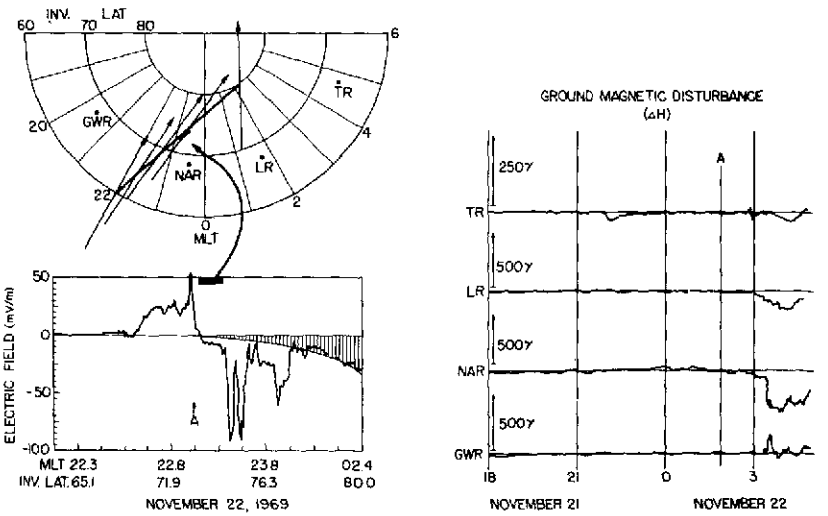


Figure 2. Electric field measured during a boundary crossing at a time of low substorm activity.

arrows in the polar plot above. Note that all four ΔH traces from the ground magnetic observatories at Tromso, Lervoguir, Narssarssauq, and Great Whale River are at zero. Station locations are shown on the polar plot.

Thus, the electric field becomes an ordering function or pattern that is present in both quiet times and during substorm activity. The Harang discontinuity or fault line in the electric field defines the location in the ionosphere and magnetosphere from which events can be triggered to create the phenomena described earlier. The triggering mechanisms and what controls the dynamics of this discontinuity are still open questions.

MAGNETOSPHERIC RING CURRENTS

Robert A. Hoffman

In 1961, Axford and Hines first proposed that the convection system of the magnetosphere could supply and energize protons for the magnetic storm-time ring current, which produces the main phase magnetic variation during magnetic storms. Their theory has been considerably developed in the intervening years by a number of workers and is now generally accepted in principle. However, until the acquisition of data from the S^3, Explorer-45 satellite, there had been almost no direct experimental measurements of the characteristics of the storm-time ring current protons which could substantiate the theory.

The qualitative aspects of the theory are displayed in figure 1, which shows the general flow patterns of particles resulting from a combination of three types of motion. The first type is the inward convection of protons from the plasma sheet due to the northward geomagnetic field in the equatorial region of the earth and the cross-tail electric field. This radial drift velocity is independent of the energy and species of the particles.

The second type of motion is the gradient drift of high energy protons ($E > 10$ keV) to the west and around the earth. The velocity of the drift is proportional to the energy of the protons. This drift in local time, and especially the diamagnetic characteristics of this body of protons, produce the westward toroidal ring current. For this current the right-hand rule of thumb indicates that the magnetic field inside or earthward of the toroid is in a southern direction, opposite that of the main field of the earth. At the onset of a magnetic storm, the toroid is incomplete, so the evening hours on the earth observe the largest magnetic field deformation. Detailed analysis of the particle motions shows that the lower energy protons penetrate deepest into the magnetosphere, or more exactly, the lower magnetic moment protons penetrate deepest. This is the first important characteristic to remember.

The third motion is due to the rotation of the earth: The magnetic field lines close to the earth also rotate, carrying particles to the east around the earth. For very low energy protons this rotation speed to the east is faster than the gradient drift to the west, so these protons will be missing from the evening hours at the onset of a storm. This is the second important characteristic.

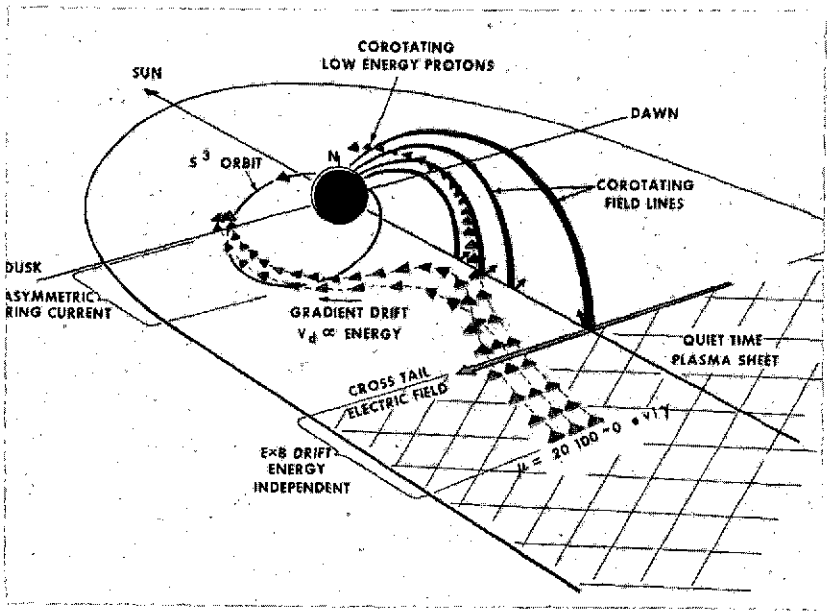


Figure 1. General flow patterns of protons.

During the early life of the satellite, the S^3 orbit intersected the region of storm-time protons on the evening side. We next look at the proton observations from S^3 in these evening hours. Figure 2 is an energy spectrogram taken at the beginning of a magnetic storm main phase. Energy is plotted on the ordinate, time, or increasing radial distance on the abscissa. The gray shading is a measure of the proton flux, with black being most intense. Note the following features which have been characteristically seen in the evening hours at the beginning of main phases.

First, protons are present with high intensities in the evening hours; in this case the observation was made an hour and three quarters after the beginning of the main phase. Next, for protons with energies above about 20 keV, the lower energy protons penetrate deepest into the magnetosphere. Finally, at a given L value, protons below some characteristic energy are absent. As I just pointed out, these features are predicted by theory.

In figure 3 we show that protons are not only carried into the ring current region by convection but are also energized as a result of this motion. One of the constants of motion for particles is the first invariant, the magnetic moment, $\mu = E/B$, which is the energy of the particle from motion perpendicular to the magnetic field line divided by the magnetic field at the location

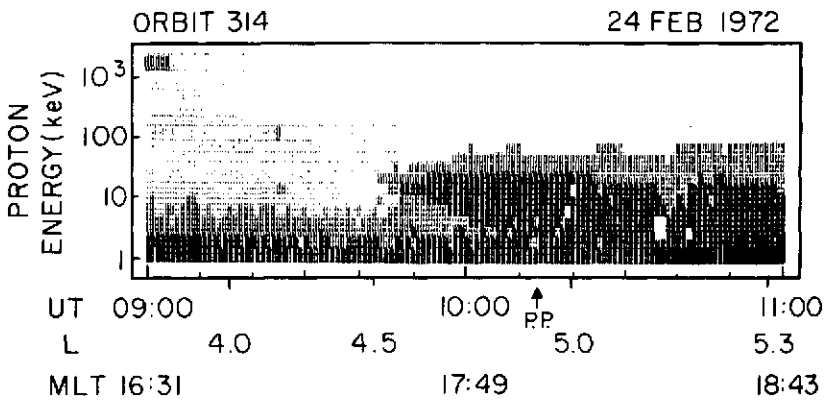


Figure 2. Energy spectrogram taken at the beginning of a magnetic storm main phase.

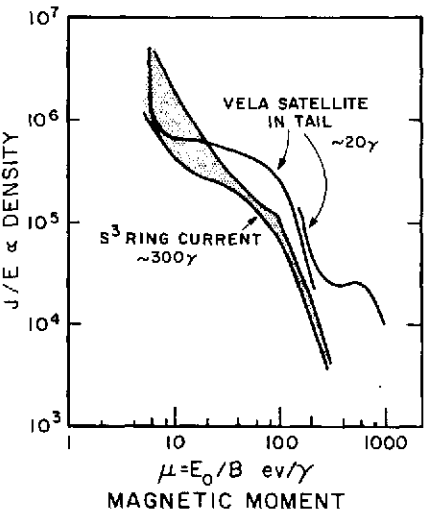


Figure 3. Density of protons as a function of the magnetic moment for ring-current and plasma sheet protons.

of the particle. This is a constant. In the plasma sheet in the tail of the magnetosphere, the magnetic field strength is from 5 to 50γ, whereas in the ring current region it is several hundred γ. Thus by convecting protons from the low magnetic field region of the tail into the fairly high field region of the ring current, energizations of at least a factor of ten will occur.

To show that there are sufficient protons in the plasma sheet to account for the ring current protons, we plot in figure 3 the density of protons as a function of the magnetic moment for both ring current protons, measured by S^3, and plasma sheet protons, which have been measured by the Vela satellite. This method of plotting adjusts the energy scale so that spectral comparisons can be made in two different regions of the magnetosphere. In both regions the densities vary considerably, although in fact there are only a couple of proton spectra available to date from the plasma sheet. However, note that, in general, there are sufficient protons in the plasma sheet to account for the ring current protons after energization from convection into the deeper regions of the magnetosphere.

In conclusion then, we see that, qualitatively, the proton distributions during the development of main phase magnetic storms have the characteristics predicted by convection theory, and that the plasma sheet in the tail of the magnetosphere contains sufficient densities to account for the ring current protons. Finally, it should be pointed out that considerable theoretical and experimental efforts are still required to solve, quantitatively, the ring current problem:

- Measurements and analysis are required to determine the dynamic characteristics of the magnetospheric electric field patterns from electric field and particle measurements.
- Calculations should be carried out on particle trajectories in the real magnetic and electric field configurations (as well as we know them), taking into account temporal variations.
- Detailed analysis still needs to be performed on the S^3 and related data to confirm the theory and calculations.

DYNAMICS OF THE THERMAL PLASMA IN THE MAGNETOSPHERE

Abel Chen

The role of thermal plasma in the ionosphere and magnetosphere system has been overlooked for some time, mainly because of the energy density being too low compared with other plasma constituents in the magnetosphere. In reality, this cold plasma which surrounds the earth in the region called plasmasphere has direct or indirect effects on the balance of energetics in the thermosphere, wave-particle interactions in the magnetosphere, and magnetospheric convection, which is a major energization process in the magnetosphere.

We will discuss the latter of these, the connection between magnetospheric convection and thermal plasma dynamics. Generally speaking, the plasmasphere is observed to change its density abruptly at the plasmapause, a field-aligned boundary which involves a density drop of one or two orders of magnitude. However, one often observes density enhancement near or outside the main plasmasphere. This irregularity, called plasmatail, has been simultaneously predicted by Grebowsky (1970) and Chen and Wolf (1970). In fact, the first observational evidence of the plasmatail was presented by Dr. Grebowsky at this meeting 2 years ago. Since then we have found several more cases in mass spectrometer data obtained from OGO-4 and OGO-6 satellites. Today I would like to present one simple model developed by Dr. Grebowsky and myself to account for the observations.

In this model, the magnetic field is assumed to be dipolar and the field lines beyond 10 earth radii are assumed to be open to the magnetopause. The convection electric field is taken from Dr. Stern's model in which the field direction is more or less from dawn to dusk and the magnitude increases with radial distance. However, the electric field also varies in step with the Kp index. The other component of the electric field is that of corotation. The flux tube at a given time and location is traced backward in time to see how long the flux tube has been closed and on dayside.

As a primary source of thermal plasma originates from dayside ionosphere, presumably the longer the dayside closure time, the higher the density would be. The steady-state flow pattern is shown in figure 1. The flux tubes inside that shady region circle around the earth all the time while the

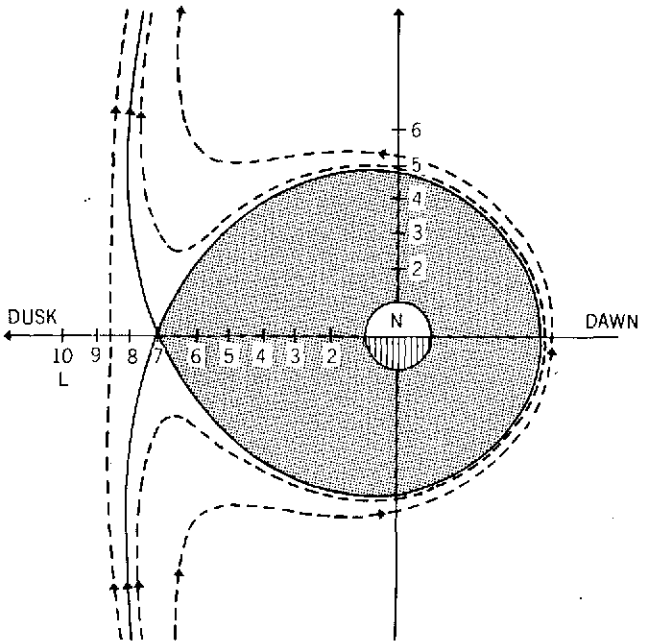
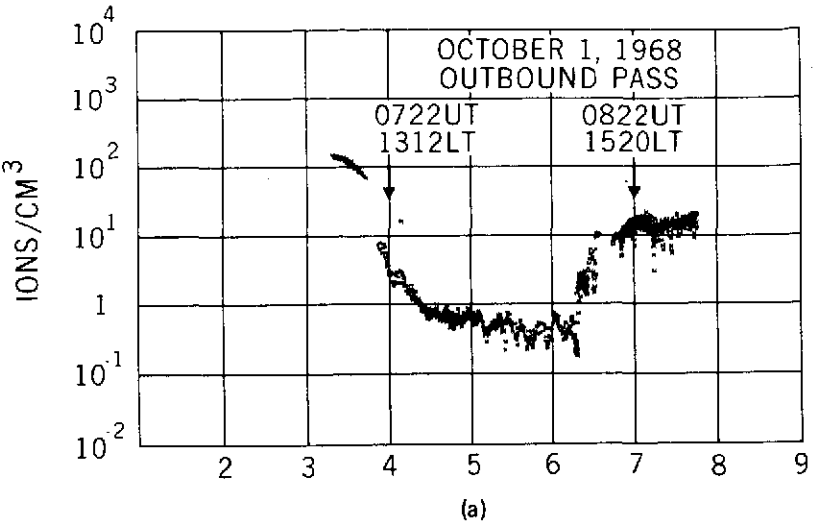


Figure 1. Stern's model, steady state flow pattern.

flux tubes outside stay several hours on dayside before they are convected to the magnetopause, losing their thermal plasma in the process. Consequently, the density inside the boundary would be much higher than that outside. The shape of the closed boundary becomes more complicated when the convection field starts varying in time. We will present two examples in the next two figures.

Figure 2a shows H^+ density plotted against radial distance. We can see that density drops sharply at about 4 earth radii, and remains so until the density increases at 6.3 earth radii. The computed dayside closure times are given in figure 2b. The region with 6-day closure time represents the densest region. The density decreases with shorter closure time. The satellite can be seen to leave the main plasmasphere near local noon, enter the partially filled region where density is expected to be low, and reenter the dense region again near about 1530 local time in general agreement with observation. The structure in this case is rather simple and a more complicated one is shown in figure 3.

Again, in figure 3a, we can see that density drops sharply at about 3.6 earth radii and density starts fluctuating significantly from 5.3 earth radii outward.



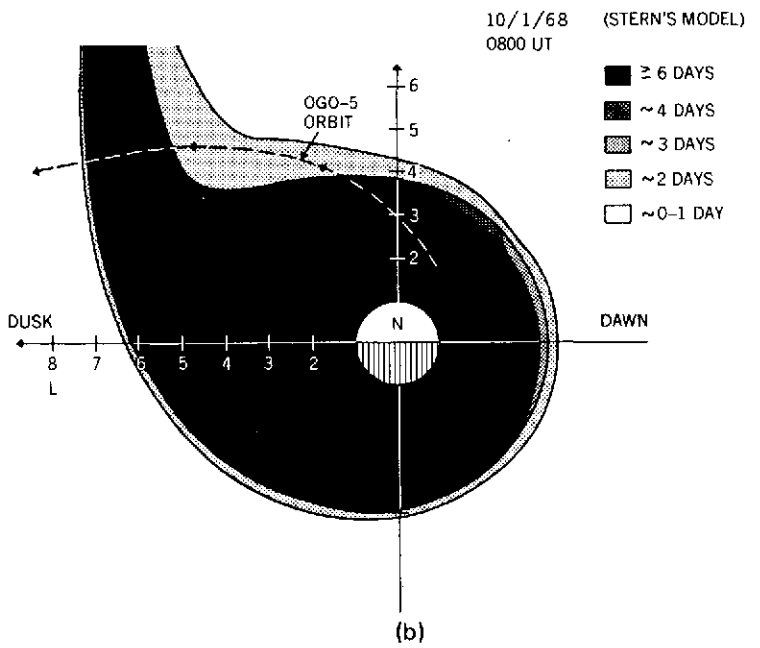


Figure 2. (a) Density versus radial distance; (b) Computed dayside closure times.

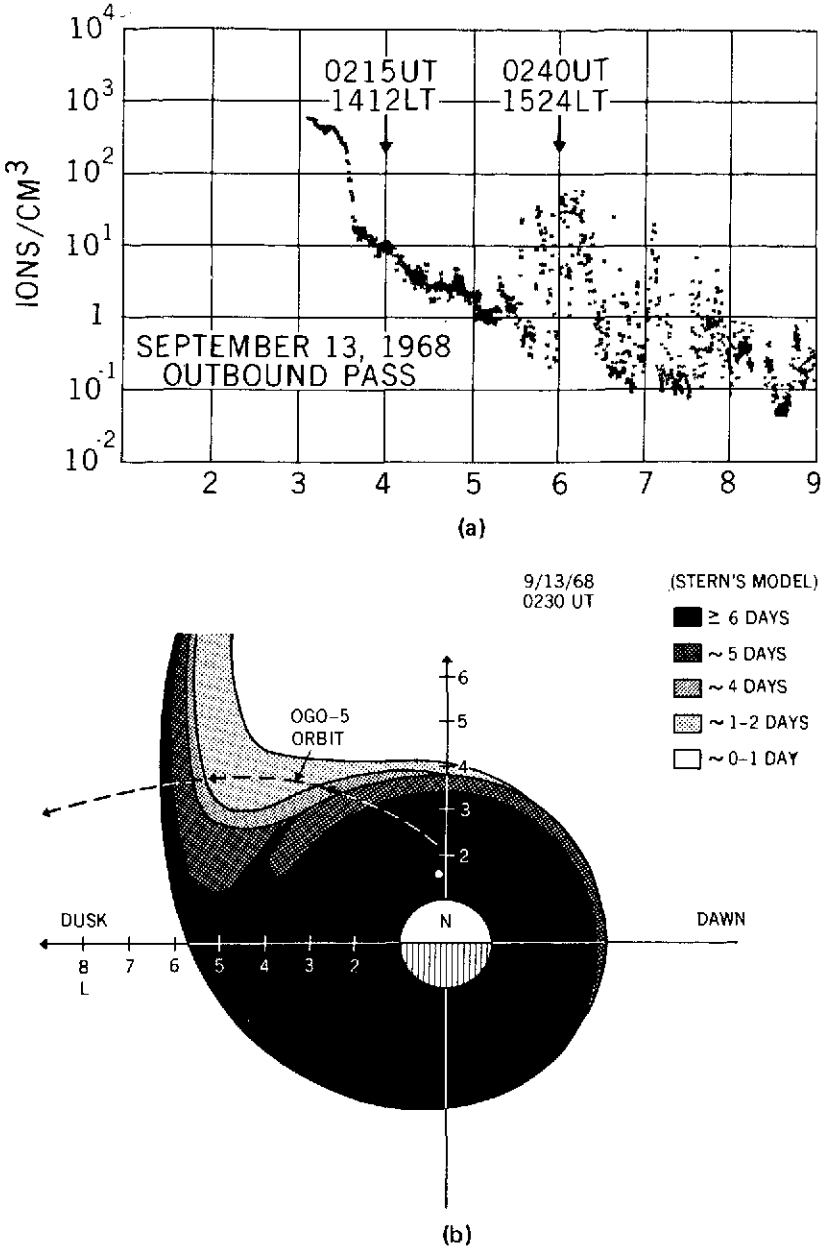


Figure 3. (a) Density versus radial distance; (b) Computed dayside closure times.

The computed closure times are given in figure 3b. We can see that the satellite leaves the main plasmasphere and cuts through regions with substantial variability in closure times. Consequently, the density would be characterized by large fluctuations.

In conclusion, it appears that the observed irregularities outside the main plasmasphere are caused by the fluctuations of the convection electric field in the form of plasmatail, leading from the main plasmasphere toward the magnetosphere.

SOLAR AND CORPUSCULAR HEATING OF THE UPPER ATMOSPHERE

Sushie Chandra

It is generally recognized that the two main sources of energy responsible for heating the upper atmosphere are solar corpuscular and solar ultraviolet radiations. Their relative importance under varying geophysical conditions have, however, been a subject of considerable discussion which has often generated more heat than light.

From the observational viewpoint, the situation has not been very helpful because of the difficulty of measuring thermospheric temperature on a global scale. The temperatures inferred from satellite drag data, which until recently has been our main source of information about the neutral atmosphere, appear to be a very poor index of the thermodynamic state of the upper atmosphere. This is because most of the satellite drag data are obtained in altitude regions where atomic oxygen is the principal neutral constituent. In deriving thermospheric temperature from the drag data, it is explicitly assumed that atomic oxygen concentration, along with other neutral constituents, is invariant at the turbopause level. This, in fact, is a very questionable assumption. Theoretical calculations indicate that both atomic oxygen and helium are subject to considerable variability due to changes in thermal and dynamic conditions in the lower thermosphere. Compared to these two constituents, molecular nitrogen is less variable. The least variable of all is argon, which is a minor constituent in the entire region of the upper atmosphere. Thus, temperatures inferred from nitrogen and argon should be closer to the gas temperature than those inferred from atomic oxygen and helium. The usefulness of molecular nitrogen in driving thermospheric temperature has already been demonstrated from OGO-6 neutral mass spectrometer measurements. I shall attempt to show here that argon can also be included in this list.

The data which I am going to present here are obtained from the AEROS NATE experiment flown by Nelson Spencer of Goddard. The AEROS satellite, as some of you know, is a joint venture of NASA and the German Space Program. The satellite was launched in December 1972 in a polar orbit with a perigee of about 230 km. During the initial phase of its lifetime, the perigee was in the winter pole. This situation was very useful from the

point of view of studying the specific role of corpuscular heating, since the polar atmosphere was usually devoid of any solar ultraviolet radiations.

We may start with figure 1 which shows the exospheric temperature inferred from argon, nitrogen, and oxygen gasses for two typical days, January 2 and 3, 1973, which are quiet days with the three hourly Kp values of about zero, and January 10, 1973, a moderately disturbed day with Kp values of about 4. The temperatures inferred from the various gasses are plotted against the invariant latitude. The magnetic north pole is indicated by an arrow which also happens to be the dividing line of the day and night conditions; the left-hand side refers to about 3 hours local time and the right-hand side to about 14 hours. In the middle region, the local time is changing rapidly from about 4 hours to 13 hours even though the solar zenith angle in this region is about 100° which means almost no solar radiation.

The first point to note here is that during quiet conditions, the neutral temperature in the auroral and polar regions is generally lower than anywhere else, with the lowest temperature in the region of 70 K invariant on the nightside. The temperatures inferred from all the three gasses show a monotonic increase with solar zenith angle on the dayside. The corresponding temperatures on the nightside are generally 30 to 40 percent less, which is a simple nighttime cooling effect. Except for some minor differences, there is no basic difference in the overall pattern of temperatures inferred from the three gasses, which is what one should expect if the basic assumption in deriving these temperatures is correct. The situation, however, changes very dramatically during the disturbed conditions. The region which was the coldest region before the onset of the storm suddenly becomes activated with the injection of energy, with temperature rising from about 600 to about 850 to 950 K. This is an increase of about 300 K.

Another point to note here is that both argon and nitrogen temperatures follow very similar behavior. The enhancement in the auroral and polar regions is very similar in both of them. This particular region becomes an entity by itself during storm periods separated from the middle and low latitude regions which still show dominant solar UV heating effect.

The situation is quite different if we infer these temperatures from atomic oxygen. The region which we interpret as the hottest spot from N_2 and Ar measurements is actually not so hot based on O measurements. The temperature changes are not significant, as we can see by comparing the temperatures from the quiet and disturbed conditions. The temperature increase is seen only in the auroral and midlatitude regions by 50 to 100 K. The important point to note here is that this is the kind of temperature one infers from satellite drag data. Any theory based on this kind of measurement will thus lead to a very different estimate of heat budget during magnetic storms.

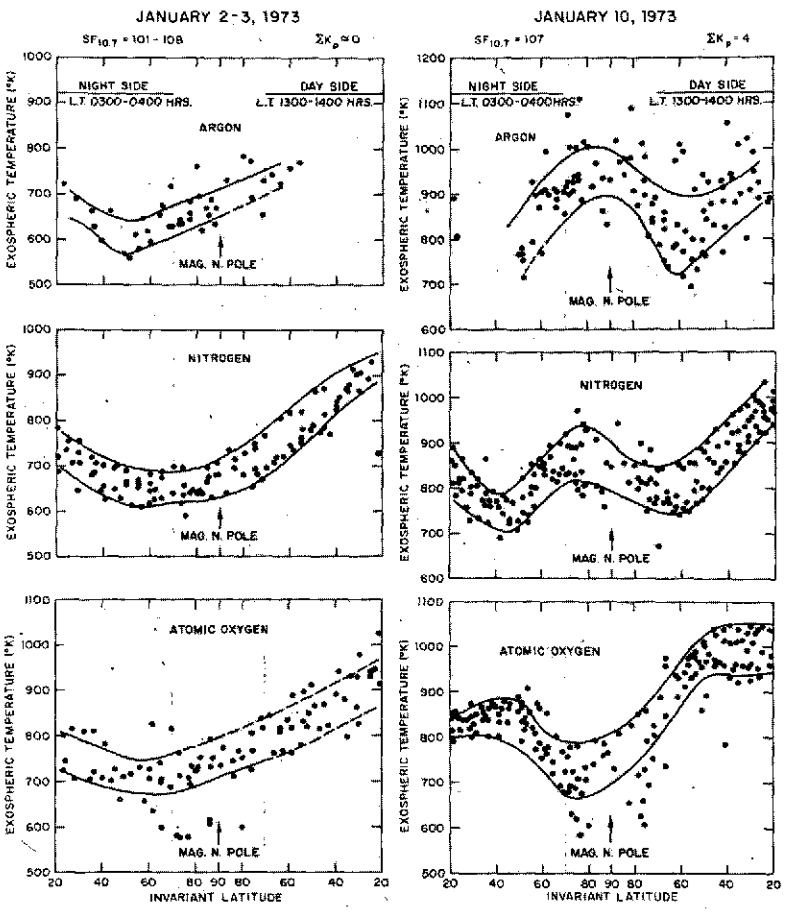


Figure 1. Exospheric temperatures inferred from the neutral constituents.

Since we believe that temperatures inferred from argon and nitrogen are closer to the neutral temperature, we can use these measurements to deduce changes in neutral composition, particularly in atomic oxygen and helium. This is shown in figure 2.

Plotted in figure 2 are the helium and atomic oxygen densities at 120 km for quiet and disturbed conditions. Compared to quiet conditions, both the gasses show considerable depression in the high latitude region. It seems as

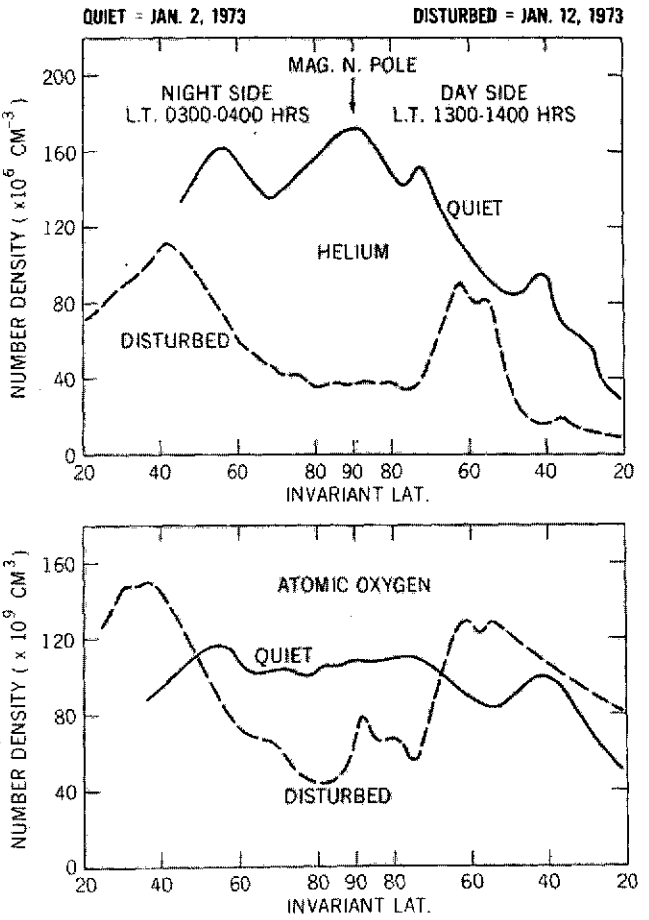


Figure 2. Latitude variations in helium and atomic oxygen at 120 km, inferred from AEROS.

if these gasses have simply been squeezed out of the polar region and deposited somewhere in the low latitude region. These variations in the neutral constituents have very important consequences in constructing the ionospheric models during magnetic storms, which we do not have the time to discuss here.

In conclusion, we note that during a magnetic storm most of the energy deposited in the neutral atmosphere is in the high latitude region, extending farther out to the polar region. The heating in the polar region seems to be of corpuscular nature since it is difficult to postulate the existence of Joule

heating in the region of the open field line. The magnitude of this energy per unit volume is comparable to the solar ultraviolet energy, though the total energy content is much less since it is confined to a narrow region. The neutral composition changes which occur during the storm are very significant and are probably responsible for a large number of ionospheric phenomena, including airglow emissions. Finally, because of the changes in atomic oxygen in the lower atmosphere, the temperatures inferred from satellite drag measurements do not give a realistic picture of the thermospheric models which were developed in the 1960's, including the recent CIRA model. These models, therefore, need considerable revision.

MAGNETIC-STORM-RELATED PLANETARY WAVES

Hano Mayr

Global scale atmospheric oscillations which are independent of local time and longitude are called planetary waves. The most prominent among these waves are the seasonal variations and here, in particular, it is the semiannual component which has fancied the imagination of scientists for some time, although such a half-year periodicity has not been found in the atmosphere. This is, in principle, understandable, if we simply consider that at the equator the sun is overhead twice a year during equinox, thus producing a semiannual heat source. The difficulty has been that the semiannual effect is, by comparison with the annual variations, much too large to be simply accounted for by solar radiation.

However, there is still another energy source for the semiannual effect in the atmosphere, and that is related to magnetic storms. It is well established that magnetic storms occur most frequently during equinox periods. And a recent theory by Russell and McPherron seems to explain this phenomenon.

Circumstantial evidence from ionosphere data, neutral composition measurements at 120 km, and meteor trail observations of mesospheric winds led us a few years ago to postulate that the semiannual component in the thermosphere should have a primary maximum at high latitudes indicative of a magnetospheric source. Recent composition measurements on OGO-6 have confirmed this, as will be shown also by the next speaker. Furthermore, estimates have shown that the energy supplement from magnetic storms could satisfy the energy requirements for explaining the magnitude of the semiannual oscillation in the thermosphere.

The following venture into the lower atmosphere should be viewed against this background of an apparently successful magnetic storm hypothesis for the semiannual effect in the thermosphere.

In collaboration with Dr. Belmont from the Control Data Corporation in Minneapolis, we have studied the semiannual component in the zonal wind field of the lower atmosphere. Figure 1 shows the magnitude of the zonal velocity derived from meteorological soundings that span 10 years of observations. Since longitudinal variations are significant, as I will show later on, the data were chosen from longitudes near 80° W. These winds

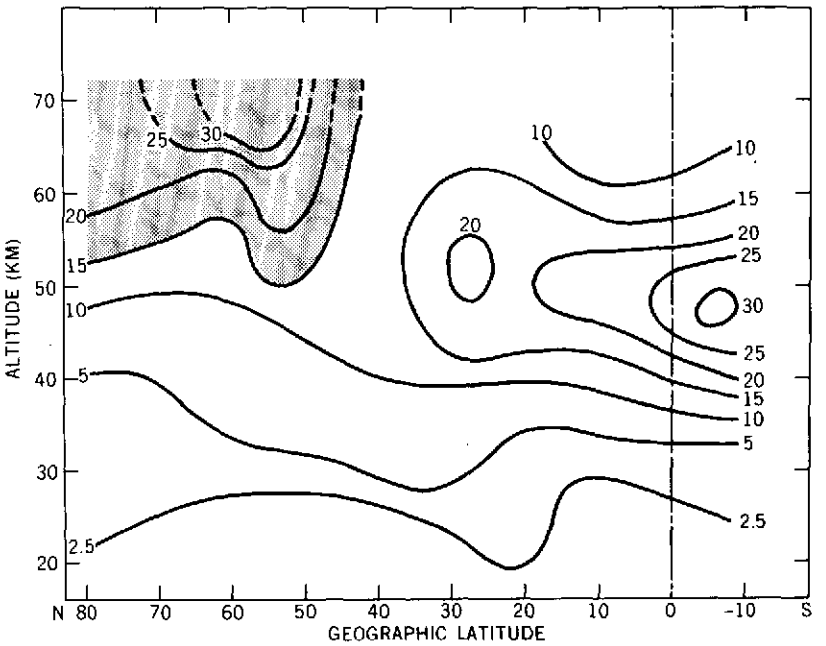


Figure 1. Magnitude of the zonal wind velocity.

are westerlies during equinox periods, and they are plotted as functions of altitude and latitude. Shown here at low altitudes is the well-known tropical maximum in the semiannual wave, which has been attributed to solar radiation, although there seems to be a problem energetically.

In addition, however, we also find this polar center in the circulation which is quite distinct from the tropical one. What we see here, we believe, is the lower edge of a polar circulation pattern that actually extends upward into the thermosphere, as is evident also from meteor tail observations near 100 km. This feature is of particular interest since it can explain the long observed weakening of the intense winter polar westerlies which, in turn, may set the stage for the well known sudden stratospheric warmings. Furthermore, we also find that this maximum is not occuring at the geographic equator but that it is displaced to the south, a feature which will be discussed further in figure 2.

Figure 2 shows what we consider to be a display of apparent magnetic field control. Data now taken from 50 km are shown in geographic and geomagnetic coordinate systems. We see that the tropical maximum does not occur at the geographic equator but seems to be centered around the geomagnetic

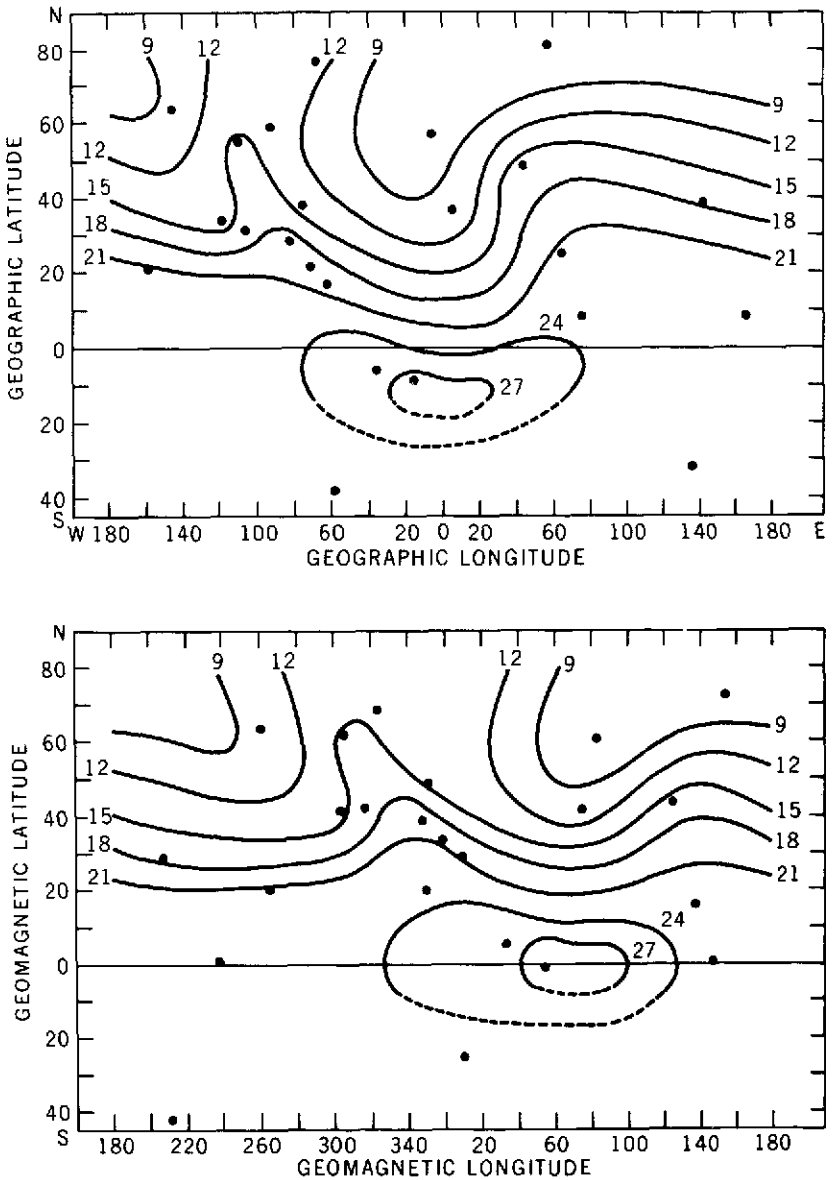


Figure 2. Apparent magnetic field control.

equator. In some cases the data coverage is not sufficient to draw contours, but in every case it seems that the highest values occur near the magnetic equator. If the magnetic field control were perfect, these contour lines would be parallel to the magnetic equator. This is not the case. But clearly, we find that in the geomagnetic coordinate systems the north-south elongations are much less significant than they are in the geographic coordinate system. The question we ask then is what is the linkage between the lower atmosphere and the magnetosphere?

Few would argue that the enormous energies involved in the lower atmosphere could not influence the comparatively tenuous magnetosphere. But for the very same reason, it is also hard to imagine that the comparatively small energies of the magnetosphere could have an impact on the lower atmosphere, except as a trigger. In the block diagram in figure 3, it is seen that the sun's radiation is absorbed in the atmosphere, increasing as we go to lower heights. In addition, the solar wind imparts energy to the magnetosphere, producing electric fields and energetic particles. These energetic particles dump their energy into the atmosphere at various heights, depending on their energy, producing heat, ionization, and chemical reactions. The electric fields produce currents and these currents are dissipated in the form of Joule heating at altitudes around 150 km in the thermosphere. They are also responsible for ion drifts and these drifts impart momentum to the neutral atmosphere, thus driving winds directly.

We believe that the processes of Joule heating, ion drift, and particle precipitation are to some extent responsible for the semiannual oscillations in the thermosphere.

In the lower atmosphere, the direct heat input from the magnetosphere is essentially negligible. There the key is probably ozone, which absorbs most of the solar UV radiation, and therefore, its density is critically important for the energetics of the mesosphere. An increase in the ozone concentration would lead to a proportional increase of the solar heat input due to absorbed solar radiation. The ozone chemistry, involving precipitating particles, is very complex. At this point, however, I would like to refer to the work of Maeda and Aikin of Goddard, who suggested several years back that the effects on the ozone concentration could be significant.

At the American Geophysical Union meeting in San Francisco, evidence was presented, at least I saw the abstract, for semiannual oscillations in the total ozone content of the Northern Hemisphere. And a recently published paper by Swider reports on correlations between the variations in the ozone concentration and polar cap absorption events. So while we do not yet have any hard evidence that the magnetosphere is a source for semiannual oscillations in the mesosphere, we feel that circumstantial evidence exists.

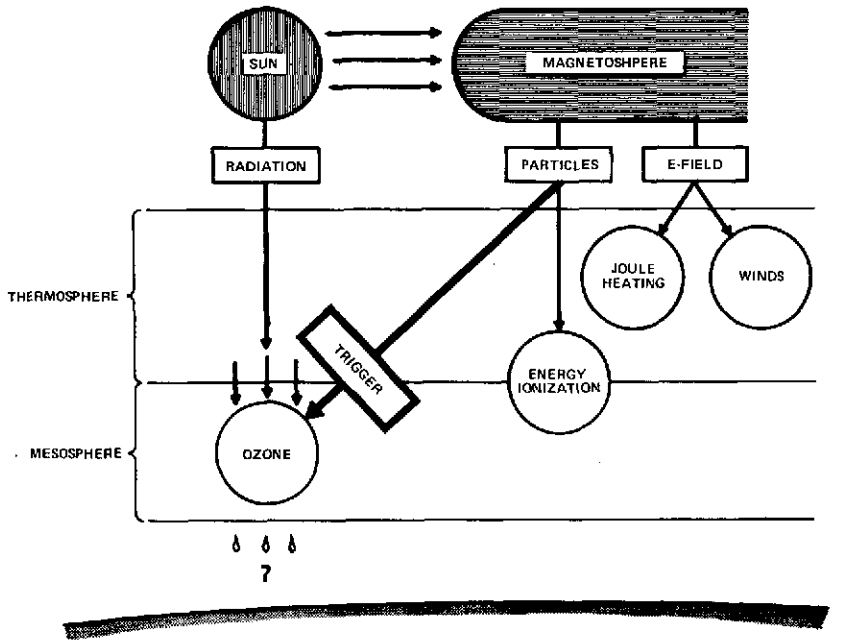


Figure 3. Block diagram showing the relationship between the magnetosphere and the lower atmosphere.

SEASONAL VARIATIONS OF UPPER ATMOSPHERE COMPOSITION

Carl A. Reber

We have heard this afternoon about various effects of solar energy on the earth's environment. This talk continues in the same vein by illustrating the effect of the seasonal variation, due to the tilt of the earth's axis, on the "weather" of the upper atmosphere. The large body of data obtained by the OGO-6 mass spectrometer has allowed the latitudinal/seasonal variations in upper atmospheric structure to be studied in great detail for the first time.

We have analytically represented the densities of the main gasses at 450 km—molecular nitrogen, atomic oxygen, and helium—in the form of a statistical model, in order to study the variations over the course of a year. These three gasses form an interesting combination, for a number of reasons, due to their molecular and atomic masses.

Solar energy deposition in the upper atmosphere has a number of consequences, among which are an increase in the gas temperature and the generation of pressure gradiants which lead to air motions or dynamics. Of the three gasses that we are discussing, the density of molecular nitrogen is extremely sensitive to temperature changes, reacting rather like a good thermometer. On the other hand, helium, because of its mass being so much less than the mean mass in the lower thermosphere is very sensitive to air motions; in particular, its density decreases at high altitudes when there is an upwelling of air beneath it. Atomic oxygen, with its intermediate mass between the N_2 and the helium, reacts both to temperature and to dynamics in a rather complicated way, and it gives some measure of the balance between the two.

The data in figures 1 through 4 are displayed in the form of logarithmic densities at 450 km as a function of geographic latitude, and represent the time period 1969-1970. In figure 1, which shows the March equinox (day 80), atomic oxygen is fairly uniformly distributed across the globe. The helium density is also fairly uniform, and down an order of magnitude or so from the oxygen. The N_2 is also almost uniform at equinox, with a slight indication of a density enhancement around the equator near the subsolar point. There is also an indication of enhancements at high latitudes in both hemispheres, indicative of some measure of high latitude heating, polar heating.

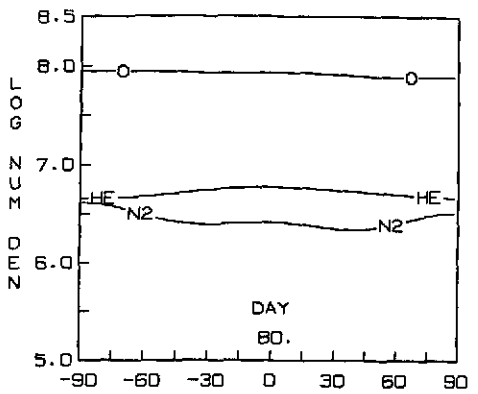


Figure 1. Gas densities as a function of geographic latitude for the March equinox, 1969 and 1970.

Figure 2, the June solstice (day 170), shows that the N_2 density in the summer hemisphere is much higher than in the winter hemisphere, indicating a large temperature differential; at the same time, the helium has increased in the winter hemisphere, indicating that the air motion is upward in the summer hemisphere, with flow across the equator and down in the winter hemisphere. The atomic oxygen at this time has moved in a direction to mimic the N_2, indicating that at 450 km, the temperature effect is dominant. If we took these same measurements lower, say at 200 km, we feel that we would see an enhancement of atomic oxygen in the winter hemisphere similar to that which we see for helium.

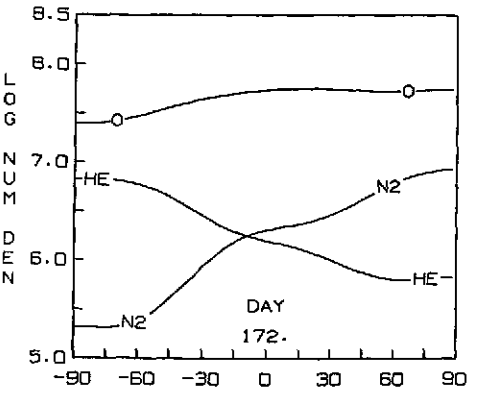


Figure 2. Gas densities as a function of geographic latitude for the June solstice, 1969 and 1970.

For the September equinox (day 264), we can see in figure 3 that the N_2 density is now increasing in the southern hemisphere. We find that once again we have a fairly uniform distribution of atomic oxygen, helium, and N_2. The absolute values are a bit different from the March solstice as there is a semiannual effect coming in. Again the N_2 shows the little bulge over the equator, indicating a subsolar warm spot there plus the slight enhancements in the polar regions.

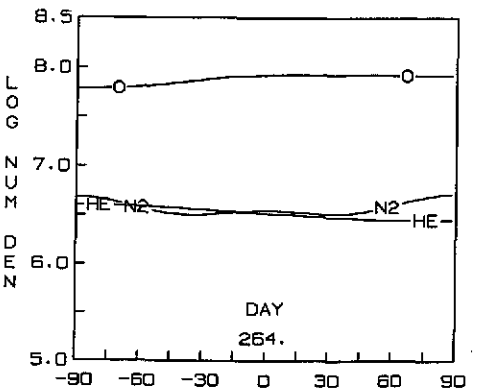


Figure 3. Gas densities as a function of geographic latitude for the September equinox, 1969 and 1970.

As we go on to the December solstice (day 355) in figure 4, we see the opposite effect taking place from what happened in June. The summer hemisphere is now in the south, the winter hemisphere is in the north, and we see a fairly large variation, nearly two orders of magnitude, in the N_2 distribution, a slightly smaller variation in helium, and once again the atomic oxygen is mimicking the distribution of N_2. If we were looking at this with a satellite drag measurement, we would see the atmosphere following essentially the atomic oxygen distribution, and we would really be missing all the interesting phenomena in the minor constituent distributions. This is what Dr. Chen was mentioning earlier. The temperature differential at solstice, which goes along with the N_2 distribution, is about 400° from pole to pole.

Now, we will go through the whole thing fairly quickly, with a whole year's worth of data compressed into about 55 seconds (a film will be shown here). We are starting at Day 1, going through the course of the year now. We are in January, getting into February, and now getting up close to the March equinox. It is continuing on, and you can see the way the variations are

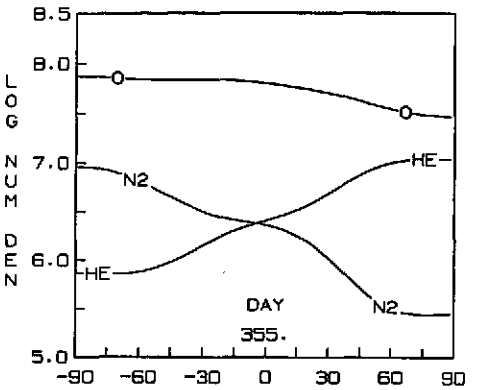


Figure 4. Gas densities as a function of geographic latitude for the December solstice, 1969 and 1970.

manifesting themselves. There always appears to be a little bit of enhancement in the N_2 near the equator, no matter which time of the year we are looking at. By looking at this sort of display repeatedly, one can pick out the phase differences between the variations in the N_2 and the helium, indicating the fact that the temperature appears to be leading the dynamics in this case. We're coming up now toward the end of the year and there we have the end of one year's look at the "weather" in the upper atmosphere.

OVERVIEW

Alton Jones

Spacecraft technology, as defined by this session, covers a wide area stretching from specific spaceflight instrument design and performance to propulsion technology that resulted in a biomedical spin-off. The central themes that pervade this program are cost, timeliness, performance, and reliability.

The effectiveness in our pursuit of these aims regarding specific accomplishments will be established through the presentations that follow. Accomplishments covered include rescue of the S^3 (Explorer-45) mission from an instability condition that was caused by an unexpected thermoelastic bending torque; development in-house of RAE-B sensor and data compression techniques that permitted determining spacecraft attitude and behavior of the 230-meter antennae; onboard data processing and data management of experiment and spacecraft functions; thermal and systems control; and a biomedical spin-off.

There have been numerous other accomplishments that cannot be covered in the alloted time period but it is felt that the sampling that follows will establish the scope and success of our technology program.

NUTATIONAL BEHAVIOR OF EXPLORER-45

Thomas W. Flatley

The Small Scientific Satellite, commonly known as S^3-A, was launched on November 15, 1971, from the San Marco launch platform in Kenya. It has a highly eccentric, near-equatorial orbit with a period of about 8 hours. The satellite is equipped with a magnetic attitude and spin control subsystem which can only be used near perigee when the magnetic field strength is strong enough to be useful. Soon after injection into orbit, the satellite was observed to have a spin rate about 1 rpm higher than nominal and a nutation angle of about 7°. The satellite had lost its passive nutation damper in a weight reduction program, but it was felt that within a few orbits structural energy dissipation would damp out the nutation.

After a complete orbit had passed with no apparent nutation decay, however, the possibility of active nutation damping was first considered. To bring the spin rate down to a nominal 3.75 rpm, despin operations were begun on the first perigee pass. Three such operations were carried out before it was noticed that the spacecraft nutation angle had begun to grow at a much faster rate than could be explained by the despin.

Figure 1 shows some of the early data from the first S^3 instability. It includes the shape of the S^3 orbit for reference. The apogee was about 5 earth radii and the perigee was in the range of 200 kilometers. This figure shows a plot of spin-rate versus the spread in the sun aspect angles being recorded, which is a good indicator of the nutation angle. The first point corresponds to the launch time on November 15, and begins a short arc covering the first three orbits during which despin operations were carried out. They were terminated on November 16. During the next 2 days, or about 6 more orbits, the nutation angle growth continued and panic set in on the project. During this time, possible causes for the instability were examined--aerodynamics, magnetics, the chance that the booms did not deploy, the booms fell off, and so on--and recovery plans were formulated. These plans included spinning up the spacecraft, pointing the spin axis closer to the sun, and accelerating the development of the active damping technique. The purpose of the spin-up was to increase the structural energy dissipation. The sun angle change was expected to help according to a preliminary hypothesis, which in principle turned out to be right, that we

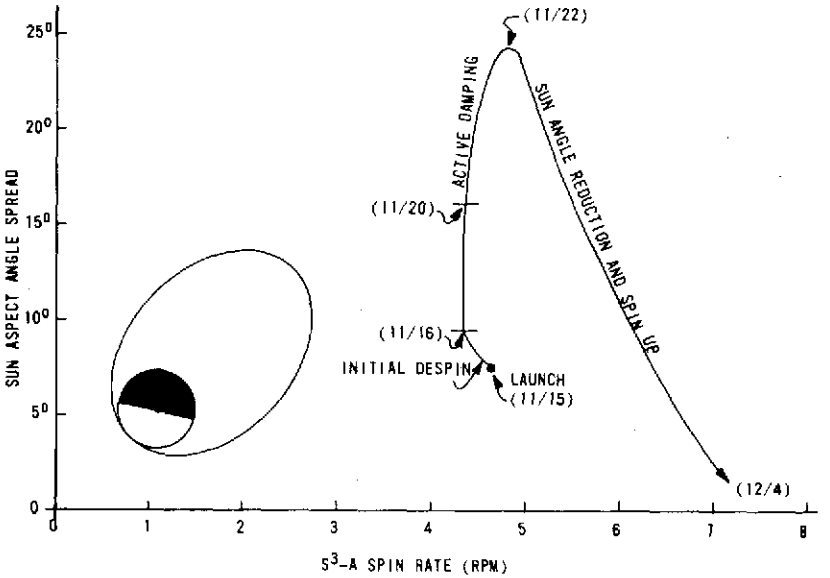


Figure 1. First S^3-A instability.

were up against a thermal-elastic instability involving thermal bending of four booms on the spacecraft. For the active damping, it was necessary to determine the relative positions of the spin axis, the angular momentum vector, the earth's magnetic field vector, and the spacecraft-sun line, and pulse the attitude control coil at appropriate times. A method was quickly improvised with outstanding help from the former Attitude Determination Office at Goddard and their contractor support, which used real-time sun sensor and magnetometer data from the spacecraft and four Goddard computers, one of which generated the necessary commands that were sent out by radio link to the Quito ground station.

Quito was not only the only ground station with which this command link was possible, but it was also the only station which could see the satellite close enough to perigee to be useful. Conditions were right for active damping only about once per day, and seldom during normal working hours. The first recovery operation was an active damping attempt on perigee number 11, shown on the figure by the date (11/20), and during the next 49 consecutive perigee passes, one or more of the three recovery techniques were employed—either the active damping, the spin-up, or the reorientation of the spin axis.

The nutation angle peaked on November 22 and started down, and recovery operations ended on December 4, with the spin rate slightly over 7 rpm and a nutation angle of about 1.5°. The S^3 experimenters did not want to spin any faster than this and the nutation angle was at a level where the original method of active damping was becoming increasingly difficult.

This brings us to a point about 20 days after launch, seen in figure 2. The nutation angle stayed nearly constant for the next 100 orbits, which lasted about 30 days, and it was decided to despin the spacecraft to 6 rpm to observe the nutational behavior. This despin caused the nutation angle to increase rapidly and when it reached 3° and showed no signs of slowing down, the spin rate was again increased to 7 rpm. The angle peaked at 3.2° and began a slow decay, which surprisingly continued until all signs of nutation had disappeared. This decay turned out to be a consequence of spacecraft shadowing which began for the first time near the second peak. For the first 60 days of the mission, the spacecraft had been in full sunlight. When these shadows started, they provided times of relief from the solar excitation, which allowed the energy dissipation onboard to damp out some of the nutation. This effect was significant enough to bring the nutation all the way down to zero.

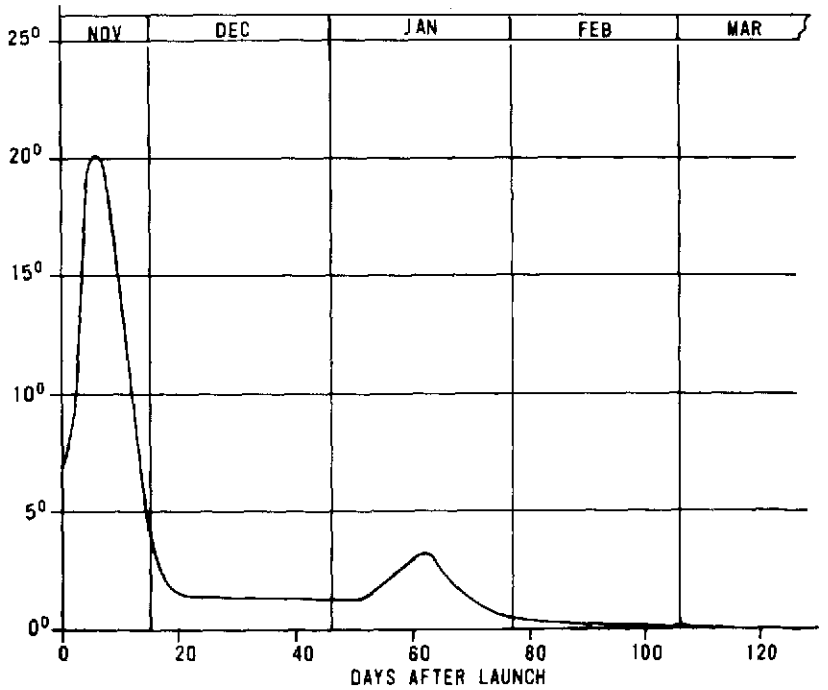


Figure 2. S^3-A maximum half cone angle history.

During the last 2 years, several minor instabilities have occurred when sun angle and shadow combinations were unfavorable. These have been controlled by an improved automated active damping method involving only one Multisat computer and capable of reducing the nutation angle to less than one-tenth degree whenever required. It was used most recently on December 13, 1973, and will be used again during the latter part of this week.

RAE PANORAMIC ASPECT SENSOR

Mclean Grant

This is a discussion of the panoramic aspect sensor, PAS, an attitude determination device flown on the Radio Astronomy Explorer-B, or RAE-B spacecraft, now in lunar orbit.

Attitude determination for RAE-B was complicated by the lack of a strong magnetic field in the translunar and lunar orbit environment, by the fact that the onboard solar aspect system could resolve only two of the angles required, and because the RAE-B spacecraft could not support the size, weight, power, and complexity of star trackers and star scanners. In addition, RAE-B had a large number of different operations (figure 1) requiring fine attitude determination following injection into translunar voyage, through the precession for midcourse correction, thermal control, reorientation for retrofire insertion into lunar orbit; and, five precessions for orbit circularization and for science. Spin rates were 50 rpm on the way to the moon and, in the early lunar orbits, 50, 12, and 4.7 rpm. Following this period, the spacecraft was despun to 0.4 rpm and oriented for deployment of the four 229-meter antenna booms, which initially were deployed to 185 meters each. Following that deployment, the spacecraft was three-axis stabilized by gravity gradient, with one revolution per orbit about the pitch axis.

The PAS was developed and conceived at Goddard to provide 4π steradian coverage in the spinning mode to give earth, sun, and moon position relative to the spacecraft's spin axes, and to give spacecraft axes relationships to the local lunar vertical in the gravity gradient mode. These functions were accomplished with a lightweight, low power device shown in figure 2.

The scanner head is composed of a housing with a 360° slit, stepper motor, and support electronics. In this figure, the tube represents the 0.7° spherical field of view of the scanner, which is rotated around through the slit, similar to a lighthouse.

In figure 3 we have the 0.7° circular field of view lying in the plane of the figure with the spacecraft spinning about this axis. The field of view is held here at an angle α from the spin axis, and maintained there for

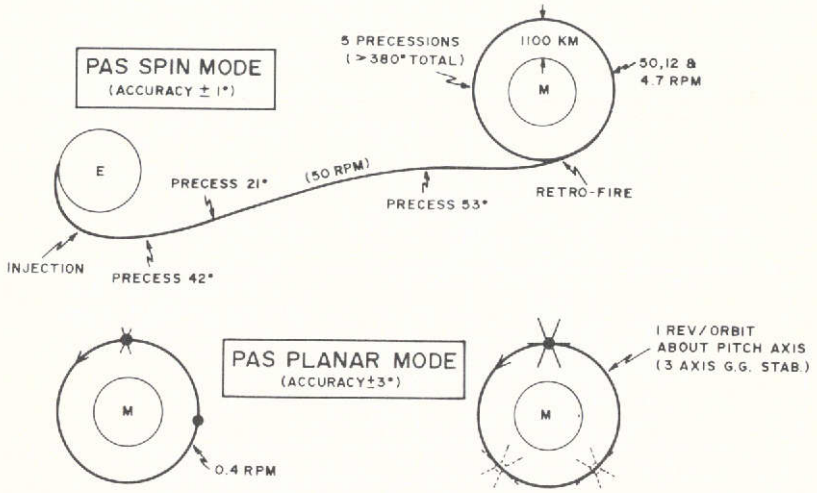


Figure 1. Panoramic aspect scanner and RAE-B flight operations.

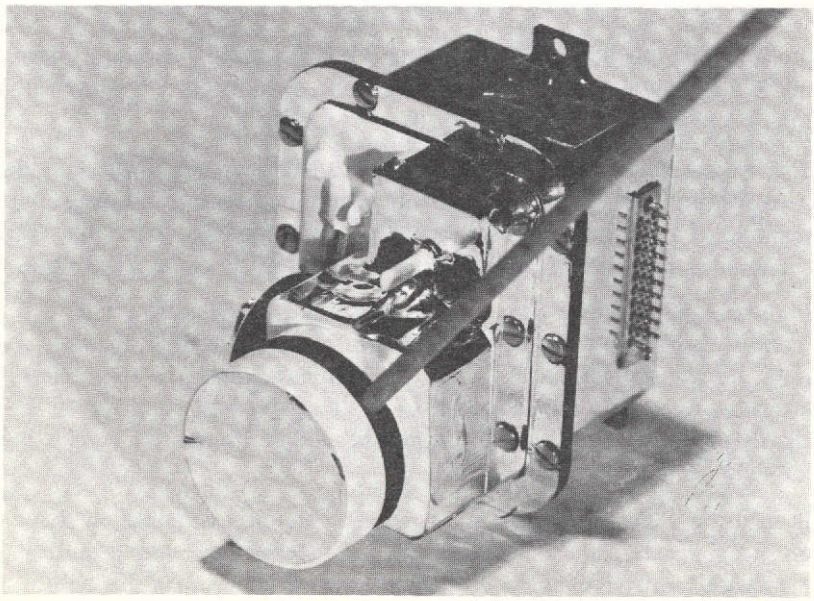


Figure 2. RAE-B panoramic scanner.

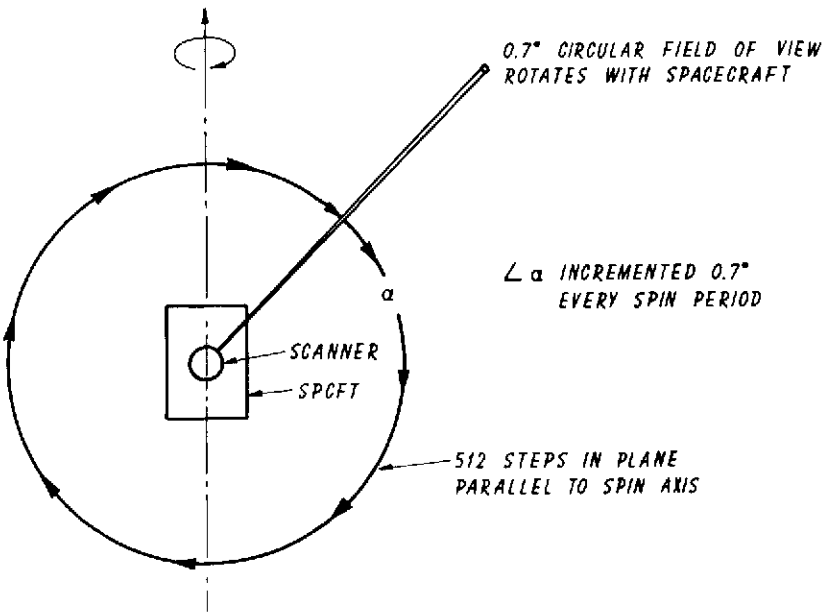


Figure 3. RAE-B panoramic aspect scanner, spin mode.

precisely one spin period, at which time it is advanced 0.7° in angle α from the spin axis and held for another spin period. The process continues, with the field of view being rotated 512 steps before repeating.

This process is described by the beehive picture in figure 4, showing complete coverage of the spherical field of view in a given number of rotations. For example, at 50 rpm the 360° spherical scan is accomplished in 5.1 minutes and reports the elevation and azimuthal positions of sun, moon, and earth.

In the nonspin mode, or planer mode, the original spin axis is relatively fixed, as seen in figure 5. This is in line with the local vertical. We wish to get horizon and terminator information from the moon. In this mode, instead of once per spin period, the 0.7° field of view is stepped very rapidly at 100 steps per second, scanning around in this fashion, completing the full scan in 5.12 seconds.

In figure 6a, looking along the axis toward the moon, we again show the original spin axis, a line along the local vertical. Starting at zero position, the scanner is stepped rapidly around as shown, picking up, in this case, a horizon and a terminator. Looking down at the moon along the Z-axis, we have the scan crossing as shown, picking up horizon and terminator.

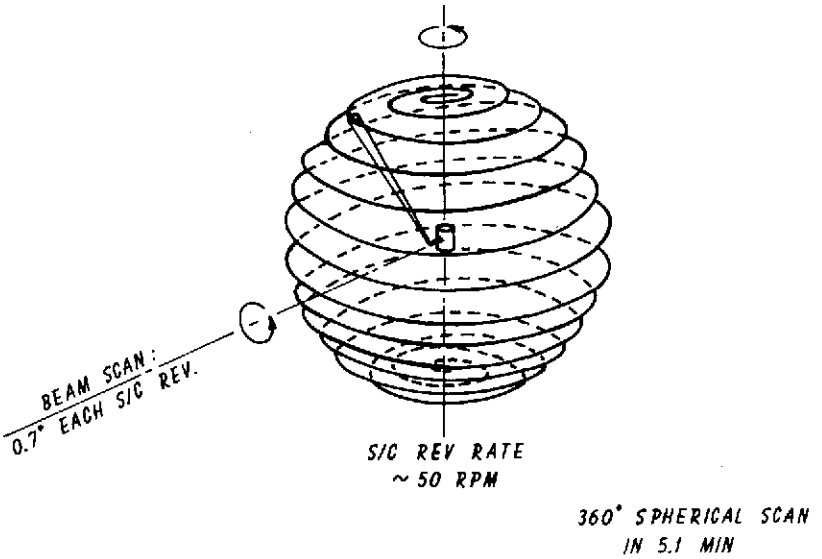


Figure 4. Panoramic aspect scanner, spin mode, spherical field of view.

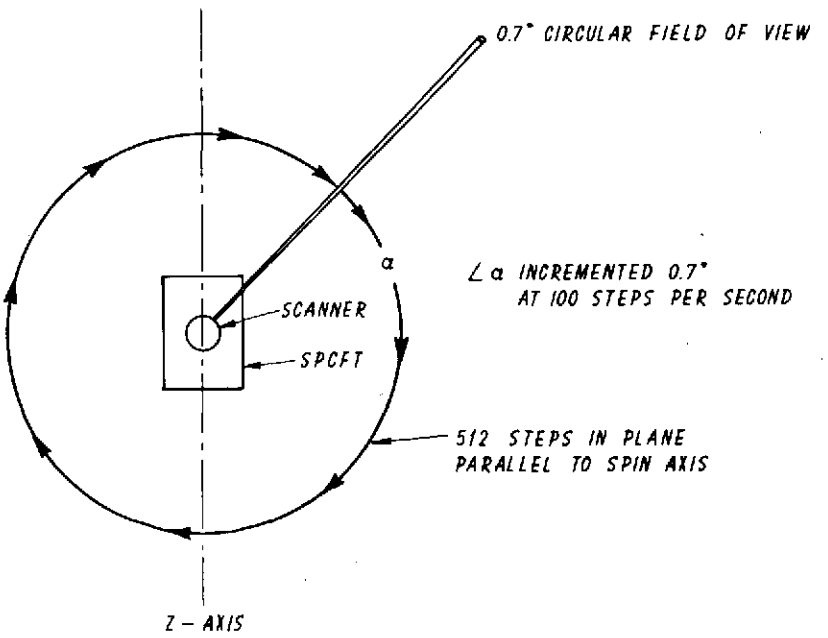


Figure 5. PAS gravity gradient or low spin mode, planar mode.

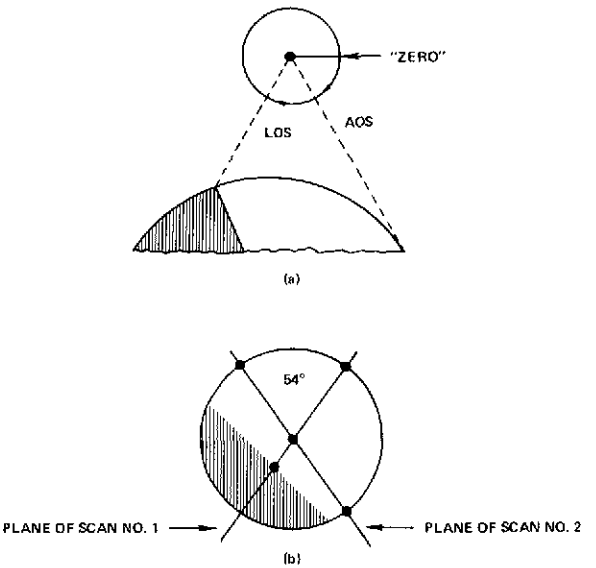


Figure 6. PAS in gravity gradient or lo-lo spin mode. (a) PAS planar scan across the moon; (b) Intersections of planar scans with the moon.

To resolve ambiguities, RAE-B flew two scanners; their planes of scan were 54° apart. Figure 6b shows the first scan, from scanner 1, and a second scan, from scanner 2. These scans were alternated.

Accuracy requirements for the PAS were ±1° in the spin mode and ±3° in the nonspin mode. The PAS has met or exceeded these requirements. Recently, a patent has been issued to NASA for the PAS, which will be flown on IUE and other projects.

RAE-B CAMERA SIGNAL PROCESSOR

Joseph Novello

My speech is a brief description of the camera signal processor in the antenna aspect subsystem of the RAE-B spacecraft (figure 1). The main function of this system is to determine the shape and the position of four 229-meter booms which make up the very large antenna and the main experiment on the RAE-B lunar mission. The reason for the processor is that the project had purchased a camera and encoder which were not originally designed for this lunar mission. At lunar distance, the RAE-B telemetry link could not support the 20-kb data rate offered by the camera encoder without introducing a large number of bit errors in the telemetry system. In order to decrease this bit error rate to an acceptable probability of bit errors (1×10^{-5}) and at the same time add some signal margin (6.8 dB), it was proposed that a signal processor be put on line between the encoder and the transmitter to achieve a bandwidth reduction of 32:1 by going to a transmission rate of 625 bps. This was an alternative proposal to a camera redesign.

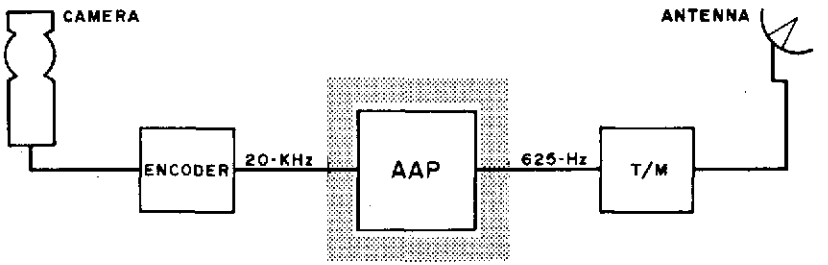


Figure 1. Antenna aspect subsystem for the RAE-B spacecraft.

The processor not only had to achieve the desired goal of bandwidth reduction and add signal margin, it also had to be cost-effective over a camera redesign. According to John Shea, the project manager, all of these goals have been met.

The camera in the system is a facsimile camera and it scans 512 lines per picture. The picture generated is a panoramic picture as seen in figure 2. From such a picture, the ground people hope to extract the following information: the general shape of the booms, the spacecraft attitude

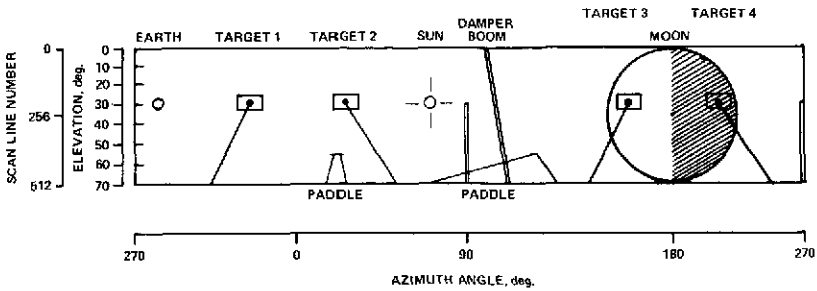


Figure 2. Typical panoramic view from camera.

relative to the moon, that is, the moon terminator, and boom tip locations so as to obtain the precise deployment of the antenna.

A computer study was made of our simulated picture, and from this study it was determined that a combination of schemes would be used to solve the problem. The technique adopted was one of data-deletion, data compression, and data storage. In real-time data processing, three of every four camera lines are deleted. Every fourth line is processed with a zero order predictor for data compression. With these two schemes a total data compression rate of 32:1 is achieved in real time. In playback and memory dump mode there is no need for data compression.

The modified zero order predictor compression scheme has three main features: a variable threshold, an adaptive run length, and an adaptive aperture. The variable threshold which is commandable from the ground is used primarily to guard against the camera noise due to age. The adaptive run length is used primarily to determine one of two data formats. By selecting formats, this scheme enhances large data changes and integrates small ones. The aperture, which is defined as the difference between two signal levels, is adaptive and is used primarily to guard against a buffer overfill as a particular line is being processed.

To extract the precise information about the boom tips location, selected raw data from the camera is stored in a dynamic P-MOS 64,000-bit memory. The stored area of the picture is determined by four sectors whose coordinates are commandable from the ground and can be moved anywhere in the picture. This stored data is transmitted at the end of a real-time picture and at the same rate as real-time data. The memory control, as well as the rest of the logic, is implemented with complementary MOS (C-MOS) devices, primarily because of their low standby power characteristic. Figure 3 shows the unit before potting.

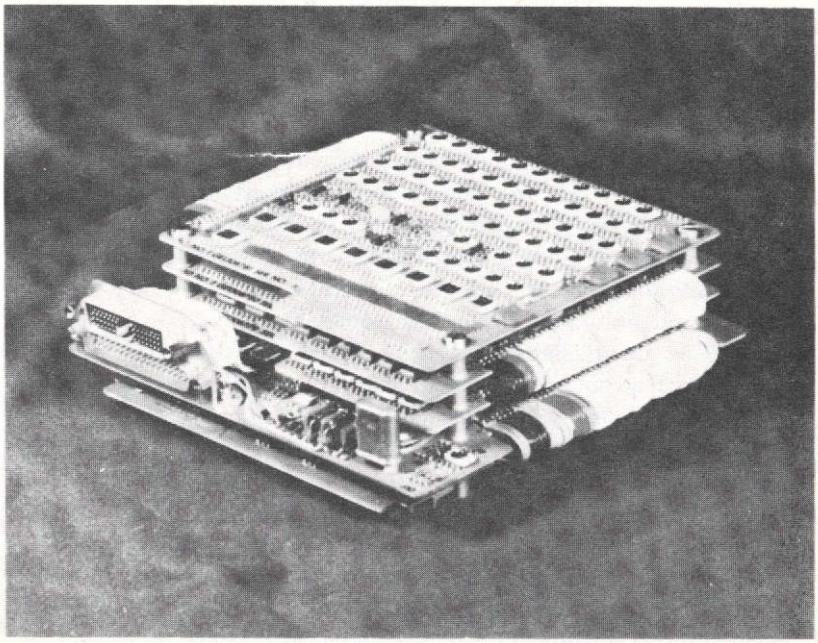


Figure 3. RAE-B camera signal processor.

One of the mechanical features of this processor is the through-the-insulation welding technique used to interconnect signals. This technique lends itself to easy modifications and quick redesign implementation. The processor is quite small. It occupies less than 2,000 cm^3, weighs less than 1.4 kg, and on the average dissipates less than 0.7 W. One technical feature worth noting is that for the first time Goddard used a dynamic P-MOS memory of this bit density. Another feature is that for the first time this particular data compression scheme was instrumented to solve a space-related problem. It is also worth noting that the processor has been flown successfully and has operated properly for the last 8 months.

ADVANCED ONBOARD PROCESSOR

Raymond Hartenstein

Our three primary goals in the advanced onboard processor (AOP) program are as follows. The first one, which is very important nowadays, is low cost. We are achieving low cost, first of all, through standardization of computer modules. This allows us, through broad usage of standard memory modules and processor modules, to purchase these units on a volume-buy basis, and it also provides for common spares from one program to the next. We are utilizing large scale integrated circuitry (LSI) logic, which has proven not only to give us the expected size, weight, and power benefits, but if properly implemented to offer significant cost savings, especially when volume buys are involved. Implementation time is being shortened, which allows us to put systems together quickly and efficiently to meet various requirements. Another item is software commonality, which is deserving of attention. We find that not only is the ground support type software, which you would expect for the same type of computer module, going to be common and represent a significant cost savings, but several of the flight software modules (for instance, the executive and the basic command handlers) will have a tremendous amount of commonality from one program to the next.

Our second major goal is long life, as we start addressing ourself to 5- and 10-year missions. This is a very important item, also. Again, we see the LSI logic showing up here, and it provides a two-pronged improvement in life. First is that it decreases tremendously the number of parts involved in any given program, and second, it also decreases the amount of interconnect, consequently improving reliability by a significant amount. Also very important in achieving long life is the use of proven modules to a large extent from one program to the next. Standby redundancy, through utilization of a redundant bus, allows us to add spare modules as a requirement presents itself. Again, proven programs (I am referring to software programs, that is, instructions that reside in the computer) will be utilized and standardized from program to program.

A third major goal is universal or wide application over many programs. We achieve this through modular growth, so that we can meet both small and large spacecraft requirements–low power is significant in many smaller

spacecraft. We think of all computers as having powerful central processing units, but many people forget this item, a flexible input-output, which is extremely important for being able to adapt computers to the many unique specific requirements of diverse spacecraft programs.

I would like to address myself to the processor area of the computers at the moment. The second column in table 1 is representative data from the Orbiting Astronomical Observatory (OAO) onboard processor which was reported on last year. I am happy to say it has now completed 15 and a half months (I do not have it down to the hours and seconds) and it is a very big success for us and we are happy with it. We achieved all of our major goals in speed, volume, weight, and power. The cost item was fairly large and reliability, of course, on a purely mathematical basis, is not what we would like (certainly not adequate for 5- or 10-year programs).

The last column in table 1 is representative data from an advanced onboard processor, which consists of an integrated central processing unit and an input-output unit. This was delivered by Westinghouse Electric last spring. It is operational and, together with a memory on which I will show you some data in a moment, has passed thermal vacuum and vibration qualification tests at the Test and Evaluation Division. It is significantly faster with some size and weight advantages and minor power improvement, but most significantly, we have decreased the cost; these are actual numbers for a typical unit. We have decreased the cost, as you can see, by almost a factor of three, and it is not going to stop there.

For reliability, I show 10^5 hours mean time between failure (MTBF). This, again, is a mathematical calculation, but it is very significant, especially when we consider that we have 14 of these devices on test out at the Jet Propulsion Laboratory at our request. They now have 6,000 hours on 14 of these devices, not only without a failure but without significant degradation. This is at 413.5 K (140°C). Now if one of those devices were to fail tomorrow, the failure rate would be one-half that used to arrive at this 10^5 number; that is, we would be twice as good as that number shows, or 2 times 10^5.

I would now like to address the other major area in the computer, the memory module. Looking at some comparative data on 8 K memory modules, in table 2, the first two columns of data, the core and the plated wire, represent today's technology, state-of-the-art, readily available memory units. The core unit is representative of the memories which were flown on IMP-1 and the OAO computer down to the cost figure. This cost represents what we could buy an 8 K module for today, and this is about half what we paid for these core modules a couple of years ago.

Table 1
Comparison of OAO Onboard Processor
and AOP Data

	OAO/OBP-1 Processor and Input/Output Unit	AOP Integrated CPU-I/O Unit
Speed (add) (μs)	10	5
Volume (dm^3)	9.1	2.0
Weight (kg)	14	1.3
Power (W)	11	7
Cost (K$)	140	50
Reliability MTBF-Hours	10^4	10^5

This is due to the fact there is a lot of activity in core, especially in the military area, and by utilizing a module which was almost off the shelf, we have been able to achieve these volume cost advantages.

The second data column lists representative numbers for plated wire. The most significant characteristic we are after here, of course, is a power reduction, because 50 W, even though it is power switched and varies from 0 to 50 as you busy the computer up, is quite a bit of power from many programs, especially the smaller ones such as IUE. The plated wire is now state-of-the-art. We have received two 4K plated wire units, in which we have also incorporated power switching to even further reduce the average power, since typically a computer is busy only 10 to 15 percent of the time, especially in a real-time environment. This cuts down your orbital average power significantly.

The last item, though not the least, is cost, and it shows that as far as we are concerned, the cost of flight-qualified plated wire memory units is still entirely too high.

The last column in table 2 is a semiconductor volatile memory development which we have underway at the moment. Again, we are utilizing a hybrid package which contains 3 bits by 1024 words and was developed for the military. We are simply repackaging it to be compatible with our stack modules and interfacing it to agree with our standard interface, so that all these modules are interchangeable and can be mixed and matched

Table 2
8K Memory Modules

	Core	Plated Wire	C-MOS
Volume	2050 cc	3280 cc	2050 cc
Weight	2.7 kg	4.5 kg	2.3 kg
Power	50 W	6 W	2 W
Cost	25K (18¢/bit)	70K (50¢/bit)	60K (42¢/bit)

from mission to mission. We may do just that on IUE; for instance, fly two plated wires, since they are nonvolatile and will contain our program area, and fly two of these Complementary Metal Oxide Semiconductor (C-MOS) modules primarily for data storage. The significant factor here, again, is the extremely low power of the C-MOS—if you can take the penalty of it being volatile. In most cases, the volatility is not really a serious drawback because with these kinds of power levels, the standby power is only a few hundred milliwatts, and most people can stand that. So you never really need to turn the power off to that unit.

The last item in that column is representative of today's prices in C-MOS. We know that this technology is in its infancy and it is going to come down significantly—it might have done so by today. This data is a month or so old, and the cost is dropping rapidly. We expect it to be comparable to core in the very near future.

Not to belabor the point, but I have assembled, in table 3, four systems of 16 K typical computers with the three significant trade-off factors: power, weight, and cost. The first column is representative of the OAO in-orbit computer, and again shows 50 W, rather heavy and entirely too expensive. Of course, this was the first time down the road. Now, combining today's available core with the basic AOP processor, we can drop the cost significantly. If you can stand the power, you can immediately have the cost advantage of the core. Looking at plated wire, there are significant power savings, but again, the cost bounces back up, not prohibitively high, but it is a penalty. This is assuming you flew all plated wire. Now, the C-MOS version—I put these in parentheses because these are the only numbers I have shown you that are not actuals; they are very nearly that because of very tight budgetary estimates, and we have procured most of the components for the C-MOS memory—would drop power down again, nominally around 9 W. It would be even less than that as standby power,

Table 3
Computer Cost Trade-off (16K Type)

	OAO/OBP-1	AOP/CORE	AOP/PW	AOP/C-MOS
Power W	50	50	15	(9)
Weight kg	28	8.2	11.8	(8)
Cost K$	280	110	200	(120)

probably 6 or 7 W. The weight, of course, is low, but again the cost based on next year's prices would be around $120,000. Based on today's prices, it would actually be around $170,000 to $180,000.

So you see, we have a wide selection of processor capability. These computer interfaces are all comparable with the standard communications and data handling efforts which are going on here under the auspices of the Tichler committee, and we think we can meet most of Goddard's upcoming onboard computer requirements.

GSFC P-MOS MICROCIRCUIT DEVELOPMENTS

Theodore Goldsmith

Goddard has an inhouse facility for producing microcircuits using the P-MOS metal oxide semiconductor process. P-MOS is slower and of higher power than C-MOS and has radiation sensitivity. However, its advantages are high density, which leads to its use in large complicated circuits, and low design and tooling costs for making custom circuits, which is our particular interest at Goddard. The primary purpose for designing new microcircuits is to reduce the cost of space flight hardware by reducing the number of parts required, although there are concurrent improvements in reliability, weight, and volume. Some of the economic considerations involved in this regard are as follows.

The unit cost of $125 is our quoted number on a per microcircuit basis for building flight hardware. This includes the cost of the microcircuit itself, an inspected, flight quality, burned-in unit, plus the prorated cost of building the printed circuit board, inspecting the board, and so on. In other words, if a black box had a hundred microcircuits in it, it would be worth about $12,500. Tooling cost is the cost to produce a new microcircuit design in-house using our P-MOS process, running from $5,000 to $10,000 depending upon the complexity of the circuit.

So, it can be seen that if by designing a new microcircuit, one can reduce the total number of parts in a given application by at least 80, then the design-cost will be recovered and a net dollar saving will result. Thus, we use the phrase "total parts count reduction" as a kind of economic merit factor for new microcircuit design, and consider this as probably the most important aspect governing whether we will design a new microcircuit to replace old parts or whether we will go with the existing parts.

Figure 1 shows, as a typical example, our type 621 microcircuit, which is designed for the HEAO-A Goddard x-ray experiment. The circuit contains two 16-bit counters and two 16-bit shift registers and provides for data accumulation, temporary storage, and serial read-out in the experiment's histogram commutator. This experiment data system was originally designed using existing microcircuits and required approximately 1100 parts. Each 621 that was then specifically designed for this experiment application replaces approximately 2.3 of these earlier circuits each so that the commutator now contains only 478 parts, yielding a reduction of about 600 parts.

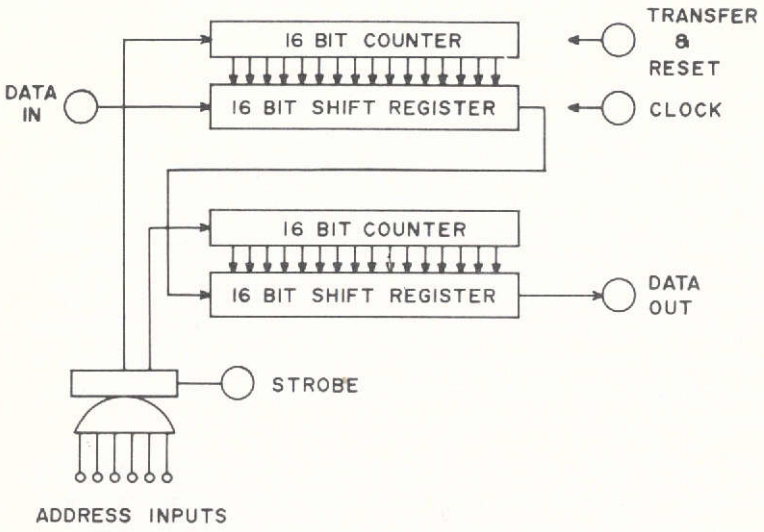


Figure 1. Type 621 HEAO-A dual 16-bit counter register.

Figure 2 shows a picture of the 621. The circuit is 5.2 mm by 2.6 mm in size, and contains 1136 transistors and resistors. It is a fairly large-scale integrated circuit.

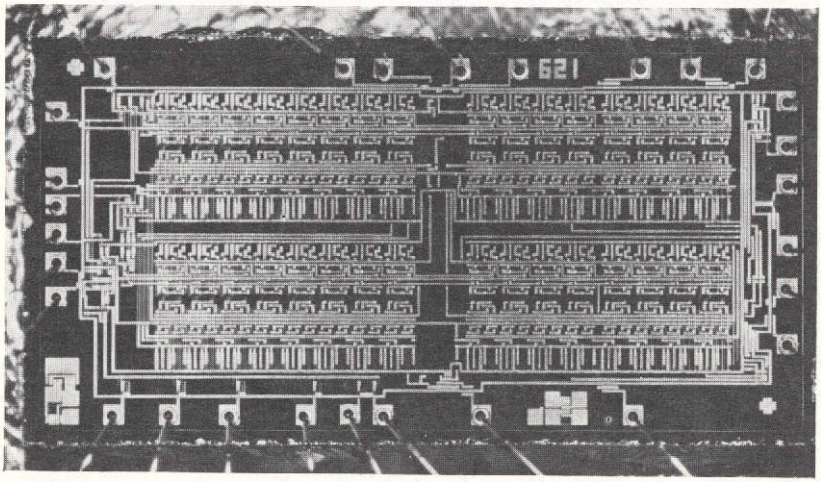


Figure 2. Type 621 microcircuit.

At the Goddard facility we use the ion-implanted depletion resistor process which is a variation of the basic P-MOS process. This variation is industry-compatible, and allows us to use outside manufacture of in-house design circuits where the volume requirements would exceed our in-house capability. We extensively use computers for automated design and for testing of the completed circuits to reduce the tooling cost. The last 70 circuit designs at Goddard have all been produced using computer-generated art work. Where we have trade-off possibilities between tooling cost and unit cost, we tend to go for the lowest possible tooling cost even though our unit cost in-house is higher. The industry is more oriented to much larger production runs and tends to have much higher tooling costs in an effort to get their unit cost down.

Some of the applications, current and planned, are as follows: The data multiplex unit, or DMU, is the standard common central data system for the planned IME, IUE, and HCMM missions. Another DMU is currently in flight as part of an engineering experiment on Explorer 50 (IMP-J) launched last October. All of the digital microcircuits in the DMU are Goddard designs. We are also supplying or planning to provide parts for experiment data handling on OSO-I, HEAO-A, and IME.

ZERO GRAVITY TESTING OF HEAT PIPES

Roy McIntosh

During recent years, heat pipes have become increasingly important to the thermal design of Goddard spacecraft. This is evidenced by their current excellent performance on OAO-C, their current use on ATS-F, and proposed applications to IUE.

The heat pipe derives its unique capability to transport large quantities of heat over long distances with very low temperature gradients in the following manner (figure 1). Heat is input to the pipe in the region called the evaporator, where it evaporates working fluids such as methyanol, ammonia, or water. The vaporization of this fluid generates a vapor pressure which tends to drive the vapor toward the colder region of the pipe, which we call the condenser. Here the vapor condenses and is returned to the evaporator through a wick arrangement, thus completing the heat pipe cycle.

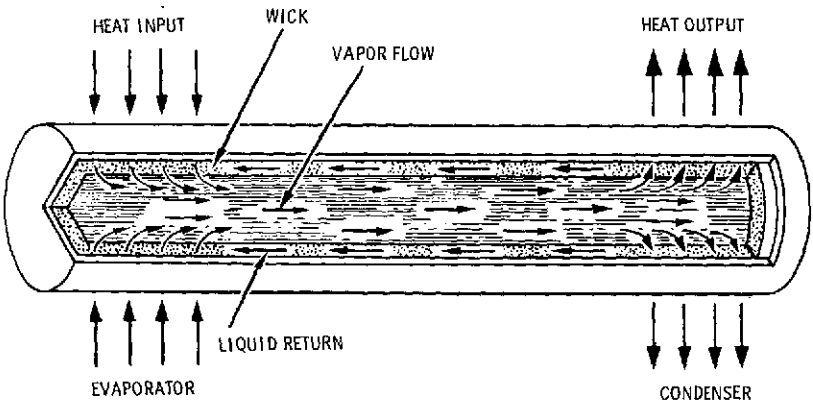


Figure 1. Heat pipe cycle.

One of the major problems associated with the manufacture of heat pipes today is the determination of how much working fluid to put into the pipe. If you put too much liquid in the pipe, it will tend to form a puddle, such as is shown in the upper portion of figure 2. This is no significant problem

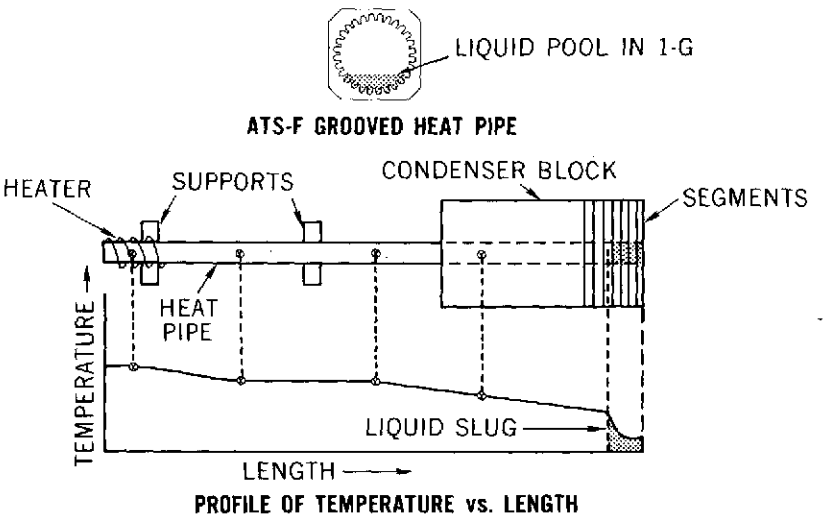


Figure 2. Sounding rocket experiment to determine liquid slugging in zero-G.

on the ground. However, it is postulated that in zero-G this puddle would form a slug as is indicated in the middle section of figure 2, where I have shown a typical heat pipe with a heater on one end and a condenser on the other end, represented by a solid aluminum block. If the liquid forms as we think it will, a solid slug will develop in the condenser of the heat pipe. If there is so much liquid in the pipe that the entire condenser is filled, the pipe will no longer operate because there is no place for the vapor to condense.

In the lower section of figure 2, I have indicated a typical temperature profile along the pipe. In the evaporator region, the temperature will be relatively high; it will be fairly flat in the adiabatic section because there is no heat transfer taking place. In the condenser, the temperature will be slightly lower. But when we come to a slug, the temperature drops off dramatically to the temperature of the heat sink, thus negating the capability of the heat pipe to maintain low temperature gradients.

Within the last 3 years, the Thermal Systems Branch at Goddard has developed a unique capability, we think, to study the performance of heat pipes in zero-G, looking at fluid management such as described above, as well as other gravity-related problems.

We designed an experiment, such as the one indicated in the middle section of figure 2, and flew it on an Aerobee 200 sounding rocket. The rocket gives us the capability of approximately 6 to 8 minutes of zero-G time, which is sufficient to study many of the phenomena associated with heat pipe operation. The data which we got for this particular experiment is shown in figure 3.

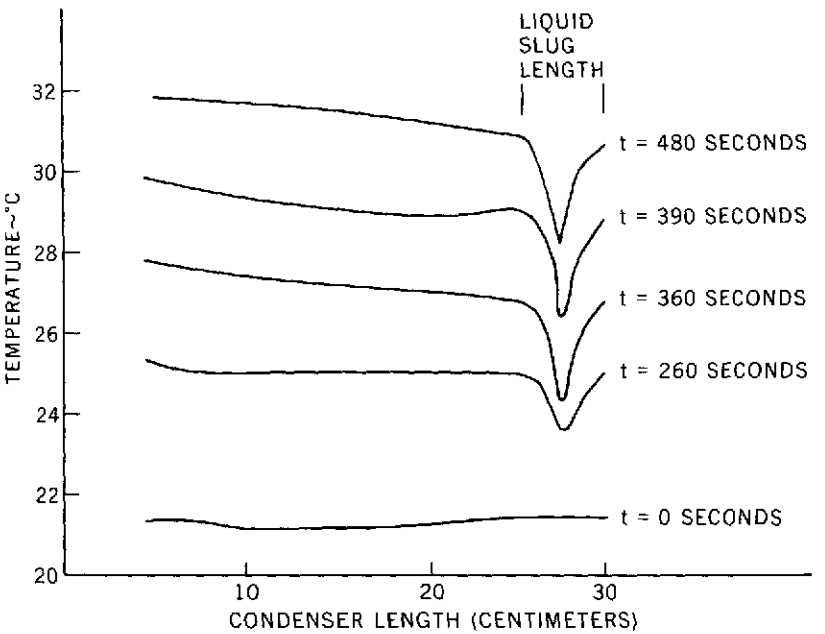


Figure 3. ATS liquid slug in condenser block.

The bottom curve in figure 3 relates the rest performance of the heat pipe where we have a relatively flat temperature profile. As the time progresses in the flight, you will notice that we developed a larger and larger slug in the condenser region. However, the temperature profile is not exactly as we had predicted. If you remember figure 2, you will note here that the temperature right near the end of the pipe has gone up again. This indicates that perhaps our theories regarding the formation of liquid slugs in heat pipes in zero-G may have been in error. Now we think that it may be possible that the fluid forms an annulus in the condenser region of the heat pipe. This would be very nice for us in some respects because we could stop worrying about the slug, and it would enable us to handle the heat pipe more effectively.

There is also a problem with a heat pipe if you have too little working fluid. If you have less than approximately 70 percent of the total fluid required to fill the wick, the heat pipe will cease to function at all and will be no better than an aluminum or stainless steel tube. An experiment was flown on our last sounding rocket to study this phenomena as well.

Figure 4 shows a typical profile of maximum power throughput for the heat pipe as a function of fluid fill in the pipe. This particular curve is for an ATS type heat pipe, and is therefore of considerable interest. The ATS is condenser limited. A slug in the condenser might cause some difficulty. The data which we received are indicated by the two x's in the graph, and came very close to the profile which we predicted.

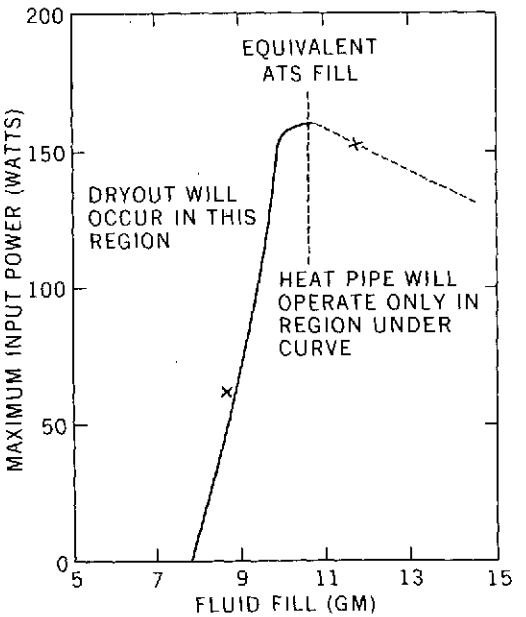


Figure 4. Maximum power handling capability versus fluid inventory.

We believe that this technique of flying heat pipes on sounding rockets offers significant advantages in terms of cost and very short lead time (which allows us to look at new designs), and it also gives a good indication of how these heat pipes are going to operate in zero-G. We believe that if this program is continued, we will gain useful information which will eventually allow us to greatly increase the cost effectiveness of future heat pipe programs such as IUE and shuttle-related activities at Goddard.

EFFECTIVE CLEANING PROCEDURES FOR HEAT TREATED ALUMINUM HEAT PIPES

Carl Johnson

The thermal control of the ATS satellite depends on the efficient operation of some 55 internally grooved aluminum heat pipes containing ammonia. When a heat pipe was cut open for inspection, a gray-black residue was found on the inside. Because of this, a full-scale laboratory investigation into manufacturing and cleaning procedures was conducted.

The top section of figure 1 shows a small heat pipe complete with an end cap and a fill tube. The bottom is an enlarged cross section of the heat pipe, showing the grooves and the teeth. Since these pipes need full strength to resist internal pressures during elevated temperature bonding into honeycomb panels, they are heat-treated after the endcaps and fill tubes are welded on. To prepare for heat treatment, an internal vacuum is pulled in the pipe through the fill tube opening. While holding the vacuum, the fill tube is pinched and welded closed. The pipe is then heat treated at 799.15 K (526°C) with presumably a moisture-free vacuum environment inside the pipe. If any water is present in the heat pipe during the heat treatment, the water, of course, is converted to steam. At the heat treatment temperature, the steam breaks down and atomic hydrogen diffuses into crevices or into the grain boundaries of the metal. At these places, the hydrogen reconverts to molecular hydrogen. Miniature explosions take place and porosity holes are created. This phenomenon is a form of hydrogen attack or hydrogen embrittlement.

Our investigation showed that the amount of water in a closed pipe during heat treatment is directly related to the degree of porosity. Figure 2 shows enlarged, photographic cross sections of heat pipe teeth from six different heat pipes. It is the large white areas with the black porosity specks that we are talking about. Two additional ATS heat pipes which were on life tests were cut open. In one (A) we found a moderate degree of porosity. In another (B) we found a more severe degree of porosity. In the laboratory, controlled amounts of water were added to heat pipes prior to heat treatment. The porosity in (C) represents one drop of water in a heat pipe 1.5 meters long. In the case of (D) several drops of water were added to a heat pipe of the same length. In the case of extreme porosity (E) there was enough water added so that the porosity holes actually linked together and formed a crack.

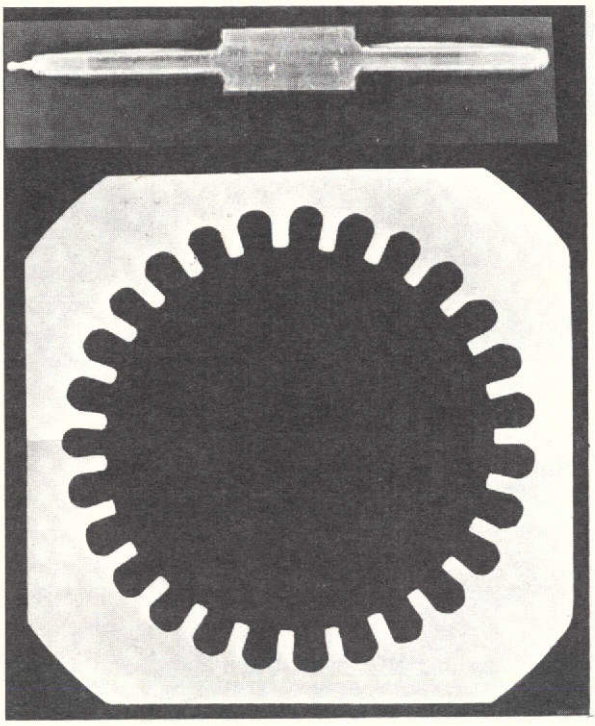


Figure 1. Heat pipe and enlarged cross section of the same.

To prepare for heat treatment, the manufacturer was pulling the internal vacuum through a very small opening for less than 5 minutes. In a series of laboratory vacuum pump-down tests at various temperatures, we determined that 4 hours was necessary and that, if you heated the pipe to 588.15 K (315°C) during that pump-down, even the water of hydration was driven off. This vacuum time/temperature pump-down technique was incorporated along with other cleaning improvements into a cleaning process which solved the insidious hydrogen porosity related problem. This process was used to fabricate ATS and IUE heat pipes, and the resulting interiors were clean and free of porosity as illustrated by tooth F in figure 2.

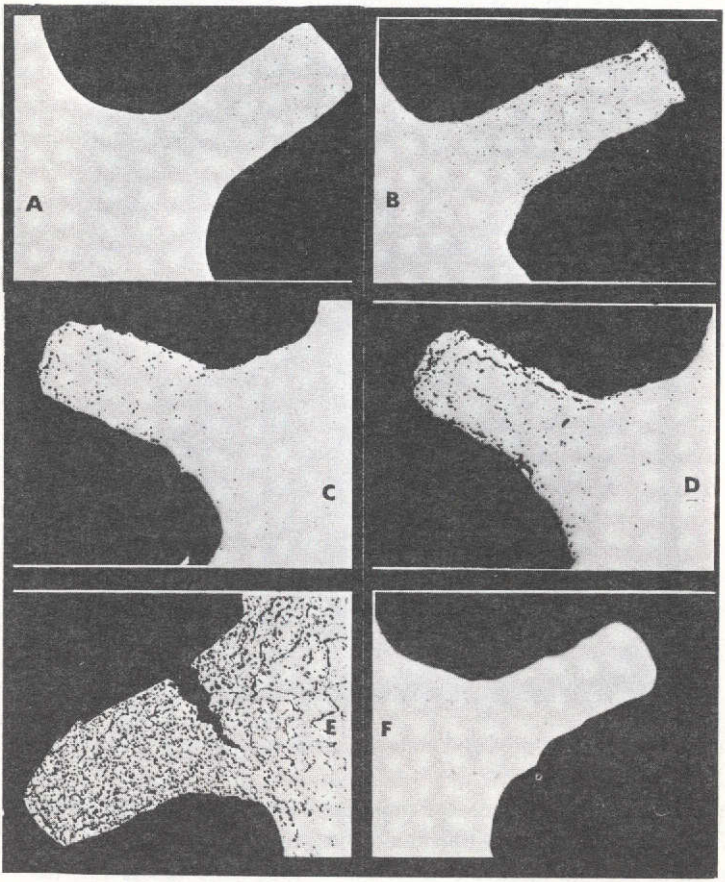


Figure 2. Cross sections of heat pipe teeth from six different heat pipes, showing degree of porosity.

OPTIMIZATION OF PARTICLE SIZE IN A DIFFUSE BLACK COATING

Thomas Heslin

As radiation detectors become more sophisticated, one of the problems encountered is reduction of optical noise from surrounding surfaces. Low absolute total reflectance and low specularity are generally properties that are required.

At low grazing angles, this type of problem becomes particularly manifest because most surfaces become specular. It was just such a problem on ITOS-D that initiated this study. We needed to paint the thermal insulation on the earth-facing panel of the spacecraft in order to reduce calibration errors in the visible channel of the very high resolution radiometer. We tried using a Black Velvet epoxy type paint, but it was too brittle and cracked off the insulation when the insulation was flexed. So we tried to combine the ease of application, good mar-resistance, mechanical flexibility, and good outgassing characteristics of a black polyurethane vehicle with the good optical properties of glass spheres as exemplified by the 3M Black Velvet paints. This marriage was successful and, in addition, we obtained some interesting and useful information with regard to the effects of particle size.

Glass microballoons were fractionated into 20- to 37-μm, 44- to 53-μm, and 63- to 74-μm particle sizes, and mixed into separate charges of the polyurethane base paint. These paints were sprayed onto aluminum substrates and measurements were made of total emmitance, absolute reflectance, specularity, and surface roughness. The most noteworthy result of these measurements was that in the visible portion of the light spectrum, absolute total reflectance and specularity decreased with increasing balloon size for low grazing angles of incidence.

Goniophotometer measurements were made using the system shown in figure 1. Monochromatic light was used to make the measurements. The instruments detector traveled in the plane of incidence which is the plane of the figure, and measured the fraction of incident radiation which is reflected into the solid angle defined by the detector's aperture. This quantity is designated $R(\theta_r)$ in figure 2.

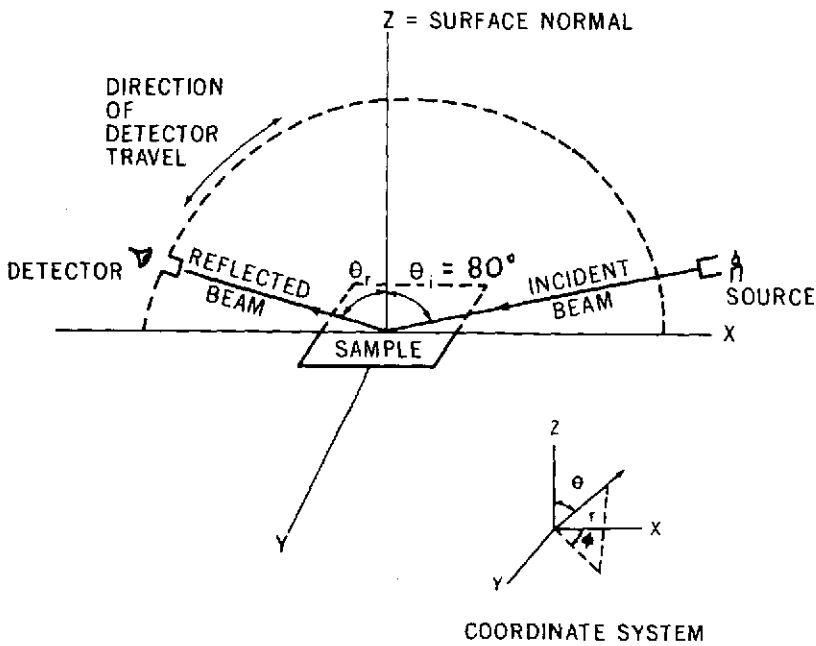


Figure 1. Goniophotometer schematic.

In figure 2, $R(\theta_r)$ is plotted versus θ_r, the angle of reflection. This plot shows the large effect glass balloon particle size has on specularity. For example, the specular peak in curve 4 is about one-third the height of the specular peak for Black Velvet 401, which is shown as curve 5 for comparison.

Absolute total reflectance was also measured at 546 nm, using an 80° angle of incidence. Qualitatively, the results of these measurements could be predicted or guessed at by integrating the curves shown in figure 2. The actual measurements show that using 63- to 74-μm balloons decreased the absolute, total reflectance by more than 40 percent relative to the unfilled base paint.

In summary, this study began by trying to marry the good mechanical properties of a black polyurethane vehicle with the good optical properties due to glass, heading for a particular application on ITOS. However, it has been found that this paint system is generally applicable to optical systems where there is a low angle first or second incidence line into a detector. Therefore, it is presently being used on the interior of the optical train of SAS-C and on the 26-in diameter sunshade of IUE.

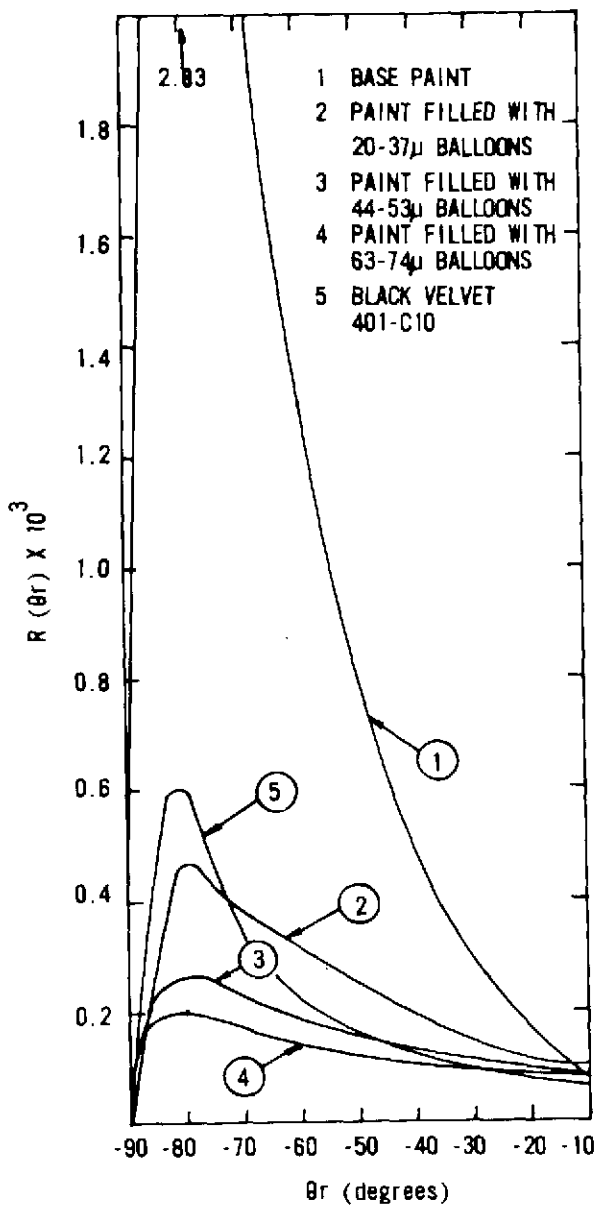


Figure 2. Effects of particle size on specularity.

FIVE-YEAR MAGNETIC TAPE FOR SPACECRAFT

Carl Powell

The program to develop the spacecraft magnetic tape was undertaken to alleviate a severe problem facing the manufacturers and users of spacecraft tape recorders: namely, the use of commercial tapes in a mode contrary to their designed purpose.

The problem was the use of commercial tapes on unattended spacecraft tape recorders. Commercial tapes are not designed for the numerous repeated cyclings required of spacecraft recorders, nor are they intended for operation without maintenance or personal attention. So the problem rooted in low-wear resistance manifests itself in tape debris and friction buildup. Whenever either or both occurs, a change of tape becomes the only answer. Thus, we have inflated recorder costs and limited life in orbit. To explain somewhat, I should say that during production, tape changes are not uncommon. A tape change, planned for or otherwise, will move the emerging recorder back several squares in the production flow. Recovery is costly in time and money. In orbit, the need to change tape can best be proclaimed in an obituary. One year plus or 20,000 passes is par for today.

Growing out of this need, the tape development program has had, as a prime objective, tape durability for 100,000 passes without destructive deterioration, starting from origin at the factory to shutdown in flight. Despite the mystique that the manufacture of magnetic tape is known only to a few mysterious figures, working their magic behind locked doors, IITRI, our contractor, formulated several binder systems and successfully coated these using the precision machinery of the industry. In the normal sweep of work, lubricating materials of different genre and class were investigated. A Dupont fluorinated hydrocarbon, trade-named Krytox, astonished everyone with unbelievable results.

The wear test results speak for themselves. As you can see in table 1 under IITRI, with Krytox as a lubricant, we exceed five million passes with both commercial and IITRI tapes, commercial tapes typified by the 3M variety, for example, 888 or 900, which are both instrumentation tapes. The question now is, "How do we fully exploit this unexpected windfall?"

Table 1
Wear Test Results of Krytox

	Commercial	IITRI
Without Krytox	< 40,000 passes	100,000 passes
With Krytox	> 5,000,000 passes	> 5,000,000 passes

First, commercial instrumentation tapes for longitudinal fixed head recorders of the types flown on Nimbus, ITOS, and the upcoming shuttle, or those for the standard recorder, if treated with Krytox could conceivably extend their lives and usefulness well beyond the lives of the mechanical transports themselves. This should also include and apply even more emphatically to the multitrack longitudinal recorder for EOS, simply because of the ultra high lineal and track packing densities required. The ERTS recorder, a victim of early tape wear, could also benefit. But more exploration is needed here to overcome the limited history accumulated to date on rotary head video tapes.

These tapes have binder systems and surfaces somewhat different from the instrumentation types we more thoroughly analyzed. There is sufficient test evidence now to warrant an intensive pace of activity if we are to capitalize on this find for ERTS-B.

A second way in which Krytox will be beneficial is in permitting a return to low cost heads. Low cost heads have been recently bypassed in favor of wear resistant, though costly, hard-faced heads. These can again safely be considered for many applications.

Third, recorders outside the special domain of spaceflight might also profit from this new found tape and head durability. For example, computer recorders may well be spared the adverse economics of short head life and the dread of errors certain to occur when the head tape magnetics deteriorate with tape wear.

Despite these most encouraging prospects, much remains to be done to fully confirm this lubricant in all aspects related to performance in tape recorders. We are currently enlarging our efforts in this area and, with the superior oxide-binder system derived in this development program, we foresee the 5-year magnetic tape for spacecraft as an unquestionable reality.

ULTRA ACCURATE INCREMENTAL ANGULAR ENCODER

Robert Fulcher

The need for this development became apparent as a final link to achieve low rpm performance of direct-drive, precision phaselock speed control systems. Applications requiring such speed control are advanced tape recorders and earth sensors. In these applications where the objective usually is to maintain one or more constant speeds with very low jitter, the encoder emerges as the critical feedback element that ultimately limits phaselock performance. An encoder's cycle per cycle accuracy translates directly into speed control precision.

The encoder design which was evolved to fulfill the needs of the applications cited is shown in figure 1. This encoder was designed and built by the Baldwin Electronics Company under Goddard Space Flight Center technical direction. It has 2^{14} cycles to accommodate direct-drive phaselock operation down to 2 rpm. Axial and radial TIR runouts at the disk mount were only 1.7 and 1 μm respectively.

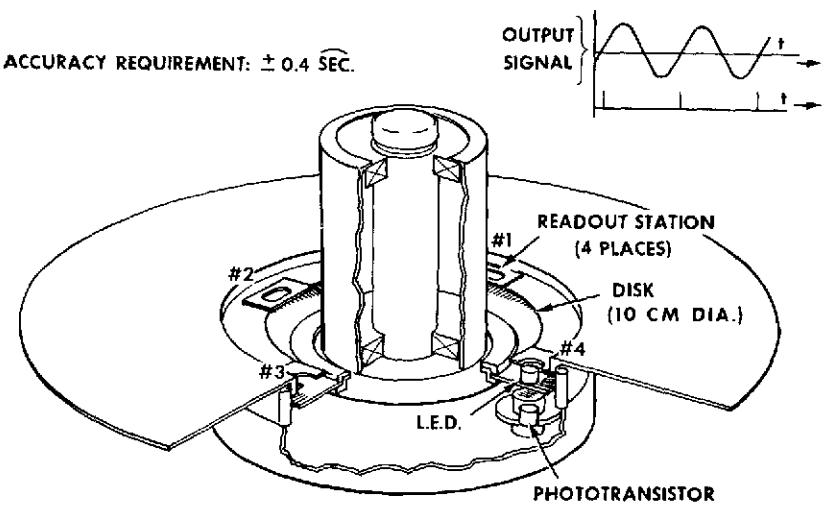


Figure 1. Incremental angular encoder with 16,384 cycles/revolution.

Precise as this is, however, the encoder's fabrication represents only the sound application of state-of-the-art techniques in the photo-optical encoder field. In figure 1 we show the photo-optical portion of the encoder, with the light-emitting diode source, the phototransistor sensor, and with the high density disk going through the readout station. But state-of-the-art techniques alone could not, in this package size, begin to satisfy the stringent ±0.4 arc-second accuracy requirement. To illustrate, when this accuracy is considered in terms of the 10-cm disk diameter, it is translated into only ±0.1 μm allowable error at the readout station locations. The novel means chosen to achieve the required accuracy of this encoder was to seek error cancellation by electrically summing the outputs of the multiple readout stations on a single track. In this case, we have stations 1, 2, 3, and 4. The upper right-hand portion of figure 1 illustrates this combined summation of the signals and indicates also how a pulse readout was achieved.

Measuremetric analysis had shown that multiple readout of a track would give the simultaneous information needed to detect harmonics of an error function. Once detected, combining such signals would then provide the means of reducing those particular error harmonics. For example, four readout stations should provide for the effective suppression of up to second harmonic errors. Figure 2 shows the effectiveness of this approach for the design in figure 1.

Figure 2 shows the amplitude modulation improvement that was available through the four-station combined approach. The top portion shows the amplitude modulation that was present in the individual readout stations; here the sinusoidal outputs of the encoder are shown on a very compressed timescale so the amplitude modulation is clearly visible. The lower portion of the figure shows the greatly reduced amplitude modulation present in the four stations' combined output. It is noteworthy also that this quality of output was achieved using light-emitting diodes (LED) light sources in the encoder design.

Figure 3 shows directly the final accuracy improvement achieved through the four-station approach. Reading down the side, one can see the effect of employing progressively more readout stations. At the top is shown the one-station measurement. I should mention that this is for one full revolution and includes all cycles of the encoder output. No cancellation of error harmonics is possible of course with the one-readout-station configuration. Combining the signal from two readout stations 180° apart eliminates the fundamental in the error function, and this result is shown in the center plot–the two-station measurement. Here we see, however, that a significant second harmonic error was revealed in the two-station measurement. The bottom plot shows the resultant error plot when all

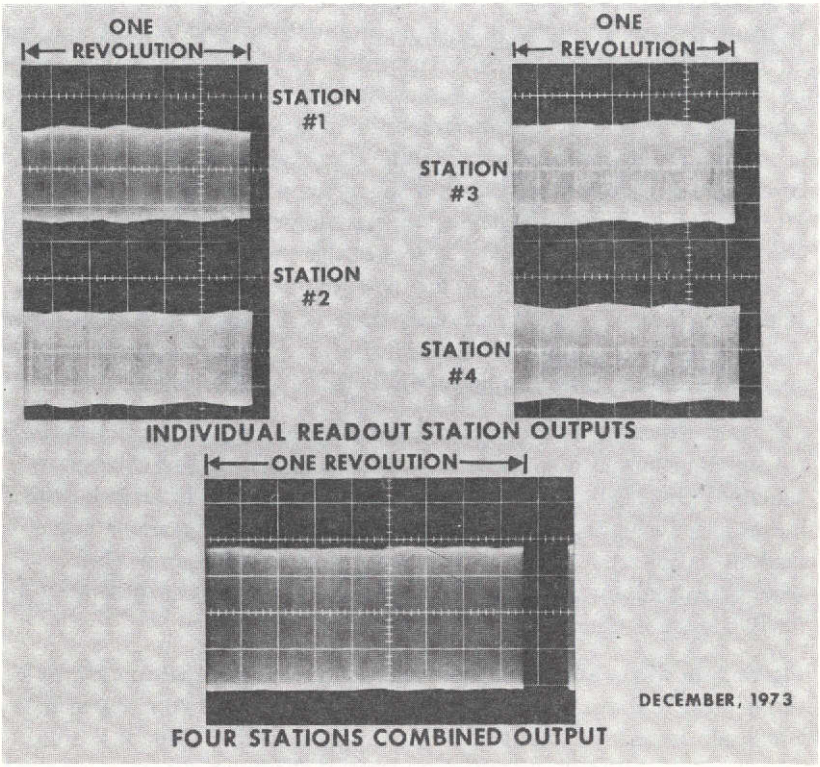


Figure 2. Amplitude modulation of encoder output.

four stations of the encoder are combined to produce the final readout. As seen, the second harmonic no longer has any significant effect in this four-station configuration. The final error measurement, ±0.4 arc-second peak-to-peak error, represents nearly an 80-percent reduction over the other two readout configurations.

In summary, it can be stated that the four-station encoder represents a practical solution to enhancing encoder accuracy beyond what has been previously available. The advancement did not occur by going beyond state-of-the-art manufacturing techniques, but rather by applying a new approach to encoder error cancellation. The choice of four readout stations has been shown to be sufficient to suppress all typical encoder errors while still providing a reasonable package size. The encoder design has direct application to the thematic mapper, tape recorders, and advanced scanning requirements.

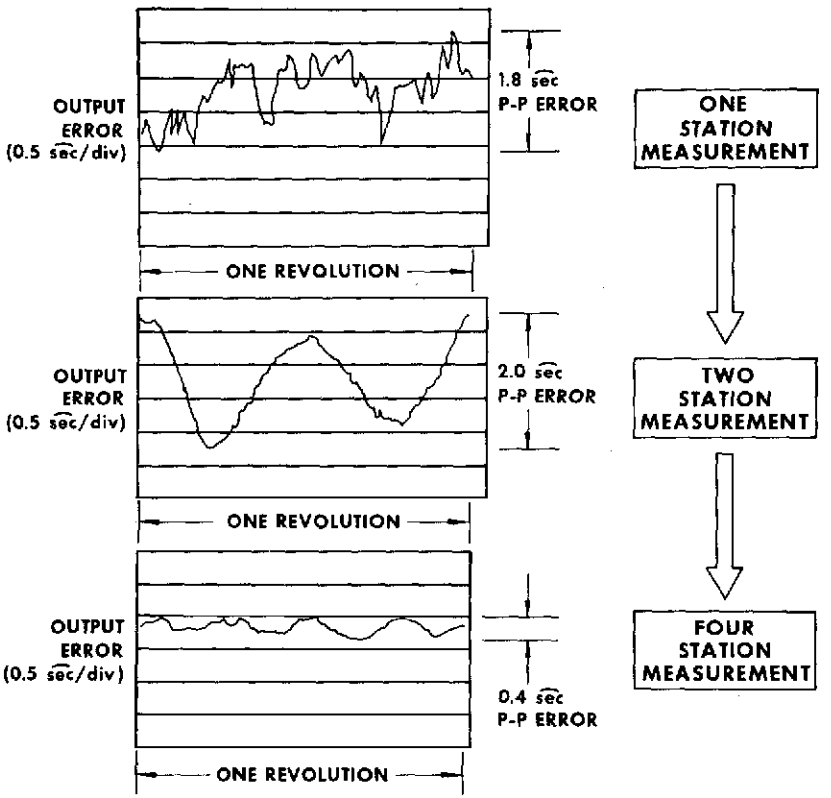


Figure 3. Encoder error plots.

BLOOD FREEZING APPARATUS FOR LEUKEMIA RESEARCH

Thomas Williams

This work was undertaken as the result of a problem statement that was generated by the National Cancer Institute in Bethesda and was circulated through the technology utilization offices. The Auxiliary Propulsion Branch here at Goddard was requested to do this work because of our experience in electronics, cryogenics, and thermal control, and also because of our proximity to NCI.

Researchers at NCI are working on techniques to freeze and store white blood cells for use in leukemia research and treatment. Tests have shown that if a cell is cooled too quickly, it is destroyed by rupture of the cell membrane, whereas, if it is cooled too slowly, the cell is destroyed by dehydration.

Presently available biological freezing equipment is not capable of providing the direct control over the blood temperature that is desired. Consequently, results shown by the solid line in figure 1 are obtained. We are showing here the temperature of the blood as a function of time during the freezing process. The main characteristics are: the supercooling, followed by a rapid rise in temperature at the onset of freezing, a plateau during the main freezing process, and then a rather rapid drop in temperature as you reach the completion of freezing. The important thing here is that the blood is undergoing large changes in cooling rate during this process. It is not known whether the slower or the faster rates are causing the cell destruction. What NCI wants is a system capable of freezing at a constant rate as shown by the dotted line. With a system capable of rates between 1 K/minute (1°C/minute) and 1 K/minute (4°C/minute), it will be possible to experimentally determine if an optimum freezing rate exists. Because of the phase changes in the fluid and its tendency to supercool, this requires a technique capable of large rapid changes in heat pumping capability.

Figure 2 shows the system that was developed here at Goddard for this purpose. The blood sample to be frozen is housed in a polyethylene bag that is sandwiched between two hardened aluminum plates. The outer surfaces of these plates are covered with printed circuit heaters. A thermocouple is mounted on the inner surface of one plate to monitor the blood temperature. In operation, the sandwich assembly is placed in a standard

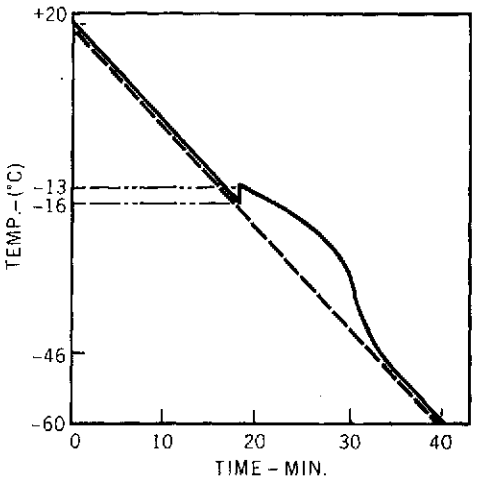


Figure 1. Temperature of the blood as a function of time during the freezing process.

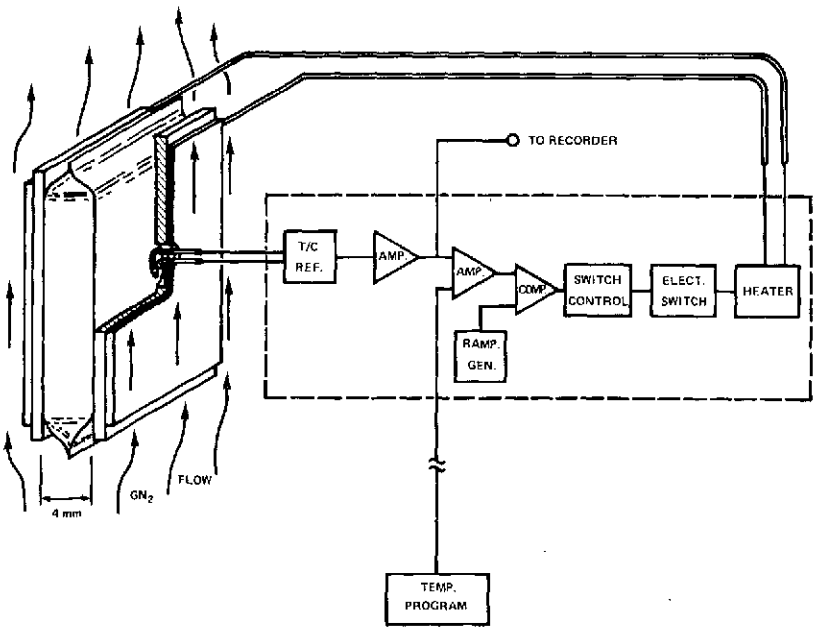


Figure 2. GSFC system capable of freezing blood cells at a constant rate.

biological freezer containing cold gaseous nitrogen as the cooling medium. The heater power is automatically controlled by the electronics so that the temperature at the thermocouple tracks the desired profile that is stored on the temperature programmer.

This design satisfies the following basic requirements: First, the thin bag cross section minimizes thermal gradients within the blood while maintaining a sterile environment. Second, the high thermal conductivity and low heat capacity of the sandwich permit the large rapid changes in heat pumping rate that are necessary to maintain a linear temperature profile. Third, the system is easily expandable so that many samples can be simultaneously frozen in a single biological freezer.

Figure 3 shows blood temperature and freezer gas temperature versus time for an actual system test. It is easily seen that the desired linear temperature profile was maintained throughout the freezing process. Similar results were obtained at rates between 1 K/minute (1°C/minute) and 1 K/minute (4°C/minute).

At the present time the system is being prepared for delivery to NCI.

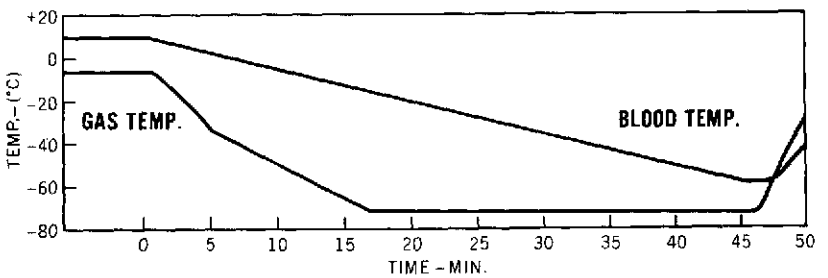


Figure 3. Blood temperature and freezer gas temperature versus time.

OVERVIEW

William Nordberg

First, I want to announce that there has been a program change from the originally printed program. We had to eliminate a few papers because we simply did not have enough time to accommodate the 20 or 25 presentations that we had originally planned. We have selected 16, and the sequence has been slightly changed.

Also, I would like to mention that the first half of the United States ERTS picture mosaic, which is being prepared by the U.S. Department of Agriculture's Soil Conservation Service in cooperation with the people of Lou Walter's group, is exhibited right outside the entrance here. It is impressive to have a view of the entire United States on a scale of one to one million. Of course, we just have the Western half as yet, and are planning to finish the rest of the United States in the next few months.

There is also a report on advanced scanners and imaging systems for earth observations, which was prepared by a working group in a meeting held just about a year ago at Cocoa Beach and was edited by the people in Code 730.

The earth observation program encompasses a number of significant elements which I have enumerated many times already, but to put into perspective the presentations which you will hear in the next 2 hours, I shall recapitulate once more the basic milestones of the program.

First, we are concerned with the understanding and assessment of the requirements for making observations of the earth and its atmosphere from space; this includes the analysis of the benefits that result from such observations and a critique of their utility and potential applications. Today you will not hear any presentations on this subject since NASA has relied in the past on the so-called user agencies to conduct such assessments and analyses. Based on inputs from NOAA, this has resulted in the definition of a number of specific objectives for NASA's meteorology program, among them the design and development of systems which permit the understanding and forecast of severe and short-lived weather hazards, the long-range forecasting of weather, the monitoring of air pollution, and the complete description and understanding of factors which affect this planet's climate.

C - 3

The era of earth resources surveys from space which was inaugurated with ERTS-1 has convinced us of the need to place much greater emphasis on requirement and benefit analyses than before, because of the vast diversity of applications that can be derived from observations of the earth's surface. We have a commitment to strengthen this particular effort and I hope that next year there will be presentations on this subject.

The second programmatic element is to develop new concepts and techniques for making earth observations from space-sometimes called signature studies. Much work has gone on in this area in the past and we shall hear four particularly recent examples of such work in the presentations by Schmugge, Webster, Wilheit, and Hansen.

The third element is to design and operate earth observing instrument and supporting techniques. This is an area where we have been traditionally strong and you will hear several presentations on this subject: two by O'Brien and Vermillion on the performance of existing systems and those by Thompson and Feinberg covering developments which are applicable to future missions such as Nimbus-G and EOS.

The final milestone is to extract useful information from the various observations, model and understand the processes and phenomena relating to the environment, and most importantly, demonstrate the utility and applications of the results. In this area, we had much catching up to do and it is therefore particularly gratifying that we have seven presentations in this area covering a wide variety of applications. Rango and Salomonson will present papers on floods, water quality, and water runoff, respectively; Anderson will cover land pollution; Gloersen's and Theon's presentations will be on sea ice and precipitation surveys; and Halem will do weather forecasting.

Please keep in mind that these are only selected examples of highlights of our work in this area and not even in toto do they constitute a comprehensive description of the overall program. Also, the sequence of the presentations will not follow the four milestones as I mentioned them. Instead, they will go from earth-surface-oriented concepts and applications to atmosphere-and-weather-oriented concepts and applications and finally to the design and operation of instruments and supporting systems.

ERTS-1 SNOWMELT RUNOFF RESULTS

Vincent V. Salomonson

Many areas of the world receive most of their water supply for irrigation, hydroelectric power, and human consumption from runoff derived as a result of the melting of snow deposited in high mountainous regions. Since the snow is deposited in areas where it is normally difficult to gain access for conventional measurements, the satellite offers an attractive means of repetitively monitoring the extent and change in extent of the snow cover. Meteorological satellite observations have shown evidence that snow cover over large regions can be monitored from space, but the recent launch of the Earth Resources Technology Satellite, ERTS-1, has provided relatively high spatial resolution observations that offer a tool with which to monitor snow cover in individual watersheds, such as those in the Western United States.

A study has been initiated which seeks to relate ERTS-1 snow cover observations to runoff in selected basins to ascertain if these observations can be used for improved runoff prediction. Since at this time ERTS-1 data exists for only one snowmelt season, several river basins have been examined in a given region to assess if well-behaved relationships between satellite-observed snow cover and runoff exist that reflect physical differences in the watersheds. The test area chosen is in the Wind River Mountains in northwestern Wyoming. Figure 1 shows this area and changes in snow cover during the first year of satellite operation, as obtained from ERTS 0.6- to 0.7 μm observations. Seven watersheds were examined, ranging in size from 1207 km^2 (466 $mile^2$) to 196 km^2 (75 $mile^2$). Snow cover was obtained from all ERTS-1 scenes over the watershed, and the percent of each basin covered by snow was measured planimetrically and plotted versus observed runoff per unit area.

In order to summarize the overall results, the average snowmelt and runoff curves from these drainage basins are shown in figure 2. The figure shows that a well-behaved relationship normally does exist between the percent of each basin covered by snow and the runoff per unit area when plotted versus time. The snowcover estimates are accurate to ±5 percent.

To ascertain if predominent physical differences in the watersheds are reflected satellite snow cover-runoff relationships, the results from all watersheds with mean elevations below 3050 meters (10,000 feet) were contrasted

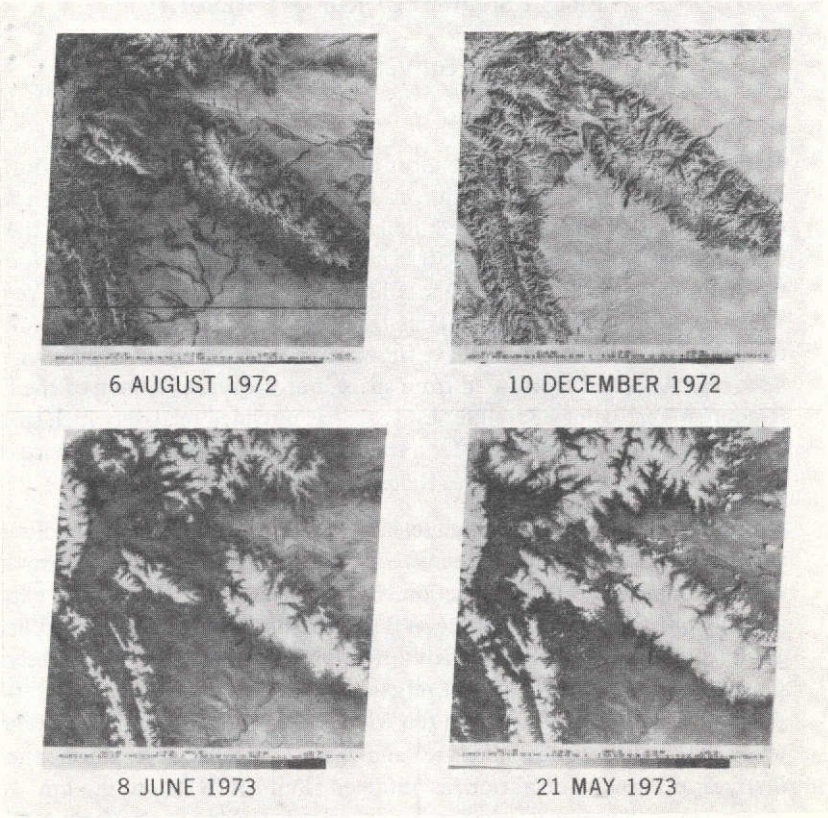


Figure 1. Snow cover changes in northwestern Wyoming.

with the average relationships derived for all watersheds with a mean elevation above 3050 meters. The results are shown in figure 3. These graphs show that the runoff per unit area is larger for the higher watersheds and peaks later in the season. Note that the peak runoff is in early June on the low watersheds and late June for the high watersheds. Since the higher watersheds have deeper snow and lower temperatures, these results seem quite reasonable.

The physical reasonableness of these results from one year are encouraging for future efforts that will attempt to composite results from several years of high resolution satellite data so as to obtain a family of curves permitting estimates of seasonal runoff. In addition, these results suggest that relationships such as these, developed on watersheds where runoff is measured, may be applied to other ungaged watersheds having similar physiographic characteristics for improved estimates of watershed yield due to snowmelt.

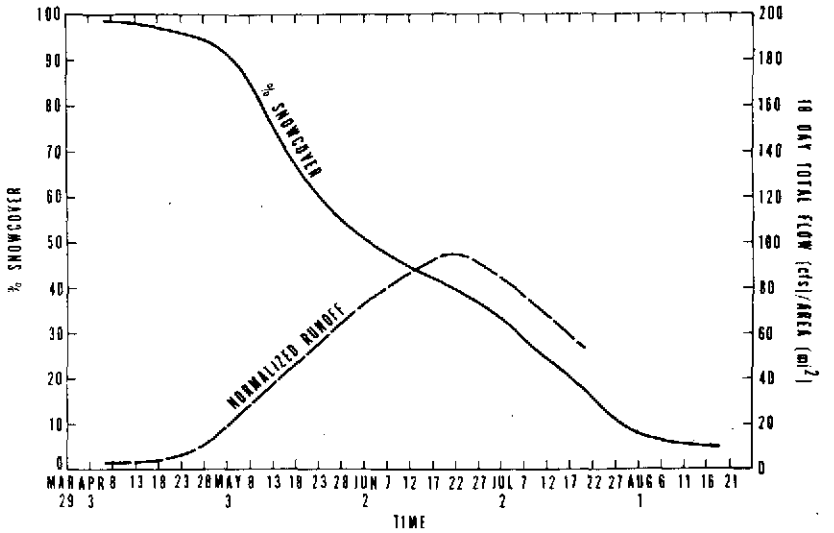


Figure 2. Average snow cover depletion and runoff curves for seven watersheds in the Wind River Mountains, Wyoming, during the 1973 snowmelt season.

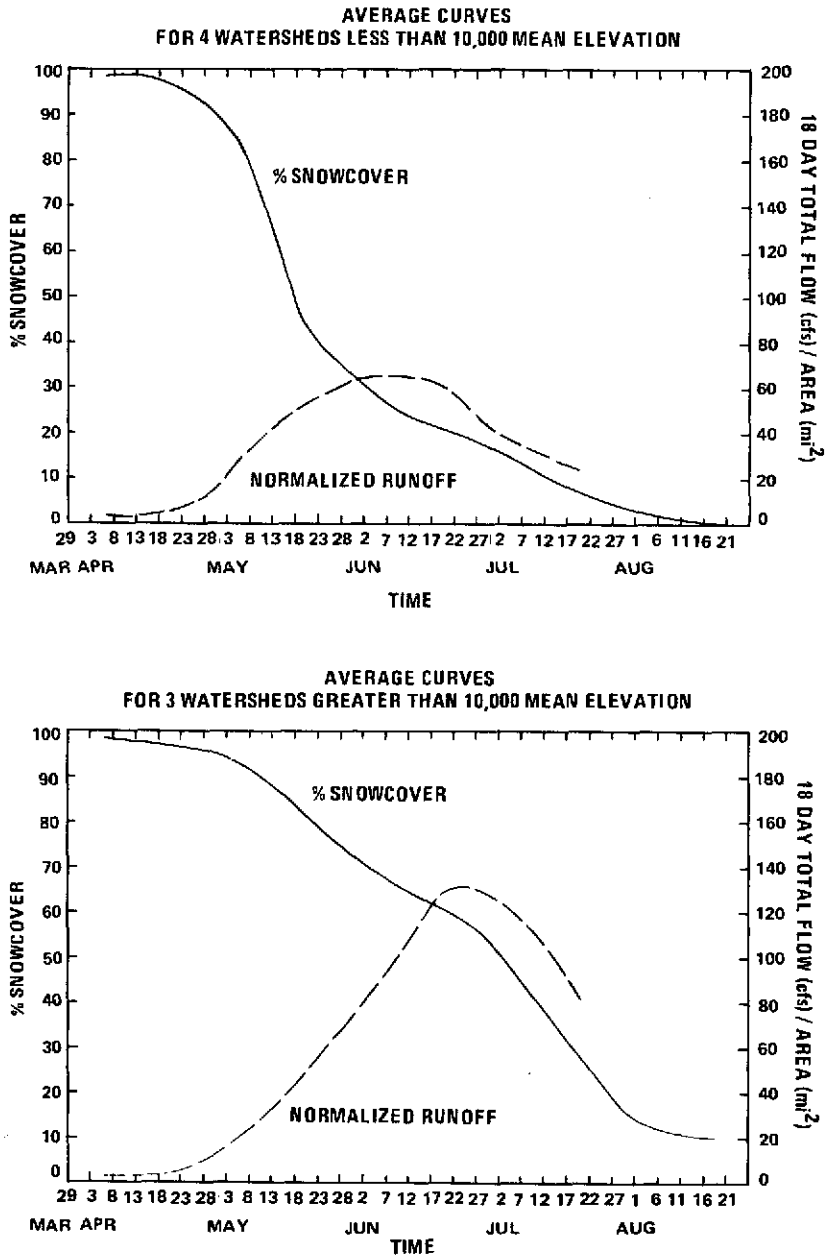


Figure 3. Comparison of average snow cover depletion and runoff curves for watersheds of different elevations in the Wind River Mountains, Wyoming.

ERTS-1 ANALYSIS OF MISSISSIPPI FLOOD

Albert Rango

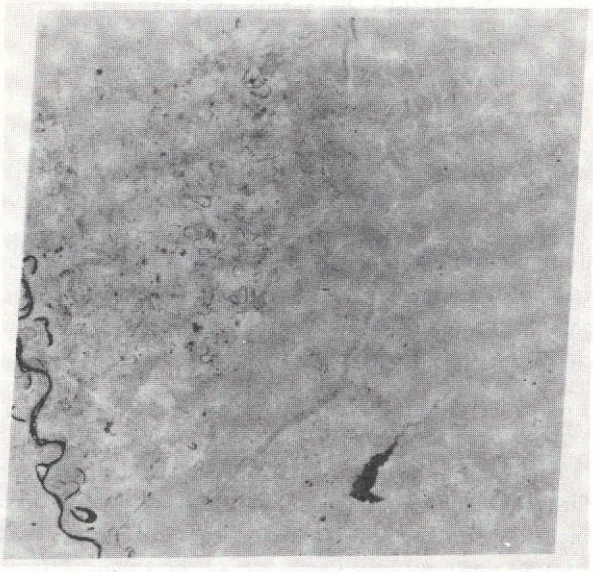
ERTS-1 photographic and digital data have been used for floodplain mapping in the Mississippi Valley and for assessment of the extent of flooding associated with the spring 1973 Mississippi River floods. In order to explore the feasibility of using ERTS-1 to delineate features on the floodplain that are indicators of flood susceptibility, and thus help provide much needed floodplain information to managers and planners, the northwest Mississippi delta region was chosen for study.

Figure 1 is a composite taken from ERTS-1 on September 13, 1972, before any flooding had occurred in this particular region of the Mississippi River floodplain. We decided to take a look at this particular scene and see if we could delineate features on here that were indicators of flood susceptibility. The first thing that we noted was a highly reflective area with associated extensions down into the lower part of the scene. After examination of soil and topographic maps and also some ground observations in this area, we determined that this particular region was a natural levee system. Because natural levees are made up of very coarse soils, we can see them from space as highly reflective materials and, more importantly, these particular areas are 1.5 to 3 meters (5 to 10 feet) above the natural elevation of the floodplain and thus, less susceptible to flooding than these lowland regions.

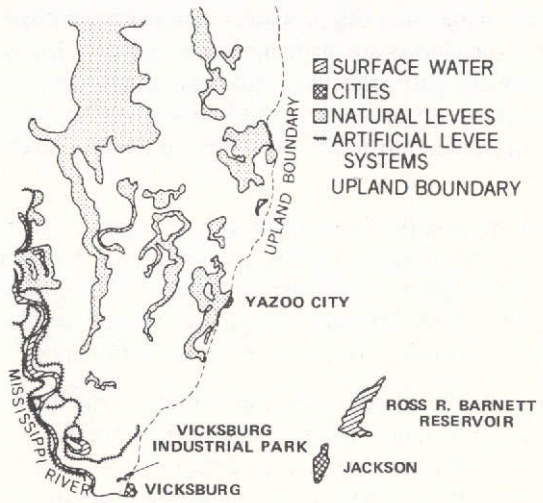
A second feature to note in the figure is the upland boundary, as shown on the overlay. The reason we can see the upland boundary so easily is that there is a marked vegetation contrast caused by a relative relief in this area of about 60 meters (200 feet). In general, it separates the non-flood-prone regions from those areas that are prone to flooding.

A third feature that cannot be seen very easily on this particular composite but was evident on the original is the area labeled the Vicksburg Industrial Park. This is a highly reflective region, and further examination indicated that this area had been flood-proofed–surrounded by dikes and levees to keep it from flooding.

The fourth feature that was delineated in this scene was the artificial levee system along the Mississippi River; it is indicated on the overlay.



(a)



(b)

Figure 1. (a) September 13, 1972, 0.8- to 1.1-μm ERTS-1 photograph showing Mississippi Valley before flooding. (b) Overlay showing floodplain features extracted from the September 13, 1972, ERTS-1 color composite.

In order to determine if this particular analysis was looking at the correct floodplain features as indicators of flood susceptibility, we compared the overlay to an ERTS-1 flood scene taken on May 5, 1973 in the same general area (figure 2). The first thing to note is the upland boundary. As we have hypothesized, it separates non-flood-prone regions from those that are flood-prone. This area is a backwater lake about 64 km (40 miles) across caused by flooding of the Big Sunflower and Yazoo Rivers. This whole area is inundated. The second thing to note is that the natural levees we had indicated on the overlay are still highly reflective and not flooded. This was confirmed by ground observations. As a result, it indicates that the natural levee system is a very good indicator of an area that would not be flooded in this type of situation.

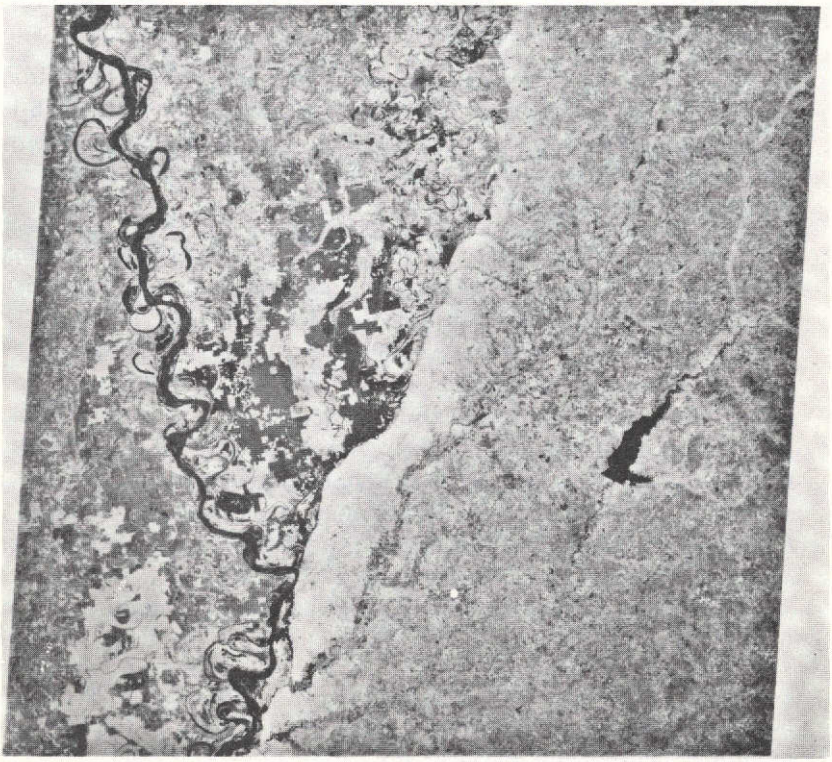


Figure 2. May 5, 1973, 0.8- to 1.1-μm ERTS-1 photograph of Mississippi River Valley flooding.

The third thing that we looked at was the Industrial Park. Again, it is hard to see here but we are able to pick it up as still being highly reflective, and it was indeed a good indicator of an area not susceptible to flooding. The fourth feature, the artificial levee system, did not turn out to be a very good indicator of flooding potential in this region. The reason was that the artificial levee system along the Mississippi was confining the Mississippi flood flow, but the backwater flow was so extensive that it was lapping up behind the levees, negating their original purpose. So, in this particular region, we decided we could not use artificial levees as indicators of whether or not a region was flood susceptible.

We think that it is feasible to use ERTS-1 data in this manner. We have constructed flood-prone area maps along the Mississippi River that are comparable to more expensively derived products of the U.S. Geological Survey and U.S. Army Corps of Engineers.

Let us take a quick look at flood area assessment. Figure 3 was taken from the Goddard IDAMS system which was used in combination with ERTS-1 digital data in change detection analysis of the Mississippi Flood in eastern Arkansas. It shows the White River and the Mississippi; the St. Francis River enters the Mississippi.

Let me explain how this composite was derived. It is a computer combination of a October 1, 1972, scene in eastern Arkansas before flooding and a March 31, 1973, scene under flood. The first thing I want to point out is that a light blue trace (in the original color photo) runs along the Mississippi River and along some of the other streams. The light blue area was determined to have water in it, both in October and March. This is the normal flow condition; through a digital pixel count, which can be done with the IDAMS system, we determined this area to have about 186 km^2 (46,000 acres) of water. Now, the second area to look at was the dark blue area (in the original color photo), which occurs primarily along the White River and the St. Francis River. This particular area (1250 km^2 or 308,000 acres) was covered by water only in March, and thus was determined to be flooded.

We think we can use the digital system to point out these areas that have been flooded or potentially could be flooded. ERTS-1 is opening new areas of information gathering and analysis for water resource managers that should allow them to reduce flood damage and increase their understanding of flood events.

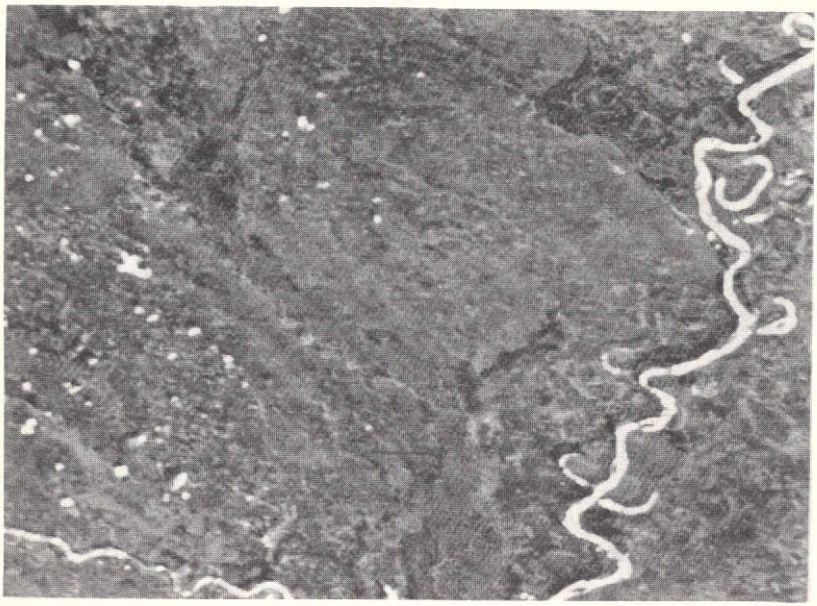


Figure 3. ERTS-1 composite of scenes from October 1, 1972 and March 31, 1973 of area in eastern Arkansas.

GLOBAL ICE SURVEYS WITH ESMR

Per Gloersen

I would like to discuss the culmination of the sea ice studies from aircraft platforms that have been described by some of you in previous years.

An electrically scanning microwave radiometer (ESMR) which measures 1.55-cm radiation emanating from the earth, similar to the one that has been used on the NASA Convair 990 and from which Tom Schmugge just showed you some results, is on board the Nimbus-5 satellite now in orbit. For the first time, it has been possible to obtain a synoptic view of the entire polar regions of the earth, both in the Arctic and Antarctic.

The north polar region is shown in figure 1 for the winter and summer of 1973. These are polar projections of the data obtained from the individual orbital strips, 5 days' worth in the case of the winter scene and 1 day's worth for the summer projection. The black squares indicate an absence of data. At the bottom of the radiometric brightness temperature scale is water, which at this wavelength is typically 130 K for calm seas. This shows up as white on these figures. Since the emissivity of the first year sea ice is close to unity, the brightness temperature is about 240 K in the winter and about 260 K in the early summer.

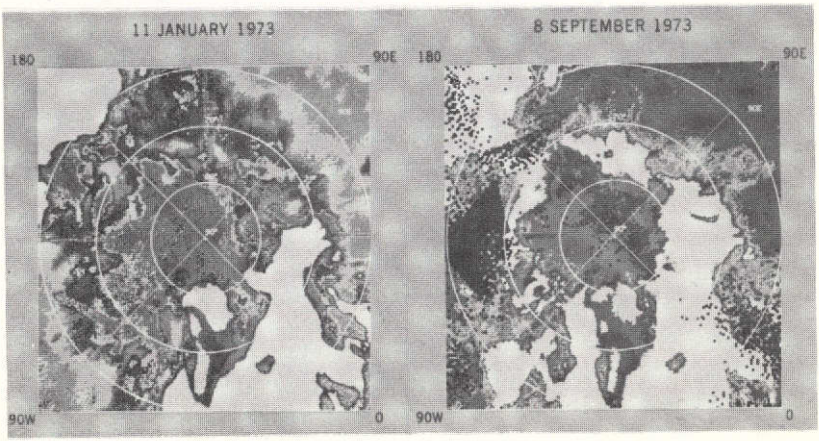


Figure 1. Seasonal changes in the Arctic sea ice canopy.

Multiyear ice, which has a lower emissivity, usually shows up some 20 K lower. These spectral properties of open water and sea ice can be used to infer ice concentration in cases where the open water areas are smaller than the 30-km resolution cell of the instrument. A good measure of the ice concentration in a resolution cell can be obtained from a linear interpolation between the brightness temperature values for completely open water and 100-percent concentration of first year ice.

Significant changes in the ice concentration within a season have been noted when comparing a number of projections, such as the one shown in figure 1. The most apparent difference between these scenes is, of course, the ice coverage in the two seasons. You can see open water all along the shoreline here in the summer around the Beaufort Sea, the Chukchi Sea, the East Siberian Sea, Laptev Sea, Kara Sea, and the Barents Sea.

The multiyear ice has an irregular distribution pattern around the pole. You can see that there are points going into the Beaufort Sea in the eastern part and the East Siberian Sea, which are quite irregular. This had been noted previously with some puzzlement in the Beaufort Sea during the analysis of our earlier aircraft results from that area.

Another interesting phenomenon that we observed while analyzing polar projections such as these is that all of the ice identified as first year ice this past winter did not melt last summer. The persistent first year ice occurred mostly in the western Beaufort, the Chukchi, and the East Siberian seas. What we may have observed is a localized formation of what eventually becomes the multiyear ice in the central polar regions.

In figure 2, the seasonal variation of the ice cover in the Antarctic region is illustrated. The actual continental outline in austral summer is inside the ice boundary on this particular shoreline. So you can see that an appreciable amount of ice persists through the summer. In the late spring scene, not at the maximum ice cover, the beginnings of the ice breakup are vividly depicted around the Ross Ice Shelf where the sea ice has been torn away from the edge of the ice shelf in a jigsaw-like fashion, exposing part of the Ross Sea. Probably the most puzzling phenomenon observed in this area is the first year signature of the sea ice in this part of the Weddell Sea, where the sea ice is known to persist from one year to the next.

These data have been used to prepare both ice concentration maps and maps showing maximum and minimum ice cover for the year 1973. Such maps are important for the future improvement of global weather prediction models and for navigation in polar waters. As examples of the latter, the U.S. Navy Fleet weather facility presently uses our quick-look orbital strip

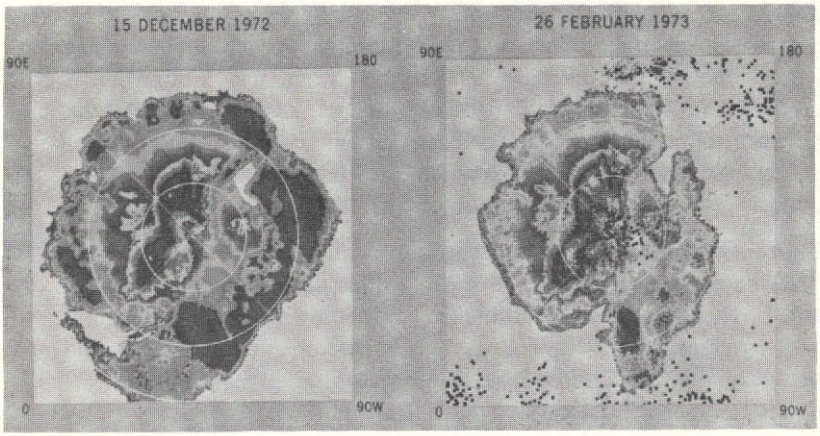


Figure 2. Seasonal changes in the sea ice around Antarctica.

data for preparing their ice edge maps. Also, the National Science Foundation is sponsoring an ice breaker mission into the Amundsen Sea to explore persistent areas of open water along the shoreline shown in figure 2 (Bellinghausen and Amundsen Seas). These polymyas are sometimes as much as 1,000 km inward from the ice edge.

In conclusion, I have illustrated how the scanning radiometer onboard the Nimbus-5 satellite has been used to study the variation of sea ice in both polar regions, both seasonal and within a season. From such studies we hope to learn how to improve our weather prediction models and navigation in the polar regions.

MICROWAVE SIGNATURE OF SNOW

Thomas Schmugge

Microwave radiometers were flown during March 1971 on board the NASA Convair 990 over several snow targets in the western part of the United States to study the variation with wavelength of the microwave brightness temperature for snow. A summary of the results is presented in figure 1.

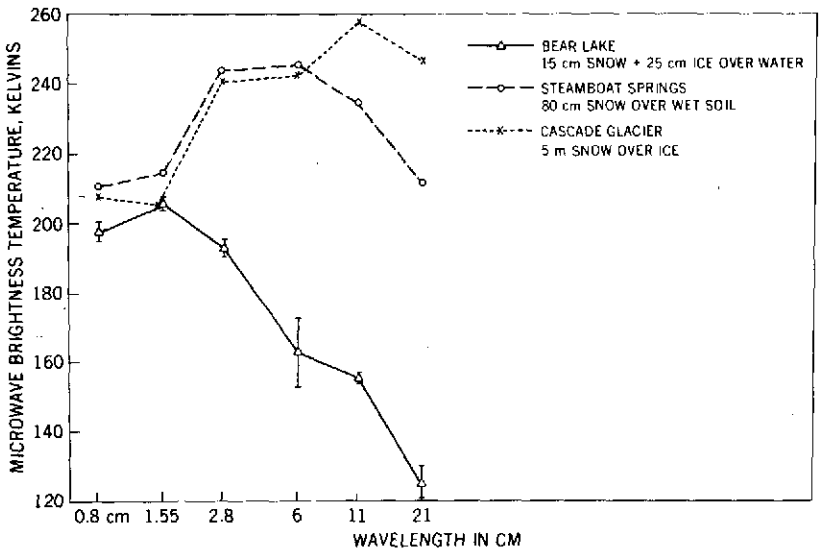


Figure 1. Variation of microwave brightness temperature for snow with radiometer wavelength.

Three sites with different substrata were overflown. The first of these is Bear Lake, on the Utah-Idaho border, where we had 15 cm of snow and 25 cm of ice over the water. The second site was a river valley south of Steamboat Springs, Colorado, where we had 80 cm of snow over a wet soil. The third site that we considered is the South Cascade Glacier in the state of Washington where we had 5 meters of snow over ice. This was at a point near the terminus or the end of the glacier.

In figure 1, we have plotted the microwave brightness temperature versus the wavelength of the radiometers on a logarithmic scale. The radiometers from 1.55 to 21.0 cm were all nadir viewing; the 0.8-cm radiometer looked at an angle of 45° to the rear and received horizontally polarized radiation.

One of the first things that we notice is the large variation in brightness temperature observed by the longer wavelength radiometers, indicating their sensitivity to the media underlying the snow. The short wavelength radiometers, the 0.8- and 1.5-cm, displayed the least amount of variation and, in general, displayed a lower brightness temperature than the longer wavelength radiometers. We believe that this is due to the fact that they were responding primarily to the surface snow, which in general was dry with a density on the order of 0.2 g/cm^3. The lower brightness temperature for these wavelength radiometers, we believe, is due to a volume scattering effect caused by the snow grains.

At Bear Lake, the 21-cm radiometer has the lowest brightness temperature, indicating that it is seeing the low brightness temperature of water through the snow. As the wavelength decreases, the brightness temperature increases until the dry snow values for the 1.5- and 0.8-cm radiometers are reached. Similarly, at Steamboat Springs, the 21-cm radiometer has the lowest brightness temperature, indicating that it is able to see the wet soil through the snow.

At the glacier, the 21-cm and 11-cm radiometers had the highest brightness temperatures, and the values observed were in good agreement with those calculated using the known dielectric properties for ice and snow. As the wavelength decreases at the glacier, the brightness temperature also decreases, until the low brightness temperature values for the 0.8- and 1.5-cm radiometers are reached, which indicates that these low brightness temperatures are in disagreement with the values calculated using the bulk dielectric properties for snow and ice.

The effect of liquid water on these low brightness temperatures is observed in figure 2, which is a false-color, 1.55-cm microwave image of the South Cascade Glacier area.

In this figure we note that the areas with low brightness temperatures (190 to 200 K) correspond to the glaciers as indicated. The point of particular interest is South Cascade Lake at the north end of the glacier. At this point the brightness temperatures are about 20 to 30 K warmer than on the glacier itself. We believe that this is due to the presence of liquid water in the snow covering the lake. The effect of the liquid water is to increase the effective loss tangent for the snow and, therefore, to wash

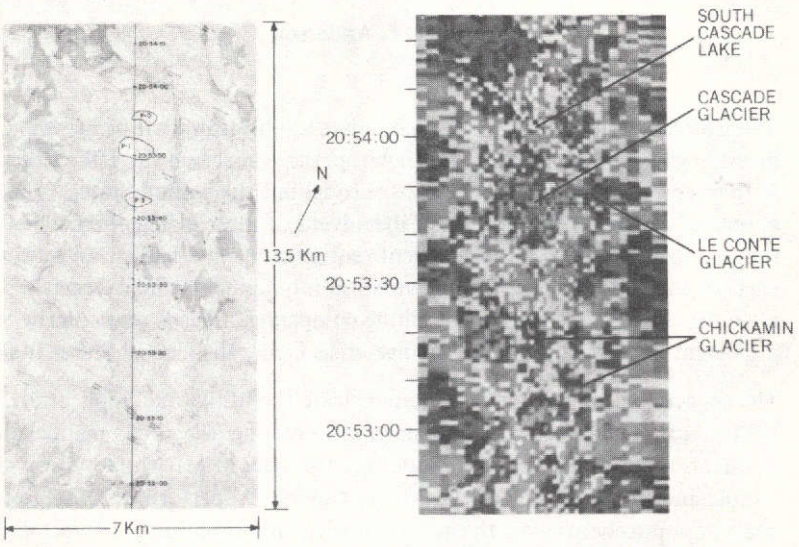


Figure 2. Microwave, 1.55-cm image of South Cascade Glacier area, March 8, 1971.

out the effects of the volume scattering in the snow which had produced the low brightness temperatures on the glacier itself. This rise in brightness temperature over the lake was also observed with the 0.8-cm radiometer.

In conclusion, these results indicate that the effects of volume scattering in dry snow are important for the free space wavelength less than about 2 cm as had been observed earlier from multiyear sea ice. At the longer wavelengths, the 11- and 21-cm, the results are in good agreement with those calculated using the bulk dielectric properties of ice and snow, indicating that scattering may not be an important mechanism at these longer wavelengths.

ERTS-1 ANALYSIS OF STRIP MINING IN WESTERN MARYLAND

Arthur T. Anderson

The ERTS system has shown an immediate application, that of mapping, inventorying, and monitoring through repetitive coverage the effects of strip mining in numerous coal basins throughout the United States. The extent of available coal basins and the advent of more efficient means of strip mining has led to more stringent requirements and legislation on the parts of the states to establish reclamation requirements and extensive planning prior to licensing strip mining companies. This is particularly important in view of the fuel shortage crisis facing the United States today.

The objective of this study is to demonstrate the utility and application of ERTS data to an integrated analysis of strip mining using both manual and automatic data processing techniques for the state of Maryland by providing a rapid and synoptic view of the affected areas. To perform this integrated analysis and to help assess the adverse effects of strip mining, aircraft data and ground verification information aided the analysis and classification of the end products.

Western Maryland has five major coal basins in which strip mining is presently active (figure 1). The focus will be on two of these basins since they contain over 90 percent of the strip mining. Evidence of these disturbed areas in the coal basin can be seen in the four spectral bands of ERTS and best observed in band 5 (0.6 to 0.7μm) and band 7 (0.8 to 1.1μm).

Figure 2 shows the two basins of the Upper Potomac and Georges Creek which lie within a northeast plunging (Georges Creek, Wellesburg) syncline bordering Allegany and Garrett Counties. Geologically, this region lies within the Allegany Plateau and contains 19 meters (62 feet) of usable coal seams in the 510 meters (1700 feet) of sedimentary formations.

The ERTS computer compatible tapes for the analysis of the entire strip mine area were found to be the most accurate method used for this analysis. It should be noted that the advantage of digital data over analog data is that the maximum spatial and gray scale resolutions are available, and the digital values can be manipulated either singularly or in combination to extract this information.

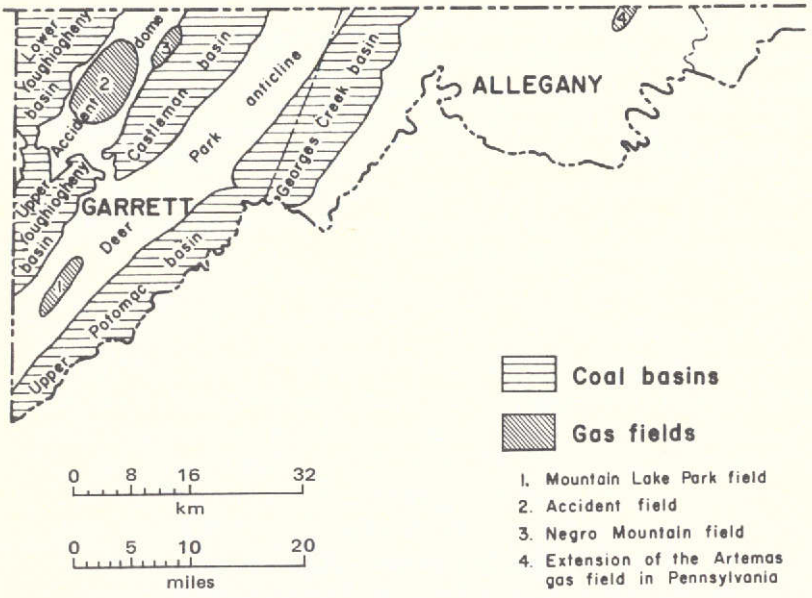


Figure 1. Coal basins and gas fields in western Maryland.

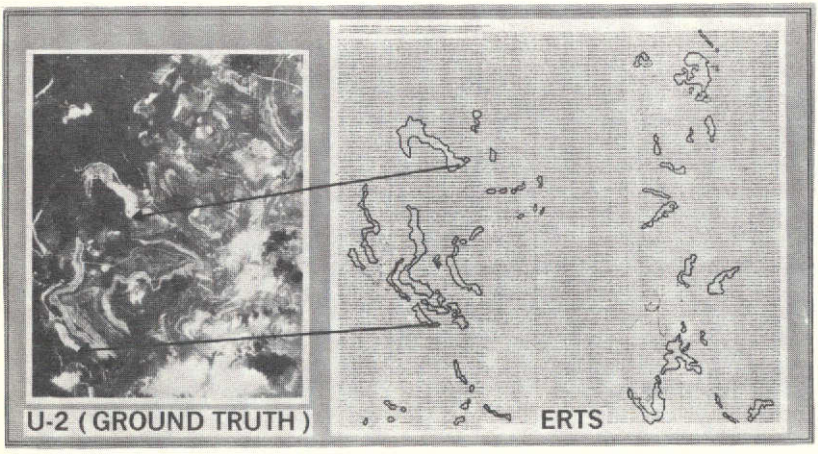


Figure 2. ERTS analysis program output.

To extract this information, an ERTS analysis program was used (figure 2). This program performed a point-by-point digital analysis of band-5 and band-7 spectral signatures based on known spectral values using the International

Biological Program to identify the affected areas. Figure 2 is a segment of the analysis of the two basins showing the computer output in the associated band-5 and band-7 derived classification. The actual spatial resolution of each sensor element is the average reflectance level over an instantaneous field of view of the 57- by 80-meter pixel or picture element. This figure also shows the location of the strip mines and the ground verification from the U-2 aircraft photography used to verify the output, along with ground truth information.

The final results with the use of ERTS data demonstrated the ability to detect and inventory strip mines in the two Maryland basins (figure 3). For example, 140 strip mines were identified in the 12.5 km^2 (3100 acres) of disturbed land. It is interesting to compare these results with the finding of the December 1972 Report of the Bureau of Mines which listed 58 active strip mines affecting 2.3 km^2 (568 acres). This does not mean that the Bureau was unaware of the other strip mines; rather, it indicates that the focus of attention has been placed on maintaining a balance of a 1:1 ratio of new and reclaimed strip mine acreage since the enactment of the stringent laws in 1971. Because of this primary emphasis, little consideration was given to the extent of necessary reclamation activities. The

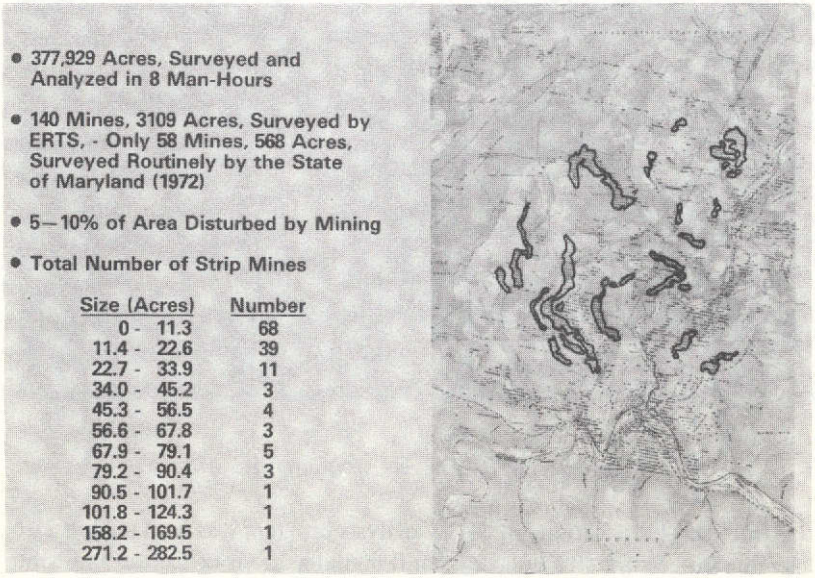


Figure 3. Results of ERTS inventory of Maryland strip mines.

possible contributions of ERTS imagery analysis in this area were recognized by the Maryland Bureau of Mines and the Geological Survey when viewing the computer output transparency product of figure 3, registered to the older 14-year old topographic base.

In summary, the reclamation figures that are now being examined will give the state a better means of detecting and monitoring past and present reclamation efforts for planning and monitoring future reclamation projects. In this context, it should be pointed out that, with the increase in the fuel needs, the 1.4×10^9 kg (1.6 million tons) of yearly production in Maryland will increase by at least 10 percent in 1974. This means that increased priority will be given to monitoring the disturbed regions to observe the increase in progress of stripped, backfilled, and planted acreage throughout the state.

SEA SURFACE STUDIES BY MICROWAVE TECHNIQUES DURING THE BERING SEA EXPERIMENT

William J. Webster

During February and March of 1973, a joint experiment was conducted by the United States and the U.S.S.R. to observe the microwave emission in the Bering Sea region. I shall report on passive microwave observations of the sea made during the course of this experiment. These observations were made in the frequency range from 1.42 to 37 GHz with instruments onboard the NASA Convair 990.

As we shall see, the observations demonstrate the importance of the fetch in determining the emission from the sea. They also show that the dependence of brightness temperature on wind speed is roughly linear and that, at least in the wind speed range from about 11 to 27 meters per second, the major cause of this variation is the change in white water coverage.

Figure 1 shows the variation of brightness temperature at one of the frequencies of observation, in this case 37 GHz made at a nadir angle of 38° while the aircraft approached the edge of the sea ice. The observations were corrected for reflected sky emission by using simultaneous observations of the sky at this same frequency. The two circled points are a region of near-total foam coverage and ice streaks. Note the decrease in the brightness temperature from about 0 hours 22 minutes toward the end.

From other measurements, we can eliminate surface temperatures, salinity, and decreasing wind speed as possible causes for this change. In point of fact, the surface winds actually rose by 3 meters per second during this period. This decrease, amounting to about 7 K in the brightness temperature, can be attributed to the change in the character of the white water coverage and the wave structure due to the decreasing fetch. The first conclusion we can draw is that the distance over which the wind acts on the water (that is, the fetch) is a measurable and important influence.

Figure 2 shows one of the plots of brightness temperature versus wind speed for various low altitude measurements made beyond the region where the fetch is important. The two lines are objective fits to the data points, and the numbers in parentheses next to each of the data points are the

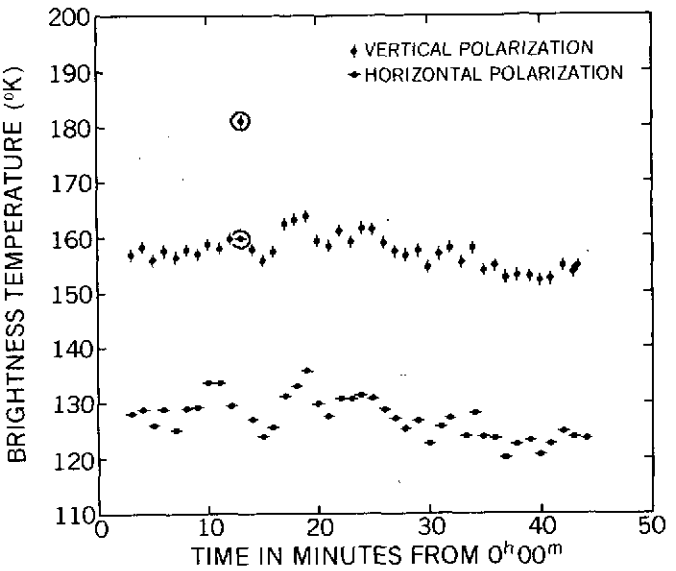


Figure 1. Variation of 37-GHz brightness temperature at 155 m altitude, March 13, 1973.

weights that went into the least squares analysis. In this case, the weights are the number of minutes of observation employed. The second conclusion we can draw is that over the wind speed range from about 11 meters per second to 27 meters per second the variation is roughly linear.

Figure 3 shows the variation of the slope of the various brightness temperatures versus wind speed curves as a function of frequency. Note that the slope rises between 1.42 and 10.69 GHz and remains roughly constant from 10.69 to 37 GHz. This variation is consistent with a thin matching layer model for the sea surface, that is to say, the foam. So a third conclusion which we can draw is that, except for some possible wave effects in the 37-GHz measurement as evidenced by the polarization, the change in white water coverage is by far the most dominant influence in the observed emission.

To summarize, the observed microwave emission of the sea is influenced by wind speed, fetch, and surface temperature, as well as a number of relatively minor effects. In the 11- to 27-meter per second wind speed range, the brightness temperature varies approximately lineally with wind speed, and the dominant cause of this variation seems to be the change in white water coverage.

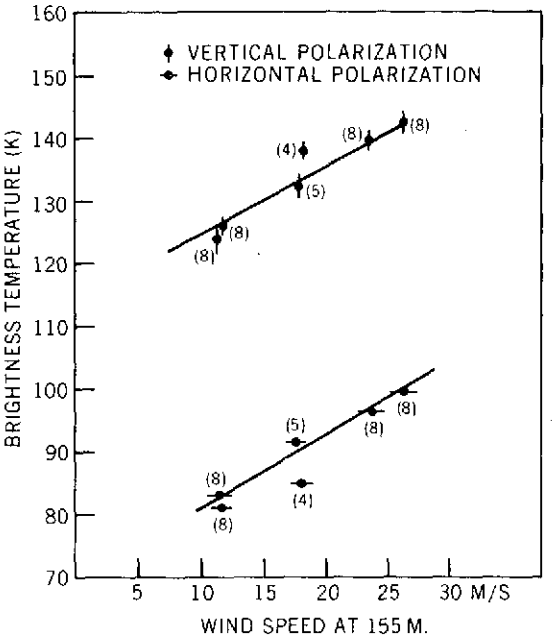


Figure 2. Plot of 10.69-GHz brightness temperature versus wind speed.

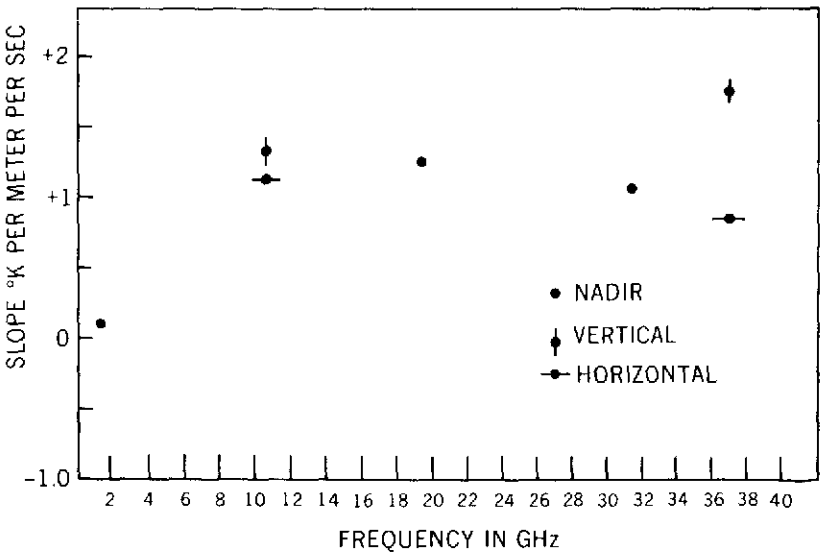


Figure 3. T_B as a function of wind speed slope versus frequency.

CLOUD TYPES FROM POLARIZATION DATA

James Hansen

Remote measurements of cloud properties are made with a variety of techniques in different spectral regions. The work that I am reporting on here is an attempt to define the information obtainable from polarization measurements of clouds.

By polarization I refer to the linear polarization of reflected sunlight and thus to the spectral region from the near UV to the near IR. The nature of the polarization of scattered sunlight is easy to understand. When light is scattered by a particle larger than the wavelength, it becomes polarized in the same way that light reflected by a paved roadway or a flat water surface becomes polarized. Thus the degree of polarization depends on the scattering angle and on the refractive index of the scattering material as indicated by the Fresnel reflection coefficients. The polarization also depends on the cloud particle shape and on the particle size.

For spheres, the polarization arising from a single scattering can easily be computed for particles of any size based on classical electromagnetic theory. For multiple scattering, which occurs in clouds, there are greater complexities, but these can be handled by means of computer programs which we have developed. Our multiple scattering computations indicate that the polarization is a sensitive measure of the particle size, shape, and refractive index.

In the last 2 years, we have used polarization observations of Venus to obtain some remarkably precise information on the nature of the Venus clouds. The cloud particles on Venus are spherical, their refractive index is 1.46 at 0.35 μm wavelength, decreasing to 1.43 at 1 μm. The mean particle radius is 1.05 μm. The variance of the size distribution is 0.07. The cloud tops occur at 50 mbar. This information from the polarization is a basis for the recent identification of the probable chemical composition of the Venus clouds as sulfuric acid.

David Coffeen and I are now extending this work to remote sensing of earth clouds. Figure 1 shows examples of observations on a deck of altostratus and cirrus clouds. The wavelength here is 2.2 μm and the strong feature in the stratus clouds has the same origin as the well known rainbow.

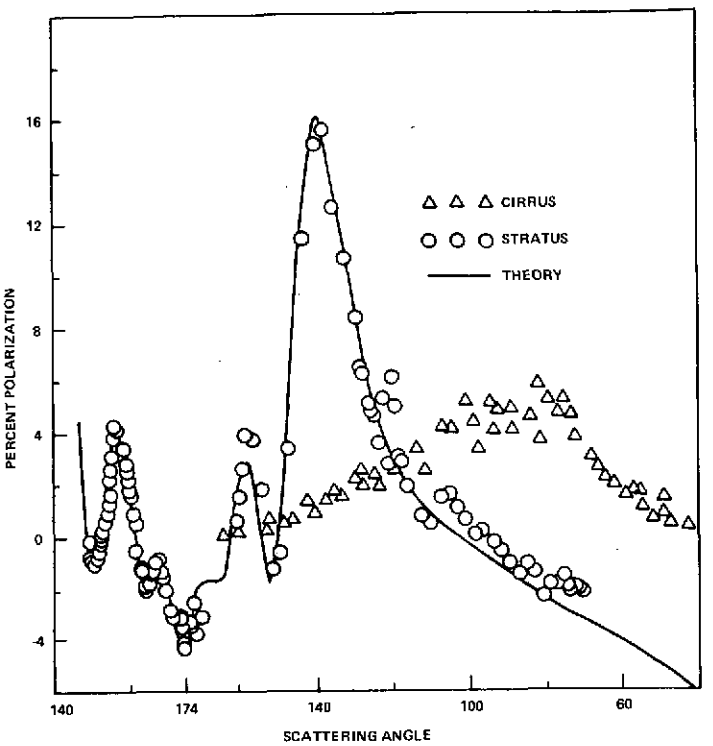


Figure 1. Remote sensing observations on a deck of altostratus and cirrus clouds.

It depends on the spherical shape of the particles. We have found this feature in all of our observations on clouds which were known to be water clouds.

From the theoretical curve and other similar curves, we deduced that the average particle radius in the stratus cloud was 14 ± 2 μm. From the limited number of observations which we have obtained, we are able to verify that the polarization can be used to distinguish ice clouds from water clouds and to measure the average particle size in water clouds. We need to take more observations. We must determine the relations between the particle microstructure and the cloud type. We would like to develop the use of the wavelength dependence of the polarization. It is inconvenient to rely on the scattering angle dependence, because with satellite

observations this would require measurements of a given area from different points in the orbit.

In conclusion, it is interesting to note that the physical properties of the Venus cloud particles are more accurately known than those of the particles in the earth's atmosphere at the same level, the Junge layer. We anticipate that the polarization techniques will be further developed and used for remote sensing of the earth's atmosphere.

RAINFALL INTENSITIES AND ESMR BRIGHTNESS TEMPERATURES

Thomas Wilheit

As you have been told, one of the instruments on the Nimbus-5 satellite which was launched just over a year ago is an electrically scanning microwave radiometer, the ESMR. This instrument is just over a year old, and as of orbit 5000 it was operating perfectly. This instrument measures the thermal radiation in a band centered around 1.55 cm or 19.35 GHz and, since it is a scanning instrument, the data may be formatted in an image such as is shown in figure 1.

The main feature you see in figure 1 is the tremendous emissivity contrast between the land and the water areas. This land has a high emissivity, typically 0.9 to 0.95, and the water is more in the order of 0.4 for its emissivity. This very low emissivity of the water gives a very good background to view the intervening atmosphere. In particular, if there were no intervening atmosphere, the brightness temperature of the water would be about 120 K. There are a number of effects that will raise this brightness temperature: If there is wind, as you have been told, this roughens the surface and creates foam to the extent that, if there were about a 25-m/s (50-knot) wind, it would give a 20° increase. Water vapor in the atmosphere raises the brightness temperature and, except in extreme cases, the maximum would be about 25° increase there. Nonraining clouds could contribute up to about 15° for a sum of about 180 K in the nonraining situation.

Rain, on the other hand, has quite an enhanced effect for two reasons: A rain cloud typically has much more liquid water content than a nonraining cloud; the particle size is larger, so that the diameter of the droplets is comparable to the wavelength divided by the index of refraction. Therefore, anything that is above the 190-K brightness threshold of this image over water should indicate rain. So we see that from Cape Hatteras, down across the Atlantic, across Florida, and into the Gulf, we have an apparent rain storm. We do have ground truth we can compare with this in figure 2.

Figure 2 is a radar summary for the same time period. Note it does indeed show a rain storm from Hatteras, across the Atlantic, and into the Gulf. The ESMR image shows the rain going somewhat further into the Gulf.

Figure 1. Nimbus-5 ESMR image
for January 22, 1973.

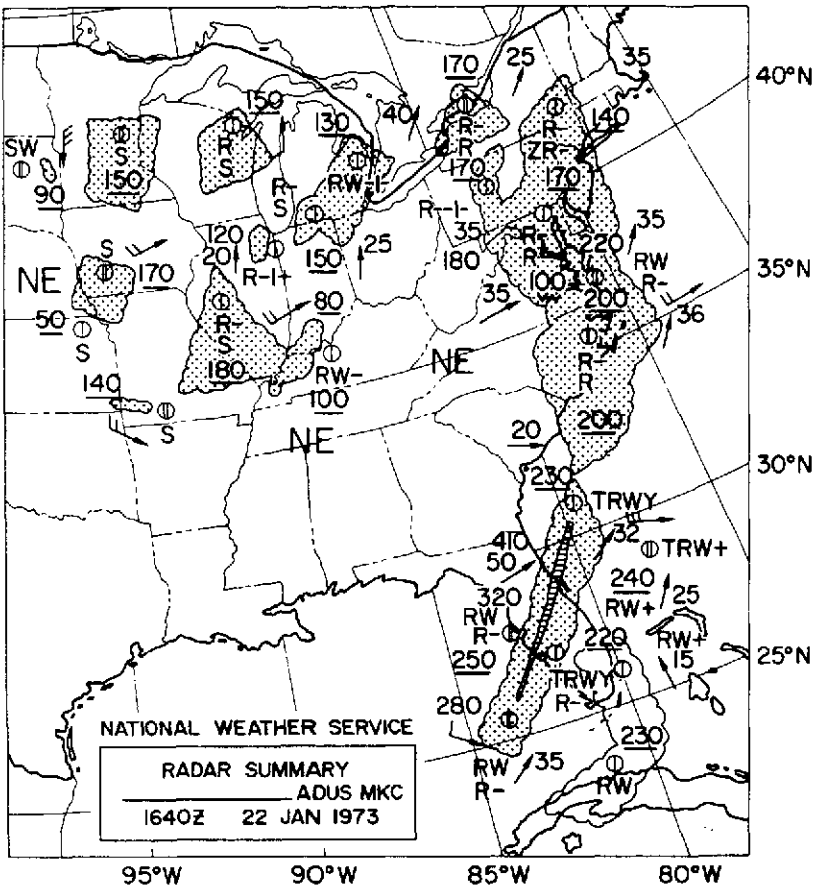


Figure 2. National Weather Service radar summary for January 22, 1973.

This is simply a range problem. You get too far from the stations. The core down the middle indicates the very heavy rain. This sort of data is available in a quantitative form that is well calibrated. An approximate rain rate related to the radar return and a relationship between the ESMR brightness temperature and the apparent radar-indicated rain rate have been derived, as is shown in figure 3.

The empirically derived relationship that we have come up with and the theoretical relationship agree rather well for low rain rates around 4 mm per hour and less, as seen in figure 3. Above that they diverge. This, we think, is primarily because of the beam filling problem; that is, an intense rain storm or rain cell is rarely as large as the 25-km instantaneous field of view of the instrument, so we underrate the rain in our measurement.

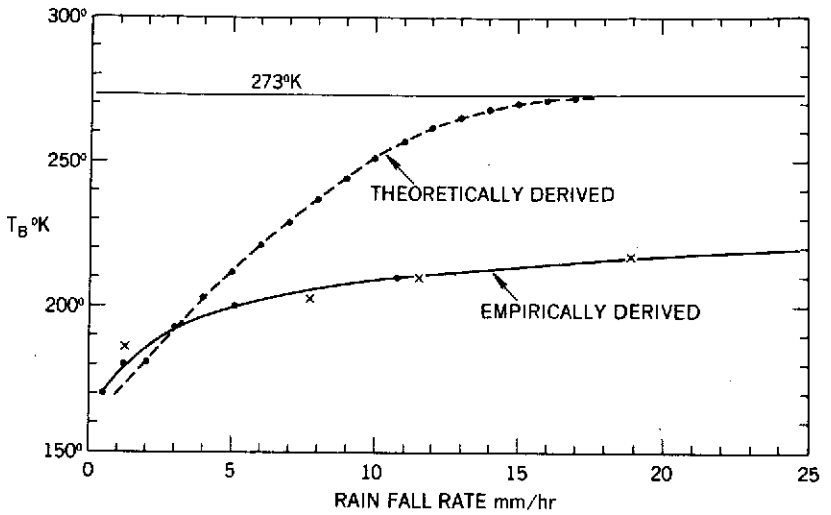


Figure 3. Relationship between ESMR brightness temperature and radar-indicated rain rate.

So, in conclusion, we have demonstrated that we can indeed detect rain over the ocean with ESMR and, in a rough way, measure the rain rate.

GLOBAL DETECTION OF RAINFALL OVER THE OCEAN WITH ESMR

Mirle Rao and John Theon

Because large quantities of sensible heat are released when water vapor condenses into droplets, and because this change of phase occurs most intensely in areas where rainfall is taking place, it is important to account for rainfall in any attempt to understand the energetics and dynamics of the atmosphere.

Rainfall is a valuable indicator of large-scale vertical motion which itself produces significant effects on the circulation. And the horizontal displacement of water vapor by this circulation is an effective mechanism for transporting large amounts of potential energy over great distances. In addition, the index of refraction of the atmosphere is highly dependent upon its moisture content. This affects radio wave propagation, making observations of atmospheric moisture useful in solving communications problems. The Nimbus-5 electrically scanning microwave radiometer (ESMR) makes it possible to estimate daily the extent and intensity of rainfall on a global basis.

In figure 1 we have a montage of ESMR images obtained during 13 consecutive orbits on January 11, 1973. This image may be familiar to you. It covers virtually the entire surface of the earth. There are significant amounts of rainfall associated with the Intertropical Convergence Zone (ITCZ), as you see across the equatorial belt, and there are also significant amounts of rainfall associated with extratropical cyclones and their frontal systems which extend equatorward from high latitudes toward the tropics.

The familiar infrared and visible satellite images of the globe would indicate vast regions of cloudiness, but would give no definitive information on the exact extent of the rainfall or its intensity.

Since the network of observing stations which report rainfall are confined to land, there are no reliable conventional rainfall data available for three-quarters of the surface of the earth. Ships cannot make reliable rainfall measurements because of the platform instability and sea spray problems, and island reports are not representative of the surrounding ocean because orographic effects modify the flow and rainfall patterns, especially in the

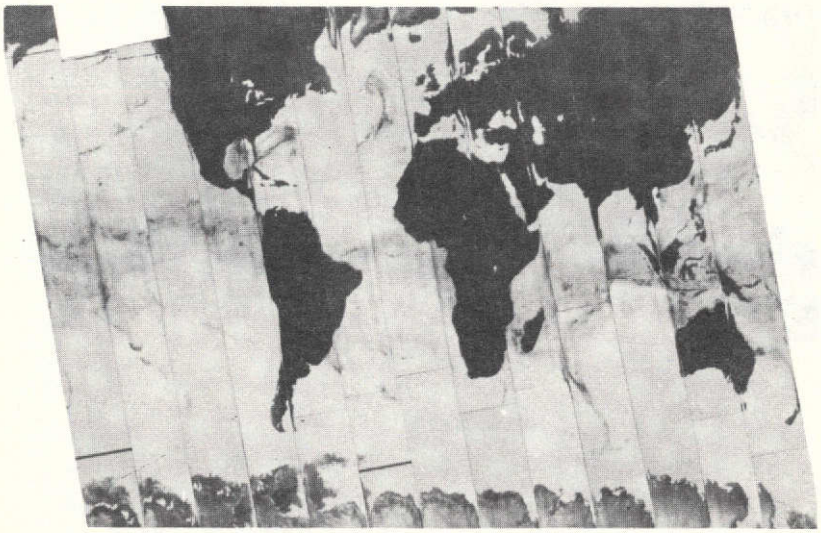


Figure 1. Montage of ESMR images obtained during 13 consecutive orbits on January 11, 1973.

tropics. In order to provide a useful input for the numerical models, we have averaged ESMR rainfall observations over a typical computer grid of the earth, in this case 4° of latitude by 5° of longitude (figure 2). This is a Global Atmospheric Research Program (GARP) grid used by the Goddard Institute.

Because of the uncertainties in the calibration curves explained in the previous paper, we have classified observed rainfall into three general categories:

- Light rainfall-rates of less than 0.1 mm/hour
- Moderate rainfall-between 0.1 and 1mm/hour
- Heavy rainfall-where the rainfall rate was greater than 1 mm/hour

Figure 2 shows the result when we numerically take the data obtained by ESMR and average them over the computer grid of figure 2. The light rain is indicated as shaded, the moderate rain is crosshatched, and the heavy rain is solid. This map gives the results for July 1, 1973.

In figure 3, if we superimpose on the map of figure 2 a summary of the conventional surface analysis with the frontal patterns and the low pressure centers, you will notice that in some cases there is very good correspondence

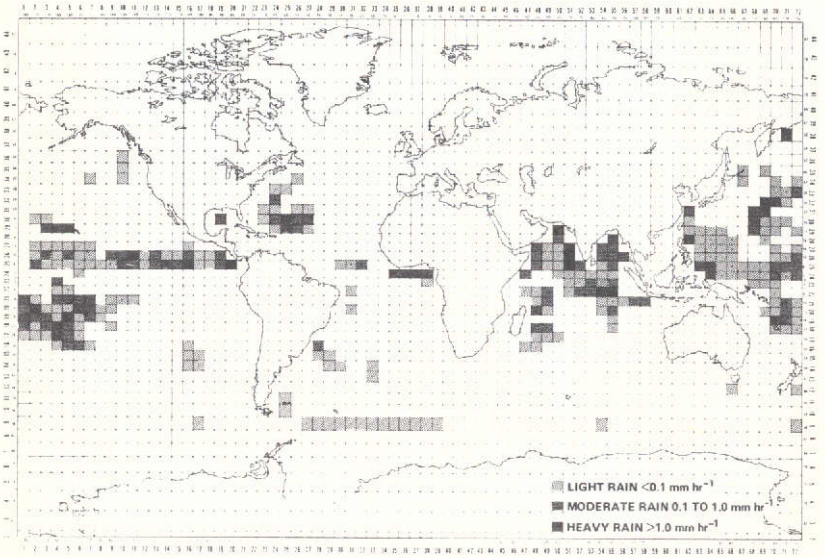


Figure 2. ESMR global rainfall map for July 1, 1973.

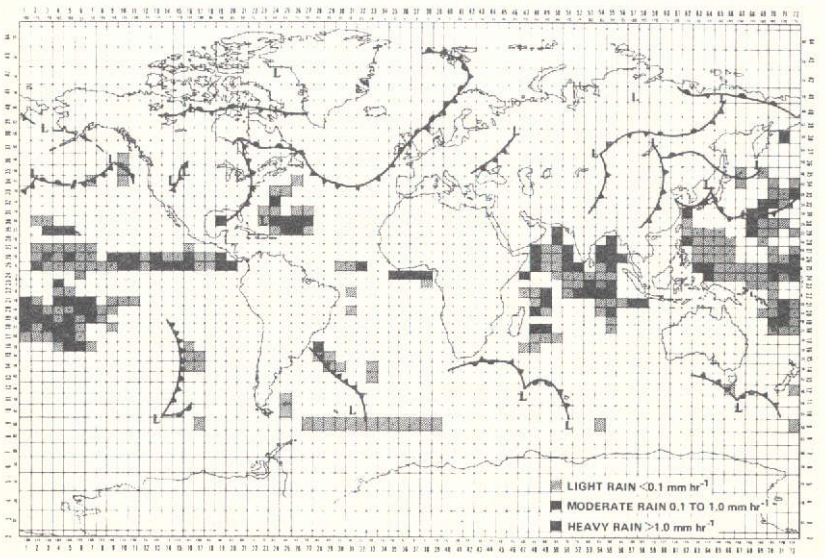


Figure 3. Summary of the surface chart for 0000 GMT July 1, 1973; ESMR global rainfall map for July 1, 1973.

between frontal systems and rainfall over the oceans, and in other cases, there is not. Therefore, we simply cannot take the conventional analysis and assume that it will give us an idea of what the rainfall is. Note that the greatest amount of rainfall occurs in the Intertropical Convergence Zone; the Indian Monsoon extends further into the Indian Ocean than was heretofore accepted; and very little rainfall occurs in the Southern Hemisphere at middle and high latitudes.

As an indication of the quantities of latent heat release involved, we have calculated the energy released by rainfall for a typical 4°-by-5° grid, in this case a crosshatched area with a rainfall rate of 0.3 mm/hour. The energy amounts to 7.2×10^{14} calories/minute, which is very nearly half the solar input for the same area. In more familiar terms, this is equivalent to burning 31,000 barrels of oil per minute in that area.

The global analysis of rainfall data made possible by ESMR is a promising new tool in a continuing effort to better understand and predict the state of the atmosphere.

EFFECT OF ADDING VTPR DATA TO GROUND-BASED OBSERVATIONS

Milton Halem

We have previously shown statistically that the average accuracy of winds is improved by 1.4 m/s by adding vertical temperature profile radiometer (VTPR) data to the radiosonde network. The next step was to see if specific forecasts are improved in the way that has significance for the man in the street.

Today I am reporting on the results of the first test of this kind that we have made. This test of the impact of the VTPR data on forecast accuracy consists of making a 3-day forecast from 2 initial states, one obtained by adding VTPR temperature profiles to the conventional observing system and the other from the conventional system alone.

The forecast was started from 00 Z on August 20, 1973. Figure 1 shows the sea level pressure fields after 72 hours. I call your attention mainly to the features in the Pacific and the Western United States because we would expect to find the strongest impact of the VTPR data in these regions since the land areas and the North Atlantic are well covered by the conventional system in the Northern Hemisphere.

We see that the forecast from the VTPR and combined conventional data predicts a high which is in better agreement with the observations than the forecast from the conventional data. It is better because the high is about 4 millibars higher, better located, and the pressure gradients along the West Coast are better represented than conventional data. Over the western portions of the United States, the satellite-derived predictions predicted a low, in good agreement with the observations. On the other hand, conventional data predicted a high where the low should have been. This difference is the difference between rain or no rain over the states of Wisconsin, Nebraska, parts of South Dakota, and southern Minnesota. The actual observations for this date were showers and thunderstorms with 5 cm (2 in) of precipitation over Wisconsin and lesser amounts in the surrounding areas. This is important weather for the man on the street or on the farm which a satellite-assisted forecast would have predicted, but which the conventional data could not predict.

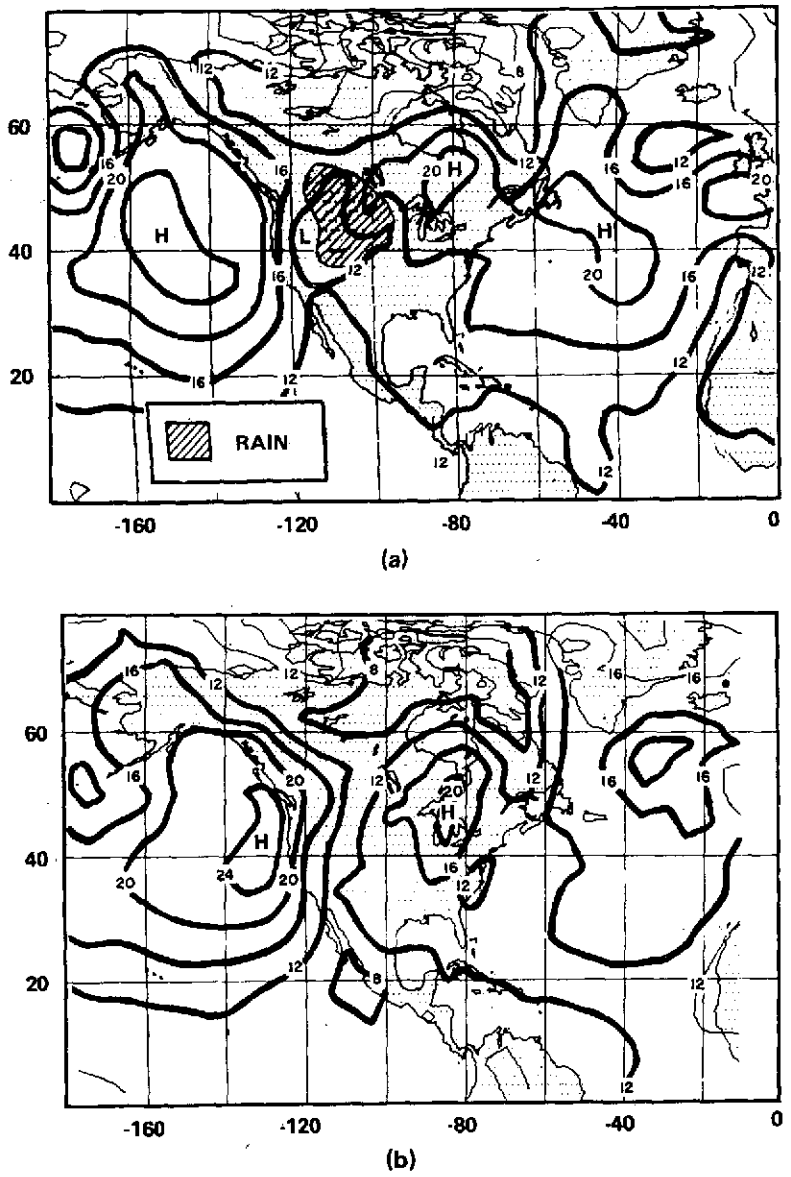
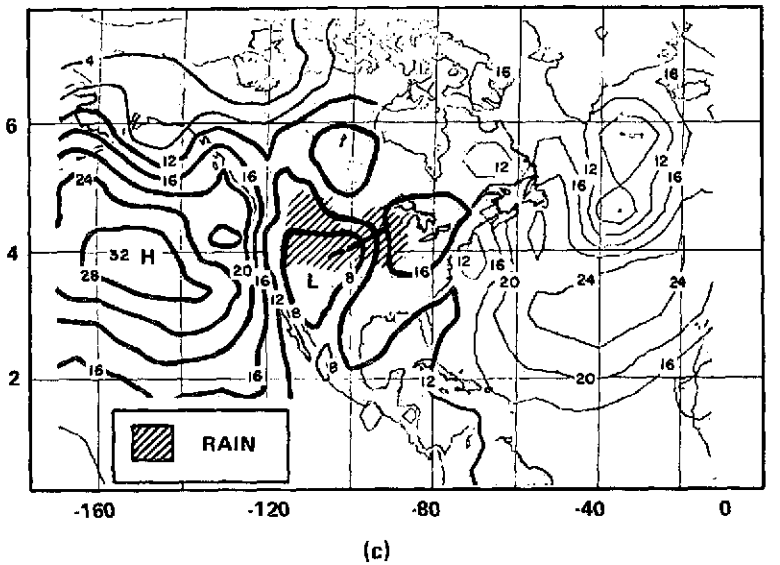


Figure 1. Comparison of 3-day forecast from (a) VTPR plus conventional data and (b) forecast from conventional data alone.



(c)

Figure 1. (c) Comparison of 3-day forecast with actual observations.

The main conclusions to be drawn from these experiments are first, that there is a significant impact from the VTPR data on the forecast, and second, the impact is favorable. These results differ from the NOAA impact tests in which Dr. Bonner's group found no significant impact of the VTPR data on operational forecasts. We feel that the explanation for the differences between the Goddard Institute for Space Sciences (GISS) results and those obtained by the National Meteorological Center (NMC) are due mainly to greater yield of the GISS VTPR temperature profiles and somewhat better accuracy. In particular, GISS processing with VTPR data yields about three times more temperature profiles over the ocean than the National Environmental Satellite Service (NESS), that is, about 2500 profiles versus 800 per day for NESS. If we had inserted only 800 profiles per day, we also would have found no impact.

I should emphasize that this is only one example. We need numerous examples to make this point. But it is encouraging that we hit a major impact on the very first case we tried. It took us a couple of months to get to this point and a long time was spent learning how to do these data tests. We are now analyzing the impact on other cases for September 1973 and December 1972.

VHRR, ONE YEAR IN ORBIT

James O'Brien

Two very high resolution radiometers were launched aboard NOAA-2 ITOS-D on October 15, 1972. To date, over 5,000 orbits later, both radiometers are operating and producing useful operational data. This paper describes the performance of these units in orbit with special emphasis on the IR channel, including systems stability and the passive radiant coolant performance. Data on later VHRRs throughout one year's history, as well as correlation of data with recent thermal vacuum tests, will be presented.

Immediately after the launch of ITOS-D, the heaters on the radiant coolers were turned on, thus inhibiting the IR channel operation of both instruments. After 10 days, the heaters on VHRR No. 1 were turned off and the IR channel sensitivity was noted to be down 15 percent from the prelaunch value. Subsequent measurements have showed continued loss of sensitivity.

When the deice heaters were turned off for radiometer No. 2, it showed a sensitivity loss of 8 percent and it, too, continued to degrade. You will notice in figure 1 that the sensitivity degradation of both instruments has leveled off and appears to be asymptotic with 35- and 28-percent loss, respectively. It should be emphasized that, even with this degradation, the system is presently exceeding specification with regard to signal-to-noise ratio.

During the testing of ITOS-E and ITOS-F in thermal vacuum, the radiometer IR sensitivity was monitored on a continuing basis while a quartz crystal microbalance monitored the accretion of contamination within the chamber. The crystal was cooled to 233.15 K (-40° C), the approximate temperature of the IR filter in the cooler.

In figure 2, the solid curve shows the degradation of the IR channel, while the dash line indicates the contaminant buildup of the microbalance. You will notice that, in this test, we broke vacuum and the IR sensitivity was restored. But, in the retest, it continued to degrade. This is after we broke vacuum. We then heated the cooler to spacecraft temperature, recooled, and the sensitivity recovered again. It must be noted that in the last heating

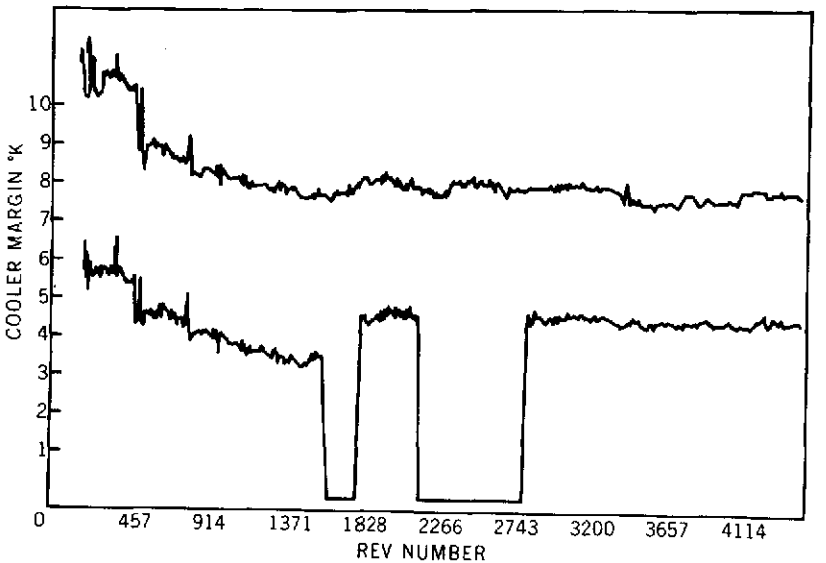


Figure 1. Channels 122 and 131 VHRR-1 and -2 proportional heater drive.

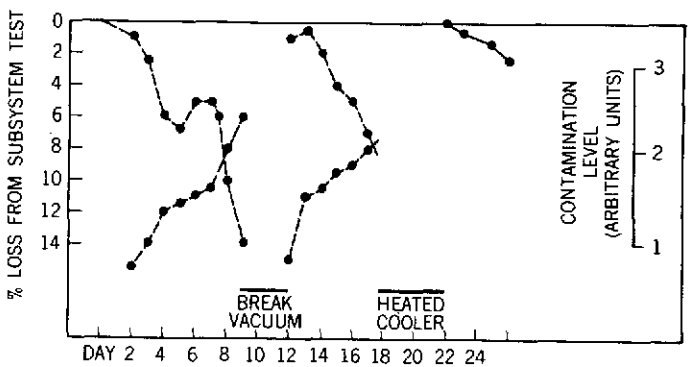


Figure 2. ITOS-F VHRR-2 infrared sensitivity versus contamination.

cycle, external heaters were used, thus applying more heat than the decontamination coolers are capable of. In this testing and the in-orbit performance, we have concluded that the cause of the degradation is a contamination buildup on the IR filter mounted on the outer frame of the cooler. We have designed a fix which amounts to a heater that will control the filter-to-spacecraft temperatures. This fix is planned for later radiometers.

A word about cooler performance. To detect a cold patch which is radiantly cooled and then heated by a closed loop temperature, a servo is required to maintain a constant temperature for the detector. The amount of heater power required to control the cold patch at 107 K is a measure of the cooler margin.

Figure 3 shows the performance of the cooler over one year's operation. As you will notice, there is no change in cooler performance over about the last 3,000 orbits.

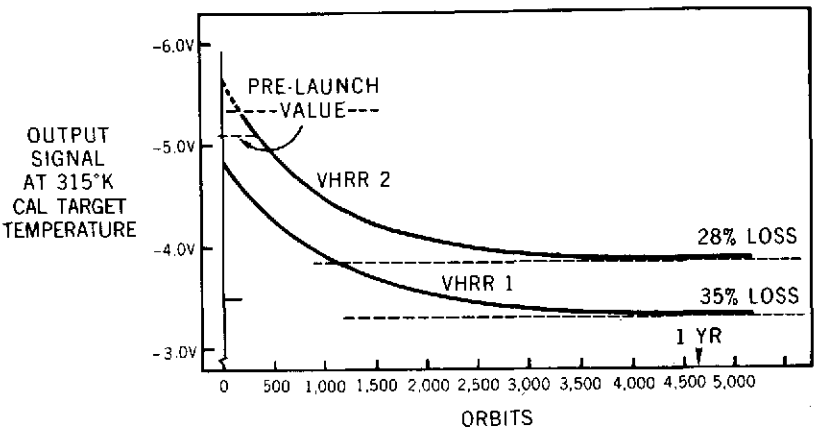


Figure 3. NOAA-2 gain change for VHRR.

WAVE DIRECTIONAL SPECTROMETER

Jerome Eckerman

The wave directional spectrometer will be one of the first active or radar systems to be flown on a satellite for the measurement of ocean waves. Its application is just as the title expresses, to measure the ocean wavelengths and directions. The need for ocean wave spectra is in the application of ship routing and global wave forecasting.

Global wave forecasting can provide very improved ship routing information to the maritime industry. With ship routing aided by wave forecasting, the ship captains can avoid regions where large waves and great hazards are anticipated. Two large economies can be realized: First, they save crossing time and increase vessel usage. Secondly, there is a substantial reduction in insurance rates anticipated with the reduced damage to large ships.

Global wave forecasting requires three elements. One is a knowledge of the ocean wave spectral distribution at some time, and another is the wave field that is changing that wave distribution. The third is a computer model that will take the initial wave field, let the wind field operate on it, modify the waves, let the waves propagate and generate. Ultimately this computer printout is delivered to the ship routing organizations, and they operate on it to provide the optimum routing to ship captains.

In the Seasat phase A study, we identified two sensors which could accomplish the measurement of the wave spectra. One is a synthetic aperture imaging radar, which is a very complex, quite sophisticated machine. This radar requires about a 10-Mb data rate datalink and complex processing on the ground. The other is a device that I am going to mention briefly this morning. This wave spectrometer is a 4-kb data rate device and provides the direct readout of the directions and lengths of large gravity waves.

The basic idea can be seen from a simple comb. This is an acoustic analog. Consider that the teeth of the comb are analogous to the ocean waves, and my finger running across the comb is like the radar pulse moving out across the ocean waves. When you move your finger across the comb it generates a tone, and this, in the radar sense, is an A scope presentation.

The frequency of the modulated tone which you hear is proportional to the speed of my finger (and to the speed of light in the radar case) and inversely proportional to the wave spacing. This is the general concept.

In figure 1 I show a schematic view of a satellite. You see the short pulse, nominally about 6 ns, traveling down to the ocean surface. When the radar pulse strikes the side of the ocean wave on the slope facing the antenna, the backscatter is going to be quite large compared with what it will be when it travels beyond the crest and reaches the trough region. This causes modulation of the backscattered signal by the ocean wave. So, this device operates by ocean wave slope modulation of a short radar pulse.

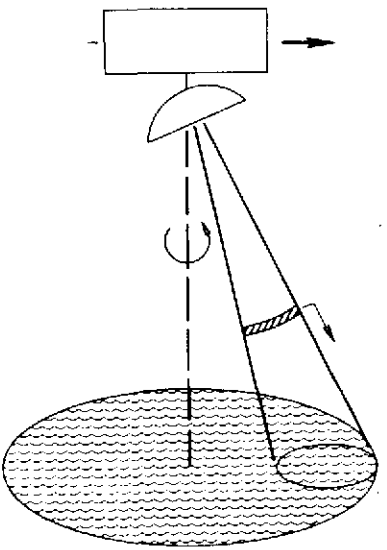
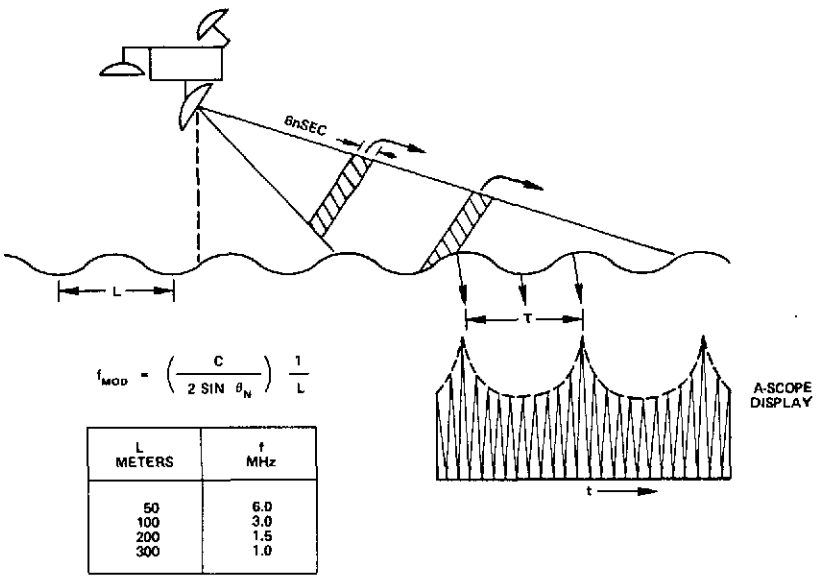


Figure 1. Satellite with short radar pulse traveling to ocean surface.

In figure 2 I show that the resultant modulation frequency varies with the speed of light inversely with the wavelength, wave spacing, and the sine of the nadir angle. For the waves of importance to shipping (greater than 50 m in length), the modulation frequency is on the order of 1 to 6 MHz.

Implementation of this system is simplified because most of the future ocean use spacecraft will be equipped with short pulse radars for topographic measurements. These will primarily be in support of these geoid studies that you have just heard discussed. We can use the transmitter of



L METERS	f MHz
50	6.0
100	3.0
200	1.5
300	1.0

Figure 2. Modulation frequency for the wave directional spectrometer.

this radar directly, with a scanning antenna and a special type processor (figure 3). Down conversion of the received signal is followed by a square law detector and then into a multifilter spectrum analyzer. The analyzer divides the modulation frequency into its incremental elements, each band of frequencies corresponding to another ocean wave grouping. Finally, the wave spectra data is digitized and relayed to the ground. This concept can be implemented on EOS or the next Seasat satellite.

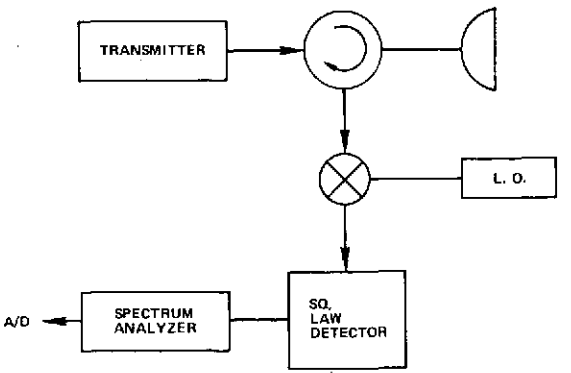


Figure 3. Wave directional spectrometer system.

HIGH RESOLUTION SOLID STATE IMAGER

Leslie Thompson

For future advanced earth observation missions, it is likely that significantly higher resolution will be required than that currently provided by the ERTS-1 imaging systems which normally have a 100-meter ground resolution. One approach to the problem is to use a sensor which operates in push-broom scan mode. The first figure illustrates this concept.

In figure 1, a system is shown which has the capability of offset pointing its field of view as shown by the three successive ground tracks. Such a sensor has been defined for the EOS spacecraft and is known as a high resolution pointing imager. By concentrating on just one of the fan-shaped areas, we can discuss the push-broom scan technique. A line array of detectors is oriented cross-track, and satellite orbital motion sweeps the line array along the track. The line array is electronically scanned in the cross-track direction so that an image is formed on the scene. It should be pointed out that the projection of the line array on the ground is one resolution element wide in the along-track direction.

The push-broom scan idea is a straightforward idea. However, a significant problem occurs if individual discreet detectors must be used to form the line array. If individual detectors were used and formed into an array of several thousand detectors (which is typically the number of detectors required for the applications we are looking at), this would mean that literally thousands of wires, individually, would have to be brought out to the other system electronics. This approach has formidable manufacturing problems and very high cost. What can be done?

The solution has been supplied by using silicon large-scale-integration technology. The detectors, amplifiers, and electronic scanning circuits are all integrated on a silicon chip. In the insert of figure 1, an array has been made of 6 discreet chips; each chip has 96 photodiodes on it. By assembling the 6 chips, an array of 576 detectors has been provided. By way of reference, the diodes are the dark line at the center of the array; the amplifiers and scanning electronics are on either side of the array. This detector array has been assembled for a feasibility demonstration program and is operating successfully. This program is supported by SRT funding.

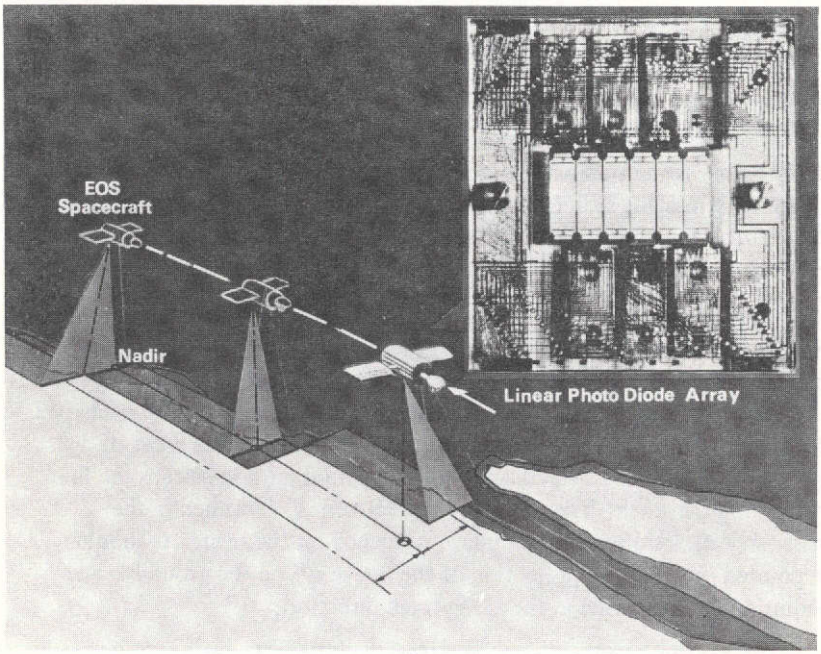


Figure 1. High resolution pointing imager.

Figure 2 illustrates, in schematic form, some details of the detector arrays and system electronics used in the feasibility program. As stated before, the photodiodes are located in the center of the chip, and the incoming photons are imaged onto this array. Associated with each detector is an amplifier which buffers the stored signals on the photodiodes from the common output busses. The shift register provides the electronic scanning which sequentially connects the amplifiers one at a time to one of the four common output busses. The output busses drive off-chip amplifying circuits which process the detector signals and prepare them for digital conversion. Following digital conversion, the signals are multiplexed into a single serial bit stream. In the paper that follows this present paper, a high speed digital multiplexing system will be discussed that would be satisfactory for a high resolution system application.

Making the feasibility program operate up to its expectations has required extreme care in the fabrication of the array and, especially, the analog processing circuits. Presently the system is being used to evaluate detector performance and to measure data in areas such as temperal stability, temperature dependence, and calibration requirements.

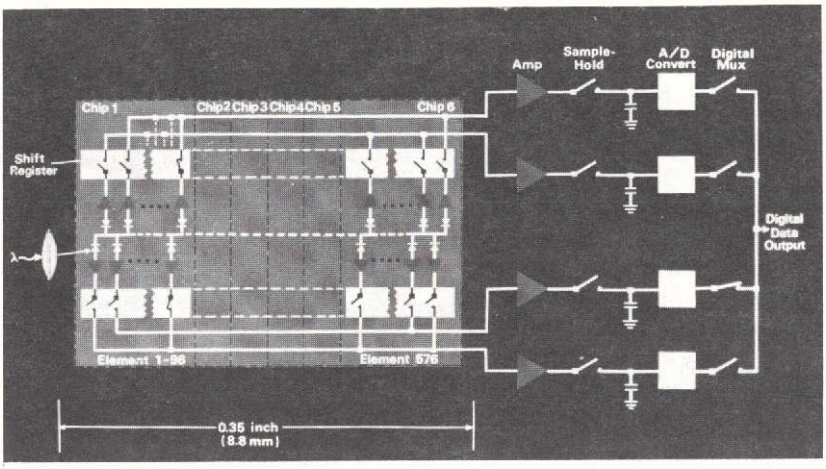


Figure 2. Solid-state linear array imager.

With regard to calibration requirements, at the present time a five-point calibration table is stored for each detector. This is used to correct the responsivity variations of each detector amplifier combination. The five points were arbitrarily selected, and we are currently looking at the possibility of reducing this to a two- or three-point calibration table.

Figure 3 shows an image made with the 576-element photodiode array. The photodiode array is oriented normal to the gray scale and sweeps across the target parallel to the gray scale. The image was made using a scene translator that moves a color transparency past a uniform source of known radiance. The color scene was imaged on the detectors through a narrow band filter centered at 0.55 micrometers.

In closing, I would like to add that recently an array with 1728 photodiodes has been assembled and aligned and is now under test. This technology offers many new opportunities with regard to high resolution multi-spectral imaging systems for earth resource applications.

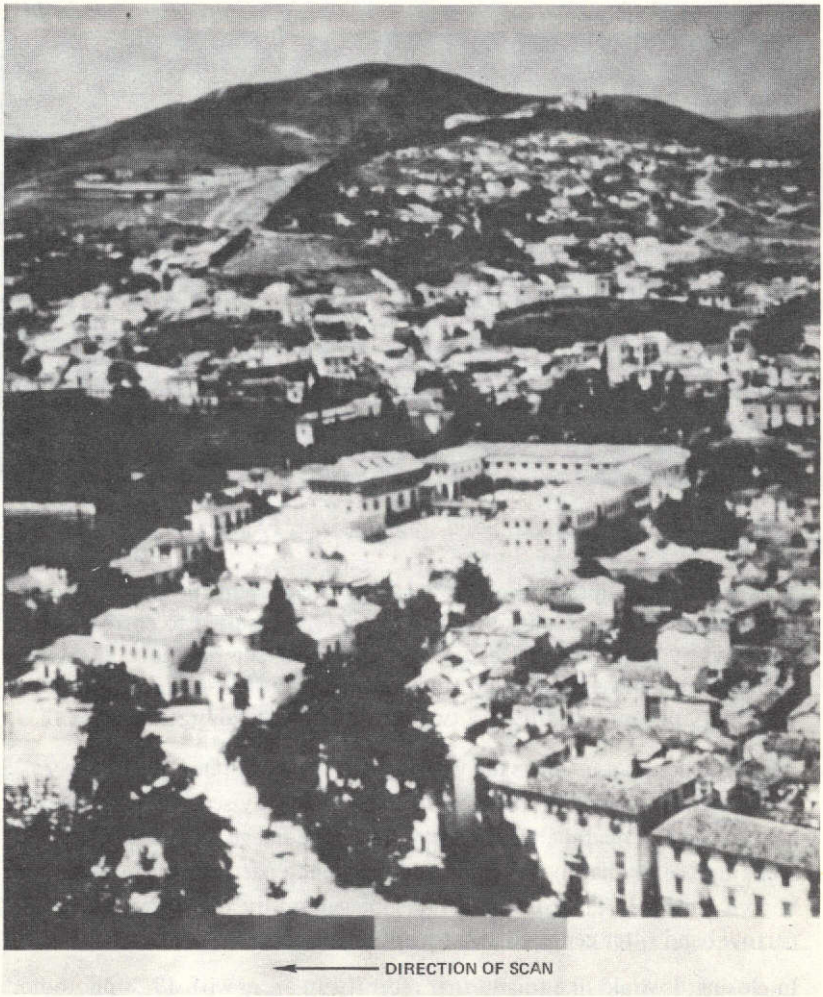


Figure 3. Image made with the 576-element photodiode array.

280-MEGABIT MULTIPLEXER SYSTEM

Paul Feinberg

As a result of the increasing use of high resolution imaging sensors for earth resource missions, a need has been created for efficient and more accurate processing and transmission of wideband data. Table 1 shows the bandwidth requirements for several existing and future imaging sensors. While present sensors require up to 25 Mbs, future sensors, such as the presently planned thematic mapper and high resolution pointing imager for EOS, as previously described, will require a total of 240-Mbs transmission.

Table 1

High Data Rate Sensors

ERTS-1	MSS RBV	4 channels 3 channels	15 mb/sec 3.2 MHz
SMS	VISSR	10 channels	24 mb/sec
EOS	Thematic Mapper	100 channels (7 bits/sample)	120 mb/sec
	High Resol Pointing Imager	80 channels (7 bits/sample)	120 mb/sec

Although limited techniques have been available to accommodate these data rates, these techniques were primarily used for single channel systems operating at high power levels in aircraft applications. Previously, system power consumption, accuracy over a wide temperature range, size, weight,

and reliability have not been compatible with spacecraft applications. Therefore, the primary objective of this Supporting Research and Technology (SRT) program was to develop and build a system that would accommodate the required data rates while satisfying the constraints imposed by the spacecraft environment. Since the baseline specification to which the system was developed was representative of future applications rather than specific known requirements, the system designed was made flexible to accommodate a wide range of sensor configurations.

To meet the future high data rate multiplexing requirements, the multimegabit operation multiplexer system, MOMS, was developed and successfully breadboard tested under a SRT program with Radiation, Inc. The MOMS configuration, as seen in figure 1, consists of two identical blocks, each operating at 20 megasamples per second, whose outputs are then combined for quadraphase modulation. This system was sized to multiplex and digitize 60 narrowband analog channels and 4 wideband analog channels. At the required 7-bit quantizing accuracy, a 280-Mbs combined output rate resulted. The MOMS data format and channel sample rates are controlled by read-only memories which provide flexibility to accommodate a wide range of data requirements so that further definition or changes to EOS data requirements can be accommodated.

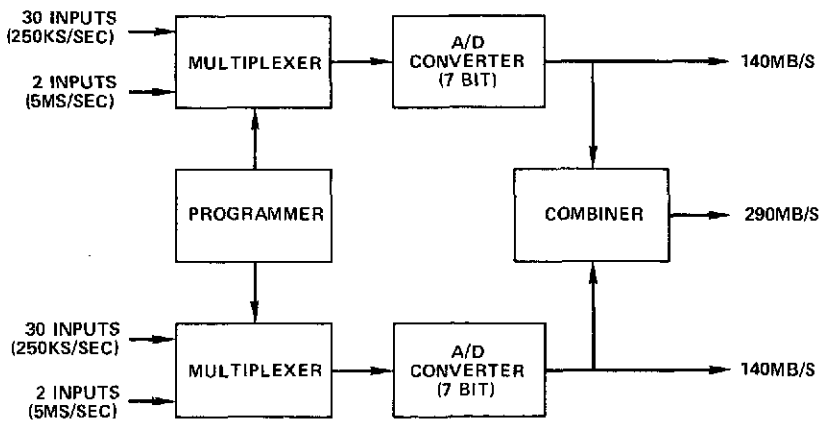


Figure 1. MOMS block diagram.

A detailed electrical and mechanical design of the system, as well as fabrication of an electrical system breadboard, has been completed. Of particular significance is the low system power consumption of 27 W while operating at the full 280 Mbs rate, thereby representing a significant

speed/power performance improvement over systems heretofore obtainable.

The key to the MOMS low power design is a high speed 7-bit analog-to-digital converter and utilization of a new low power logic element technology. Using this logic technology, it is possible to achieve 29-ns, 2-mW logic gate which represents an order of magnitude improvement over the closest commercially available logic element. Other MOMS characteristics are as follows: input channel capacity—60 channels, dc → 100 kHz (250 ks/s), 4 channels, dc → 2.5 MHz (5 ms/s); format—programmable; weight—11 kg (24 lb). The entire MOMS redundant system is estimated to weigh 18 kg (40 lb) and occupy 10,300 cm^3 (624 $in.^3$). Breadboard testing of the basic MOMS system has confirmed design predictions. The MOMS program has led to a detailed designed, breadboarded, and tested system which could be utilized for future EOS missions.

DIRECT READOUT GROUND STATION FOR HIGH RESOLUTION METEOROLOGY IMAGES

Charles Vermillion

The direct readout ground station for receiving, recording, and displaying the meteorological images, both visible and infrared from operational satellites (such as VHRR from ITOS, and VISSR from SMS), was designed and built under the operational satellite's improvement program. Special emphasis and significance are the advance features of this system. A high resolution, high power laser recorder, and the use of dry silver film provide essentially an instantaneous readout. A high precision drum mechanism is necessary to achieve the high resolution, automatic operation, and affordability needed.

The objectives of the development program for the direct readout ground system were to design, fabricate, and test a complete ground system for receiving, recording, and displaying the meterological data from the VHRR and VISSR sensors.

Direct readout is the only way the regional and local world users have to receive high resolution meteorological data. Onboard storage cannot provide data within their time requirements. These improvements can only be realized if the ground terminals to which the data are transmitted are capable of recording and rapid display of the resultant imagery with a resolution, geometric fidelity, and dynamic range inherent in the sensor design. At the beginning of this program, no ground system existed that would satisfy the sensor demands, and the direct readout ground system was developed parallel to the sensor developments.

As seen in figure 1, the direct readout ground system consists of three major elements: the antenna pedestal assembly, the receiving and demodulating electronics, and the recorder processor. The antenna pedestal is a 7.3-meter parabola mounted on a trailer. The system temperature of less than 270 K is achieved by using a parametric amplifier-down converter combination. The receiving and demodulating electronics are located in a van some 450 meters away. There the down-converter signal is received, formatted, and converted to analog signals for use in the recorder processor.

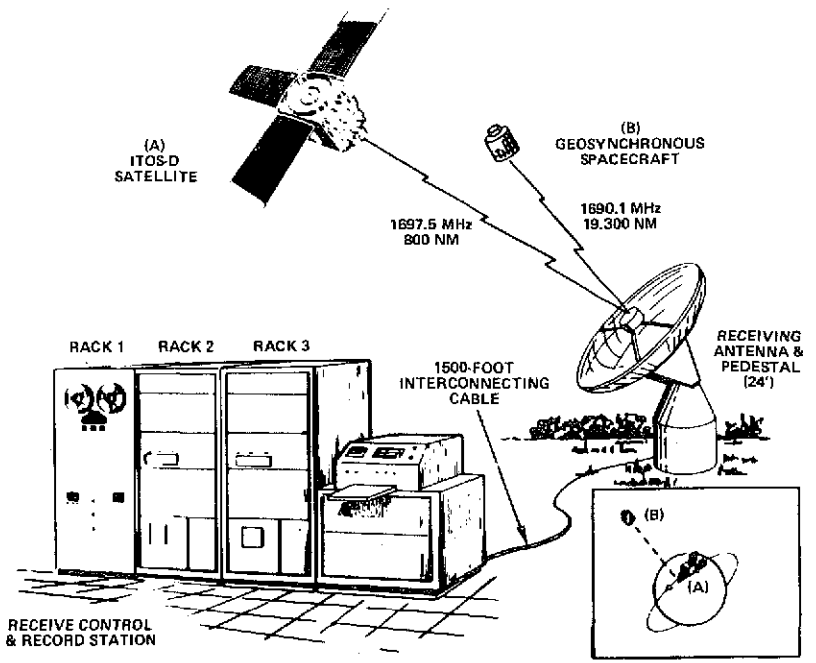


Figure 1. Direct readout ground system dual purpose.

In the recorder processor (figure 2), the analog signal modulates the laser beam in proportion to the scene brightness as viewed by the sensor. The laser beam is focused on rapid access film held by a rotating drum. Scanning is accomplished by using a drum lead screw approach. After exposure, the dry film is automatically fed into a heat processor where the resultant image is available within minutes after the satellite transmission ends.

The following list gives the important parameters of this system. The most significant is the optical spot size as imaged on the film, 38μm or 1.5 milli-inches. This is required for the VISSR full earth disk.

Recording paper/film size	56 cm by 56 cm (22 in. by 22 in.)
VISSR recording modes	0.5-, 1.0-, 4.0-nm resolutions
VHRR recording modes	0.5 nm (1 X), 0.5 nm (2 X blowup)
Optical spot sizes	38, 57, 76, 114, and 304 μm (1.5, 2.25, 3, 4.5, and 12 mils nominal)
Recording times/picture frame	10 and 20 minutes (nominal)

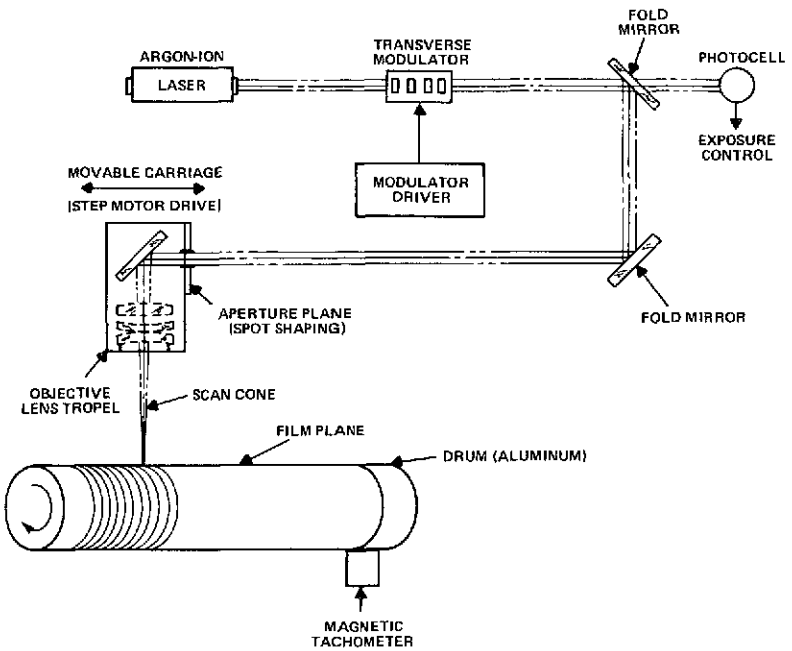


Figure 2. Laser recorder.

Processing time/picture frame	30 to 60 seconds
Maximum interval between recordings	2 minutes
number of $\sqrt{2}$ density levels	16 (film) 12 (paper)
Scan rate	50 to 1100 lines/minute
Resolution	14,000 elements/line (VISSR) 3,300 elements/line (VHRR)
Scan jitter	1/2 element (maximum)
Linearity	0.2 percent (maximum)
Electronic bandwidth	165 kHz (VISSR) 35 kHz (VHRR)

The direct readout ground system was defined by a committee of users, representing NASA, NOAA, and DOD, resulting in a requirement for a portable system. The system is completely automatic, and to change from VHRR to VISSR one only has to press a button. The direct readout ground system successfully operated for the checkout of ITOS-D, NOAA-2,

and the SMS-A bench test. The unit will now be used at Wallops for the post-launch checkout of SMS-A.

In summary, the most significant achievement was the development of the laser recorder processor. No other device exists to date that will meet the user demands as established by the committee that defined the original effort. They are: 56- by 56-cm picture format, no wet chemicals used in film processing, rapid access of the imagery, full resolution, dynamic range and geometric fidelity, a fully automatic system, and two times the selected area blow-up of a single channel of VHRR for precise definition. Last, and most significant, is that the direct readout ground system is the only machine in existence that will provide imagery of the full earth disk from the VISSR at over 14,000 elements resolution.

SEA SURFACE TOPOGRAPHY

Peter Argentiero

The spacecraft-borne altimeter has been recognized as an excellent instrument for determining the fine structure of the marine geoid. However, in attempting to employ standard parameter estimation techniques to reduce the data and estimate the geoid, one encounters a problem; namely, the numbers of parameters that one must use to adequately describe the fine structure of the geoid is very large, and it is not possible to simultaneously estimate all of them.

So, in practice, one is forced to freeze a large number of the parameters at their a priori values and estimate the others. But unless the parameters have excellent orthogonality properties in the data, the net effect is that the uncertainty in the unadjusted parameters corrupts the estimate of the adjusted parameters. This effect is called aliasing and it was a serious problem in our case. It was necessary for us to obtain a parameterization of the marine geoid which had good orthogonality properties in altimetry data.

With that object in mind, we developed a so-called geoid analysis computer program, which in essence was a parameter estimation tool that had incorporated within it several candidate parameterizations. After numerous simulations we concluded that the most promising parameterization was that afforded by mean free air gravity anomalies and Stokes' formula.

In figure 1, I show one of the simulations that we performed with our program. This is the Puerto Rican trench area. We obtained actual observations of 1°-by-1° mean free air gravity anomalies to reconstruct a geoid profile for a 20° latitude arc, shown by the solid line, which passes directly through the most anomalous portion of the trench. The result is the solid line shown on the graph. Next, in order to simulate altimeter observations, we differenced this profile with one obtained by using a standard 16-by-16 Goddard Earth Model 4 (GEM-4) field. We obtained, I believe, 340 such pseudo-observations, processed them in our program, estimated 2°-by-2° mean free air gravity anomalies, and then, by means of Stokes' formula, we projected this back into an adjusted geoid profile which is shown by the dotted line. You can see that the two lines cohere quite

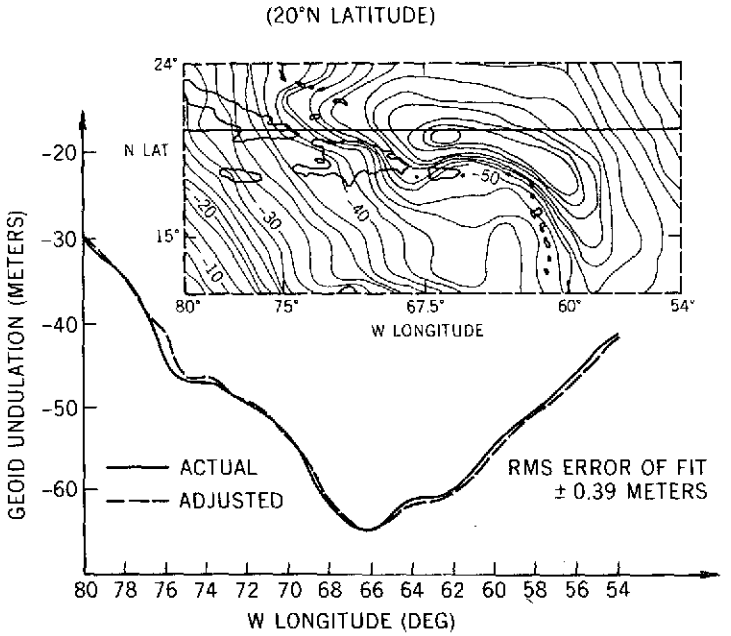


Figure 1. Geoid height profile Puerto Rican trench.

closely. In fact, the rms distance is less than 0.4 meters. We feel that this shows, at the very least, the correctness of the computational logarithm incorporated within the program. We also feel that it shows that this technique is a feasible one for reconstructing the fine structure of the marine geoid.

As yet we do not have large amounts of altimetry data, so our efforts to date have been spent at trying to obtain optimal estimation strategies with regard to these parameters and in determining how well we should expect to do when these strategies are, in fact, implemented. With that object in mind, we added covariance analysis capabilities to our programs, so we could study in finer detail the actual orthogonality properties of our parameters.

One of the results of our covariance techniques is called our aliasing map. In the situation shown in figure 2, we simulate in a covariance mode a case in which we describe the marine geoid by means of 3°-by-3° mean free air gravity anomalies. We assume we have 9 altimeter observations per grid, each with an uncertainty of 5 meters, and we map the RSS contribution in

12	13	20	26	30	32	31	20	12	8
20	21	30	37	42	44	42	37	30	67
27	28	38	47	55	57	54	47	38	15
31	33	44	56	68	77	68	55	43	19
33	34	46	59	77		77	59	46	20
31	33	44	56	69	77	68	56	44	19
28	28	39	48	56	60	56	47	38	15
20	21	31	38	44	46	44	38	31	11
12	13	20	28	32	33	33	29	12	7
5	7	11	16	20	21	20	17	12	8

Figure 2. Aliasing map (in mgal) for 3°-by-3° mean free air gravity anomalies.

milligals to the uncertainty in the estimate of this parameter due to the adjacent parameters when the adjacent parameters are assumed to be in an unadjusted mode and have an a priori uncertainty of 1 mm/s^2 (100 mgal). I think you can see that the aliasing contributions tend to drop off as the parameters get further from the estimation point, thus showing that these parameters do have a certain measure of the orthogonality property that we are looking for. The only way that I know of to overcome this aliasing problem is to simultaneously estimate adjacent parameters along with this parameter. For instance, we could estimate all the parameters within a 3° radius, a 6° radius, and so forth. In the process, we would be pushing back from the estimation point the parameters in the unadjusted mode, lowering their aliasing effect, and increasing the quality of the estimate. Unfortunately, the computational load also goes up with the estimation radius. In fact, it goes up as the fourth power of the estimation radius. This is quite a steep trade-off, and, consequently, in order for us to obtain intelligent estimation strategies, it was necessary for us to know the precise nature of the trade-off.

With that in mind, we used our covariance capabilities to determine the standard deviation of this parameter, when, in fact, it is simultaneously estimated along with all the parameters within a 3° radius, a 6° radius, and so forth. The results are shown in table 1.

In the first column we have the resultant standard deviation of the estimated parameter in milligals. In the second column we have the related estimation

Table 1
Quality of Gravity Anomaly Estimation
versus Estimation Radius

Standard Deviation of Gravity Anomaly	Estimation Radius	Number of Estimated Parameters
392 mgal	0°	1
34 mgal	3°	9
28 mgal	6°	25
27 mgal	9°	49
17 mgal	12°	81
11 mgal	15°	121

radius, and in the third column we have the number of estimated parameters that one must simultaneously obtain when a given estimation radius is used. The computational load is proportional to the square of the number of these parameters.

We are interested in about a 10-mgal resolution of the mean free air gravity anomalies, because this propagates into about a 2-meter uncertainty in the marine geoid, which we feel is a reasonable goal for our project. We know now from this table that in order to accomplish that goal we need about a 15° estimation radius, which would involve the simultaneous estimation of at least 121 parameters.

There is much more work to be done and we need, for instance, to investigate the orthogonality properties of these parameters when smaller grid sizes are used and the aliasing effect becomes more severe. But we hope that by the time we are in possession of large amounts of altimetry data, presumably during the geoid sea mission, we will be in a position to extract the maximum information from the data.

GODDARD EARTH MODELS 5 AND 6

Francis Lerch

This discussion is on the Goddard Earth Model, GEM-6. This model has recently been derived from a combination of satellite tracking data and surface gravity data. The model provides a gravity field of the earth and is given by spherical harmonics complete to degree and order 16. In addition to the geopotential field, the model provides a global network of 135 tracking site locations.

The data employed in the combination solution consists of dynamic satellite data, geometric satellite data, and surface gravity data. In the dynamic data, there were some 27 satellites and 90 tracking stations employed. Approximately 400,000 observations were used from optical, electronic, and laser tracking systems. This data was processed on 350 weekly orbital arcs, among the 27 satellites, and provided a separate model of the geopotential field. This model was designated as GEM-5.

The geometric satellite data was taken on the Pageos satellite with BC-4 cameras and provided for the BC-4 world triangulation network of 45 stations. Altogether, there were some 8,000 points on Pageos that were reduced from photographs taken simultaneously by two, three, or four stations.

For the surface gravity data, there was a global collection of 1654 blocks of 5°-by-5° mean gravity anomalies. The main feature in this new solution, not present in our previous models, was the addition of the BC-4 geometric data. The BC-4 network greatly enhanced the distribution of our station locations.

Figure 1 shows the distribution of the station locations. They were divided into two groups: the dynamic stations indicated by the circles, and the geometric stations associated with the BC-4 world triangulation network, which are indicated by the triangles. The dynamic stations are those that were processed in the orbital mode and are oriented in the center mass reference system. The lines connecting the BC-4 stations correspond to simultaneous observations and show how the entire network is interconnected throughout the world.

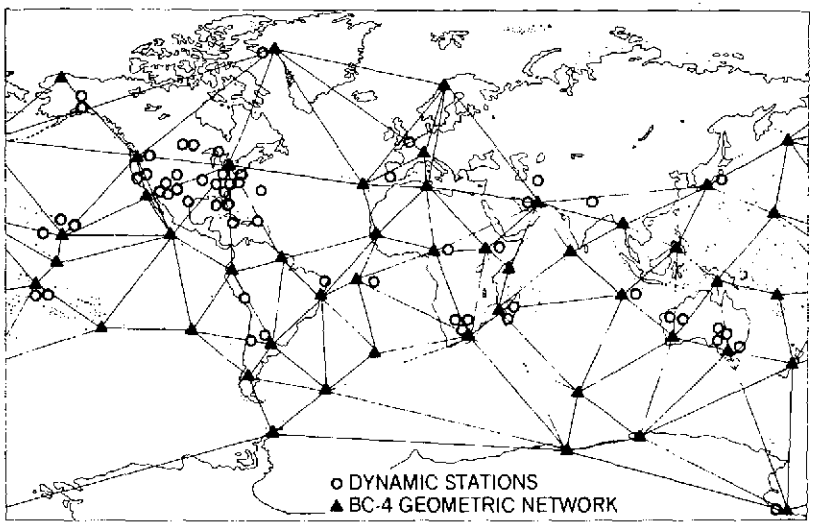


Figure 1. Tracking station locations.

There were some 20 side-by-side BC-4 geometric and dynamic stations throughout the world. The local datum coordinates at these sites were used to tie the BC-4 geometric stations to the dynamic stations. Thus, in the combination solution, the entire BC-4 network would be tied into our center mass reference system.

Some geodetic results are as follows: We have a global geoid, based upon our spherical harmonic solution; the harmonics are complete to 16 by 16 corresponding to the broad features of the gravity field, and provide a wavelength resolution of 2500 km on the geoid. This resolution has been estimated to correspond on the average to about 85 percent of the total geoid height, which would be obtained from the GEM-6 model.

We have a geodetic world datum based upon our center of mass station coordinates, where all stations on major and local datums throughout the world may be placed in this common reference system. Derived parameters for the earth reference ellipsoid consisted of a mean equatorial radius of the earth of 6,378,145 meters, a value of flattening (1/298.256), and a value of equatorial gravity (978032 mgal). Accuracy estimates for the station coordinates were 6 meters on the average and for geoid heights they were 4 meters, which corresponded to a global root mean square error. Mean gravity anomalies for 12°-by-12° squares provided about 4-cm/s^2≡(4-mgal) accuracy.

This work was done partly for the National Geodetic Satellite Program. Certain accuracy goals, set up in this program, were 10 meters for station coordinates and 4 cm/s^2 (4-mgal) for the broad features of the gravity field. Our results given here for the GEM-6 model satisfy these accuracy requirements. However, much refinement is needed in the solution in order to satisfy more stringent requirements of the EOPAP program.

GLOBAL DETAILED GRAVIMETRIC GEOID

James Marsh

During the past year, we have computed a global detailed gravimetric geoid based upon a combination of satellite and surface gravity data. Today I would like to present the results of that computation.

The data that was used consisted of the GEM-6 earth model, which was discussed previously by Mr. Lerch, and a refined set of surface gravity data. The GEM-6 model is a spherical harmonic expansion of the earth's gravity field to degree and order 16, and is based upon 400,000 observations from 27 satellites and a set of 1654 5°-by-5° surface gravity anomalies. To this, we added a set of 25,000 1°-by-1° mean surface gravity observations.

The GEM-6 model provides information on the long wavelength undulations of the geoid. At satellite altitudes the fine structure tends to be smoothed out and we see only the broad features. The surface gravity data provide information on the shorter wavelengths, that is, wavelengths on the order of 2500 km or less.

Figure 1 presents a map of the geoid height contours. Where the geoid heights are negative, it indicates that the geoidal surface is below the reference ellipsoid. For instance, in the areas off the tip of India, the geoid height is -112 m. This means that the geoid is 112 m below the reference ellipsoid. Where the geoid heights are positive, the geoid is above the reference ellipsoid; for example, the height is 70 m off the coast of Australia.

The precision of this geoid varies according to geographic area, depending upon the density of the surface and the satellite data. For example, in North America, Europe, and Australia, where we had plenty of data, the accuracy is on the order of 2 m. In the ocean areas (for example, the Atlantic and Eastern Pacific areas), the precision is on the order of 2 to 5 m. In the Southern Hemisphere, in the ocean areas especially, the precision is, we feel, on the order of 10 m, reflecting the sparsity of data. We need more data in these areas, and in this regard we are looking forward to the implementation of the new satellite-to-satellite tracking technique. This technique will provide extensive satellite coverage both geographically and in time, which heretofore has not been available from the geodetic satellites.

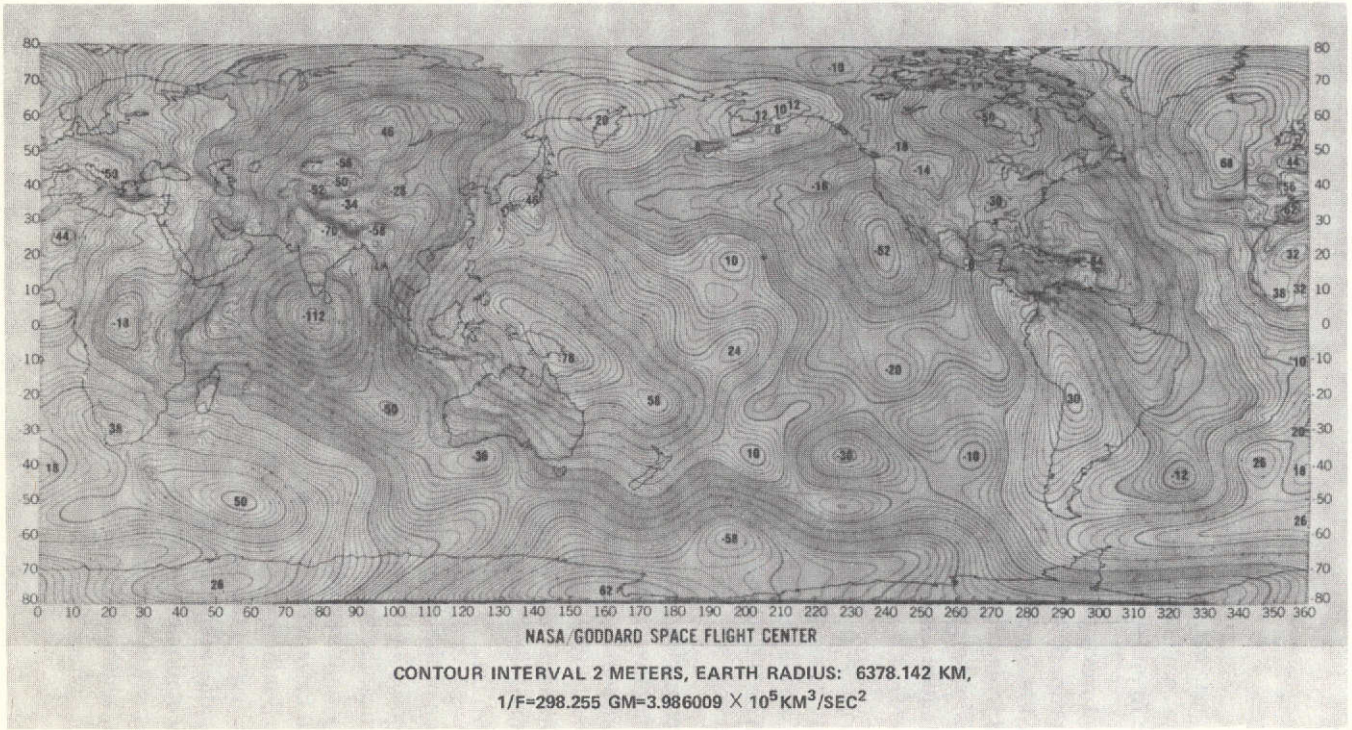


Figure 1. Global detailed gravimetric geoid based upon a combination of the GSFC GEM-6 earth model and 1°-by-1° surface gravity data.

One of the most important applications for this geoid map is in the area of satellite altimetry. With this independent data, we can now test and evaluate the new satellite altimeter data. Furthermore, in some areas of the oceans, the geoid surface and the mean sea level surface deviate by several meters due to tides, currents, and other phenomena. Comparisons of the two independent data types, that is, the gravimetric geoid and the altimeter data, are expected to provide more information on these deviations.

Preliminary comparisons of this nature have already been carried out with the Skylab altimeter data at NASA/Wallops Island. Figure 2 presents the results of such a comparison for a pass in the area of the Atlantic Ocean off the coast of the United States. The geoid as derived from the Skylab altimeter data is represented by the dots. The solid line represents our detailed gravimetric geoid. The rms difference is less than 2 m. I should emphasize that the results here are not from a least-squares fit. There has been no adjustment of one data type to the other. They are totally independent. We are very encouraged by the excellent agreement here and we are looking forward to future analyses of this nature with the upcoming altimeter data from other satellite missions.

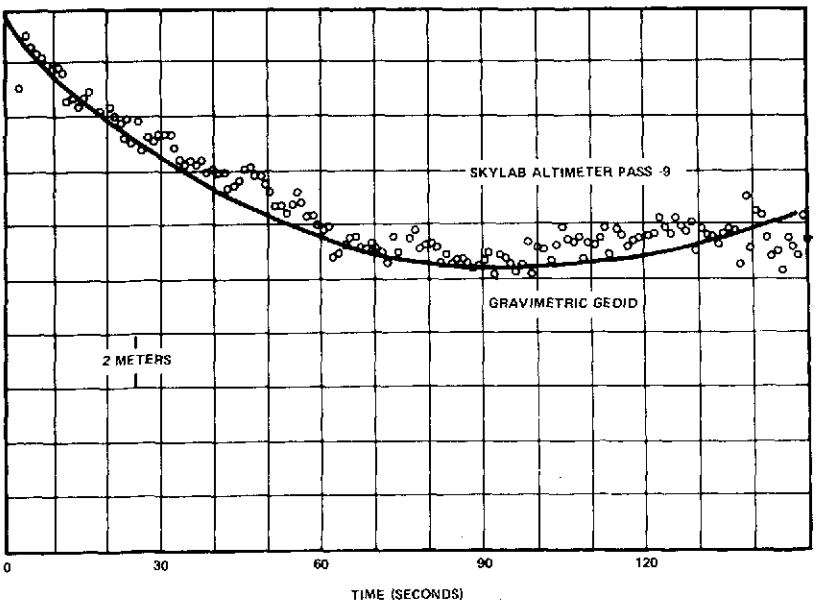


Figure 2. Comparison of Skylab altimeter data with the gravimetric geoid.

OCEAN TIDES AND SATELLITE ORBITS

Theodore Felsentreger

Today I would like to present some initial results that we have obtained toward the derivation of a global ocean tide model from satellite data. The discovery that the effects of the solid earth and ocean tides on satellite orbits can be observed has opened up the prospect of actually solving for global tidal models from satellite data. However, the fact that the same frequencies are present in both the solid earth and the ocean tides makes the separation of the two effects difficult in the solution. Efforts have already been made to solve for solid earth tidal models alone. These efforts have resulted, however, in recovered values for the love number, K_2, which in some instances were considerably different from the generally accepted value of about 0.3. These discrepancies are due to the unmodeled ocean tide effects which are still in the data.

In order to solve for either a solid earth tide model or an ocean tide model from satellite data, a model for one should be assumed and the other solved for. Since the existing ocean tide models are rather inadequate for this purpose, the most expeditious method is to assume a solid earth tide model and solve for an ocean tide model.

Since it is the long period effects that we can observe in the data, only the long period part of the ocean tide potential will be considered. It is a spherical harmonic expansion in terms of orbital elements of the satellite,

$$U = g\frac{\rho}{\rho_\oplus}\sum_{k=0}^{\infty}\sum_{\ell=0}^{k}(1+K_k')\left(\frac{3}{2k+1}\right)\left(\frac{a_\oplus}{a}\right)^{k+1} C_{k\ell}\sum_{p=0}^{k} F_{k\ell p}(I)\, G_{kpq}(e)\cdot \begin{bmatrix}\sin\\ -\cos\end{bmatrix}\left[(k-2p)\,\omega+\Omega-\alpha_\ell+\epsilon_{k\ell}\right]$$

where g is the acceleration of gravity, ρ is the mean density of sea water, and $\rho_\oplus$ is the mean density of the earth. K_k' is the load deformation coefficient which is analagous to the solid earth tide love numbers, $a_\oplus$ is the mean radius of the earth, a is the semimajor axis of the satellite orbit, $C_{k\ell}$ is the spherical harmonic coefficient , F is the function of the inclination of the satellite, and G is the function of the eccentricity of the

satellite's orbit. The trigonometric terms are sine or cosine, depending upon the particular combination of indices. The argument of perigee of the satellite is ω, Ω is the node of the satellite, and α_ℓ is an argument depending on a particular constituent. There is a potential for each tidal constituent; $e_{k\ell}$ is a phase angle. A word about the constituents. In the ocean tides there are many constituents, but there are essentially three principal ones: the M_2 tide, which is the lunar, semidiurnal, or twice daily tide; the S_2 tide, which is the solar, semidiurnal tide; and the K_1 tide, which is a combination lunar-solar diurnal or daily tide.

If you can isolate the long period ocean tide effects, then you can use this potential and solve for some of these coefficients. We took a 340-day arc of Doppler data for GEOS-1. We processed it, first of all, without a solid earth tide model or an ocean tide model, so that the variation left in the inclination reflects the influence of both the solid earth and the ocean tides (figure 1). The variation is expressed in arc-seconds and in meters. You can see, very definitely, a 160-day-period effect which is due to the lunar and solar and diurnal tide. There is a lesser 55-day period visible which is due to the solar semidiurnal tide. The lunar semidiurnal tide is of a lesser amplitude, has about a 12-day period, and is not really discernible in this data.

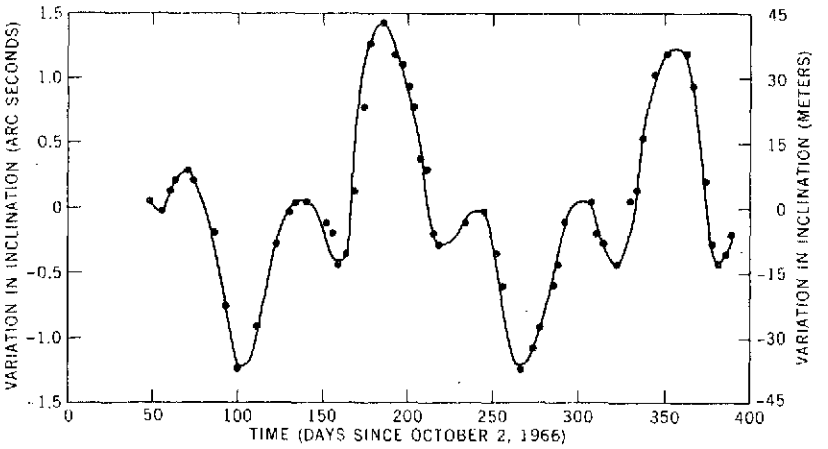


Figure 1. Influence of solid earth and ocean tides on the inclination of GEOS-1.

We then took the data, processed it with a solid earth tide model in the orbit determination program, and we got the results shown in figure 2. The variation left is simply the influence of the ocean tides alone. The

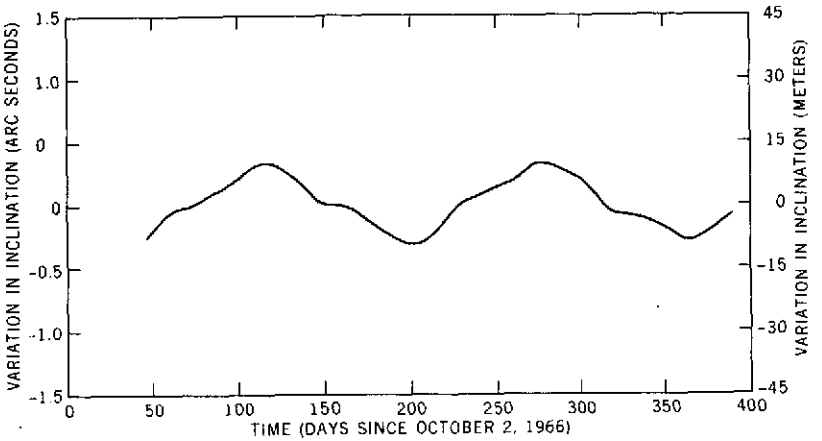


Figure 2. Influence of the ocean tides on the inclination of GEOS-1.

solid earth tides have been taken into account. You can still see the 160-day period, which is the effect now of the K_1 ocean tide, and there is a lesser 55-day period, which is the S_2 ocean tide effect. We took this data and solved for one harmonic coefficient for the K_1 tide and one harmonic coefficient for the S_2 tide. The results are as follows.

We solved for a C_{21} coefficient for the K_1 tide and got 11.3 cm. We obtained a phase angle of 190°. For the S_2 tide, we obtained a C_{22} coefficient of 2 cm and a phase angle of 23°. The coefficients are expressed in centimeters because the actual spherical harmonic expansion is an expansion of the global tidal elevation. It is incorporated into the ocean tide potential and this is why the coefficients are in centimeters. We are continuing the analysis with additional data from other satellites and we hope to come up reasonably soon with a more extensive tidal model.

ON THE BREAKUP OF TECTONIC SHELL OF THE EARTH

Han-Shou Liu

The modern concept of tectonic plate motion explains many geophysical and geological problems in a very simple way. However, it is still considered a hypothesis because one basic question has not been answered: "What was the driving mechanism which had broken the outer shell of the earth into tectonic plates?" The purpose of my presentation is to answer this question.

Paleomagnetic studies have shown that the axis of rotation of the earth has been shifted over a large angle θ, and the pole of rotation has been moved from p to prime (figure 1). Under these circumstances, the equatorial plane of the earth must move from the solid line to the dashed line because the earth must adjust its shape to the change of elipticity about the new axis of rotation. The latitude before the movement of the pole is Φ, and $-\theta$ is the latitude after the movement of the pole. A point e on the surface of the earth will move to point f. We define the distance ef as the displacement U_r, which has positive sign in the AD and BC regions. It has negative sign in the other two regions of AB and CD.

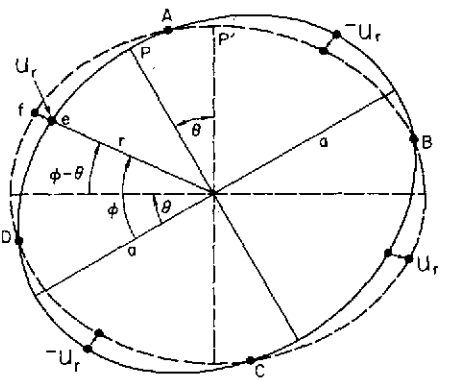


Figure 1. Effect of the shift of the earth's axis of rotation.

According to the theory of the shape of the earth, we can calculate the displacement U_r in terms of the elipticity of the earth. By using the result of this displacement, the stresses and the deformation in the outer shell of the earth can be estimated.

For the element in the outer shell of the earth, the three components of normal stress are $\sigma\alpha$, $\sigma\beta$, and σ_r (figure 2). The shear stress is τ. The three components of displacement are $U\alpha$, $U\beta$, and U_r. By prescribing the displacement U_r due to the shift of the axis of rotation of the earth, the stresses in the outer shell of the earth can be completely determined. In order to simplify our results, we have transformed these stresses into a new coordinate system. Numerical calculations show that the magnitude of these stresses is on the order of kilobars, which is sufficient to produce fracture in the lithosphere.

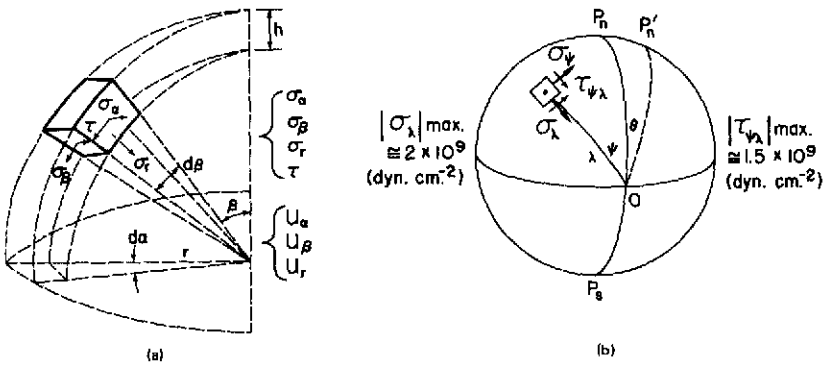


Figure 2. (a) Stresses in spherical shell; (b) The coordinate system ψ and λ.

The physical property of the outer shell of the earth may be considered as a plastic material. According to the theory of plastic deformation of shells, the pattern of plastic deformation is described by the equation of stress intensity.

$$I = 1.3 \left[\mathrm{Cos}^2 \lambda + \frac{7}{12} \mathrm{Sin}\,(2\psi - \theta)\,\mathrm{Sin}^4 \lambda \right]^{1/2} \mathrm{Sin}\,\theta$$

In figure 3 we have plotted this equation in a stereographic projection. The solid curves for I = 1 are fracture lines. The dashed curves for other values of I are stress intensity. These curves can be transformed into another projection. We have transformed them onto a map shown in figure 4.

For a large shift of the rotational poles, the fracture pattern follows directions in the NE and NW directions on the Northern Hemisphere and in the SE and SW directions on the Southern Hemisphere. It is noted that the global imagery of ERTS-1 reveals that geological lineaments are associated with this pattern of fracture. Therefore, it appears that polar wandering may produce fractures in the outer shell of the earth, and that the pattern of the fracture lines seems to agree with the boundary system of these six tectonic plates.

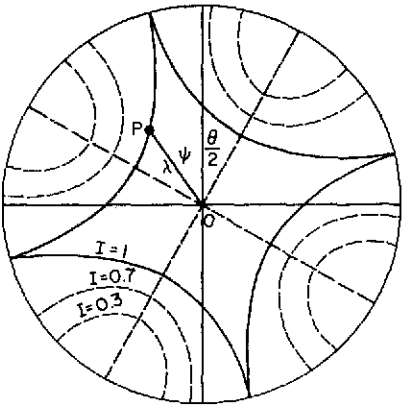


Figure 3. Stress intensity I in stereographic projection of the half-sphere for shift of the poles over θ.

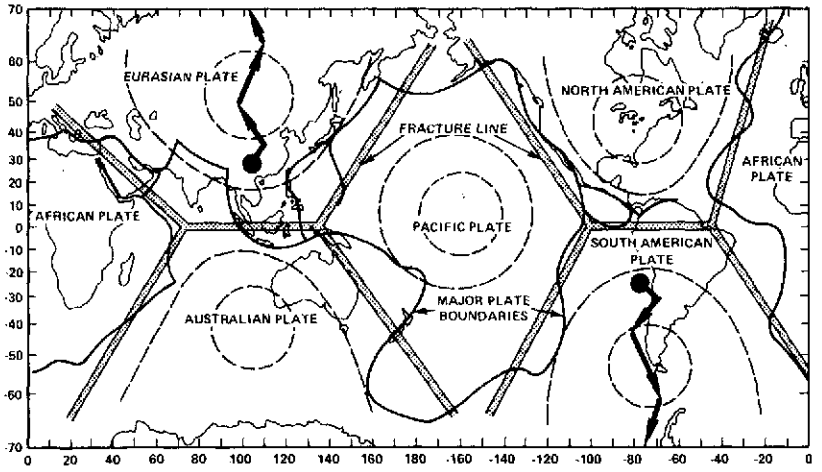


Figure 4. Fracture lines for a shift of the poles over 70° along the meridian of 75°W.

Finally, it is noted that the findings of this paper have been scheduled to appear in the *Journal of Geophysical Research.*

FM TV DISTORTION ANALYSIS

John Chitwood

I would like to describe a task whereby a computer simulation has been used to quantitatively measure the distortion on a standard video test signal that undergoes an FM modulation, filtering, and FM demodulation process. This task was undertaken as part of a study to select various system parameters for FM television systems such as those to be used in various experiments to be carried out through the ATS-F spacecraft.

Of these parameters, the predetection bandwidth is very important in that for any given carrier-to-noise power density ratio, the smaller the predetection bandwidth, the greater the carrier-to-noise power ratio in that bandwidth. However, if the predetection bandwidth is chosen to be too narrow, severe distortion will result on the video waveform.

For some time, Carson's Rule has been used to determine the predetection filter bandwidth, given the deviation and information bandwidth of the video signal. Carson's Rule is valid only in situations where the modulating wave is sinusoidal in nature, which is hardly the case with a video waveform.

The standard T-pulse-and-bar test waveform that was used is shown in figure 1a. Excluding the sync, the signal consists of two transitions from black to white to black. The first is the pulse portion of the signal which consists of one cycle of a 4-MHz sine wave. The second is the longer bar portion of the waveform, the rise and fall of which is identical to the rise and fall of the pulse portion of the waveform.

A block diagram of the system simulated is shown in figure 1b. The T-pulse-and-bar test signal is applied to an optional preemphasis network, which is the International Radio Consultative Committee (CCIR) recommended network for FM TV systems. The preemphasized video waveform then frequency modulates the carrier which is then passed through the RF channel. In this case, however, no noise is added in the channel as we are only interested in the distortion on the video waveform and not the system's performance in the presence of noise. The predetection filter previously mentioned is now inserted in the RF signal path, the output of which drives the FM demodulator. The demodulated output can then be deemphasized, and the received video signal may be analyzed.

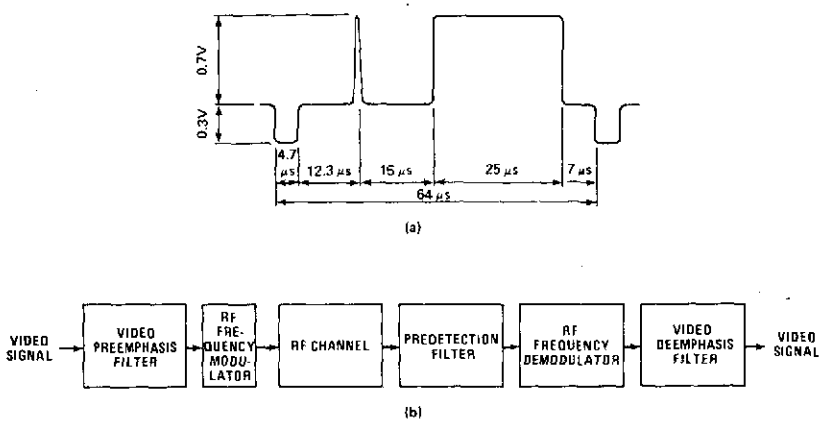


Figure 1. (a) T-pulse-and-bar test signal; (b) Block diagram of the FM video transmission system being modeled.

One of the reasons for using a computer rather than a laboratory simulation of this system was the versatility in changing system parameters. The preemphasis and deemphasis networks could be added or removed. We could change the type of filter, such as Butterworth or Chebyshev, change the filter bandwidth, and also the number of polls of the predetection filter. We could also change the amount that the video signal deviated the RF carrier, or the modulation index.

In figure 2, some of the results of the analysis are presented. The top graph shows the output for the pulse portion of the signal. There is some undershoot after the rising portion of the pulse. The waveform should rise to +0.5 V. There is also a good bit of overshoot and ringing after the falling edge. The same thing is true for the bar portion of the signal. There is some overshoot after the rising portion, and also some overshoot and a slight amount of ringing after the falling portion.

These results can also be presented in another format, as shown in figure 3. Here we are plotting the total distortion in percent on the video signal as a function of the ratio of the RF bandwidth to the peak-to-peak deviation.

Now, Carson's Rule would normally have us operating in the area where this ratio is about 1.5; as you can see, the distortion is very low, on the order of 1 percent. We have found, however, due to this analysis, that we can operate where the RF bandwidth is only 83 percent of the peak-to-peak deviation and achieve distortions in the neighborhood of only 6 and 7 percent. This is very important in that if we operate here we can achieve about a 2-dB increase in signal-to-noise ratio using a small predetection bandwidth.

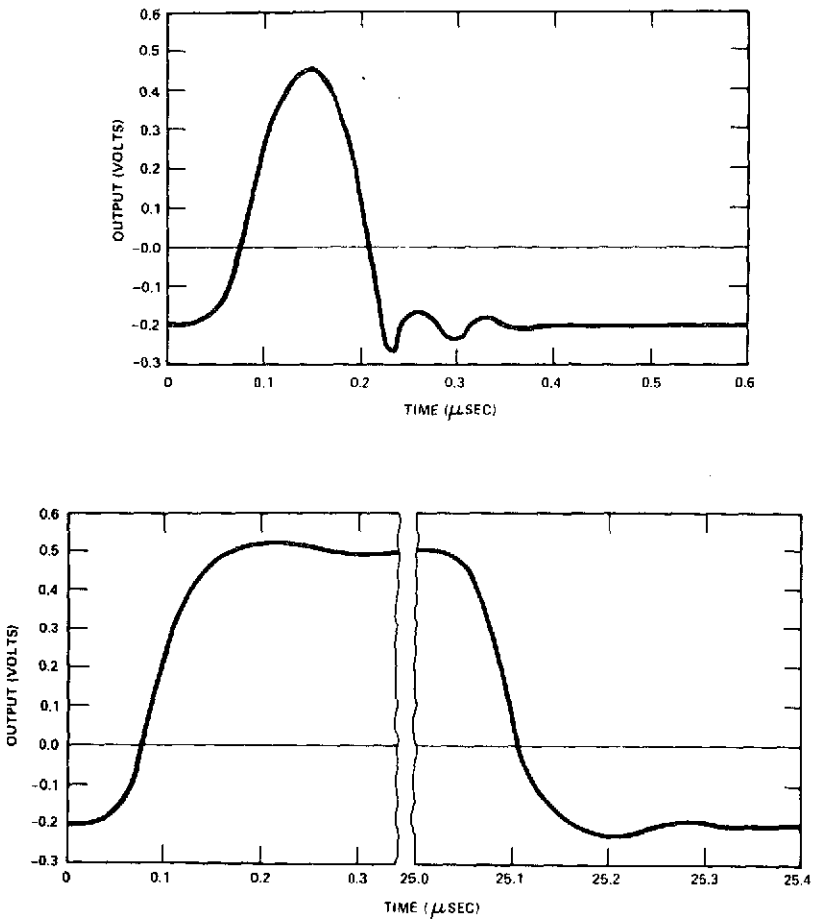


Figure 2. Results of the FM TV distortion analysis.

It is interesting to note that the CCIR recommendation 421-2 states that the maximum distortion for long haul links (2500 km) shall be less than 13 percent. And it can be seen that by operating with a smaller RF bandwidth here, one of only 83 percent of the peak-to-peak deviation, we are well within this 13-percent requirement.

I might also point out that this technique is good not only for analysis of video signals that undergo this FM modulation-demodulation process, but it can be applied to other types of data as well.

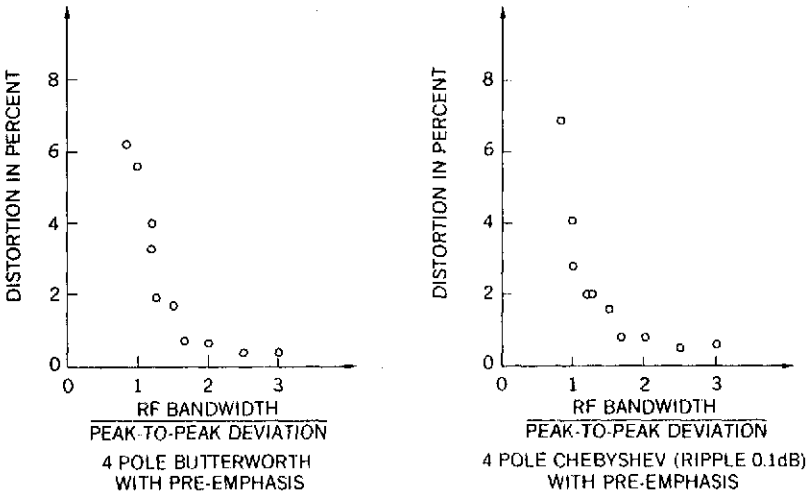


Figure 3. Total distortion on the video signal as a function of the ratio of the RF bandwidth to the peak-to-peak deviation.

MILLIMETER WAVE SPACE COMMUNICATION SYSTEM TECHNOLOGY

John King

Today I would like to present four pieces of millimeter wave spacecraft technology which have been developed for the ATS-F satellite. These components were developed under Goddard contracts with Hughes Aircraft Corporation and Comsat Corporation to provide transponder systems to conduct propagation experiments in the millimeter wave frequency band.

First, I would like to give a short description of the experiments. The first is a 13- and 18-GHz experiment which utilizes 25 small ground terminals located in the eastern half of the United States which transmit continuous wave (CW) carriers to the satellite and measure attenuation by signal level measurements. In the satellite, the signals at 13 and 18 GHz are converted to 4 GHz and then transmitted back to the ground where a large ground terminal picks up these signals and records them for later analyses. The experiment has, as one of its objectives, to take and measure the attenuation statistics over various meteorological conditions, such as those that are located in the Eastern United States. It also has the objective of looking at site diversity improvements and includes three areas where four terminals will be located from 12 to 38 km (8 to 24 miles) apart. These locations will be in the Boston area, in the Columbus, Ohio area, and in the Mississippi area.

The second experiment, a 20- and 30-GHz downlink experiment, utilizes 2-W traveling wave tubes that are designed to measure large fade margins which occur at these frequencies during rain storms. The experiment is also designed to measure coherence bandwidth using a modulation on the carriers which produces lines over a 1.5-GHz bandwidth spaced at 180 MHz. This particular experiment also has a communications capability for transmitting signals to the satellite at 6 GHz and converting them to 20 or 30 GHz for retransmission to the ground. This link has a capability of about 40 MHz of bandwidth, which means a video transmission link or up to 40 megabit data link.

These pieces of technology are being presented because they indicate the present state-of-the-art in space-qualified hardware development which has taken place over the last two and a half to three years. These technologies

will form the basic parts or groundwork for the future communications systems such as the Tracking and Data Relay Satellite (TDRS) or advanced Comsat systems which will have to use these frequencies.

In figure 1 you see the traveling wave tube amplifier at 30 GHz. It produces 3 W of RF power, the tube efficiencies run 24 to 13 percent, depending on the frequency (20 GHz has a higher efficiency), and the tube weights are relatively low at 650 g. The total weight of both the tube, which is around 23 cm (9 in.) long, and the power supply is on the order of 3.5 kg. These tubes have wide bandwidths which make them useful for multiple channel FM or very wide bandwidth data transmissions. The total power dc to RF efficiencies, when you include the power supply, are on the order of 10 to 15 percent.

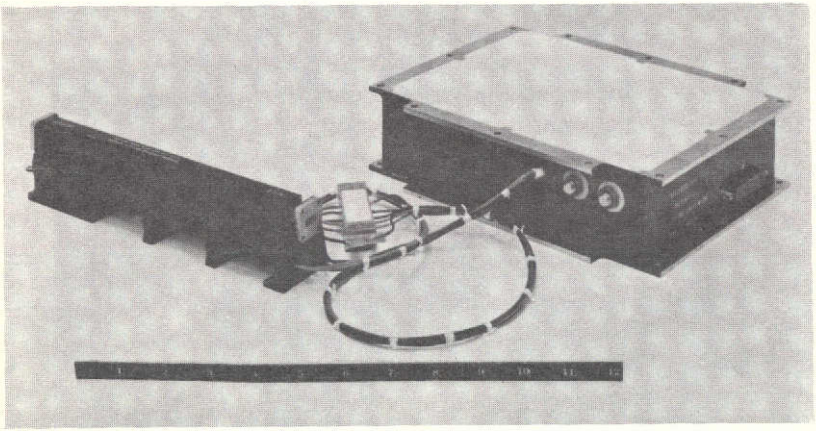


Figure 1. 30-GHz traveling wave tube amplifier with power supply.

The following list compares the characteristics of the 20-GHz tube with those of the 30-GHz tube.

Tube Characteristics	20 GHz	30 GHz
RF power output (W)	2.5	3
Tube efficiency percent	24	13
Weight (grams)	630	650
Bandwidth (MHz)	1500	1500
Maximum dc input power (W)	17	24

In figure 2 you see the modulator amplifier as installed on the ATS-F. This is a 20- and 30-GHz unit. You see here the United States coverage horn which at 30 GHz is around 10 cm (4 in.) on a side at the aperture. In this area you see the wideband phase modulator which is used to generate the multitone spectrum out to 1.5 GHz. This modulator was developed

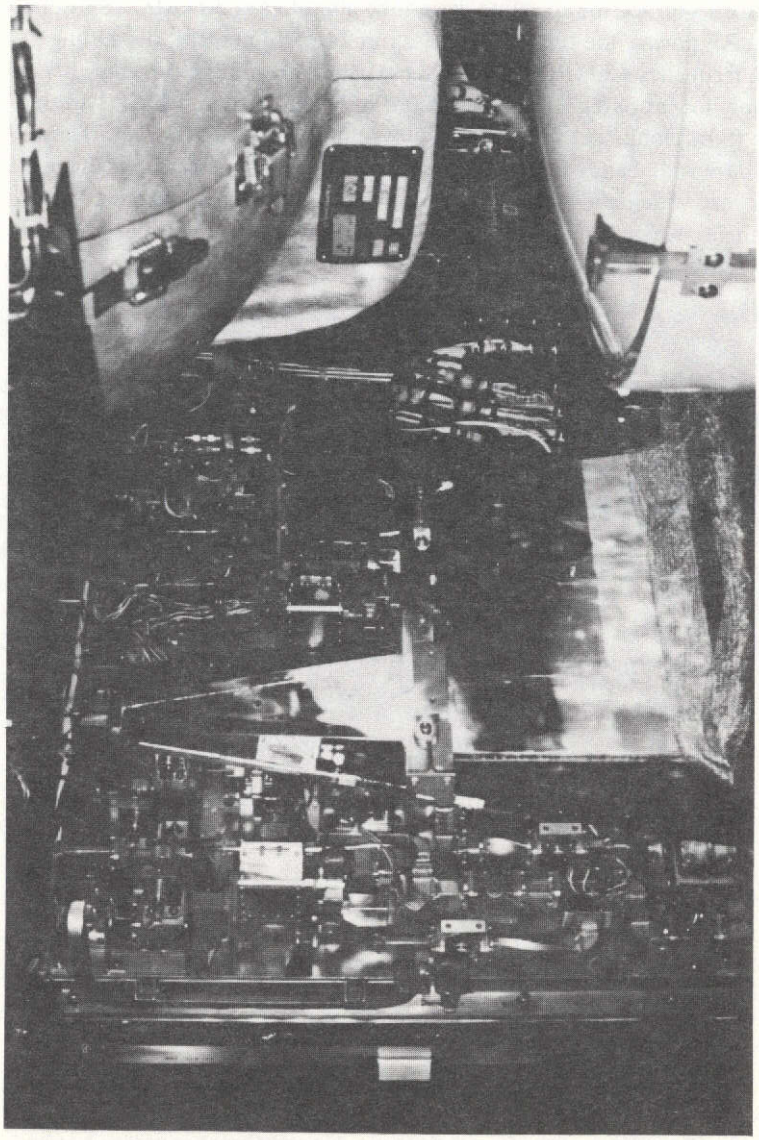
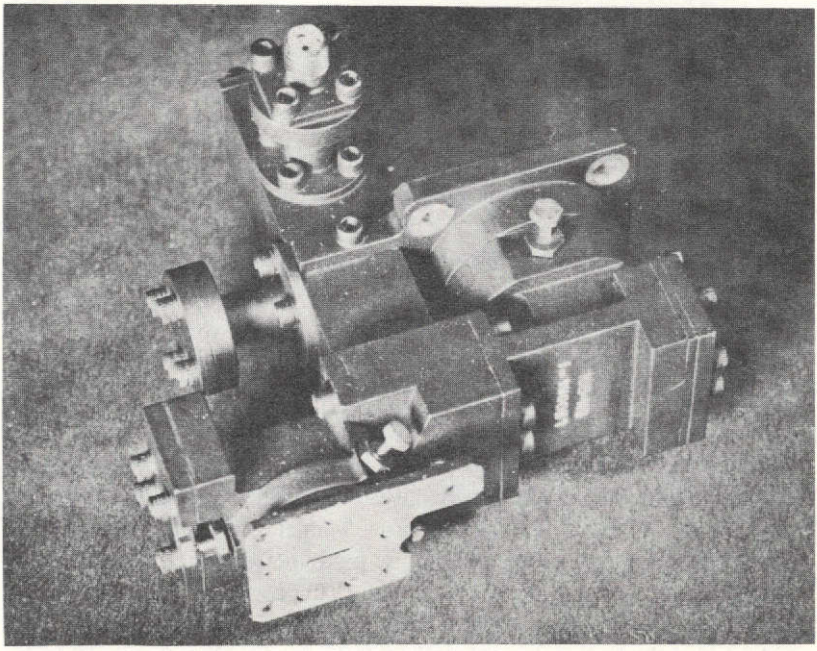


Figure 2. ATS-F 20- and 30-GHz millimeter wave experiment in the spacecraft.

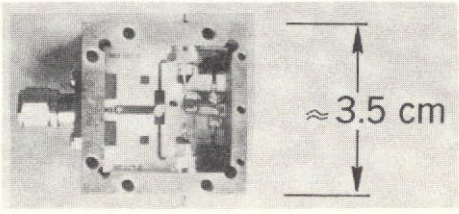
specifically for this experiment, but the technology developed should also have applications for a quadraphase modulator for a very wide bandwidth, high data rate system of the future. It has phase shifts on the order of ±90° and has a bandwidth of 500 MHz at 10 GHz. It has an insertion loss of 2.5 dB.

In figure 3a we have the image enhance mixer developed as the front end for the 13- and 18-GHz uplink propagation experiment. It has a 200-MHz bandwidth. It utilizes cylindrical cavity filters to provide the image enhancement characteristics in a small volume by using a single cavity for two separate modes. The mixer utilizes a standard silicon diode; it has a conversion loss at this frequency (18 GHz) of 3 dB, which is very good for a standard packaged diode. This unit has been through full qualification testing, that is, vibration, thermal vac, and has had no problems and is performing very well.

The second piece of technology (figure 3b) is a 4-GHz microwave integrated circuit tunnel diode amplifier which was built as the IF amplifier for the transponder; it has a 5.5-dB noise figure. Tunnel diode amplifiers have existed for some time with similar noise figure performance, but the use of the microwave integrated circuit techniques improved the gain stability over temperature. It also improved the mechanical ruggedness of the device and enabled the circuit designer to install thermisters directly on the board to be able to temperature compensate the gain over a wide temperature range. It is a very small unit and has less weight than the standard coaxial packages that have been flown on Intelsats and ATS satellites.



(a)



(b)

Figure 3. (a) 18-GHz image-enhanced mixer; (b) 4-GHz MIC tunnel diode amplifier.

12-GHz IMAGE AND SUM FREQUENCY ENHANCED MIXER

John Miller

We have been active for the past 6 years in the development of low cost ground terminal techniques and technology. To differentiate between the low cost ground terminals and the International Telecommunications Satellite (INTELSAT) or Space Tracking and Data Acquisition Network (STADAN) types of earth stations, I would like to mention some of the typical applications that we are concerned with.

This equipment that has been developed is applicable to such applications as the Communications Technology Satellite (CTS) experiment and the educational and health care experiments that are being planned and will be conducted on ATS-F. It is also applicable to some of the applications considered in the 1971 initiatives study for the President's Domestic Council, for example, resources sharing, computer networking, and a new one that is coming along at a very appropriate time, telecommunication as an alternate to transportation.

This particular development is a low noise, low cost, 12-GHz image and sum frequency enhanced mixer. It is a device which we feel will be competitive performance-wise and cost-wise with conventional tunnel diode amplifiers. The technical objectives for this task were that it cover the full band allocated to the broadcast satellite service, at 11.7 to 12.2 GHz, that the noise figure be competitive with the tunnel diode amplifier, that is, 6.2-dB band center and 6.7-dB band edge, and that it have an IF frequency in the region of 1 to 1.5 GHz and a conversion gain of approximately 25 dB.

Our cost objectives were that it be less than $500 in quantities of 100; this in contrast to a tunnel diode amplifier with a 6-dB noise figure. It has been estimated that the cost would be $1200 in quantities of 100.

This development was performed by the Westinghouse, Baltimore, Maryland facility under Goddard technical management. The completed unit (figure 1) is approximately 13 cm by 8 cm by 2 cm (5 in. by 3 in. by 0.75 in.). The two major components are the image and sum frequency enhanced mixer and the IF preamplifier. There are two input ports, one for the local oscillator and one for the signal. The local oscillator and signal are combined in the resonant ring type of filter and are impressed on a 6-μm junction width

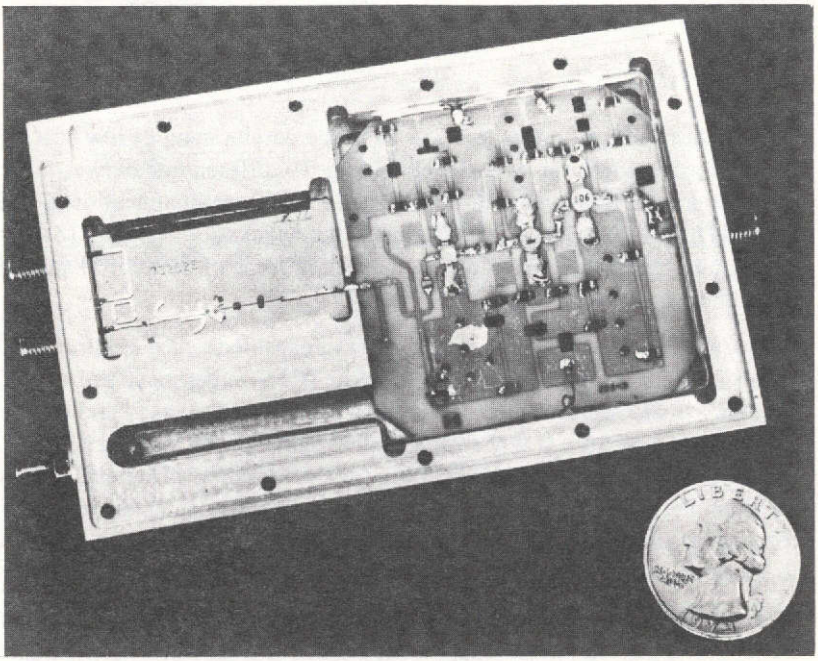


Figure 1. 12-GHz image and sum frequency enhanced mixer.

Gallium Arsenide Schottky Barrier diode, where mixing is accomplished. These sticks are the image and sum frequency shorts which reflect the signal power back into the diode for mixing down to IF. The IF, again, is 1 to 1.5 GHz and is a 3-stage silicon bipolar transistor amplifier. The losses, including conversion loss from input port to output port of the mixer, is approximately 3.5 dB in the band center, while the noise figure of the IF amplifier is approximately 2.5 dB in the band center.

The best way to judge the success of any development wherein you are looking for low cost in quantity deliveries is by looking for uniform performance of the units, unit to unit. Five units were delivered. The hatched area in figure 2 shows the design objective, that is, the noise figure, to be less than 6.2 dB band center, rising to 6.7 dB band edge. The measured results are very close to the design objectives. The maximum departure is approximately half dB at the band center and a similar amount on the edges. But, more importantly, as stated previously, there is a very good degree of repeatability, unit to unit.

In summary, then, I think we have successfully demonstrated that low cost, high performance mixers can be built and delivered.

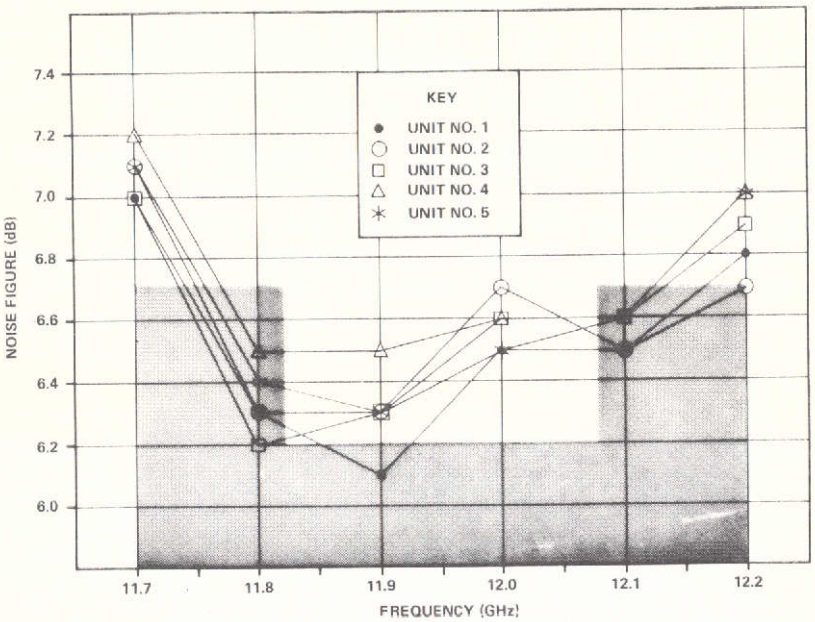


Figure 2. Noise figure versus frequency of the image and sum frequency enhanced mixer.

OPLE PLATFORM

Leonard Roach

This afternoon I would like to discuss the search and rescue package that I have displayed here. Before I begin, I would like to explain that OPLE stands for the Omega position location equipment, where Omega is the Navy's ground-based worldwide navigational system. The hand-held, lightweight, portable advanced OPLE transceiver as shown in figure 1 was redesigned, built, and successfully tested, and can be used in a search and rescue type experiment such as the global rescue alarm net (GRAN), that is shown in figure 2.

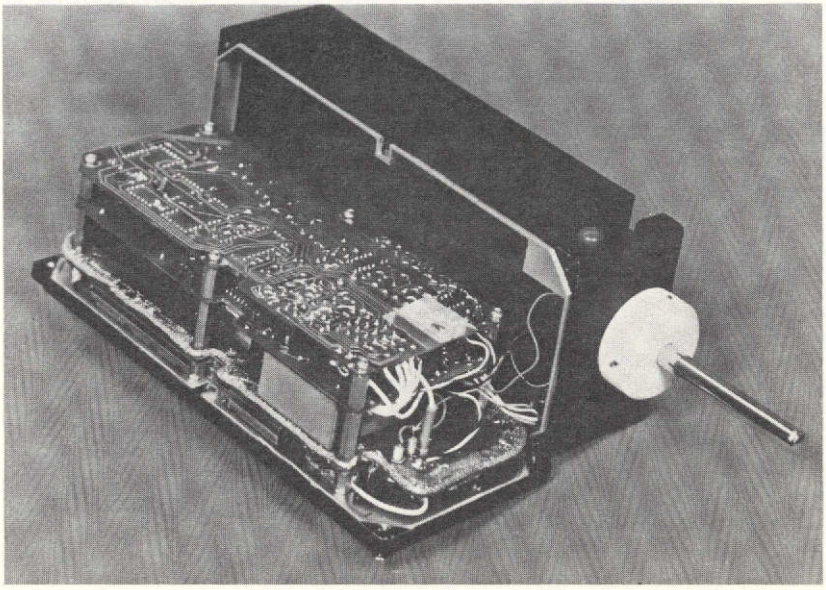


Figure 1. Advanced OPLE transceiver.

A typical operation presumes that in some sort of accident or mishap, the person involved would activate this unit by pressing a button. The unit would then receive Omega tones from at least three or more of the Omega stations. These tones would be up-converted to UHF, relayed by satellite

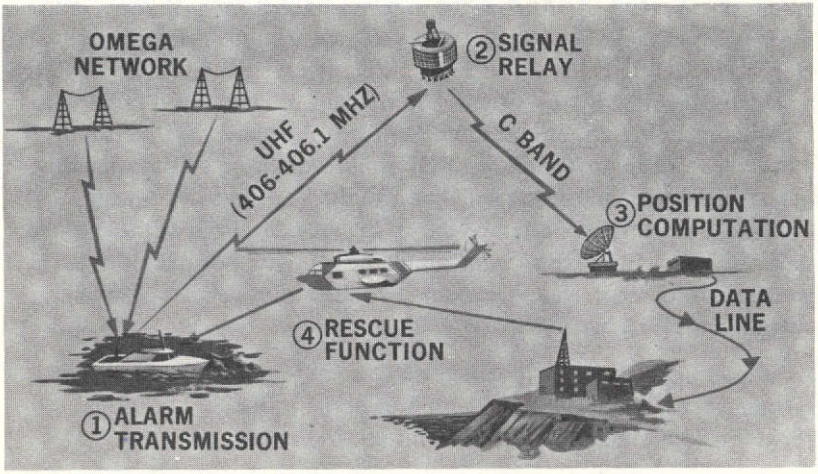


Figure 2. Global rescue alarm net (GRAN) concept.

to the Omega ground station, where the position location would be computed. Data would then be sent to a rescue facility where appropriate action could be taken.

In field testing, the satellite link was bypassed. We received the Omega tones, sent them directly to the ground station where nominal OPLE locations on the order of 3.5 to 5.5 km (2 to 3 n.m.) were computed. As a final field test, we modified our breadboard unit and used the Navy's LES-6 satellite, relaying the Omega data up to the Navy's LES-6 satellite, down to our ground station where typical Omega OPLE positions were obtained.

The unit that you saw previously in figures 1 and 2 weighs approximately 3.2 kg and occupies a volume of 1804 cm^3, including batteries. The original Omega OPLE units weighed on the order of 9.98 kg and occupied a volume of 12,290 cm^3. This unit was a prototype unit. If production units were built, it has been estimated that we could reduce the weight from 3.2 kg, down to 1.8 kg, keeping approximately the same volume. The weight reduction would be due to a lighter case and the newer lithium cell batteries as compared to the nickel cadmium batteries that are presently used in the unit. The remainder of the specs is that the unit has a 1 microvolt per meter sensitivity, which is slightly better than the original OPLE equipment, a dynamic range of 90 dB, and receives the 3 Omega tones of 10.2, 11.33, and 13.6 kHz. The transmitter is at 402.05 MHz at a 10 W power output for 2, 3-minute interrogation cycles. The unit operates from 253.15 K (-20° C) to 328.15 K (+55° C).

As the unit contains both a VLF received and UHF transmit antenna, both antennas were studied. Also, as the unit was to be hand-held in operation, the weight, size, and form factor of the unit were studied along with its electrical characteristics. In VLF operation, the system sensitivity or antenna length is limited by the VLF noise. As the Omega received phase data is omnidirectional, the antenna cannot distort the phase data. Since the unit has both VLF reception and UHF transmission, we do not want any of the UHF transmission to feed back into the VLF system. A whip, 0.97 meters, centrally located on top of the unit, meets all of these above specifications. The UHF antenna should be centrally polarized, hemispherical in coverage, and once again it should not transmit any of its energy to the VLF whip.

As seen in figure 3, a breadboarded UHF crossed-loop VLF whip antenna system was built and does tend to approach the above specifications. However, additional work would be needed in this area.

In conclusion, a hand-held, lightweight, portable OPLE was redesigned, built, and successfully tested. It is capable of operating with a satellite for a search and rescue type experiment. With minor modifications, this unit could be used as a data collection package on boats, buoys, ships, aircraft, and so on.

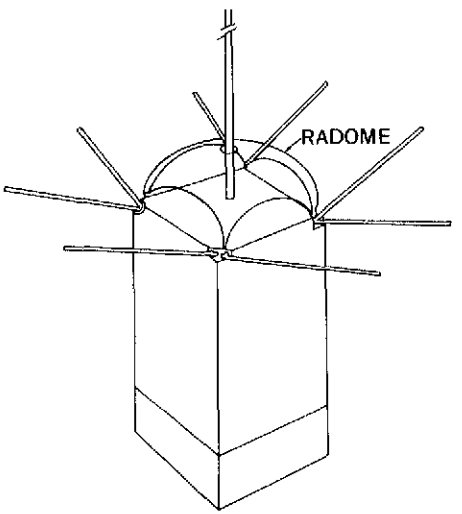


Figure 3. Breadboarded UHF crossed-loop/ VLF whip antenna system.

SHIP RADIO DETERMINATION AND COMMUNICATIONS EXPERIMENT

Richard Waetjen

The ship radio determination and communications experiment was conducted by our division during the spring of 1973 when the Vanguard sailed from its port in Cape Kennedy to Argentina in support of the Pioneer and Skylab missions. The experiment involved the Rosman ground station, ATS-3 and -5 satellites, and the Vanguard ship. The objectives of the experiment were to demonstrate the position and location of aircraft communications equipment (PLACE) in the maritime environment, using the Vanguard; to use the ATS-3 and -5 synchronized satellites to position fix the Vanguard; and to demonstrate two-way voice and digital data communications and operations control capability.

The overall experiment objectives were achieved as a position accuracy of approximately 1.85 km (1 n.m.) was demonstrated. The results are particularly significant as this was the first experiment which successfully demonstrated dynamic real-time ship position fixing and two-way communications using two satellites in synchronous orbits. The experiment was conceived and implemented in the Communications and Navigation Division. Between the concept in November 1972 and completion in May 1973, a concentrated team effort coordinated and directed the preparation, equipment installation, checkout, operation, and data collection between the ATS project, networks, contractors, the Rosman ground station, and the Vanguard.

PLACE was designed to provide the following: real-time position location of the user by means of two-way side tone ranging through two geosynchronous satellites using aeronautical L-band frequencies; position location with a nominal accuracy of 1.5-km; to provide duplex or two-way data communications; duplex voice communications; and finally, data and voice multiple access and operations control.

For the Vanguard experiment, some rather extensive modifications had to be implemented before the experiment could be conducted. These were to convert the equipment to interface with the ship's C-band equipment and to modify the receiver so it would lock onto the short signal bursts received from ATS-3 which are of the duration of 30 to 50 milliseconds, occurring approximately once every 1 second.

In conclusion, it can be reported that good voice communications were achieved despite the fact that the ATS-5 satellite is spinning and there were disturbances caused by the ATS-5 omnimode antenna not being exactly omni. It causes the signal-to-noise ratio to vary by as much as 4 to 10 dB. The signal variation is also the reason for the evaluation of the digital data channel not being meaningful. It is extremely difficult to derive quantitative data, that is, bit error rate statistics, under nonstationary conditions. However, despite these disturbances, the digital data channel could provide channel access and operations control which could be maintained with good reliability, and a user position determination accuracy on the order of nautical miles or better could be achieved.

Figure 1 shows the latitude and longitude of the Vanguard plotted against time. The position accuracy scale is indicated on the upper left as one nautical mile. It is apparent that an uncorrected position accuracy of approximately 3 miles down to less than 1 mile could be achieved. With computer processing and smoothing and corrections primarily for the ephemeris uncertainties of ATS-3 and ATS-5 or the satellites in use, this could be reduced to one nautical mile or better.

So, in summary, I would like to state that it was the concentrated team effort, initiated and implemented by the Communications and Navigation Division, which resulted in a major contribution, as for the first time a ship radio determination and communications experiment, using two synchronous satellites was successfully conducted. Experiment data and results are directly applicable to future experiments with ATS-F, and to the development of aeronautical and maritime services satellite systems.

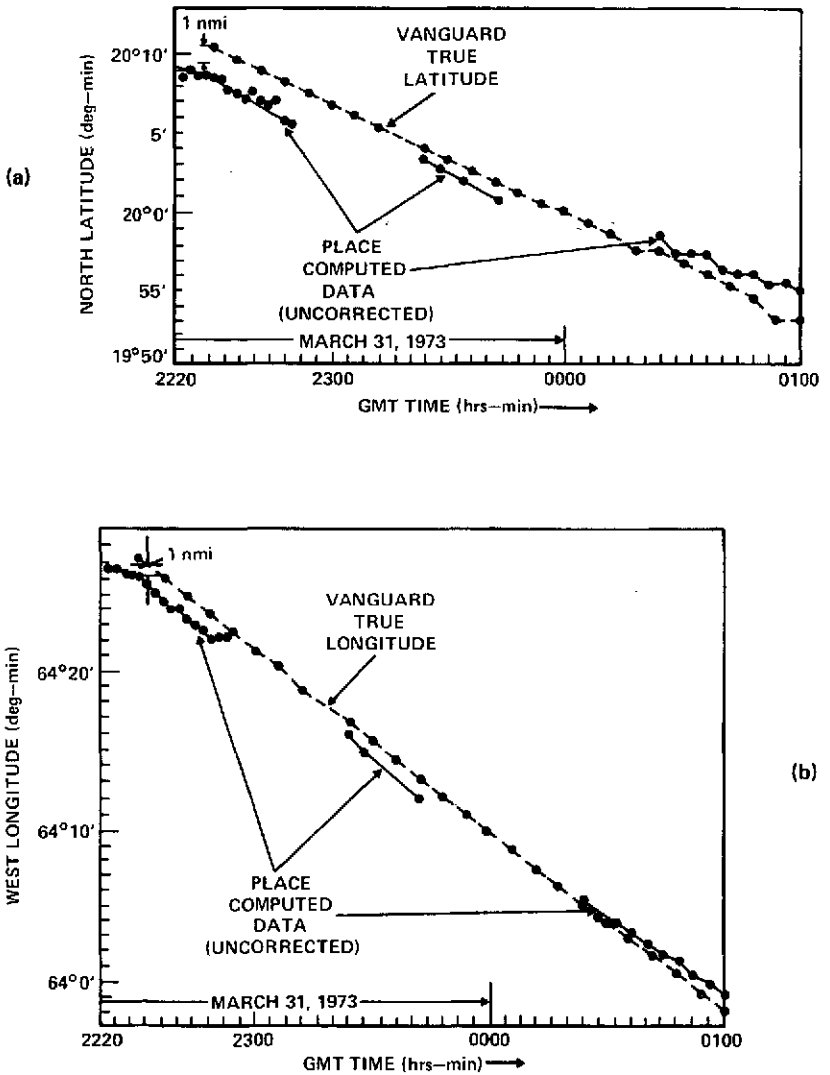


Figure 1. Ship radiodetermination and communications experiment. (a) Latitude versus time; (b) Longitude versus time.

A HIGH SPEED PHOTODETECTOR FOR THE Nd: YAG LASER SYSTEM

Dominick Santarpia

To provide some background, let me say that the Nd:YAG laser communications system is capable of operating at its fundamental wavelength of 1.06 μm and at a wavelength of 0.53 μm by frequency doubling. The advantages of operating at the 0.53-μm wavelength are higher antenna gains, reduced beam divergence, and a good selection of off-the-shelf detectors. However, the doubling process is inefficient and potentially unreliable and can cause a degradation of system performance with time. The fundamental wavelength offers simplicity, reliability, and more available transmitter power. However, this wavelength has lacked a good detector.

A major portion of our effort at Goddard has been to develop a suitable detector at this wavelength. Table 1 lists the available 1.06-μm detectors and some of the detector parameters which determine system performance. The photomultiplier tube has been the standard for obtaining ultimate sensitivity in the visible and near-infrared region because of its very low dark current, its high stable gain, and wideband widths. However, the low quantum efficiencies of photocathodes at this wavelength seriously compromise system performance. The silicon avalanche photodiode detector (APD) has too low an optical absorption coefficient to yield high quantum efficiencies. The narrow bandwidths associated with silicon APDs preclude its use in a high data rate system. Germanium avalanche photodiode detectors are not acceptable because of extremely high dark currents and because of the excess noise which is generated by the multiplication process.

An effort was begin about a year ago by North American Rockwell to develop an avalanche photodiode detector which uses III-V materials and has the characteristics of high quantum efficiency, bandwidths wide enough to accommodate a 400-Mbs data stream, high uniform internal gains, and low internal noise currents.

Figure 1 shows some of the progress that has been made. Shown in this figure is a picture of a prototype detector which was delivered to us this year. It measures approximately 16 cm^3 (1 $in.^3$) in volume and requires less than a milliwatt of prime power. The oscillogram you see in the figure

Table 1
Available 1.06-μm Detectors and Detector Parameters

Available 1.06-μm Detectors	Quantum Efficiency (percent)	Bandwidth	Internal Gain	Internal Noise Currents (A)
PMT	0.1	2 GHz	10^4	5×10^{-15}
Si APD	10-15	100-150 MHz	300	5×10^{-11}
Ge APD	60-80	1 GHz	300	5×10^{-6}

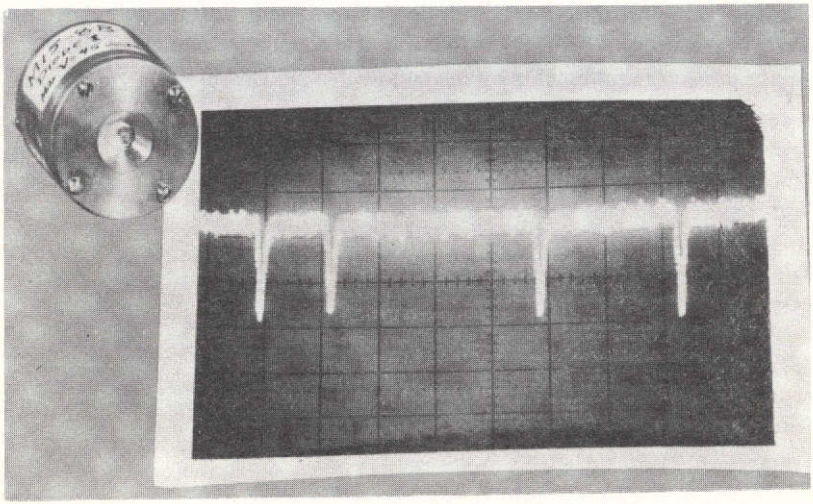


Figure 1. Prototype high speed photodetector.

is a picture of the response of this detector to a 1.06-μm laser beam which was modulated by a 400-Mbs data stream. The full width at the half voltage point measures about 200 ps; this gives you some idea of the speed of the device. Quantum efficiencies of 95 percent have been attained. Bandwidths in excess of 4 GHz have also been obtained. Leakage currents of 5 nA have been maintained. Internal gain, however, has been limited to about 17 dB.

The limitation in gain is due to microplasma type defects which have persisted. These defects are small regions in the active area of the device which break down over low bias voltage. We have already met three of the four requirements for communications systems being developed: Quantum efficiency, bandwidth, and leakage current. An internal gain of 34 dB is required if the detector is to be followed by a low noise, 50-ohm preamp. For low internal gains, on the order of 17 dB, detector performance is limited by the following preamplifier. Therefore, if a lower noise contributing preamplifier can be built, the existing detector with a gain of 17 dB can also meet the space mission requirements.

Our efforts are therefore being directed along two paths: Research is continuing to improve the materials technology so that higher uniform gains can be obtained; in addition, work is being carried out to hybrid integrate a MESFET preamplifier on the same substrate as the detector which is matched for low noise operation.

Having two approaches to the solution of a problem, and we do believe we have two solutions, we feel confident that we will have met all of these space mission requirements by the end of this next fiscal year.

MEASUREMENT OF ATMOSPHERIC ATTENUATION AT 10.6 μm USING AN INFRARED HETERODYNE RADIOMETER

Steve Cohen

A solar infrared heterodyne radiometer (IHR) has been used to measure atmospheric transmission at carbon dioxide laser wavelengths. These measurements are part of a comprehensive program of laser tracking and communication and are built upon the receiver technology reported in a previous SRT review. Prior measurements of IF transmission of the atmosphere have been made at spectral resolutions of about one wave number, or about 30 GHz. This broadband resolution is insufficient for determining the propagation characteristics of the atmosphere for monochromatic signals. On the other hand, the spectral resolution of the IHR is about 7 milliwave numbers and does provide the requisite data on line absorption by the atmosphere.

In figure 1 we show the optical front end of the solar heterodyne radiometer. Sunlight is collected by telescope and relayed through a viewing device to a mechanical chopper. The mechanical chopper serves as an optical Dicke switch. Mounted on its housing are a penlight source and photocell which generate a pulse for synchronous detection. In addition, the chopper blades are mirrored so that in that portion of the Dicke switch cycle in which the solar radiation is blocked, radiation from a reference blackbody is reflected to the photomixer. At the mixer surface, heterodyning or photomixing occurs between the local oscillator and either the solar or blackbody radiation. The output from this radiometer is the difference between the power collected in the solar channel and the power collected in the reference channel. This device can be operated either in a null-balanced, gain modulated mode or in an unbalanced mode.

Figure 2 shows an example of the type of measurements that have been made with the radiometer. The vertical deflection is proportional to the intensity of the received solar radiation at the P(18) CO_2 laser wavelength. The time duration of the scan is of no physical significance since it corresponds to opening and closing a shutter in front of the telescope. During the entire data collection process the radiometer telescope is pointed at one particular spot on the sun. The rise and the fall of the signal corresponds to the 5-s integration time constant for the system. For other measurements, the integration time constant can be varied from 0.1 to 30 s. The decrease

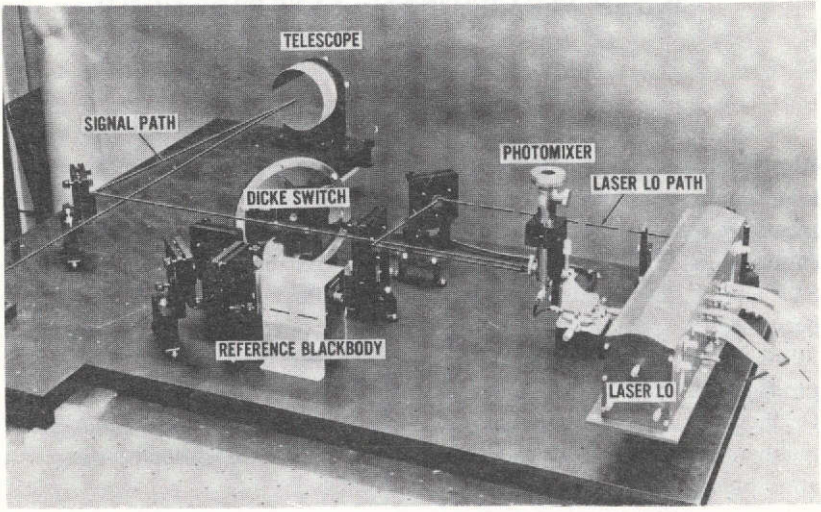


Figure 1. Optical front end of the solar heterodyne radiometer.

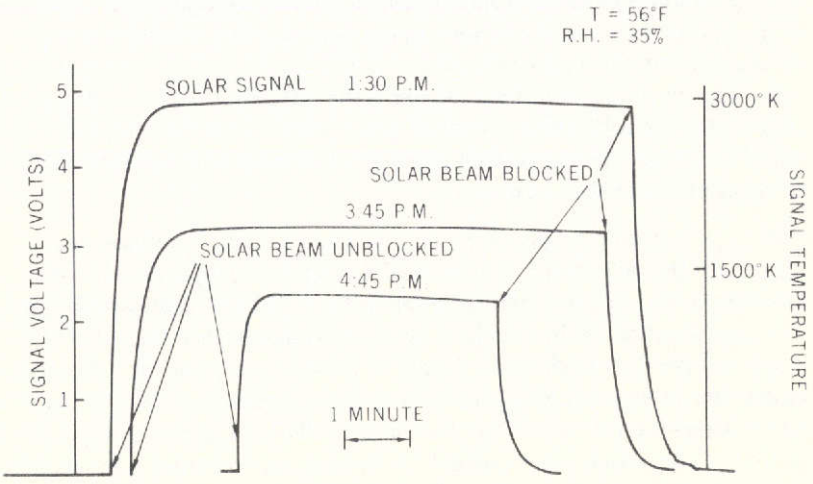


Figure 2. Atmospheric attenuation data.

in the received solar signal as the sunlight propagates through longer optical path length late in the day is evident.

In figure 3 we analyze the data just shown along with the other data collected on the same date, by plotting the logarithm of the received solar energy as a function of the secant of the solar zenith angle. The slope of the resulting straight line gives the information needed to determine the vertical path atmospheric transmission through the atmosphere. In this case the vertical path transmission was 59 percent at the P(18) CO_2 wavelength, with an atmospheric relative humidity of 35 percent and temperature of 286.15 K (56° F). This very favorable result indicates that CO_2 laser radiation can be efficiently propagated through the atmosphere. The difference between this narrowband measurement and broadband measurements is significant; under comparable weather conditions, the broadband transmission would be about 80 or 90 percent.

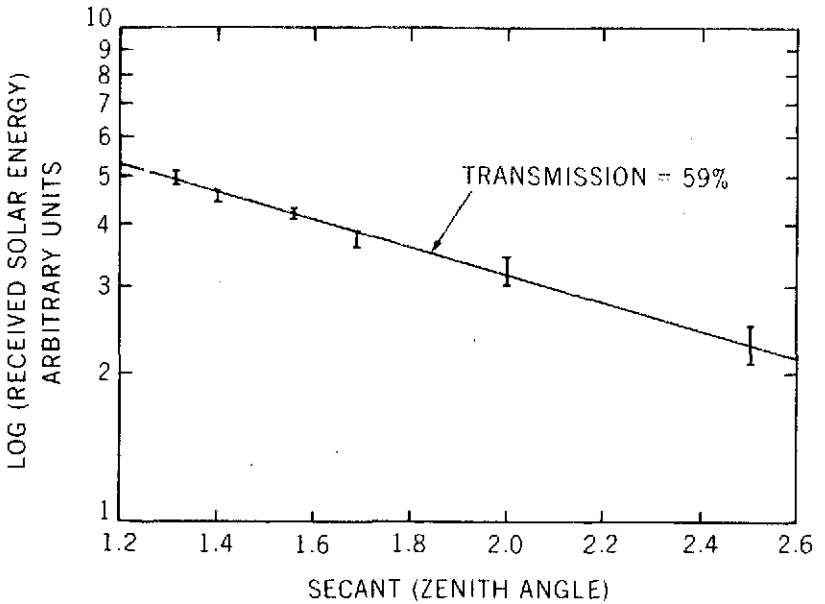


Figure 3. Analysis of atmospheric attenuation data at 10 μm.

In summary, I would like to mention other data that have been obtained with the radiometer. We have already mentioned 59-percent transmission under clear skies, this corresponding to an attenuation of 2.3 dB. Under heavy cloud cover, we have seen attenuations in excess of 25 dB, and in a summer haze we have seen up to 6-dB fluctuations over 2-minute intervals.

In addition to being used for these atmospheric transmission measurements, the radiometer also finds potential application in pollution measurements, a variety of spectroscopic applications, infrared mapping of the sun, and astrochemical measurements. A particular example of such astrochemical measurements are measurements of infrared emission from the moon, the planets, and Comet Kohoutek, which we are now performing.

OPTICAL ANTENNA GAIN

Bernard Klein

Optical communications is coming to a point of sophistication and it must have design data along the lines of what is available for more established technology. To this end we have provided the optical engineer with comprehensive designer-oriented graphs which allow the rapid and accurate evaluation of optical transmitting and receiving antennas in terms of antenna gain.

Figure 1 shows a typical optical antenna first as (a) a transmitter and then (b) in an optical heterodyne receiving configuration. The figure shows the central obscuration which restricts the efficiency of both transmitter and receiver. The ratio of the obscuration radius to the primary aperture radius is referred to as the obscuration radio, γ, and is a convenient term to depict different antenna configurations.

Some of the major points of interest which we have considered in our design curves are as follows: For antennas used as transmitters we assume Gaussian illumination and consider the effect of the central obscurations and finite aperture sizes on both the near and far field gain. In particular, we have considered how to match the incident radiation to the antenna to maximize the far field on axis gain. A detailed discussion has also been held on the effect of defocusing as an aid to acquisition.

The antennas, when used as the collecting and focusing optics, are considered to be uniformly illuminated. We again considered the obscuration by the secondary element, the effect of energy spillover at the detector for direct detection, and the effect of the local oscillator (LO) distribution on the detector element for homodyne and heterodyne detection.

The necessity of the analyses which we have conducted stems from the fact that optical antennas have different characteristics in some respects than the conventional microwave antennas. Foremost is the fact that the relation of laser wavelengths to the aperture diameters are on the order of two or three magnitudes larger than for microwave systems. This will have an effect on the diffraction effects and as a consequence, the gain. Also, for optical heterodyne communications the LO field may be focused before being injected into the optical train.

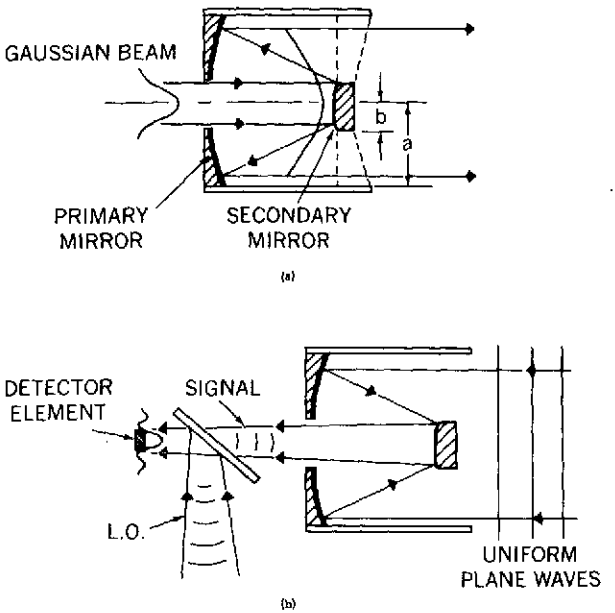


Figure 1. Typical optical antenna. (a) Transmitter antenna; (b) Heterodyne receiver antenna.

Figure 2 shows some results unique to optical receivers in connection with heterodyne or homodyne detection. Plotted here is the heterodyne detection loss in dB as a function of the ratio of the optical receiver antenna F-number, F_s, to the local oscillator F-number, F_{LO}. The signal is assumed to be focused onto the detector by the centrally obscured antenna in such a way that only the central disk of the modified Airy pattern falls on the detector. We consider a particular value of the obscuration ratio and allow the ratio of F_s to F_{LO} to vary. The limit as F_S/F_{LO} goes to zero corresponds to uniform illumination of the detector by the local oscillator. As the local oscillator is focused on the detector, the Airy pattern begins to fall on the detector. After this point, that is, as the local oscillator is focused still further, partial cancellation of the IF signal takes place and the receiver gain decreases.

In summary, what we have presented is a method of optical antenna analysis which allows the conventional range equations of microwave theory to be applied to optical communications and radar links in a simple and efficient fashion. We have developed a set of designer-oriented plots which expressly take into account the truncation of the incident radiation by a

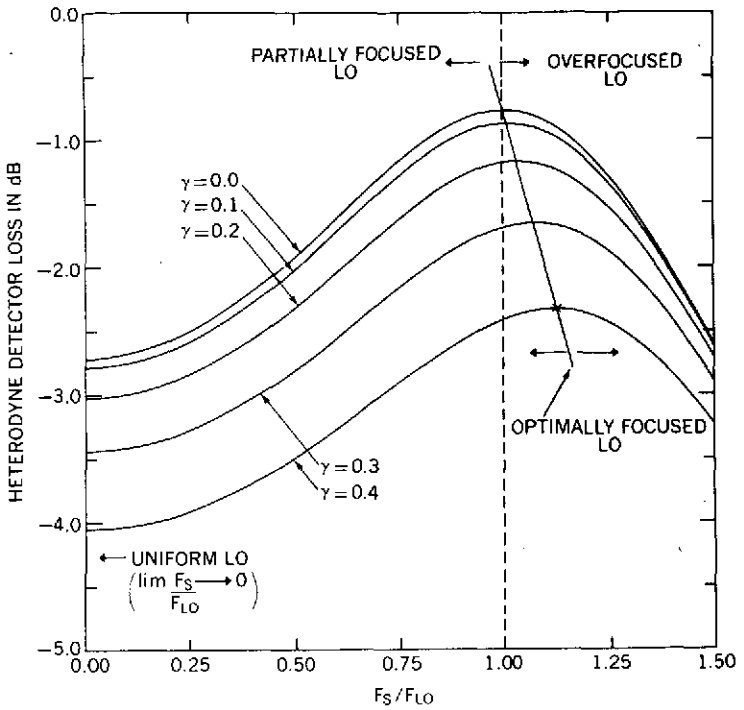


Figure 2. Heterodyne detection loss as a function of the ratio of the optical receiver antenna F-number to that of the local oscillator.

finite aperture and central obscuration, their effect on transmitting antenna gain, and the effect of central obscuration and finite detector size on receiver gain. The accurate calculation of system signal-to-noise ratios, we feel, can now be performed rapidly and without the aid of a computer. For a more complete discussion on optical antenna gain, a NASA document, TM X-524-73-185, is available from the author.

THE EFFECTS OF GAS PRESSURE ON WAVEGUIDE LASER TUNABILITY

John Degnan

Before I discuss the effects of gas pressure on waveguide laser tunability, let me take a moment to review the motivations and concepts behind the waveguide laser program here at Goddard. At last year's review I mentioned that high data rate communications systems, using CO_2 lasers, require a highly tunable laser local oscillator if spectral foldover and its resultant signal distortion is to be avoided. This is especially true if one of the receivers is on a low earth-orbiting satellite. The doppler-shifted information bandwidth must then be tracked at the receiver by the laser local oscillator. Since conventional CO_2 lasers are extremely limited in their tuning ranges, the waveguide laser is being developed to serve as a highly tunable laser local oscillator.

The waveguide laser (figure 1) is characterized by a narrow capillary, typically 1 to 2 mm in diameter, which contains a high pressure laser gas mix and is placed between two reflectors which provide feedback to support the oscillation. As with any other type of oscillator, the laser only oscillates if the round trip gain is greater than the round trip loss. At last year's review, I indicated that we had developed a theory of waveguide laser resonators which effectively determined the optical losses in this system. The optical losses, of course, depend on how effectively the mirrors couple the radiation emerging from the waveguide back into the guide. This is a function of the various mirror parameters shown in the figure. The predictions of that theory with regard to laser power output and laser far-field pattern as a function of resonator geometry have since been verified by an extensive in-house experimental program and also by experiments at other laboratories.

The next logical step was to consider how gas pressure affected waveguide laser tunability. As I said before, the laser only oscillates if the round trip gain is greater than the round trip loss. As we increase the pressure, the more frequent molecular collisions result in a broadening of the molecular line width; therefore, greater tunability will result provided the round trip gain is sufficient to overcome the optical losses.

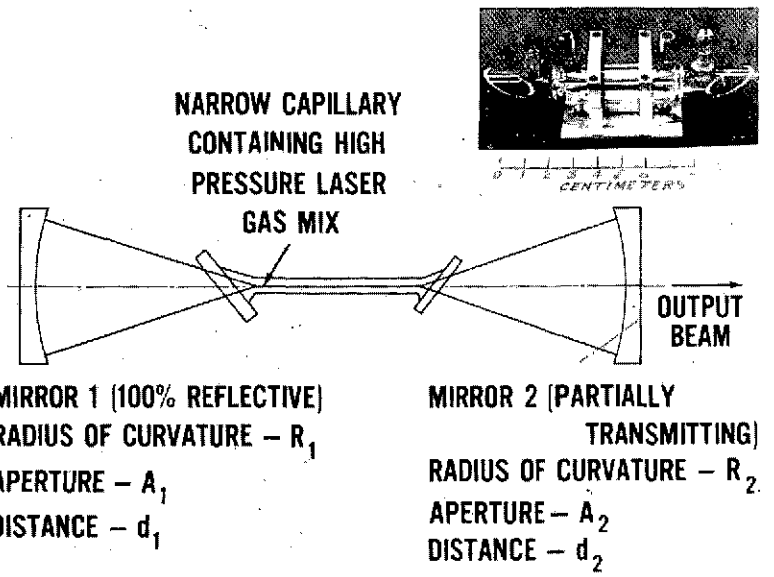


Figure 1. Schematic of a waveguide laser.

If we imagine beginning with no gas in the capillary, the laser gain is, of course, zero. As we increase the gas pressure, the laser gain must initially increase, as we provide more and more molecules to contribute to the population inversion. At still higher pressures, however, the same collisions which are broadening our molecular line also tend to depopulate the laser levels with the net result that the gain of the laser decreases beyond a certain pressure.

The effect of pressure on the gain can be seen in figure 2. For the assumption of an 8.8-cm capillary length, we see that the round trip gain is a maximum of 30 percent at a pressure of 9.3×10^3 N/m^2 (70 torr). The linewidth at that pressure is 175 MHz. As we increase the pressure to 2.9×10^4 N/m^2 (220 torr), the linewidth has increased to 550 MHz, but the round trip gain is down to about 11 percent. At 4.1×10^4 N/m^2 (310 torr), we have a molecular line halfwidth of 775 MHz, but the round trip gain is now down to 6 percent. If we assume that we have a round trip loss in the resonator of 3 percent, which includes the output mirror transmission, we find that the laser tuning range at 9.3×10^3 N/m^2 (70 torr) is ±485 MHz, at 2.9×10^4 N/m^2 (220 torr), ±850 MHz, but at the higher pressure, 4.1×10^4 N/m^2 (310 torr), the laser tuning range actually decreases to

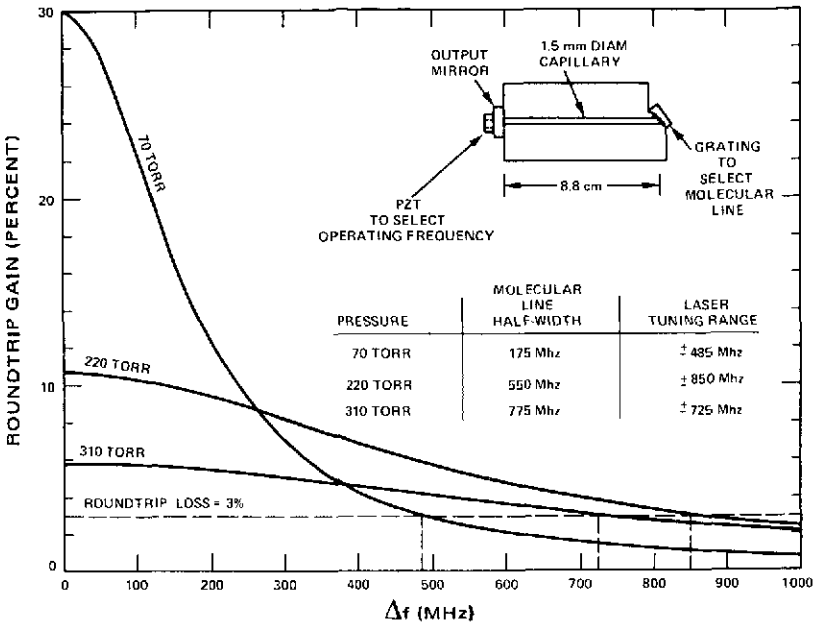


Figure 2. Effect of pressure on round trip gain.

±725 MHz. And so we can see from figure 2 that there is no direct relationship between molecular line halfwidth and the laser tuning range. It is a function not only of the molecular characteristics, but also of the resonator losses.

At Goddard we have developed a phenomenological model of a general high pressure gas laser which allows predictions to be made with regard to the optimum discharge length and the optimum mirror reflectivity, and allows a prediction of output power and laser tunability. Figure 3 is an example of the type of calculation that could be performed using this model. For a 1.5-mm diameter capillary and an 8.8-cm length discharge tube, we find that the optimum operating pressure for maximum laser tunability is 2.9×10^4 N/m^2 (220 torr) for this particular configuration. This leads to a maximum tunability of ±840 MHz which would be sufficient to track a worst case doppler shift with a 300-MHz information bandwidth. You will notice here that I have used a grating to select the molecular line. (There are various molecular lines in CO_2 that can oscillate, so we have to include some sort of frequency-selective device.) We have a piezoelectric transducer (PZT) which is controlled by a voltage that selects the operating frequency within that molecular line.

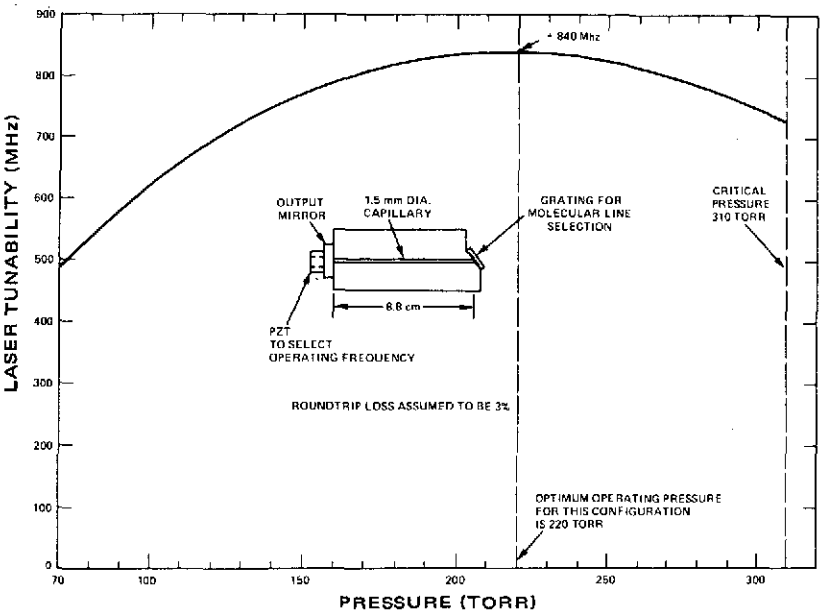


Figure 3. Example of calculation that could be performed using Goddard's general high pressure gas laser.

Our present efforts are directed toward determining whether or not the grating is actually the optimum frequency selection technique for our applications. We are presently studying techniques such as internal etalons and also three-mirror laser cavities.

WIDEBAND LASER MODULATOR

Nelson McAvoy

Unlike the case with microwave transmitters, the modulators of laser transmitters take a large fraction of the total transmitter power. For this reason, we have had an extensive SRT program in space-qualified modulators for lasers. We selected from the many techniques the most efficient one. We went into the laboratory and tested this modulator and generated all the engineering parameters. After this, it was given under the AAFA program to a contractor to be put in a space-qualified form. This presentation discusses the results of the program.

The technique selected was coupling modulation, called this because we modulate the coupling out of the resonator. In the top half of figure 1, we have two concave mirrors, which constitute the resonator. We have a glass plasma tube with windows which are cocked at an angle such that the best vertical polarization transmission is obtained and so the field inside the resonator has a vertical polarization. In this situation, there is no coupling out of the laser at all. In the bottom picture a modulator is included in the resonator. When you put a voltage onto this modulator, which is a crystal, the radiation going through the modulator and back gets rotated in polarization and the component (which in this picture is out of the plane of the figure) gets reflected off the window. This constitutes the output beam of the transmitter, and will give you something very close to double sideband suppressed carrier. If you want to insert some carrier, you simply rotate the crystal slightly, or you can put a dc bias on the crystal, which gives you as much carrier as you wish depending on the application.

Figure 2 gives you a feel for the modulator itself--it is a cadmium telluride crystal. It is 1 mm^2 in cross section and about 5 cm (2 in.) long. The two metal plates are the plates for putting the voltage across the crystal. This particular crystal, which is the finished product, has about a 1-percent insertion loss. This parameter is very important because the crystal is inside of an optical resonator and you have to maintain your high Q in the resonator. I show the relationship between the transmitter total power and the modulated power in figure 3.

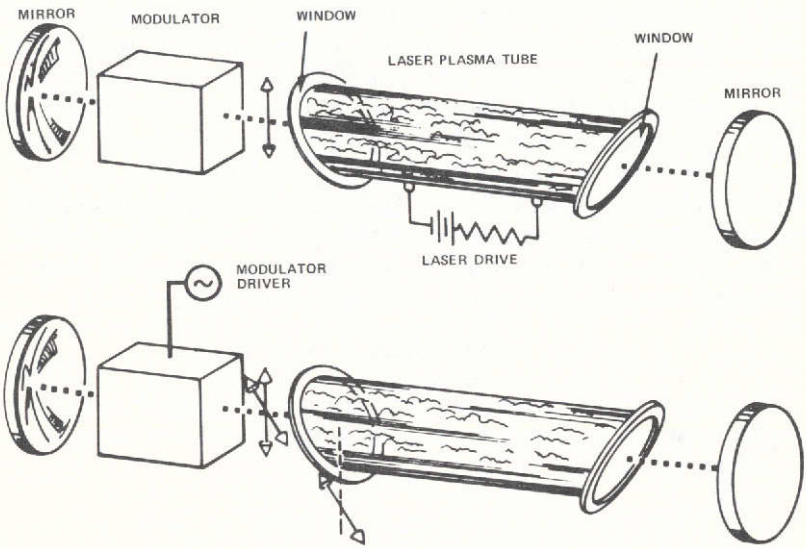


Figure 1. Laser coupling modulation.

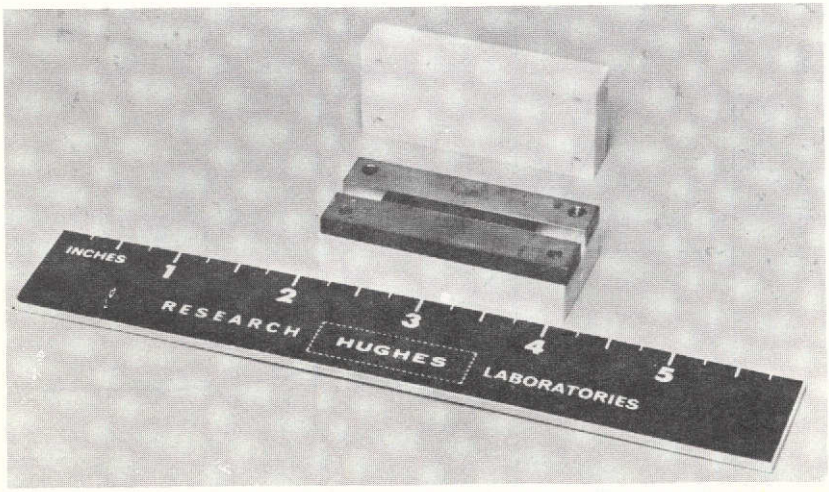


Figure 2. Wideband laser modulator cadmium telluride crystal.

We have plotted in this figure the length of the plasma tube in centimeters and plotted against that the total modulated power for a transmitter which has 1-W output and 400-MHz bandwidth. Along the bottom, we also have plotted (a bit parenthetically) the actual phase shift in radians which the beam experiences as it goes through the modulator. As you see from the figure, the longer the plasma tube, the more power it takes for the transmitter. But since you have higher circulating power in the resonator, it takes less coupling to get 1 W out. The point that we wish to make is that about 20 W of modulator power is used with about a 25 cm long plasma tube. As you will see from figure 4, these parameters turn out to be the optimum situation for this particular crystal and modulator.

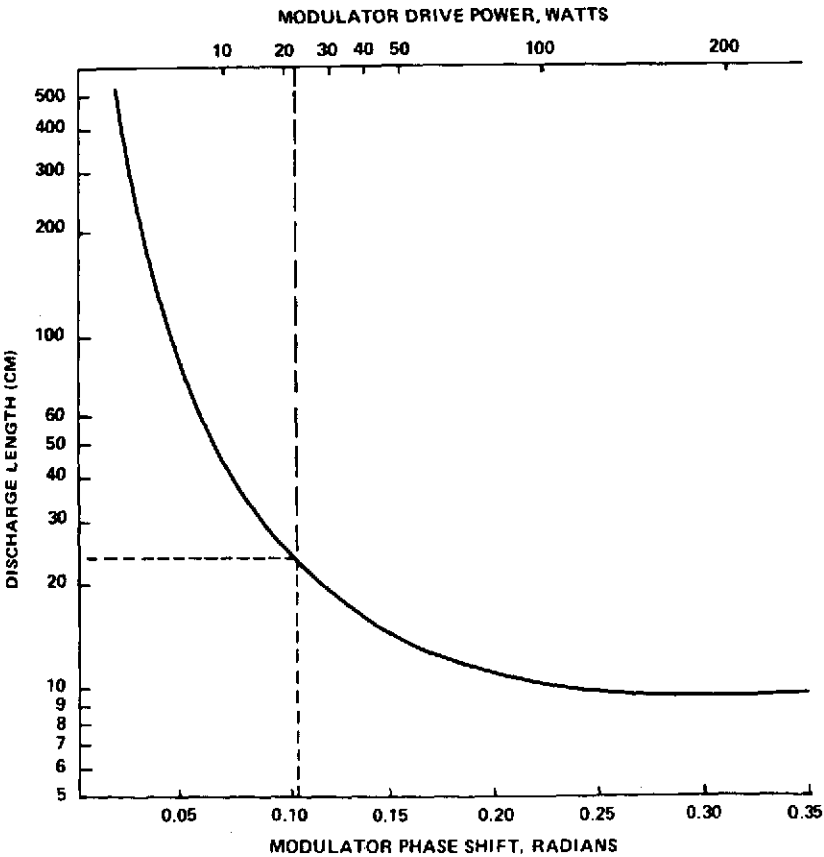


Figure 3. Length of plasma tube versus total modulated power and actual phase shift for a transmitter which has 1-W output and 400-MHz bandwidth.

In figure 4 we have plotted the same thing on the x-axis that we had before. That is, the modulator power for a 1-W transmitter with 400-MHz bandwidth; only it is plotted this time against the total prime power of the transmitter which includes both the laser and the modulator. The point on the curve at which the modulator power is 20 W for a 25 cm long plasma tube turns out to be 60 W total transmitted power. This represents the optimum situation.

The status of the program is that we have the modulator in the space-qualified form. The driver electronics for the modulator is still in the breadboard situation. The next steps will be to take different modulation formats where, for example, difference in linearity becomes important; take particular coding methods that you would want to use in a particular spacecraft application; and then design a proper driver for the best coding and modulating scheme.

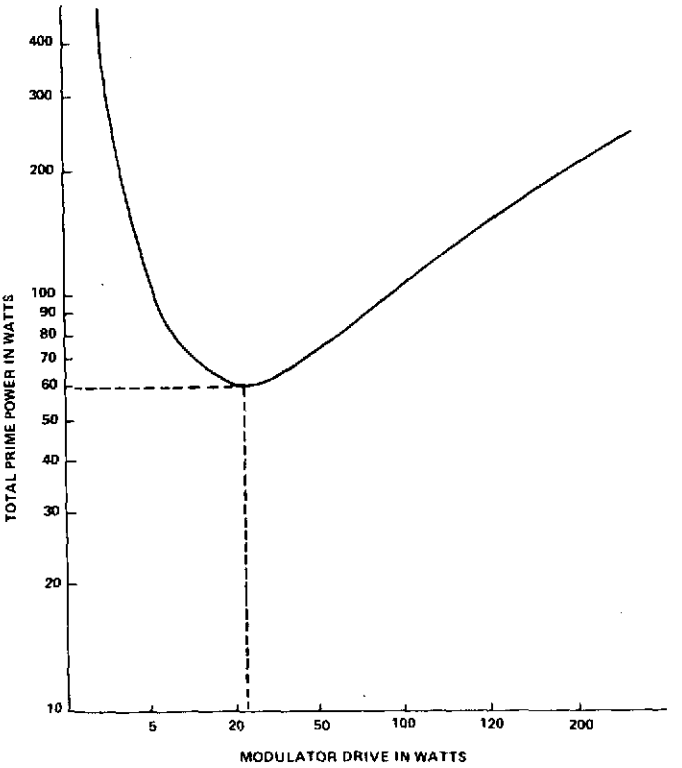


Figure 4. Modulator power for a 1-W transmitter with 400-MHz bandwidth versus prime power of the transmitter.

LONG-RANGE LASER TRAVERSING SYSTEM

Louis Caudill

This talk is sort of off the beaten track--essentially technology application more than anything. It deals with a joint effort we have had with the U.S. Department of Agriculture Forest Service to develop a surveying system which is basically designed to determine the azimuth bearing between two mutually nonintervisible points located up to 1.6 km (1 mile) apart.

The basic requirement for this system comes from the fact that present land values and land usage are becoming so important in the United States that it has become a requirement in the Forest Service to better mark or better control their borders. To have some idea of the magnitude of this problem, you should know that the Forest Service itself has 757,000 km^2 (187 million acres) of land located in 44 states including Puerto Rico. Around this land are 435,000 km (272,000 miles) of property lines, with approximately three-quarters of a million different neighbors. The lands are often intermingled in complex patterns and also most of the lands are in private ownership. So it presents quite a legal problem. Of this, many of the surveys of this land are as much as 150 years old. Property corners have not been located or marked for at least 150 years. Less than one-tenth of the Forest Service's property lines and corners are well enough located and marked so that land owners can manage land up to the lines with complete confidence. Using present surveying methods, the Forest Service is able to survey up to about 3200 km (2000 miles) a year, which at the present rate, just to reestablish existing control lines, would take an estimated 24 years to bring the surveys up to date. And this is only updating the 10 percent that the Service currently controls. They currently estimate, with the new system we just completed, that they will be able to update the 10 percent in approximately 12 years, at a cost savings estimated in access of $60 million to the Forest Service.

Our approach to the problem was first to develop an inhouse prototype which we evaluated in a field condition. We thought it was necessary to come up with an operational system, and the best way to do this was to come up with the correct engineering number so that we could go out to industry with a working model. We did not want to go out with an R&D type program. Figure 1 shows our inhouse prototype. On the left, you see

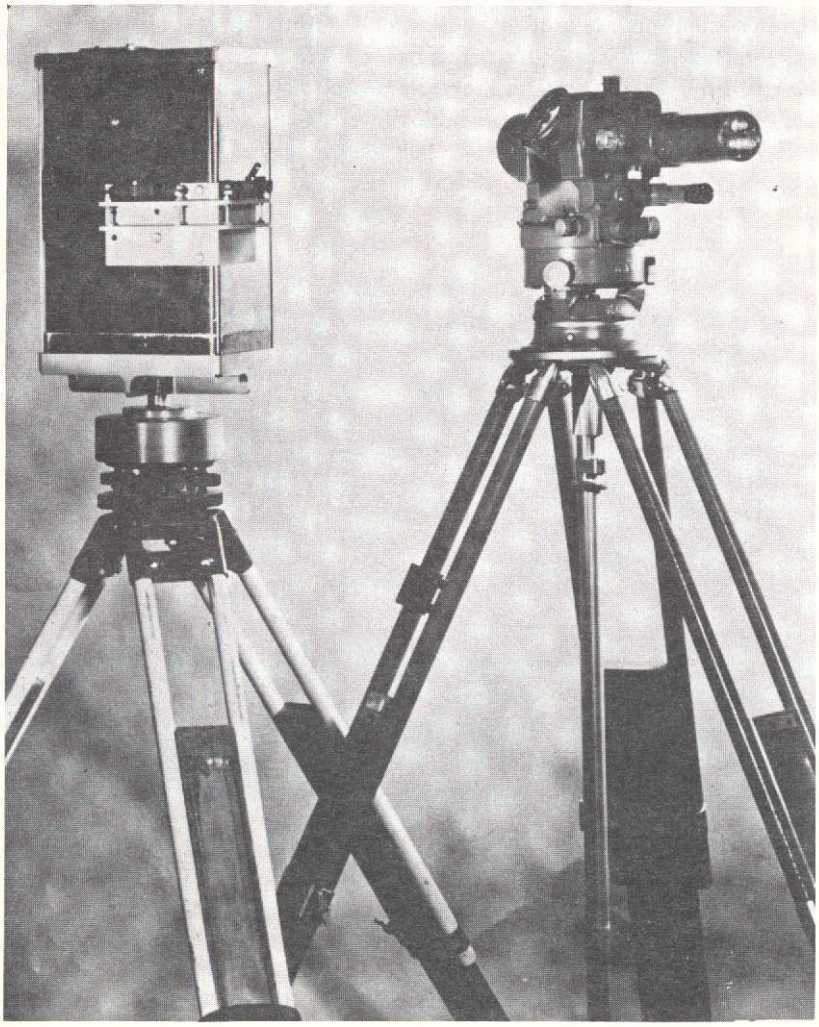


Figure 1. Inhouse prototype of long-range laser traversing system. On the left is the laser system and on the right, the electro-optical receiver system.

the laser system and on the right, the electro-optical receiver system. It is basically, as you see it there, a surveying prototype. Using this system in a field evaluation program, we generated the necessary engineering numbers to go out on competitive bid to industry on a fixed-price contract for the fabrication of a portable system to turn over to the National Forest personnel.

What you see in figure 2 is the finished product, and you will notice that it represents our inhouse prototype quite a bit. On the right you see the laser transmitter, the operator with a backpack for carrying the system. On the left is the receiver system, the operator's display, and also included with this is a similar type backpack. The total weight of this system you see here is 15 kg (34 lb) on the man's back. The total weight of the receiver system is approximately 22 kg (50 lb) on the man's back.

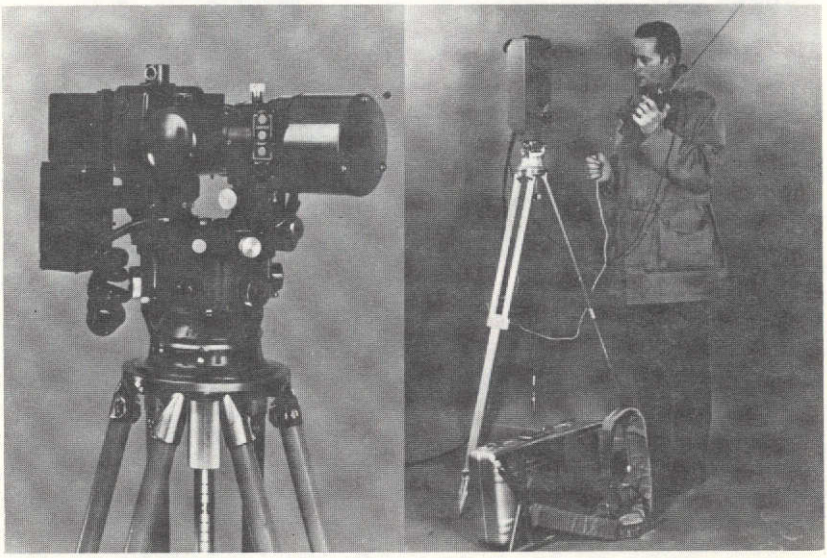


Figure 2. The finished laser transmitter is seen on the right, and on the left is the receiver system.

In figure 3 you will see an artist's concept of how this system is actually operated in the field. The system is set up on corners located approximately 1.6 km (1 mile) apart. The only reason we say 1.6 km (1 mile) apart is because National Forests are located in sections which are laid out on 1.6-km (1-mile) centers. So all the property corners of National Forests, most of them, are on 1.6-km (1-mile) centers. The transmitter system is set up on a property corner which is not visible from this point. The whole purpose is nonintervisible corners. The laser beam is oriented vertically using bubble levels located on the instrument. The levels are good to about 5 arc-seconds. The laser beam is pulsed vertically. It is not a continuous beam as shown in figure 3. It is pulsed approximately once every 5 seconds. The receiver system located here elevates its elevation axis until it clears

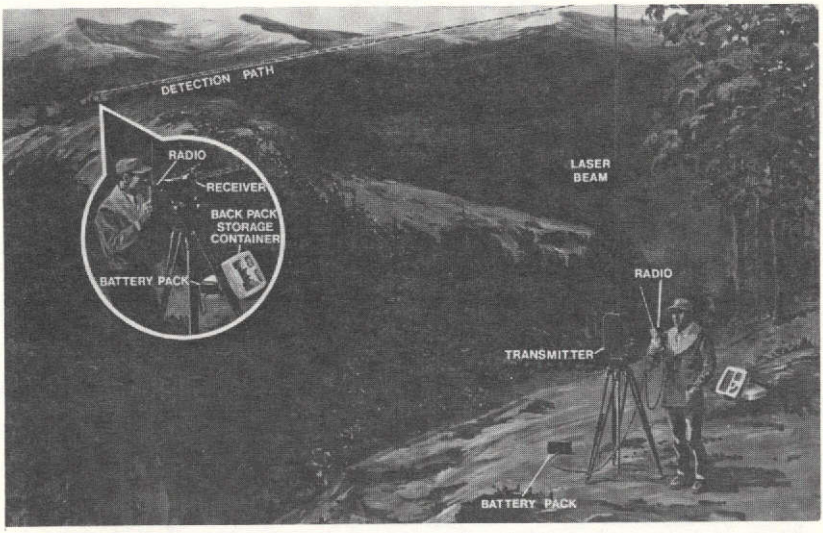


Figure 3. Artist's concept of field operation of long-range laser traversing system.

any intervening obstacles, and searches in azimuths until it locates the pulsed vertical beam. It is essentially searching for scattered optical radiation in the earth's atmosphere, both rayleigh and mie. Once detecting the beam, the receiver is electronically able to determine whether the beam falls to the left or the right of the optical line of sight of the instrument; by the operator moving the instrument in azimuth, the operator's display indicates when it is aligned with the beam, and at that point, the azimuth bearing between these points is accurately known.

The Forest Service has been operating this system for about 3 months in the field. It has been operated in Colorado, up to altitudes of 3600 m (12,000 ft) and has met all its performance specs. Its performance specs are to determine the azimuth bearing between two points to an accuracy of ±0.3 m (1 ft) in 1.6 km (1 mile). All field tests up to this point have demonstrated the ability to consistently determine the azimuth bearing to an accuracy of ±5 cm (2 in.) in 1.6 km (1 mile). So we have exceeded the spec in all phases of the operation.

Currently, the Forest Service is contractually buying seven to ten more systems to put in the field with an expected buy of one hundred units. We have also had requests from all over the world for information on availability and price of the instrument. So it has turned out to be quite a popular system.

OVERVIEW

Robert Coates

As Les indicated, the previous sessions have dealt with accomplishments related to spacecraft, both the experiments and the technology of the spacecraft. Now we are going to shift over to the ground end of the act, and go into the mission operations and data systems. These systems provide the means to operate the spacecraft to get the results in response to the experimenter's desires.

Figure 1 depicts the three basic functions we are dealing with: mission operations, orbit attitude, and data processing. Of course, these are the tools, the links between the project and the spacecraft. I am not showing the network, which I consider as a transparent pipe in the system. Of course, the requirements for operations come from the projects, through the control centers in mission operations, and get transferred to the spacecraft. The data come back to mission operations in real time to provide projects with their feedback. Likewise, data go to the orbit/attitude determination people, both from telemetry and from orbit tracking. The scientific data go to data processing where it is put in a form to go to the experimenters for their detailed analysis.

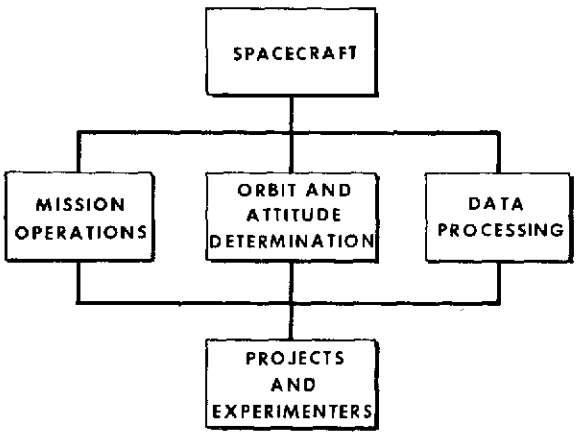


Figure 1. Mission operations and data systems are the links between the project and the spacecraft.

One of the different features of this whole area is that the three functions require rather extensive computational data handling facilities for each project. In addition, all projects have requirements for this kind of capability. So each one of these areas really constitutes rather major facilities for these data handling functions.

The main goal of the operations in this area, of course, is to be very efficient and responsive to the needs of these projects. Because it is such a large, costly facility, we have the goal to make it the most cost effective operation possible, in order to best utilize this facility for high quality project support at low cost.

In mission operations, we are pushing the multiproject use of our control centers in order to make maximum use of our facilities, while at the same time, developing new capabilities that are needed to meet unique project requirements which always exist. The first two papers that we will have in this session are samples of these two aspects of mission operations technology.

In orbit determination, cost effective operations, accuracy, and efficiency stand out as our goals. The requirements trend toward greater accuracy, and we strive to achieve this greater accuracy with more cost effective operations. We have four papers which touch on the technology for greater orbit accuracy, more cost effective operations, more rapid orbit determination, and more efficient orbit prediction.

In attitude determination, we are working toward making multiproject attitude systems; this is one area where we expect to have great profits in the future. We originally had planned to have a paper on the attitude determination system that is in use for the AE satellite, but because of the launch last weekend, the speaker was pretty busy. So, we had to remove his paper from the program and let him work on the AE. We will have to pick that up next year.

Data processing is predominantly a high data volume operation. Thus the efficiency and high quality of the operation are very significant factors. The last two papers in this session will show an aspect of improving the efficiency of our big computer operation, and will show some techniques being applied to image processing to achieve a high quality.

SOFTWARE OPERATING SYSTEMS FOR THE ATS-F SATELLITE TESTING AND GROUND SUPPORT

Jerold Hahn

We have a control center called the Multi-Satellite Operations Control Center that was established to control satellites whose life or operational activity is insufficient to require an exclusive control center. The control center includes three XDS-930 computer systems that are used for sending commands and processing telemetry data.

The key to the operation of the software in these three computer systems is an operating system that was developed in 1966 for the multi-satellite control center. This operating system was recently rewritten for the ATS-F PDP-11 computers and is the basis for the common software used to test the satellite before launch, monitor the satellite from the control center and the ground stations, and control it after launch.

The original operating system was developed to make maximum utilization of the company standard software to minimize cost. It was to be applications independent. Up until now we have had about 15 separate satellites handled by the Multi-Satellite Operations Control Center and, in order to avoid changing the operating system from satellite to satellite, we have built a system to be completely applications independent. It was to allow real time and time sharing in off-line programs all to be executed under the same operating system. The programs could either be resident or overlaid, and the overlay programs, provided they were introduced into core simultaneously, could be executed simultaneously. The system was brought together by a systems control that utilized prompting, which put the burden of the operation on the computer rather than the computer operator. There was a program rule book written that would allow the programs to be written in such a way that they would dovetail neatly into the operating system without interference among themselves.

Development of the software included first taking the problem and breaking it into separate components (figure 1). These included resident and nonresident programs, and specifications were written for each.

Then they were flow-charted, programmed, debugged, and added to a single job tape. This job tape was then selectively loaded through job control cards into either the disk for the class 2 nonresident overlaid programs, or into the core for the resident programs. In core we had common system data, resident programs, the operating system, and enough space to execute up to six nonresident programs simultaneously.

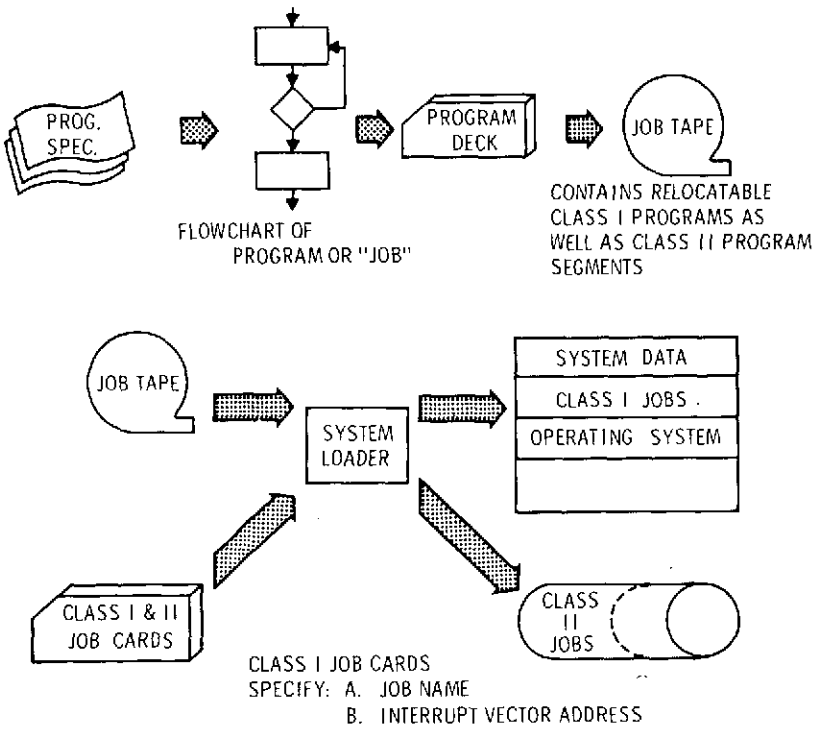


Figure 1. Integrated application software development.

Figure 2 gives an overview of the cost involved in the ATS-F integrated software development. The original multisatellite operating system cost approximately $200,000. We rewrote it for ATS-F, for the PDP-11 computers, for $100,000, and added to it $900,000 worth of software that was used to test the spacecraft and could also be used in the control center and in the ground stations. There was roughly $100,000 worth of unique test software that could not be used elsewhere. At the first level, the spacecraft testing level, we have a cost to the government of $1.1 million, but a value, since there was a $100,000 benefit here, of $1.2 million. A similar sort of thing

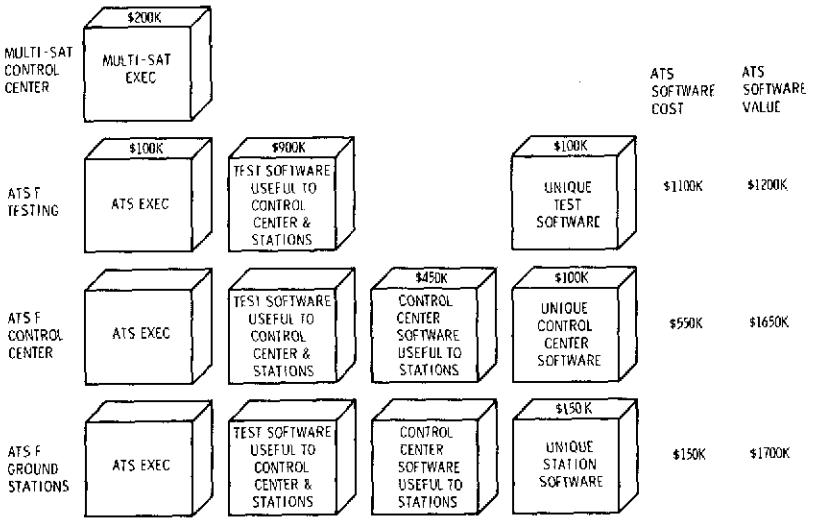


Figure 2. Cost of ATS-F integrated software.

exists in the control center. The exec is used without change. The integrated software can be used. Software worth $450,000 has to be added to the system primarily for real-time attitude computations, and some unique software is needed because of the unique data links and displays involved. The ground stations require $150,000 to be added.

To summarize, a common operating system can actually allow software to be developed for both testing the satellite, utilization in the ground stations, and in the flight control center. There is also a cost savings in software development. Finally, and I think possibly the most important aspect of this, the flight operations personnel gain experience with the common software long before launch.

INVERSION OF THE RAE-1 SPACECRAFT IN ORBIT

Harvey Walden

After more than 4 years of productive radio astronomy investigations from a gravity-gradient stabilized earth orbit, on October 31, 1972, the Radio Astronomy Explorer (RAE-1) spacecraft was rotated 180° about its pitch axis. The RAE-1 spacecraft possesses four highly flexible tubular antenna booms extending radially outward from a central spacecraft hub in a double-vee configuration measuring 460 m (1500 ft) from tip to tip. The ambitious inversion operation was conducted in order to provide an opportunity for use of the original earth-directed, 229-m vee-antenna booms in viewing radio emissions from celestial sources. A complementary objective of the inversion maneuver was to gain additional dynamical information on the stabilization and control of large gravity-gradient spacecraft arrays. The enhanced understanding of passive, three-axis, gravity-stabilized satellites that resulted proved to be valuable in contingency planning for the lunar mission of RAE-2, which was subsequently launched into orbit about the moon in June 1973.

The inversion of the RAE-1 spacecraft was a natural sequel to a series of six gravity-gradient dynamics experiments which were successfully performed in earth orbit during 1970 and 1971, utilizing the same RAE-1 satellite (refer to the published proceedings of previous symposiums in this annual series held at the Goddard Space Flight Center). Several of the original dynamics experiments involved partial single or double boom retractions and redeployments, but the objective at that time was only to disturb the quiescent dynamical state and generate large angular attitude motions temporarily and ultimately to restabilize about the original local orientation. The techniques successfully perfected during the performance of the series of dynamics experiments, including the prediction of spacecraft dynamical behavior in orbit with the use of computer model simulations, the generation of desired angular motions through active boom operations, and the control of the spacecraft attitude perturbations through use of a passive on-board libration damper mechanism, were all utilized during the spacecraft inversion maneuver. In these respects, the inversion operation could well be characterized as being deliberate, precise, and well calculated.

The RAE-1 spacecraft is in a gravity-stabilized orbit about the earth with a period of 225 minutes, such that its pitch rotational rate is equal to its orbital rate, and the result is that the same vee-antenna remains directed toward the earth. That is, the spacecraft array rotates once about its central hub each orbital period as viewed from a stationary point outside the earth orbit system. The method used for inversion, or interchange of the upper and lower vee-antennas, is dependent upon the physical principle of conservation of angular momentum. By retracting all four vee-antenna booms simultaneously to a sufficiently short length, the in-orbit rotational pitch rate of the spacecraft was increased from its usual orbital rate to a rate at which the spacecraft completed a 180° rotation in less than one-half an orbital period (see figure 1). The partial retracted length of the antenna booms, chosen on the basis of the results of numerous pre-inversion computer simulations of the spacecraft dynamics, was 160 m (525 ft), and this parametric value resulted in a total inversion time of approximately 100 minutes. After the boom retractions were completed, a carefully timed rotation coast interval of approximately 70 minutes followed, during which period the spacecraft pitch attitude was monitored in a real-time procedure. In order to reduce the satellite rotational rate to match the orbital rate so that the spacecraft would remain in an inverted position rather than continue to spin about the pitch axis, the four antenna booms were then redeployed to full 229-m lengths just prior to completion of the 180° rotation. This method of timing boom deployments such that spacecraft attitude motions are reduced to virtually negligible oscillations is known as a deadbeat technique. The spacecraft is thus captured in the inverted position (in figure 1, note the use of open and filled circles at the boom tips to indicate relative orientation), and the gravity-gradient mode of stabilization is assumed once more.

The pitch attitude of the RAE-1 spacecraft central hub during the inversion operation is displayed in figure 2, in terms of measured real-time pitch angle data. The original pitch equilibrium position is at 0° prior to the inversion. The on-board spacecraft libration damper was initially deactivated, or clamped, prior to any boom retractions, to assist in the pitch rotation. The four simultaneous boom retractions required an interval of 21 minutes to achieve the partial lengths of 160 m. From the start of the retractions, a nearly linear growth in the pitch angle, from the original 0° equilibrium position to the final 180° inverted equilibrium position, is observed. The time for initiating the boom redeployments was adjusted slightly from the predictions of the computer model simulations on the basis of the real-time pitch angle data. Initiation of deployments occurred at a pitch angle of approximately 173°, after a coasting interval of 71 minutes. Redeployments to full lengths required an additional 14

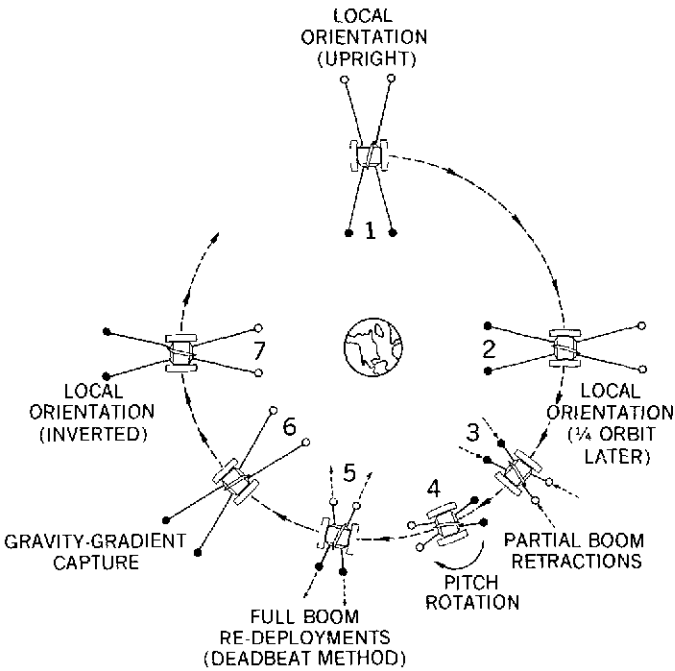


Figure 1. Method used for inversion of the RAE-1 spacecraft on October 31, 1972.

minutes and, almost immediately after full boom lengths were achieved, the on-board damper mechanism was activated to assist in the attenuation of residual attitude oscillations.

The spacecraft attitude librations about all three axes during and after the inversion maneuver are shown in figure 3. Note that the ordinate axis used to display the pitch angular data is nonlinear; the scale used during the boom retraction and coasting intervals is compressed in order to accommodate the full range of data points shown previously in figure 2. The post-inversion pitch oscillations attain maximum amplitudes of approximately $\pm 8^\circ$ about the inversion equilibrium position. The motions about the roll axis are fairly minor (generally within $\pm 5^\circ$ of the 0° equilibrium position), but there is some coupling of the pitch motions into the yaw axis motions, resulting in yaw maximum excursions of approximately $\pm 18^\circ$ about the negatively biased equilibrium position. The negative bias for the yaw equilibrium is due to the fact that the on-board libration damper is skewed out of the plane of the vee-antenna booms, and the pitch-yaw coupling results from the decreased effectiveness of the libration damper mechanism about the yaw axis, as was anticipated.

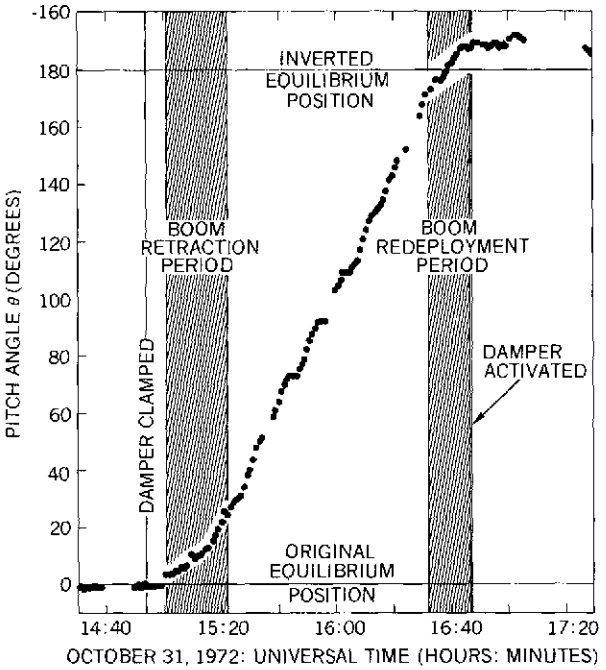


Figure 2. RAE-1 spacecraft central hub pitch attitude during inversion operation.

The three-axis attitude librations are continued in time without lapse in figure 4. It is evident that oscillatory motions about all three axes rapidly attenuated to near steady-state condition values, which are characterized by nonperiodic, fairly random scattered data points near the equilibrium positions. It is significant that the maximum amplitude oscillations which occurred about all three attitude axes were observed immediately following the boom redeployments. The motions thereafter were stable and well bounded about the new local inverted orientation. It may fairly be stated that gravity-gradient capture about the inverted position was reestablished immediately following the attainment of full boom lengths by the deadbeat technique.

The results of the inversion of the RAE-1 spacecraft may be summarized as follows. First, the entire inversion operation was completed as planned, with the re-attainment of steady-state quiescent dynamical conditions in only several hours following the inversion. The mechanical and electrical systems on board the spacecraft performed quite satisfactorily, particularly in view of the fact that certain of these systems had not been exercised in more than 4 years of exposure to the hostile environment of space.

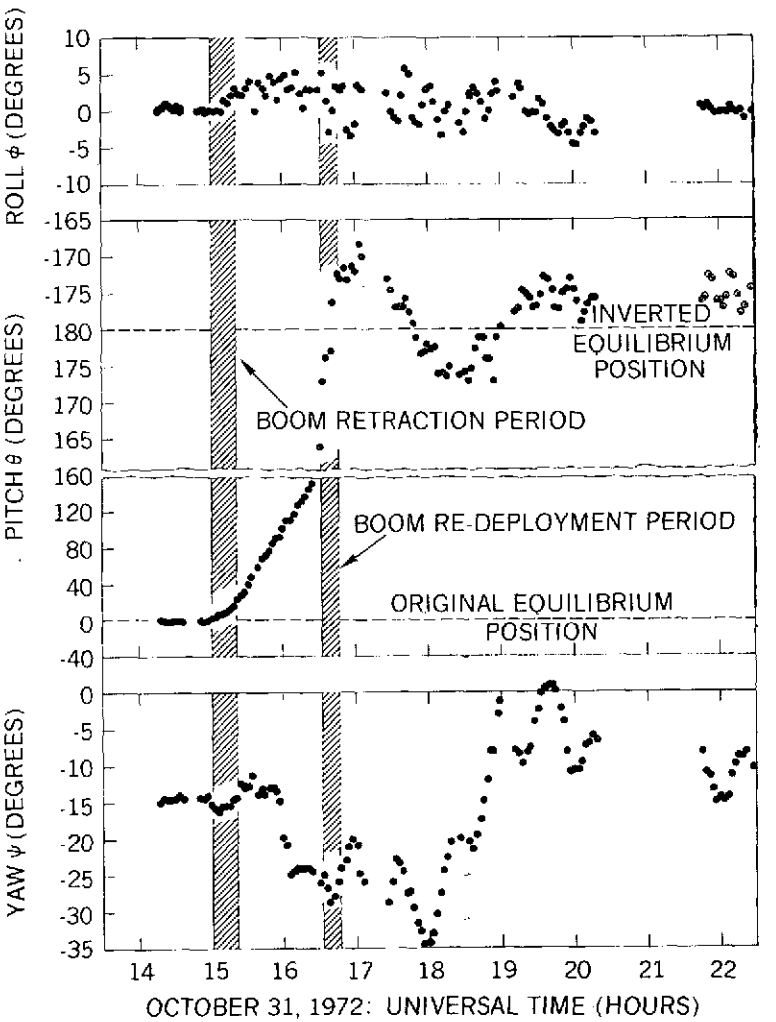


Figure 3. RAE-1 spacecraft central hub attitude librations during and following inversion operation.

Secondly, the feasibility of gravity-gradient spacecraft control by active ground command has been amply demonstrated by the inversion operation. The predictive capabilities of the computer model of the spacecraft dynamics, in conjunction with analytical techniques, which were developed prior to the RAE-1 launch in 1968 and have been improved and refined

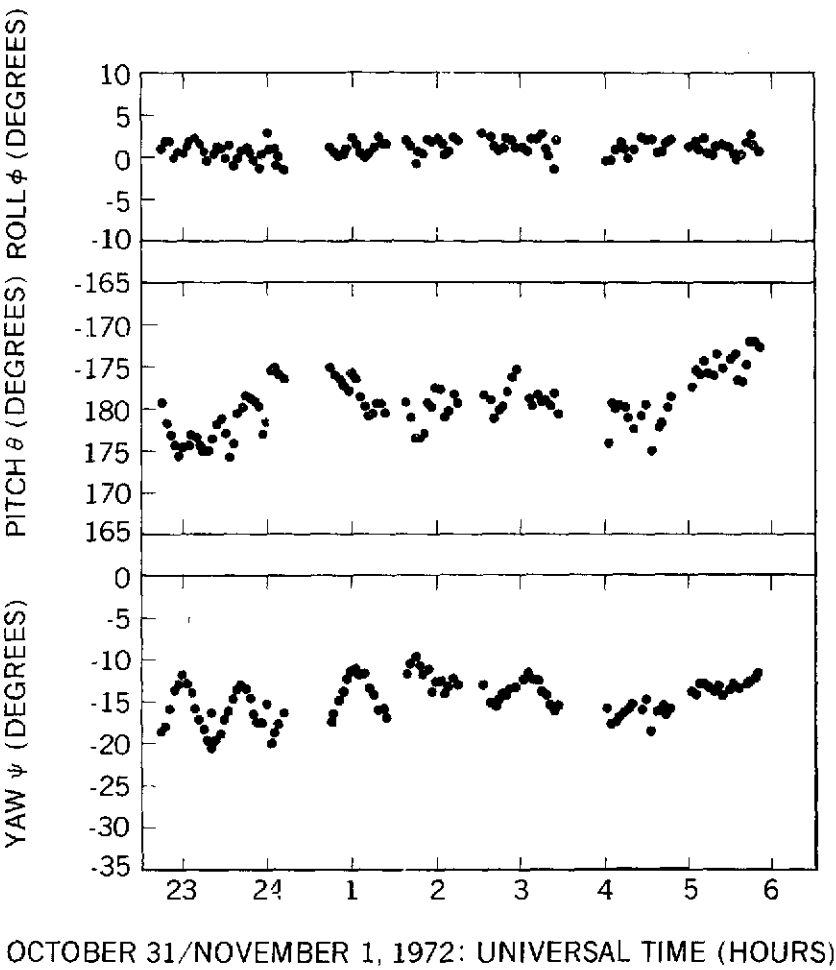


Figure 4. RAE-1 spacecraft central hub attitude librations following inversion operation.

since that time, were further validated and confirmed. The method of spacecraft inversion was shown to be possible if necessary in the event of a contingency in lunar orbit with the RAE-2 spacecraft. Finally, the useful mission lifetime of the RAE-1 spacecraft was extended several months beyond its original termination date as a result of the successful inversion in order to use the antenna receiver, initially relegated to monitoring the earth's background radio noise, for verification of radio astronomical observations from the cosmos.

ALSEP DIFFERENTIAL DOPPLER TRACKING

Joseph Ryan

For the past year and a half the Goddard Space Flight Center in conjunction with the Massachusetts Institute of Technology has been carrying out an effort to test the feasibility and gain operational experience with a new tracking technique called differential Doppler tracking. As shown in figure 1, the concept is really quite simple. A ground station receives the Doppler-shifted carriers from two transmitters in space. It then mixes the two signals, forming a differential frequency, and counts cycles of this differential frequency to generate a differential Doppler measurement. This measurement is sensitive to the relative motion between the two transmitters in space along the line to the ground station. The great power of this measurement derives from the fact that some of the most serious sources of tracking data error are largely eliminated when the differential measurement is formed. Specifically, tropospheric and ionospheric refraction effects and ground station frequency instabilities are almost entirely cancelled.

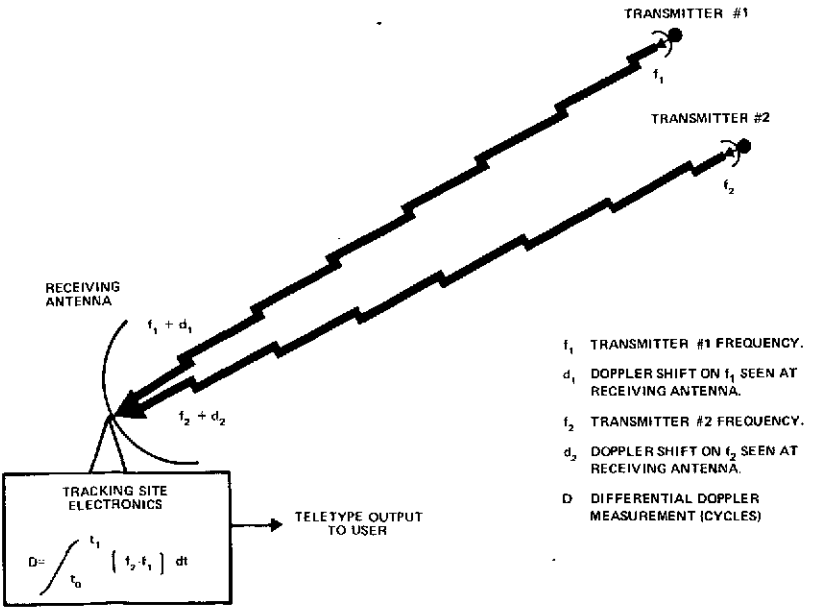


Figure 1. Differential Doppler tracking concept.

This technique was first used operationally at the end of the Apollo Program when it was used to track the path of the lunar rover relative to the landing module. In this application, motions as small as a few meters on the lunar surface were correctly observed. The present ALSEP experiment is a direct result of this very successful Apollo exercise.

During Apollo and in the beginning with our ALSEP tracking we did not have the equipment to actually form the differential Doppler measurement in the tracking site hardware. Because of this we had to acquire two streams of normal Doppler tracking data at each site using the conventional unified S-band (USB) dual Doppler mode. We then differenced the Doppler measurements taken simultaneously at each site in the central computing facility here at Goddard to form a differential measurement. Figure 2 is a plot of phase residuals as a function of time for one pass of data taken in this mode. The key things to be noted here are that the noise on the data is approximately one-half cycle of S-band phase, and that there is a large phase drift amounting to about 8 cycles over the 7 hours of this pass. This is a serious degradation of the data, however, it is not a result of any problem with the tracking technique per se, but really resulted from the way we had taken the data using the dual Doppler mode. When we realized this, MIT proceeded to design and build a differential Doppler receiver which could be installed at the sites and that would eliminate the need for using the dual Doppler mode. It brought both carrier frequencies down through one receiver and local oscillator chain, and actually formed the differential measurement onsite.

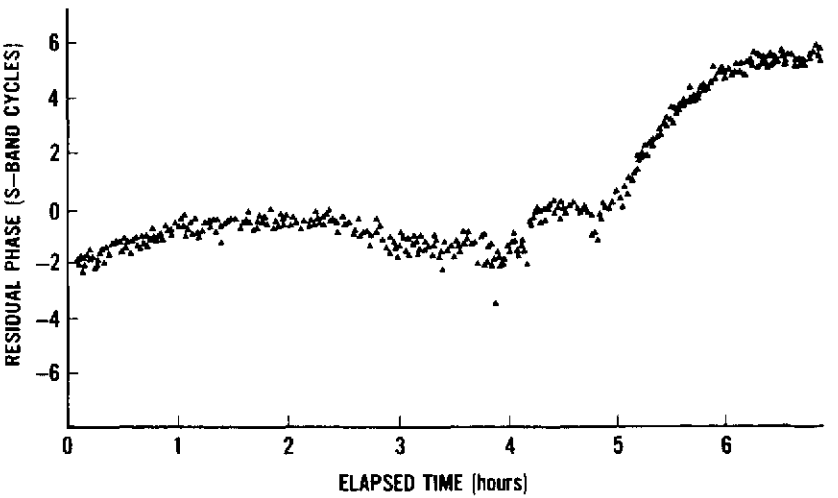


Figure 2. Post-fit residuals, dual doppler mode; differential measurements constructed in MIT computer.

Figure 3 is a plot of the phase residuals taken with our new receiver. Note that the noise on the data has been reduced by an order of magnitude—to about 0.05 cycles—and that the phase drift, which was so prominent in figure 2, has been almost entirely removed.

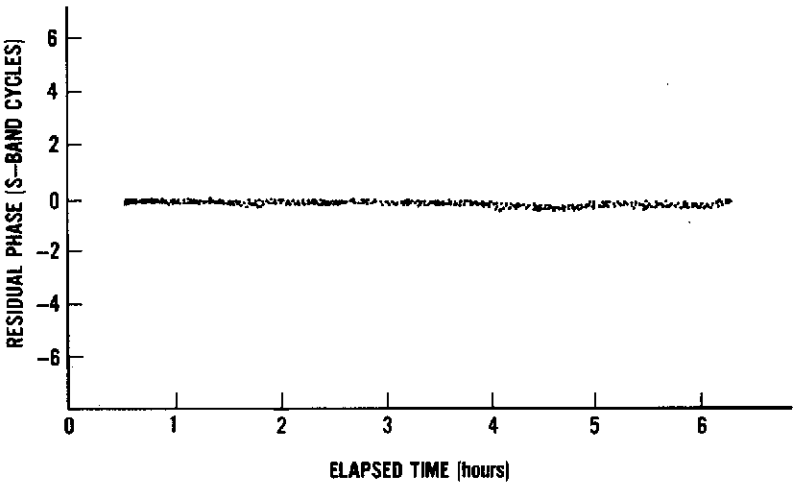


Figure 3 Post-fit residuals, differential doppler receiver mode; differential measurements made in site electronics.

Last spring, after we had convinced ourselves that our new hardware was working properly, we proceeded to outfit six unified S-band sites with differential Doppler receivers (DDRs) and began a rather intensive tracking program. To date, we have acquired about 1500 hours of station contact time, and as a matter of fact we have a tracking exercise scheduled today.

What then is the value of acquiring this type of data? ALSEP differential Doppler data is sensitive to lunar orbit parameters, lunar libration parameters, and the selenodetic position parameters of the ALSEP packages. In fact, the lunar ranging experiment (LURE) and the ALSEP-DDR experiment are the only two experiments presently being carried out within NASA which have the potential to significantly improve these parameters. Table 1 gives uncertainties in key lunar parameters. One column contains the uncertainties based on 16 months of DDR data; another column contains uncertainties based on 3 years of DDR data. The 16 months of data contain the 9 months of data already in hand plus 7 months of data which we foresee no difficulty in obtaining next year. The 3 years of data represent our current projected lifetime of the ALSEP-DDR experiment. The

most significant contribution of this program will be the improvement in the values of the lunar libration parameters. The uncertainty in β and γ resulting from traditional optical observations is in the range of three parts in 10^{-6}; the DDR observations have the potential of determining these parameters with an uncertainty two or three orders of magnitude smaller. The LURE data will also be determining β and γ with significantly smaller uncertainty in the next few years; DDR data will be the only check on the LURE results, and may also determine some lunar parameters with greater accuracy than can the LURE data.

So, to date, we think that we have shown quite conclusively that our concept was correct, that our hardware works; moreover, we feel we have acquired a very substantial data set containing valuable scientific information.

Table 1
Standard Deviations in Recovered Lunar Parameters
(Error Analysis Predictions)

Moon's Orbit	16 Months of DDR Data	3 Years of DDR Data
Inclination (Arc-Second)	0.04	0.025
Long. of Asc. Node (Arc-Second)	0.08	0.06
Long. of Perigee (Arc-Second)	0.08	0.06
Moon's Moment of Inertia		
β	4.7×10^{-9}	2.9×10^{-9}
γ	1.9×10^{-7}	1.0×10^{-7}
Apollo 15-Apollo 14 Separation Vector (cm)	30	30

MINITRACK CALIBRATION USING SATELLITE DATA

James Jackson

In this discussion of minitrack calibration, cycles or fractions of a cycle will be used as the unit of measurement. For the sake of this discussion, one cycle can be considered to be 1° in angular measurement.

Changes in the zero set bias over several months between the aircraft calibrations are shown in table 1. These changes have been as much as 0.432 cycles in this east-west fine polar data from Alaska. Drifts or changes in zero set biases can cause serious errors in orbit determination and in defining the position of a satellite. These drifts must be determined, then applied to the data in preprocessing, so as to minimize their effects. A reliable calibration technique to monitor zero set changes and to recommend new estimates is the problem for consideration.

Table 1
Changes in Zero Set between Aircraft Calibrations

Station	Change in Zero Set Bias (Cycles)		Time Differential (Months)
	Polar East West Fine	Polar North South Fine	
Alaska	- 0.432	0.037	8
Johannesburg	0.041	- 0.078	13
Quito	- 0.013	0.009	10
Orroral	not available	not available	not available
Santiago	0.002	- 0.001	15
Tananarive	0.030	0.006	13
Winkfield	- 0.005	0.000	5

Figure 1 illustrates the zero set bias determination process. The satellite calibration method is statistical in nature as several parameters are derived from a least-squares fit of the data. The best available estimates of the monthly zero set biases are used as an input to the orbit determination program. Any bias in a pass of satellite data is considered to be a change in the zero set bias whenever that bias is not pass dependent or satellite dependent. The monthly average of the monthly zero set for any antenna pair is the average of these monthly biases. The standard deviation and the noise of these pass biases are also computed to provide estimates of the confidence that can be placed in the zero sets.

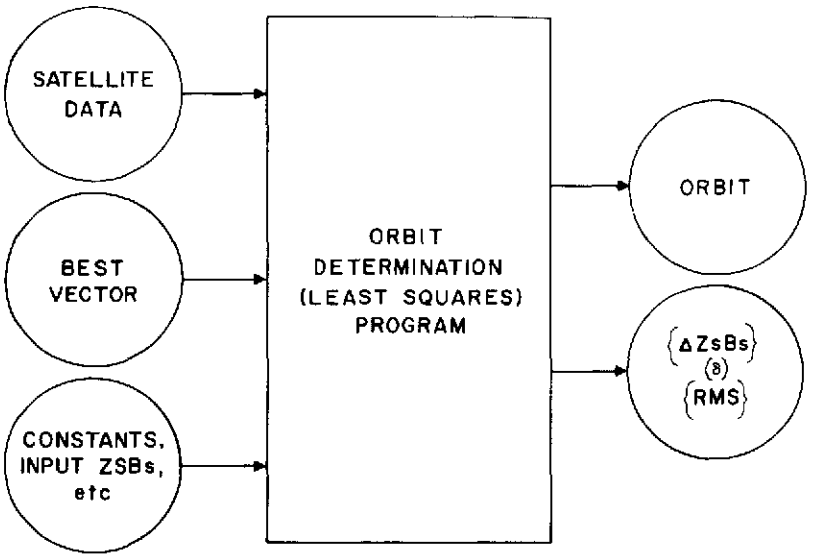


Figure 1. Satellite calibration zero set bias (ZSB) determination.

The end product is a monthly collection of zero set biases for each antenna pair, each mode, and each station in the minitrack network. Figure 2 is a graph of some monthly fine zero set biases for Alaska between November 1971 and October 1973. The horizontal bar represents the zero set, the vertical bar is the associated uncertainty. Note the change in the zero set bias from month to month.

Aircraft calibrations have been conducted once or twice a year to determine zero set biases and coefficients. The last aircraft test was scheduled for Orroral, Australia, in December 1973. I would like to interject here that this test was completed within the past few days. After the results are derived from this test, all subsequent zero set biases will be determined from satellite data.

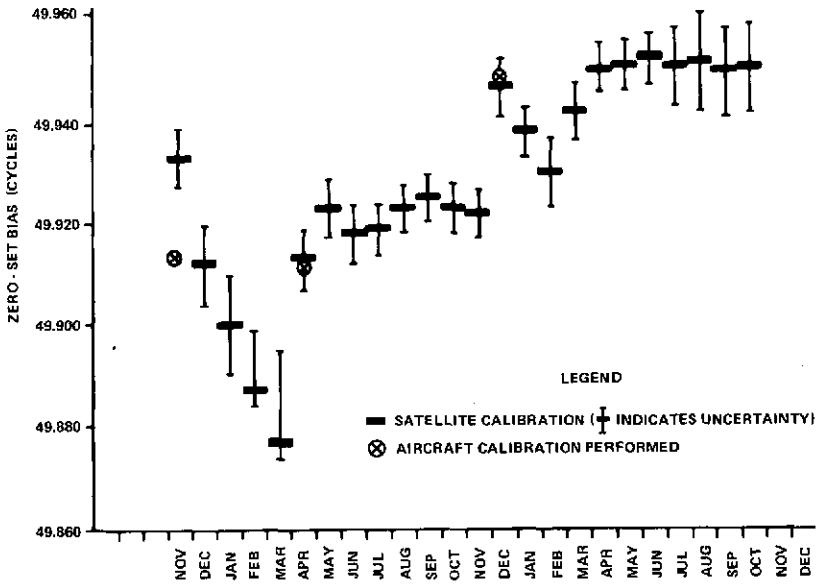


Figure 2. Minitrack station performance in the Alaska polar mode between November 1971 and October 1973.

Figure 2 also shows zero set biases from aircraft calibration as represented by the circle with the "x." The accuracy of the aircraft zero set bias is on the order of three-thousandths of a cycle, whereas the accuracy of the monthly satellite zero set bias is on the order of five-thousandths of a cycle.

The aircraft zero set is the best estimate of the zero set immediately after aircraft calibration. However, due to the dynamic behavior of the zero sets, the satellite calibration is a better estimate about 90 percent of the time. So for aircraft results we have short-term accuracy 1 or 2 months of the year, but with satellite zero sets we have greater knowledge and longer term accuracy.

The zero set biases from the two independent methods are quite similar. Table 2 shows that these numbers compare to within 0.007 cycles of one another when referenced to the month of the aircraft calibration. This statement is true for this fine data. For ambiguity data, the zero set biases from the two independent methods would compare to within 0.040 cycles.

There are several advantages to satellite calibration. Generally, satellite calibration is a byproduct of orbit determination. Frequent changes can be monitored and taken into account on a monthly basis, thus providing greater accuracy and also longer term accuracy. However, the most significant advantage of satellite calibration is that it is adequate to determine the zero sets of the minitrack network without the use of an instrumented aircraft and its support crew. Hence it is cost effective since it eliminates these costs. Finally, satellite calibration is an efficient method of zero-setting the minitrack systems in the network.

One disadvantage of satellite calibration is that the source of data is restricted to available satellite crossings. Some special arrangements are being made to provide full beam passes of tracking data to define antenna field distortion effects.

In conclusion, the minitrack network can be calibrated using satellite data with greater awareness of the dynamic behavior of the zero set biases at any given time. Satellite calibration is efficient and cost effective. Satellite calibration meets the requirements for zero setting. And, finally, satellite calibration will be used exclusively to determine the zero sets of the minitrack network beginning in January 1974.

Table 2
Comparison of Satellite-Calibrated and Aircraft-Calibrated Fine Zero-Set Biases for May 1973

Station	Polar Difference in Zero Set Bias Satellite Data–Aircraft Data (Cycles)		Date Calibrated (Aircraft)
	East-West Fine	North-South Fine	
Alaska	0.004	0.002	December 1972
Johannesburg	-0.005	0.003	June 1972
Quito	0	0.001	March 1973
Orroral	0	0.002	September 1972
Santiago	-0.003	0.007	May 1972
Tananarive	0.004	-0.002	June 1972
Winkfield	-0.006	-0.005	October 1972

ANGLES-ONLY ORBIT EXTRACTION

Donald Cowan

The block diagram in figure 1 shows the steps involved in arriving at a predicted orbit and it shows how the orbit extraction program fits in. In the boxed area are shown several different ways of arriving at the initial estimate of the orbit. I might emphasize here that this initial orbit must be very, very, precise, and this requirement has been brought out most clearly in situations where a trial and error method has had to be used and has taken many hours of effort and almost continual use of the computer in order to arrive at a satisfactory initial estimate of the orbit.

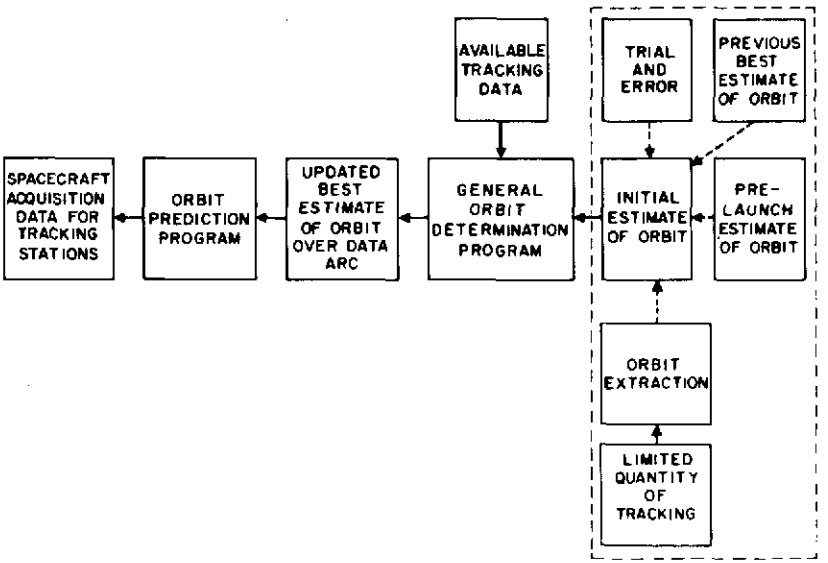


Figure 1. Steps involved in arriving at a predicted orbit.

A situation where the orbit extraction program is particularly important is a non-nominal launch situation. In that case, we have, of course, no previous best estimate of the orbit, and the prelaunch estimate of the orbit is not satisfactory to use as an initial estimate of the orbit. The trial and error method, in this instance, is too cumbersome and too unpredictable

to use to arrive at a quick and sufficiently accurate solution for an initial estimate of the orbit, which is, of course, advantageous in a non-nominal launch situation.

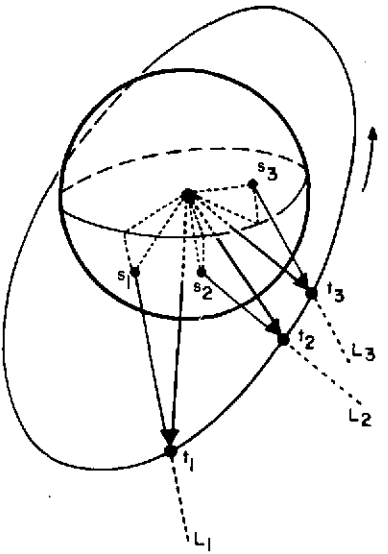
Most orbit extraction programs which are available right now require the use of range data in order to operate at all. For tracking systems where there is no range data–Minitrack being the prime example–those types of orbit extraction programs would not work at all. Also, if the range data is invalid, those types of orbit extractions will not work.

Figure 2 shows the basics of orbit geometry for the angles-only orbit extraction showing the positions of stations at three times and the positions of the spacecraft at three times. What the orbit extraction program does is to estimate ranges along lines L_1 and L_2. Once it has these estimates, the range along line L_3 is completely determined. But there is a drawback in just choosing ranges arbitrarily–that is, the three points that we end up with may not even define a valid orbit. So what the orbit extraction program does is to go through a controlled, systematic adjustment of the range estimates along lines L_1 and L_2 until it comes up with a valid orbit which is also satisfactory to start a differential correction process.

In this differential correction process a series of orbits is generated, each of which passes through lines L_1, L_2, and L_3; the computed time it takes the spacecraft to go from line L_1 to L_2 and from line L_2 to L_3 is made to agree with the actual observed time differences, T_2 minus T_1 and T_3 minus T_2, respectively.

Table 1 shows an actual launch situation in which the orbit extraction program was used. The launch was a Minitrack launch occurring in August 1972, and as you can see from columns 1 and 4, the orbit was very definitely, very significantly, non-nominal. Focusing now on the single-pass solution, we can see that the results here are really disastrous. In fact, this is one situation which shows why we do not send out site acquisition data on satellites based on only one pass of Minitrack data. The solution here is based on three points which are taken about a minute apart, and the fact that the solution is this poor is really not due to the orbit extraction technique itself but rather to small but normal imperfections in the data. When you have a very limited geometry, as in this case, the solution can vary widely with even normal imperfections in the data.

This single-pass solution can be obtained from another angles-only orbit extraction method called the Gauss method. The Gauss method has been used to determine things like planetary orbits, but it requires a very small arc of data in order to work at all, and it is completely unsatisfactory in finding quickly satisfactory solutions for earth-orbiting satellites.



t_1, t_2, t_3 = SPACECRAFT POSITION AT 3 DIFFERENT OBSERVATION TIMES.
s_1, s_2, s_3 = STATIONS (NOT NECESSARILY DISTINCT) TRACKING SPACECRAFT.

Figure 2. Orbit geometry for angles-only orbit extraction.

Looking now at the two-pass solution which was obtained using the new method, it can be seen that the first point was taken from the first pass after launch. The next two points were taken roughly two and two-thirds hours later, which is equivalent to about one revolution later, at the same station. The second and third points were about 1 minute apart. The orbit which the orbit prediction program arrived at was satisfactory for use as an initial estimate for the orbit determination program.

This program has been used in actual launch situations with great success and it is currently being implemented into the Goddard Real-Time System. I might mention in closing that, for this actual launch in August 1972, the orbit extraction program was not available, and even 6 or 7 hours after launch a satisfactory initial orbit had not been determined. In contrast, with the use of this program, only a few minutes of setup time are required, and it takes only 2 or 3 seconds of computer time in order for the program to operate to completion.

Table 1
Angles-Only Orbit Extraction Results
(REXS Launch Phase-19 August 1972)

	Prelaunch Estimate of Orbit	Single Pass Solution*	Two-Pass Solution†	Actual Orbit
Apogee (Km)	2605	967	6552	6566
Perigee (Km)	708	-6284	235	243
Period (Min)	119	37.6	160.2	160.5
Inclination	31°2	32°5	31°0	31°0

*Obtainable from Gauss orbit extraction method or from expanded orbit extraction method.

†Obtainable from expanded orbit extraction method only.

THE AVERAGE DYNAMIC TECHNIQUE AS APPLIED TO ORBIT DETERMINATION

Arthur Fuchs

Orbit determination may be said to consist of three basic processes: namely, the data gathering process, whereby tracking data is collected from the worldwide tracking network; the trajectory generation process, whereby a trajectory is generated according to some physical model of the universe and by solving some mathematical equations; and finally, a filtering process, whereby the tracking data is matched to the trajectory using some filter whether it be a least-squares batched filter or a sequential filter. I am going to concentrate today on the second of these processes—the trajectory generation process.

Classical methods of trajectory generation can be classified into two basic categories: the general perturbations technique and the special perturbations technique. An example of the general perturbations technique, which is heavily used at Goddard, is the Brouwer Theory. It is an analytical technique and it is very efficient for long-term predictions. The basic disadvantage is that it is restricted in the force model that it can accommodate.

The special perturbations procedures are numerical in nature (Cowell Equations). They can accommodate any force model as long as the accelerations can be written down, and they are extremely accurate. The basic disadvantage is that they are comparatively inefficient in that since they numerically integrate the actual motion of the satellite, small step sizes must be used, and it takes a relatively long time on the computer to integrate reasonable arcs of data. It would be advantageous if we could find a technique which had the advantages of each of the previous two methods while minimizing the penalty that one would pay in accuracy. An attempt at such a technique is classified under the semianalytical technique or the average dynamic method.

An example of the average dynamic technique may be seen by looking at the typical element history shown in figure 1. We are representing the motion of a typical element, a, by its three basic components: its secular component, which may be increasing or decreasing (that is, the long period motion of the satellite whose frequency is greater than the orbital period which is shown along the abscissa of the graph in figure 1.) and the short

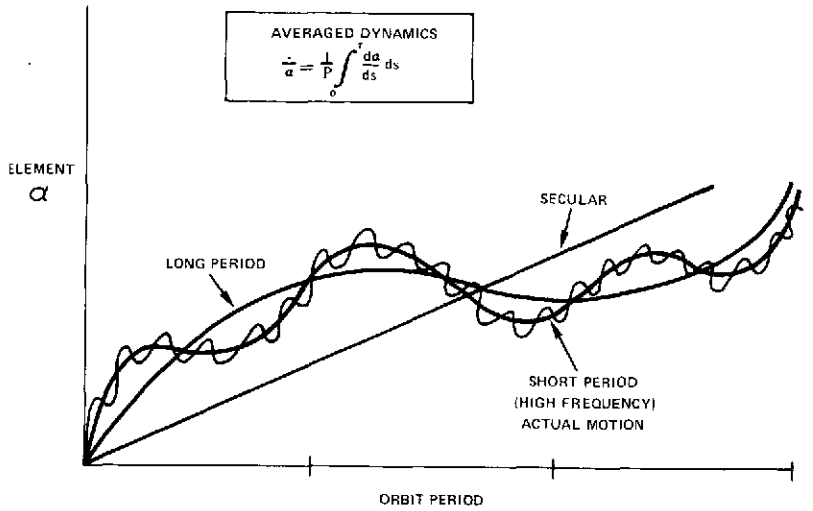


Figure 1. Typical element history.

period high frequency terms. I might point out that the actual motion of the satellite would follow this type of behavior.

Now, if we could somehow filter out the high frequency motion of the satellite, then we would be able to take much larger step sizes in our numerical integration process. This is what the average dynamic technique does. The method of filtering is shown in the box at the top of the figure. I might point out that this technique is by no means new. It has been used quite extensively for mission analysis problems for many years, and also for geodetic recovery problems where the longterm motion of the satellite orbits are of primary interest. What we are attempting here is to apply this technique directly to tracking data in the orbit determination process described previously.

Sample results of this application were gained with the ESSA-8 satellite, and are given in table 1. Here, actual tracking data was used in the orbit determination process. This particular satellite is an extremely stable one. It is approximately a 1400-km circular orbit with a 102° inclination. We utilized tracking data arcs of 2 to 4 days in the orbit determination process and then predicted these solutions forward for a period of 2 weeks. We then measured the accuracy at the end of the prediction period by doing another orbit determination.

Points to note in table 1 are first that all the techniques were extremely accurate for this satellite. Accuracies of less than 1 second were achieved,

Table 1
ESSA-8 Prediction Errors

Orbit Characteristics: 1400 km circular, 102 deg. inclination

Data arc: 2 to 4 days
Prediction period: 2 weeks

Main perturbations: J_2, J_3, J_4

Method	Along Track Error* (seconds)	Computer Time (seconds) Orbit Determination	Orbit Pred.
Cowell	.23	74	25
Averaged Dynamics	.10	21	$<<1$
Brouwer	.08	35	$<<1$

whereby the stated prediction accuracy is 10 seconds. Also, as would be expected, the Brouwer technique was extremely efficient, as measured in computer time, compared to, for example, the Cowell technique, and as was desired, the average dynamic technique was competitive with the Brouwer technique. In this particular case, it actually surpassed the technique in efficiency.

In closing, I just might reiterate that, although for this satellite a large perturbation model was not necessary because it is in a very stable orbit, the average dynamic technique can accommodate any force model.

IMAGE PROCESSING SYSTEM PERFORMANCE PREDICTION AND PRODUCT QUALITY EVALUATION TECHNIQUES

Bernard Peavey

A mathematical model has been developed to objectively describe and estimate processing errors introduced by an image processing system such as that used for the Earth Resources Technology Satellite (ERTS). This model is sufficiently general to include electronic, optical, and photographic processes, both linear and nonlinear as the case may be. It provides the capability to determine performance and operational parameter sensitivity on the basis of subsystem characteristics, and to establish operational performance tolerances in the processing chain to ensure production of high quality products. Product characteristics such as resolution, geometric, and radiometric fidelity can be estimated by a three-parameter edge target technique included in this model.

The development of this model was prompted by the complex nature of designing, implementing, and operating an image processing system required to produce extremely high quality products. Figure 1 depicts a typical image processing system (IPS) which is required to reconstruct images (or scenes) viewed by an image sensor, and display these images on film in pictorial form or on computer tape in digital form. The fidelity with which this reconstruction can be achieved depends on the particular performance characteristics of the various subsystems comprising the IPS. Thus, it is advantageous to determine the final performance and resulting product quality of a given IPS design before it is actually built and "cast in concrete." In addition, it is important to evaluate operational procedures and quality control (QC) tolerances at the various processing steps and estimate their combined effects on the output product quality prior to initiating such an operation, or in the course of system modification to meet new requirements. This can be accomplished by applying the techniques described herein.

Two basic computer programs have been developed and tested in response to the questions posed previously: image data system simulation (IDSS) shown in figure 2, and edge gradient analysis (EGA) shown in figure 3. The IDSS simulates targets representing images viewed by given sensors,

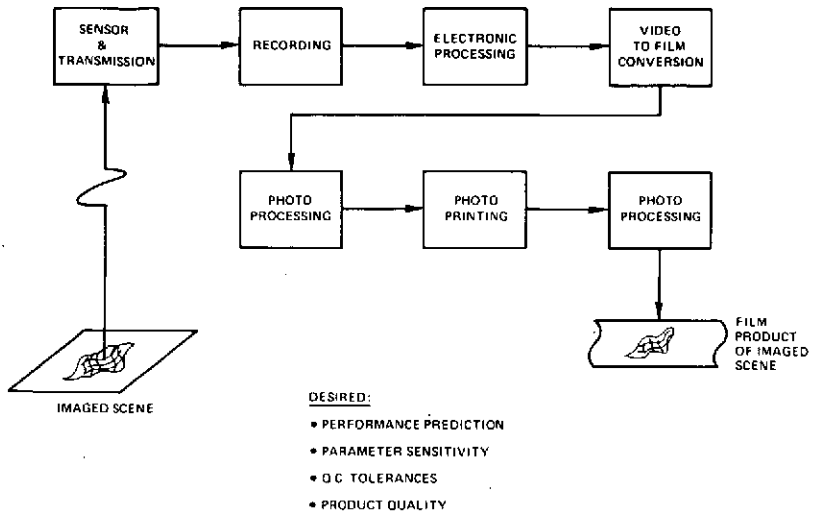


Figure 1. Image processing system.

performance characteristics of various system elements, and image sensor performance. These targets are then propagated down the chain of simulated subsystems where the target signal is convolved with the subsystem characteristics or multiplied by the gain and filter responses which may represent nonlinear processes such as photoprocessing or detector noise. As the target is propagated through the chain of elements, its radiometric and geometric fidelity (or accuracy) is degraded in accordance with the simulated subsystem characteristics. The quantitative measure of errors introduced by this process is then computed by applying the Cramer-Rao Bound which yields the variance of an unbiased estimator. Referring to the sample target in figure 2, the program computes the errors associated with the densities of regions L_1 and L_3 and location of edges in region L_2.

The Cramer-Rao Bound was found to be applicable to systems that are linear, or those that are nonlinear with a high signal-to-noise ratio. If neither condition is met, the resulting estimates will be less accurate.

The EGA program, shown in figure 3, is designed to determine system multitarget frequency (MTF) as well as system linearity. Edge targets (simulated or scanned) are converted from density to exposure domain. The program takes the derivative of the edge and then its Fourier transform. The modules of this result is the MTF shown in figure 3. The MTF curve represents the spectral response of the system. Thus, a target processed by the IDSS is subjected to EGA which produces amplitude attenuation

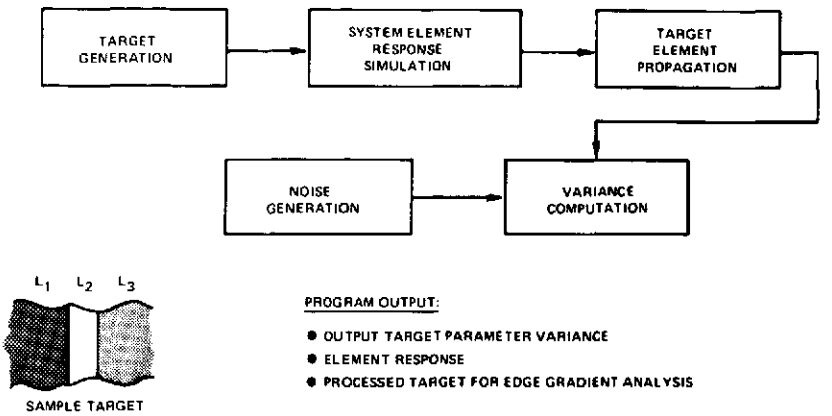


Figure 2. Simulated image processing system model.

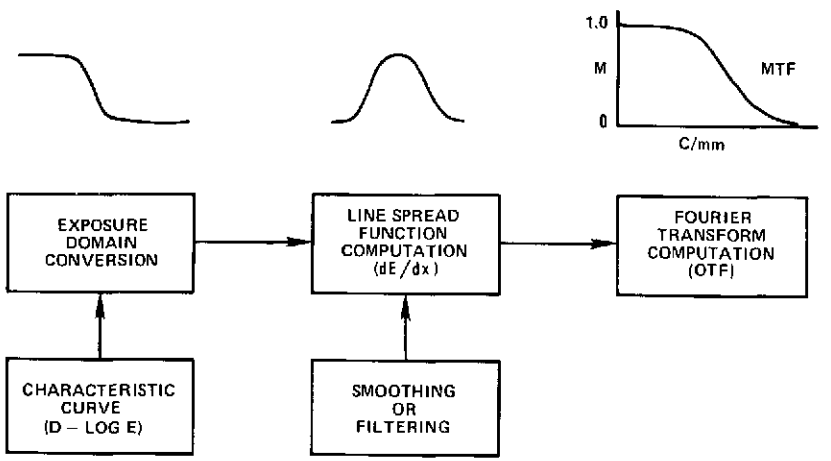


Figure 3. Edge gradient analysis program.

versus frequency. System linearity can be observed by subjecting to EGA targets of varying parameters or those processed by elements of differing characteristics. Constant MTF resulting from changing system or target conditions indicates linear system characteristics. Conversely, degree of nonlinearity can be related to the variation in MTF for given system and target situations.

Several sample results obtained by applying these techniques to ERTS IPS are illustrated in figure 4. One case considers resolution of the output product (film) after the input target (T) has been propagated through the sensor, electronic, and photographic subsystems. It is seen that the original target can barely be resolved. The other case is parameter sensitivity. It is observed that the system is less sensitive to changes in spread function of photographic printers than to the gamma of film processing. These quantitative measurements are indicative of the final product quality that would have resulted had the subsystems involved in this operation been changed and operated under the prescribed conditions.

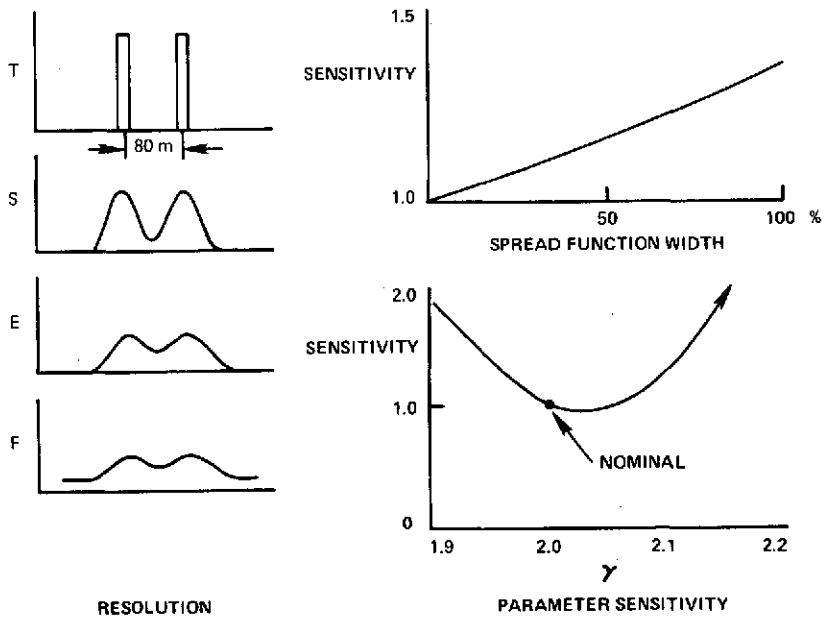


Figure 4. Performance results of the image processing system.

It should be emphasized that these techniques permit quantitative evaluation and prediction of system performance. The simulation program enables one to determine the impact of sensor and IPS design parameters on output product quality and system performance. They can be used to establish operational procedures and QC criteria for an operational facility such as the ERTS NASA Data Processing Facility.

OVERVIEW

Laverne Stelter

In the interest of saving time, I am going to accept Bob Coates' original statement that the network is a transparent pipe. So I am not going to tell you how it got that way in the past year and what we are going to do in the future to keep it that way. Rather, I am going to limit all my remarks to an introduction of the first paper.

The application of technology in networks usually ends up in hardware in one form or another in our particular application, and the source of this is two-fold. First of all, we get it from others who are developing things that are applicable to us, and by others I mean either other governmental agencies or industry in general. The second source is our SRT program.

Today the opening paper is going to be an example where industry developed a particular system. We took a look at it and said, "You know, we can use that if we just twist it around a little bit." In the process, we have taken something and applied it and it is going to result in much less expenditure of money. It is also going to result in a better technical operating system.

Before I steal any more of Bill's thunder, I would like to introduce Bill Dickenson, from the NASA Communications Division, who is going to describe a new application of wideband communications systems.

WIDEBAND DATA COMMUNICATIONS

William Dickinson

Our big business in the NASA Communications Division is communicating, and recently a number of the spacecraft projects have been generating requirements for data transmission at rates faster than we can accommodate on the classical voice communications or 4-kHz bandwidth circuit over which we have been communicating at a rate of 7200 bps. Typical user projects are the Atmospheric Explorer Project, which was just launched, and the Mariner Venus Mercury (MVM) mission which was launched earlier.

I would like to show you first the way we initially implemented these systems. This system is basically an extension of the 50-kbs transmission system which is used domestically in the United States as originally engineered by AT&T and used by them and others. Figure 1 shows a Goddard to Madrid, Spain, hookup, although we have one similar between here and Canberra, Australia, and involves what we call a group band, a continuous bandwidth circuit of 48 kHz which displaces 12 voice frequency circuits in the frequency division multiplex plan. By the way, the cost of such a circuit is nearly that–namely, 12 times the cost of an individual VF. This particular link shows a satellite earth station at Edam, West Virginia, and a satellite earth station at Buitrago, Spain. There is a straight-through connection at each location to Goddard where we have a 50-kbs modulator/demodulator (modem) and similarly a 50-kb modem in Madrid, Spain. Although this system can be used at 50 kbs, and that is what it was originally engineered for, the more typical application of the system is shown in the bottom of figure 1, in which we provide a mix of services which is more optimum to our needs and the project requirements. In this configuration we provide a 28.5-kb data transmission rate for project use and simultaneously provide some voice services, some alternate voice and data services, and voice coordination. You can see in the spectrum plan between 60 and 108 kHz how the original 12 voice channels in the system were displaced.

Figure 2 shows you how much equipment is required at each end of this system to do all these good things. It is essentially a five or six rack system with the wideband data equipment on the left, control system in the middle, then voice channel multiplexing and control systems, and so on.

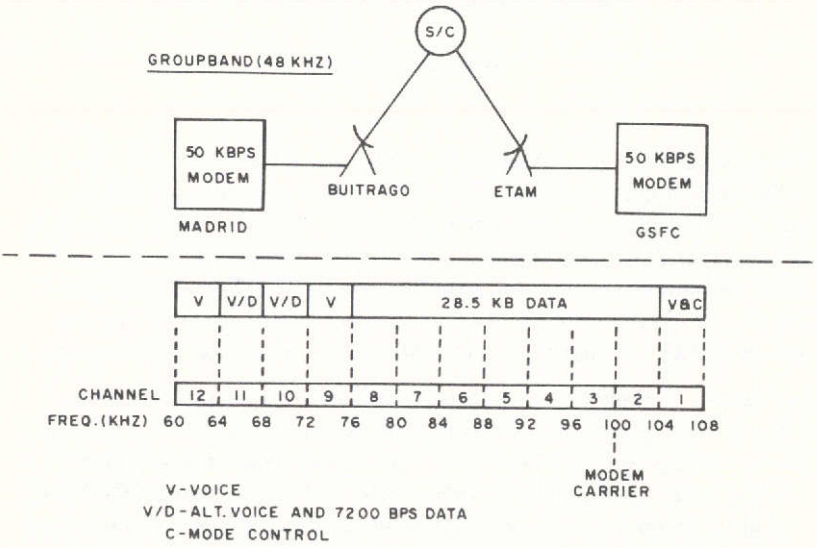


Figure 1. GSFC to Madrid communications hookup, involving a group band.

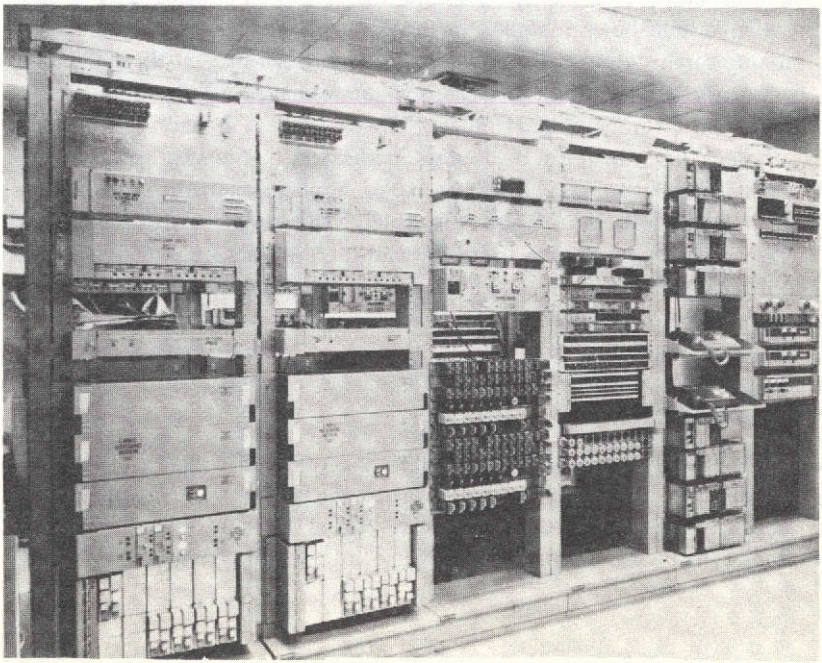


Figure 2. A five- or six-rack communications system.

What then happened in the international satellite business happened in the voice transmission area and it was a system designed to provide dial-up voice transmission through the spacecraft which is called demand assigned voice communications (figure 3). In order to accomplish adequate control and supervision over such a system, it was implemented digitally. Each channel uses a data modem which typically operates at 64 kbs for voice use, and with the application of suitable analog-to-digital and digital-to-analog converters at the end, provides the voice path and the digital bits required for supervision.

The application that came out of this in the data area, through our interest and motivation and urging of the people involved to come up with such a thing, was the twisting of this technology to provide a wideband data service. The way this was accomplished was to use a slightly higher data rate modem and a three-quarter rate, forward acting, error correcting, convolutional coding system, since the original error rate specified through the spacecraft link was only sufficient for voice transmission (only about 1 in 10^4 bit error rate). This system, with the three-quarter rate coding, provides an error rate on the 50-kb interface of between 1 in 10^7 and 1 in 10^9. This equipment will then exist at each of the earth station locations.

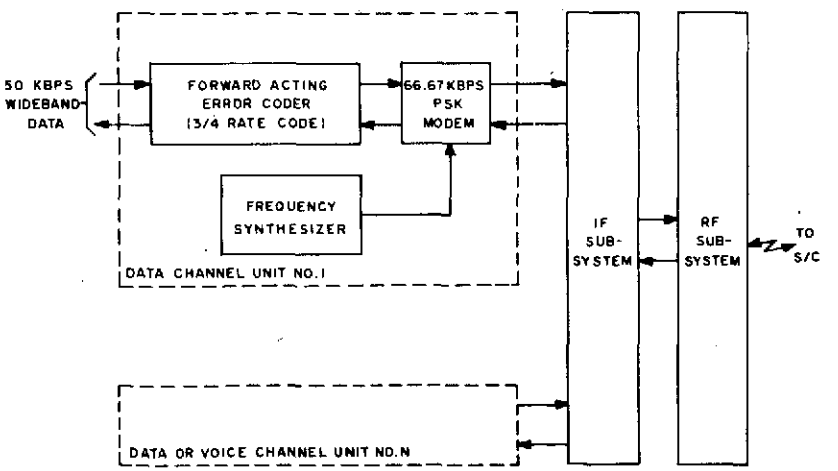


Figure 3. Demand assigned voice communications system.

In figure 4 we can see how the system was glued together. We now have a separate little 48-kHz link in the United States and a separate link in Spain, and we have digital interfaces here at the earth stations in which the data is essentially regenerated, retimed, and cleaned up. This system is now in operation and has been in operation to Spain since September 1973. We now time-division-multiplex this system, or divide the bit stream up, and provide our 28.5-kbs data channel and multiple slower speed channels in a fashion not too different from our other system.

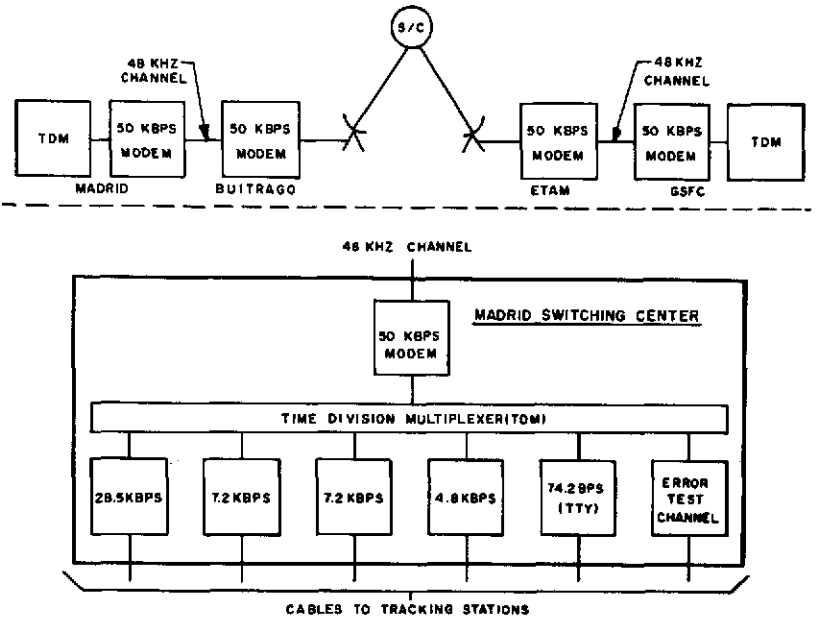


Figure 4. Complete GSFC to Madrid assigned voice communications system.

Figure 5 shows the amount of equipment that is necessary to do this job at each end, but there is a little bit of license used here. Physically, this replaces about half of the system that we saw earlier and the cost is about one-quarter that of the other system.

In summary, we have taken this technology that was developed in industry and applied it to a situation for which we had a real application. The monthly cost of the new system using the time division multiplex approach is about one-third that of the older system, because the rate structure of these satellite channels is based on a single voice circuit. Considering that

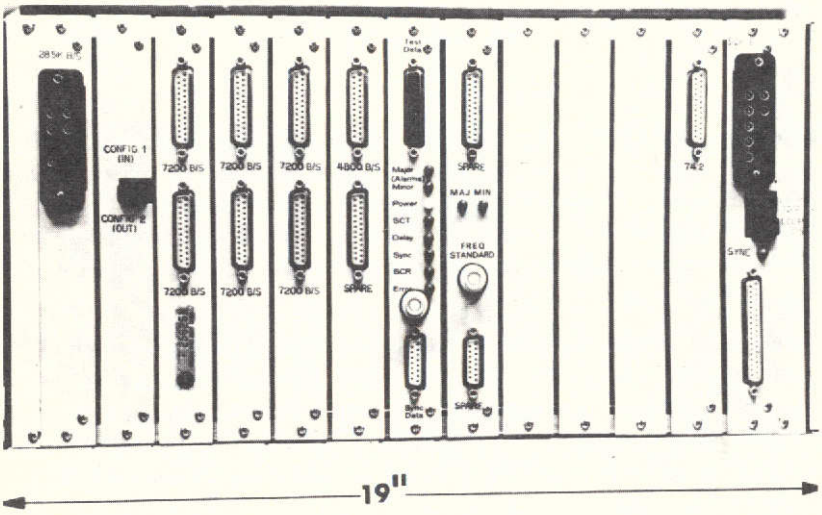


Figure 5. Equipment needed at each end of the demand assigned voice communications system.

we have to extend it within the countries on more expensive broadband facilities, the overall monthly leased cost is about one-third. The equipment cost is less, and the performance is demonstrating itself to be substantially better than what we achieve end-to-end on the conventional 48-kHz facility.

SPACE-QUALIFIED PARAMPS

Pio Dalle Mura

The development of prototype integrated spacecraft paramps at S- and C-bands has been presented at previous science and technology reviews. The extension of these results to the successful development of the first complete space-qualified paramps can now be presented. This paramp technology was developed for specific application on the Space Shuttle and TDRS programs.

Figure 1 shows a photograph of the two space-qualified paramps in their completed form. The S-band unit is on the left, and the C-band unit is on the right. Both of these units are completely self-contained and require only the application of dc power at jack J-3 on both units and the attachment to a spacecraft heat sink by means of the mounting holes around the periphery of each unit.

The S-band unit is a combination of strip line, coax, and waveguide circuitry and consists of a single stage of parametric amplification followed by a transistor second-stage amplifier. The C-band unit consists of microstrip and waveguide circuitry and is only a single paramp amplification stage.

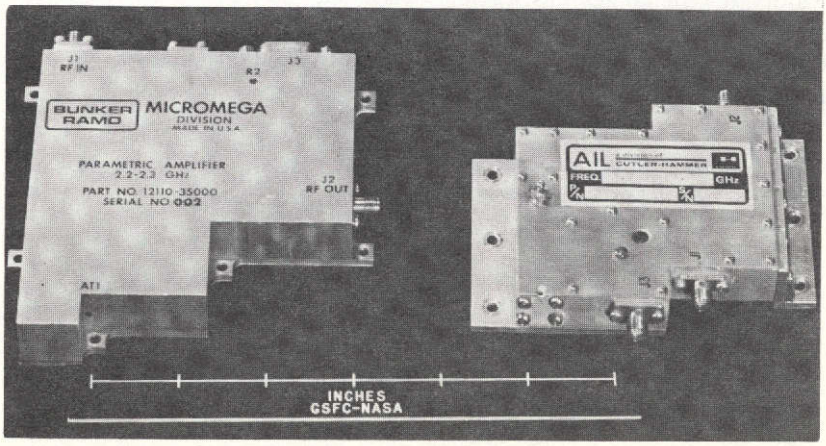


Figure 1. Two space-qualified paramps. The S-band unit is on the left, the C-band unit is on the right.

Table 1 shows some of the measured test results on the two space-qualified paramps. The noise temperature of the S-band unit, 125 K, also takes into account the second-stage transistor; the C-band noise temperature of 140 K is the paramp alone. The gain of the S-band unit is higher than the C-band unit due to the addition of the second-stage amplification. The S-band paramp is a fixed tuned unit which covers the 2.2- to 2.3-GHz telemetry band, while the C-band unit is a tunable unit which can be tuned across the 3.7- to 4.2-GHz frequency range by varying the varactor bias. The volume and the weight of the S-band unit is appreciably higher than the C-band unit because the S-band circulator is larger and also because of the addition of the transamp in the S-band unit.

Table 1
Test Results for Space-Qualified Paramps

	S-Band	C-Band
Noise Temperature	125 K	140 K
Gain	30 dB	18 dB
Bandwidth	120 MHz (1 dB)	110 MHz (3 dB)
Frequency Range	2.2 to 2.3 GHz	3.7 to 4.2 GHz
Volume	168 cm^3 (10.5 $in.^3$)	112 cm^3 (7 $in.^3$)
Weight	0.4 kg (14.5 oz)	0.2 kg (7 oz)
dc Power	20 W maximum	10 W
Gain Stability	±0.3 dB	±1.5 dB
Type Stabilization	Proportional heater	Compensation
Thermal-Vacuum	268.15 to 323.15 K, $1.3 \times 10^{-3} N/m^2$ (-5 to +50°C, 10^{-5} torr)	
Vibration	3 axes, 10 to 2,000 Hz, sinusoidal and random	
Acceleration	6 axes, 10 G	

The type of gain stability employed in each paramp is different. The S-band unit uses a proportionally controlled heater which maintains the temperature of the paramp at about 341 K (68° C). To achieve this, it requires 20 W maximum of dc power at the low temperature end of the range. The C-band unit uses a compensation circuit which varies the voltage to the Gunn oscillator and the varactor bias in order to compensate for temperature changes. This system uses 10 W maximum of dc power.

The S-band paramp provides better gain stability over the temperature range, varying only ±0.3 dB compared to ±1.5 dB for the C-band unit.

Both of these paramps were tested to the same space qualifications specifications. For the thermal vacuum tests, the temperature was varied from 268.15 to 323.15 K (-5 to +50°C) in a vacuum of $1.3 \times 10^{-3} N/m^2$ (10^{-5} torr). The units were operated continuously during thermal vacuum testing and, in addition, went through cold starts several times satisfactorily. For vibration and acceleration testing, both the paramps were operating, and performance was measured both before and after to observe any changes in operating characteristics. No changes were recorded. The vibration test was along three perpendicular axes, the frequency varying from 10 to 2,000 Hz, both sinusoidal and random. The acceleration testing was a 6-axis measurement with 10 g minimum force.

At present we are operating both of these paramps in our laboratories in order to get some feel for the long term stability and reliability of the units which is required by applications such as TDRS.

In addition to the S- and C-band units, we are also developing a Ku-band spacecraft paramp. Figure 2 shows the completed prototype of this unit. This paramp employs some advanced concept varactors which enable it to achieve some truly outstanding results at this frequency. The paramp consists basically of: a stripline circulator which has very low loss due to the type of construction used; an electroformed cavity which serves as the varactor mount and within which there are mounted a pair of very high cutoff frequency, unpackaged chip varactors in a balanced configuration; a Gunn diode pump oscillator operating at 35 GHz; a pump ferrite isolator; and a varactor doubler circuit which doubles the frequency from 35 GHz to the 70 GHz pump frequency ultimately applied to the varactors. The doubler also uses a high cutoff unpackaged chip varactor of the same design used in the paramp itself.

The Ku-band paramp was tested both at atmospheric pressure and in vacuum over a temperature range of 268.15 to 323.15 K (-5 to +50°C). Some of the results obtained on the prototype unit are shown. The noise temperature is less than 150 K over the entire 500-MHz bandwidth, the gain is 18 dB, and the frequency coverage is 14.7 to 15.2 GHz. The dc power required by this unit is 11 W. This paramp, like the S- and C-band units, is intended for ultimate application on both the Space Shuttle and TDRS programs.

A space-qualified model of this prototype is now being fabricated. It should be ready for delivery about the middle of 1974.

MICROWAVE ZOOMING ANTENNAS

Richard Schmidt

The subject of this paper is the dual parabolic-cylinder monopulse zoom antenna. The television camera zoom lens is a familiar example where a wide field of view is picked up before a specific object or target is singled out, as seen in figure 1. The object here is to do very much the same sort of thing at microwave frequencies.

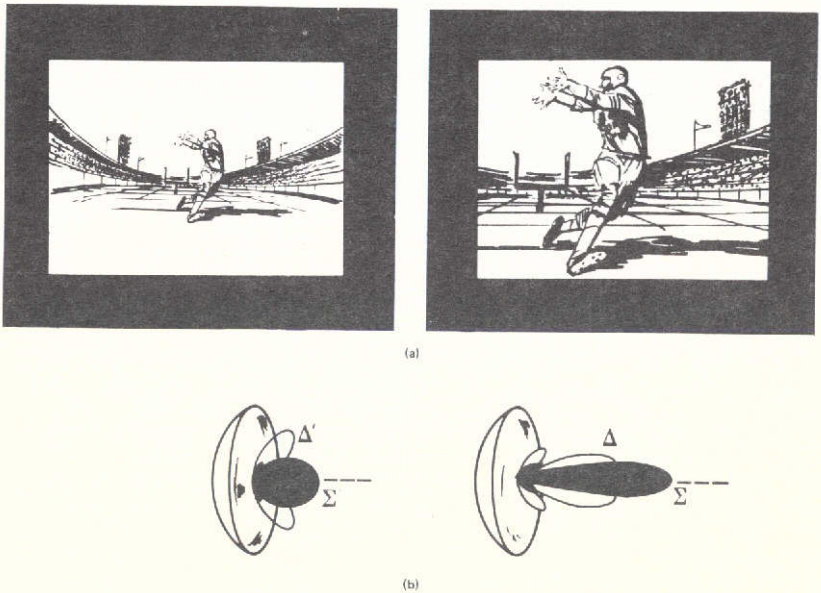


Figure 1. (a) Optical zooming; (b) Microwave monopulse zooming.

Spacecraft antennas, particularly those such as on the tracking and data relay satellite system (TDRSS), have very narrow beam, high gain antennas, and these are completely unsuited for acquisition. One would, therefore, like to have within the same antenna structure a means for generating a set of acquisition patterns as well as a set of tracking patterns. One difference between the optical case and the microwave case is that a circular polarization state needs to be maintained for the microwave antennas.

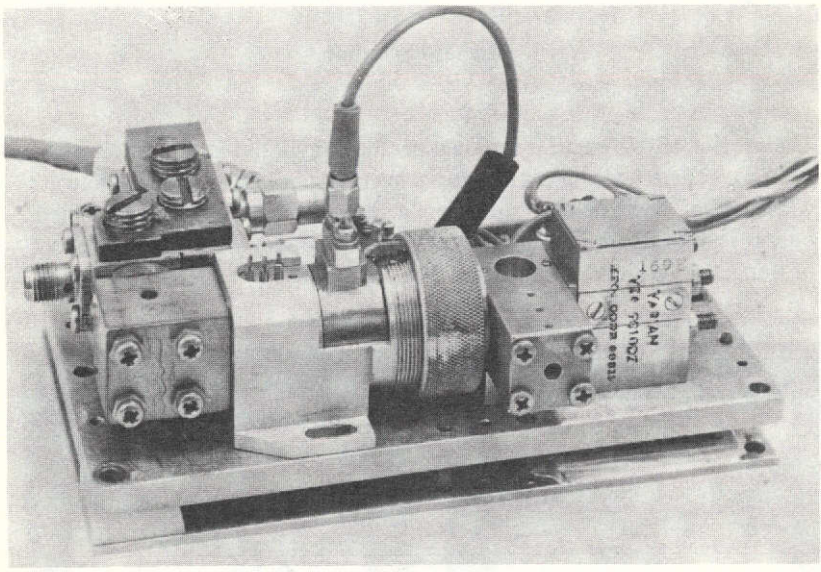


Figure 2. Ku-band spacecraft paramp prototype.

The bidirectional continuously variable monopulse zoom antenna (figure 2) is one embodiment of five original techniques developed here at Goddard during the past year. The beamwidth zooming here is about 100 percent. The geometry is that of Spencer, Holt, Johannson, and Samson of Cambridge Research Center (circa 1954). The geometry is unique because line feeds are coalesced to point feeds by means of parabolic cylinders. The entire subsystem was rotated 90° so that both the feed and the subreflector presented a minimum obscuration to the main reflector (gamma 1). The adaptation of this geometry to amplitude sensing of monopulse was conceived at Goddard, as was the zooming of this antenna. The zooming can be performed in the XZ plane by telescoping one edge of the subreflector (gamma 2). It can be zoomed in the YZ plane by telescoping the two ends of the main reflector (gamma 1). This can be done independently or simultaneously.

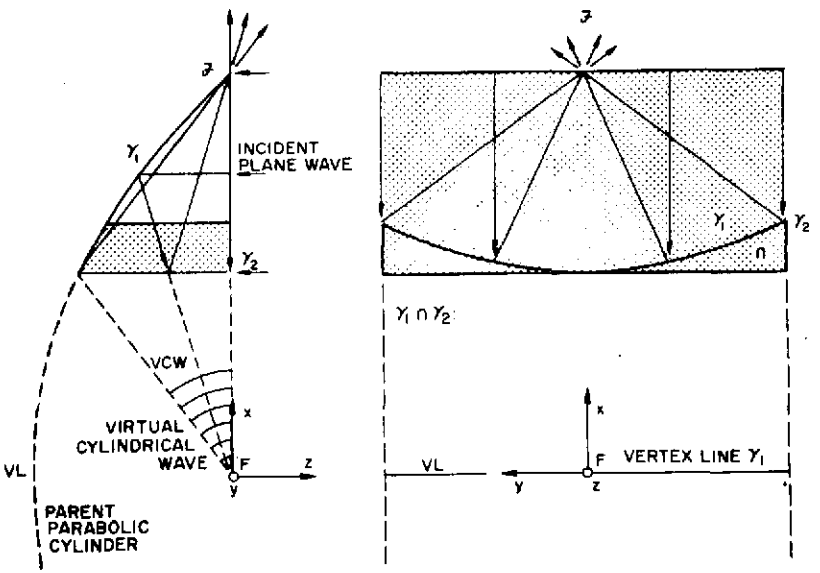


Figure 2. Dual parabolic-cylinder monopulse zoom antenna.

The positioning of the monopulse feed is not altogether obvious, although ray optics give an indication that the feed should be inclined with respect to the z-axis. A much better way to determine how the feed should be positioned is to rely on a focal region plot. Figure 3 was developed from the Goddard Kirchoff-Kottler diffraction program. The loop curves are the isophotes or constant intensity lines; the open-line curves are the wave

fronts or constant phase contours. One immediately recognizes the depth of field of the antenna, the width of field or airy disk, and the bright ring airy structure. One also sees a converging spherical wave which becomes plane in the focal region at the airy disk and then, if the energy is not intercepted by any feed, proceeds on through the focal region as an expanding spherical wave once again. There is no difficulty—one simply arranges to have the four phase centers of the monopulse feed cluster on a common wavefront.

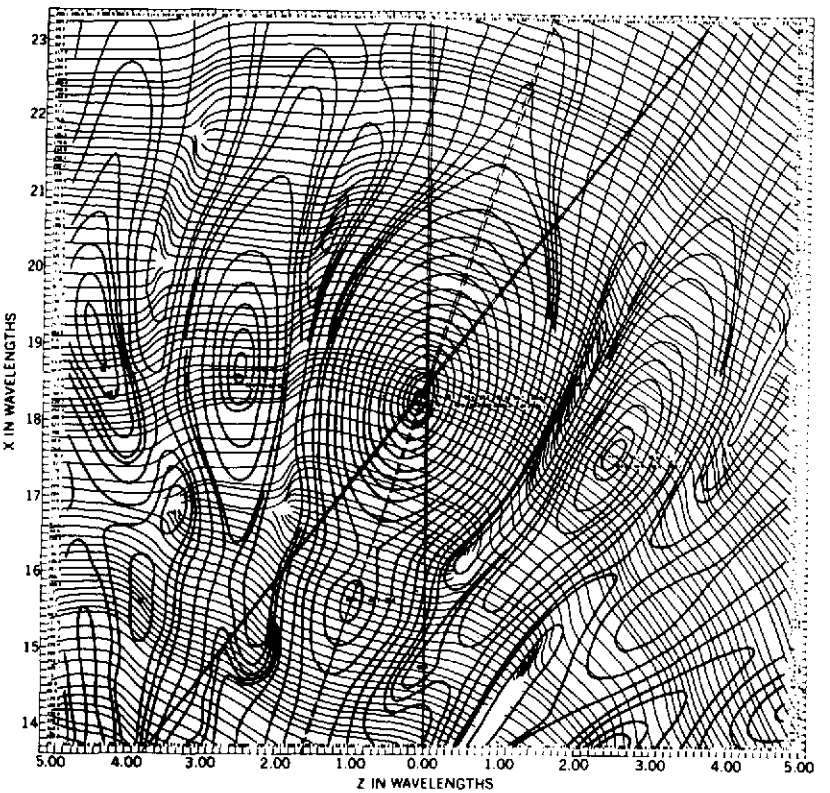


Figure 3. Dual parabolic-cylinder zoom antenna.

The GSFC diffraction program was also used to compute the far-field tracking patterns. In figure 4, the solid curves are the narrow angle set; the dashed curves are the wide angle set. It can be seen that the monopulse postulates are well preserved, the null is deep, and the error slope has been relieved for the broad set of patterns. The asymmetry seen in the diffraction structure is due to the inherent asymmetry of the reflectors themselves, not due to any computational problem.

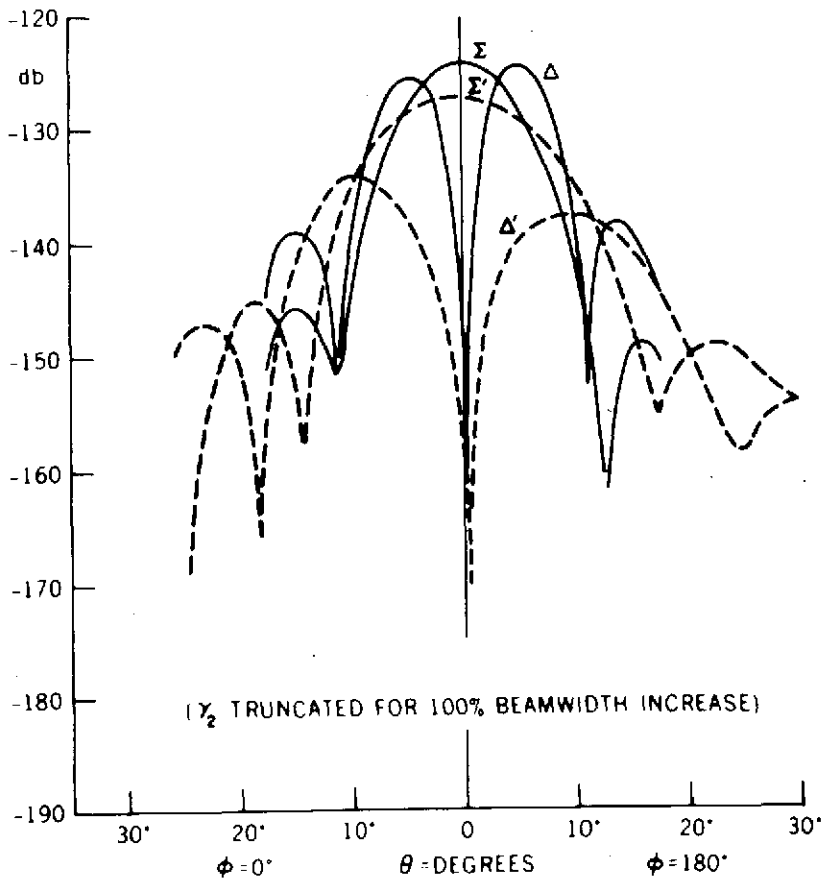


Figure 4. Channel 1 sum and difference patterns of a diamond monopulse antenna system.

Only channel 1 of a diamond monopulse system has been shown in figure 4. It has been zoomed 100 percent in this computation. Channel 2 is very similar, except that the diffraction structure is perfectly symmetrical. This is an interim effort directed toward understanding and developing monopulse zoom antennas in the microwave region for TDRSS and Space Shuttle. We have explored both electrical and mechanical means; three patent disclosures were submitted. The literature, incidentally, gives very few examples of microwave zooming. Interest in microwave zooming is somewhat recent. One finds practically no references to cases where monopulse must be obtained and the circular polarization state must also be preserved.

One point of difficulty is that we not only look for general techniques for zooming at microwaves, but we endeavor to come up with designs that represent a minimum impact on the TDRSS, since considerable hardware has already been selected (principally the reflector) and any techniques that are evolved must take this into account. The antenna discussed here today is covered in greater detail in a Goddard document, TM X-70631.

UNIFIED SPACECRAFT RADIO FREQUENCY SUBSYSTEM DEVELOPMENT

Arthur F. Block

The development of a universally applicable transponder at S-band has progressed in this past year on two paths. The first provides equipment for existing programs where the requirements are well defined, such as IUE, HCMM, and IME. The second path is for a more universal application and it will be used on all earth orbiting spacecraft.

The interim design is seen at the top of figure 1 and the universal is at the bottom. These are the major subassemblies of the S-band transponder. As you will see, many of those in the bottom are labeled the same. When the interim design is completed and evaluated there will be a great deal of fallout, and with a little updating we hope to be able to use many of the same subassemblies.

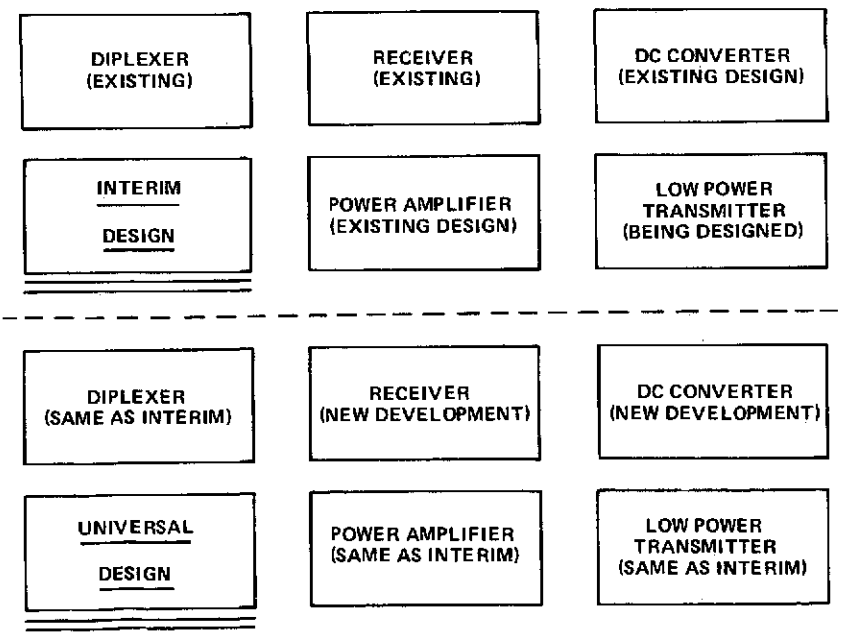


Figure 1. Major subassemblies of the S-band transponder: top, interim design; bottom, universal design.

The interim design uses subassemblies which exist. Some are used on other programs. For example, the diplexer and power amplifier are used on the tracking and data relay experiment on Nimbus. The receiver is one which was developed on SRT funding in 1970 and 1971. The prototype of the interim design transponder will be completed in the first quarter of 1974.

The universal design will have a wide application and diversification. We will have substitutable modules which can, for example, detect in the receiver anything from the simplest noncoherent signal to the most complicated Costas loop. The power amplifier is also selectable over the range from a quarter of a watt to 20 W at S-band. We have developed one portion of the receiver at this time.

Figure 2 shows one side of this two-sided module. The module is 12 cm by 4 cm by 2 cm. This figure shows the input filter to the receiver. In addition to performing the normal functions of filtering and passing the band of the receiver, this particular filter also gives 50 dB of transmitter rejection which will lighten the burden on the diplexer.

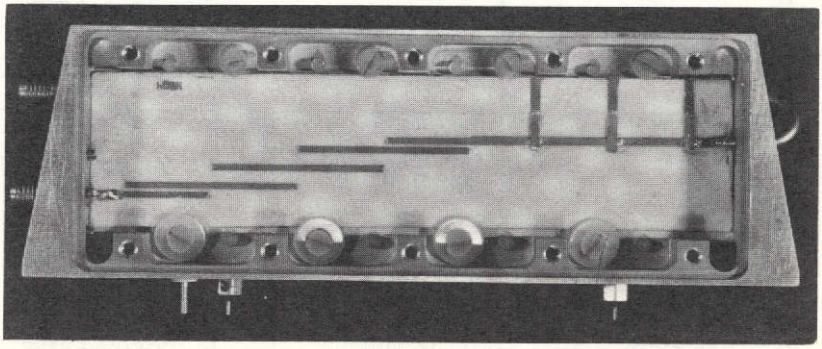


Figure 2. Receiver module, showing the input filter to the receiver.

The other side of this module is shown in figure 3, and contains the active circuitry. From right to left are two stages of RF transistorized amplification, a mixer, and a first IF stage. This has all been completed and tested in prototype form.

The RF amplifier has 20-dB gain and by itself has a noise figure of 4 dB. When combined with the filter and diplexer, the receiver noise figure will be about 6 dB. This is the first of probably six modules which will make up the transponder. This transponder, in prototype form, will be completed

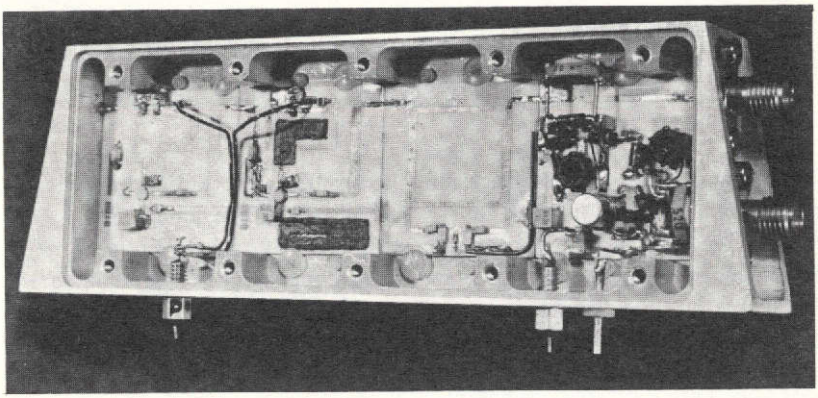


Figure 3. Receiver module, showing the active circuitry.

about a year from now. On the reverse side are four individual, 2.54 cm^2 (1-$inch^2$) substrates. The power amplifiers, which will be used on the universal transponder, will take advantage of NASA's power amplifier transistors, which have been developed on SRT funds over the past many years.

NETWORK TIMING AND SYNCHRONIZATION TECHNOLOGY

Andrew Chi

This is a preliminary report concerning the results of a recent experiment on time synchronization using Omega transmissions. The Omega transmissions are used because they form a worldwide navigation system. The transmitting frequencies are in the very low frequency (VLF) band and are thus stable for long distance transmissions. The Omega transmission format is ideally suited for transmitting precise time, and, more importantly, it is a system which has been implemented.

Figure 1 shows the eight stations of the Omega navigation system. The station in Trinidad is scheduled to be moved to Liberia. The numbers, 10.2, 13.6, and 11.33, are the navigation frequencies in kilohertz. Each station transmits the three navigation frequencies in a 10-s commutation format as shown. Since there are eight stations, eight time segments are needed.

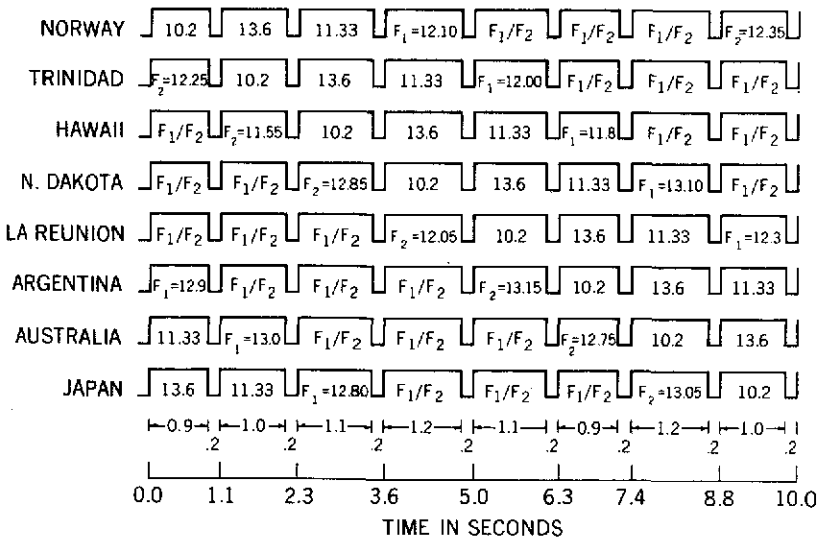


Figure 1. The 10-second Omega transmission sequence for stations A to H.

When a station is not transmitting the three navigation frequencies, it transmits, in the five remaining time segments, two side frequencies separated by 250 Hz. These side frequencies are phase controlled at the emission and can be synchronized to a time scale such as the coordinated universal time, UTC. All stations, except North Dakota, which has been in operation since 1972, are scheduled for operation in 1974 or 1975. All Omega stations are designed to radiate 10 kW of power. The North Dakota station transmits the two side frequencies at 13.1 and 12.85 kHz. These frequencies are phase controlled at the transmitter to one-tenth of a microsecond and time synchronized to the UTC time scale of January 1972.

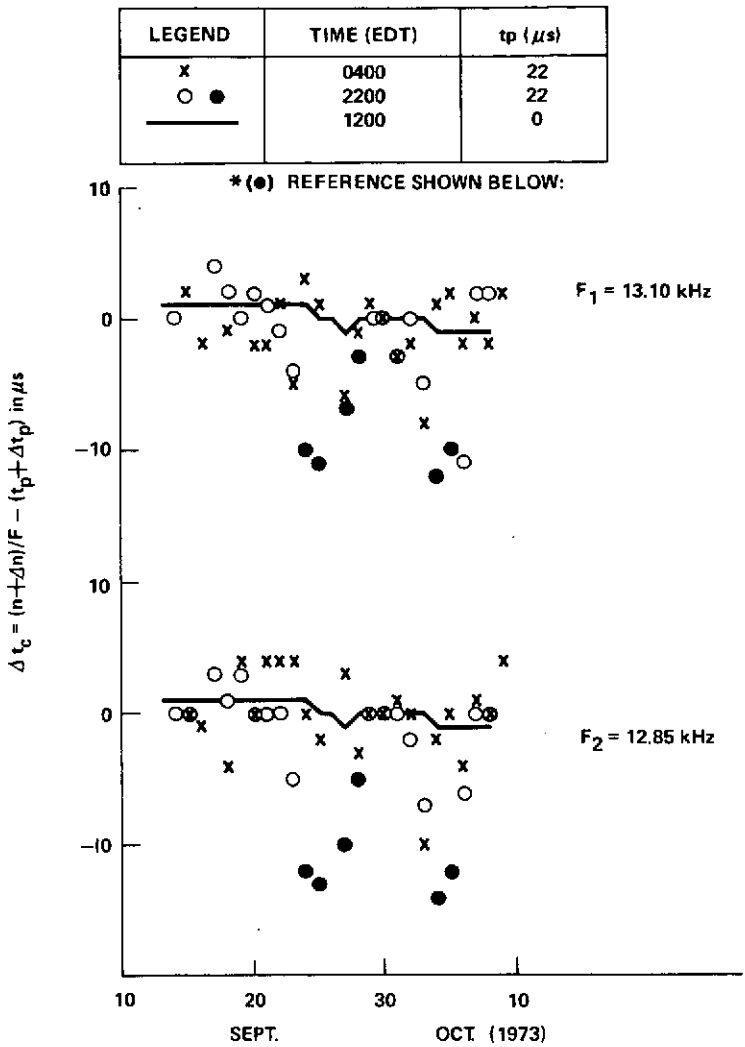
Table 1 shows the results of cycle determination of the two side frequencies from North Dakota in 1972 and 1973. These cycles were determined from the VLF data for the four sites shown. The subscript 1 is for the 13.10-kHz and 2 is for the 12.85-kHz frequency. If the local clock time is known, one can calculate the propagation delay by multiplying the cycle data by the corresponding period of the side frequencies. As one can see from the numbers given in the last three columns, the agreement between the measured and the predicted propagation delays is, indeed, very good.

Table 1
Preliminary Results of Propagation Delay Measurement
Using Omega North Dakota Transmissions
and Known Local Clock Time

Location	Measured				Predicted
	$n_1 + \Delta n_1$ (cycles)	$n_2 + \Delta n_2$ (cycles)	$(n_1 + \Delta n_1)\tau_1$ (μs)	$(n_2 + \Delta n_2)\tau_2$ (μs)	t_p (μs)
NELC, San Diego, Ca.	96.795	94.924	7388.9	7387.1	7388
GSFC, Greenbelt, Md.	84.589	82.957	6457.1	6455.8	6456
USNO, Washington, D.C.	84.100	82.488	6419.8	6419.3	6420
STDN, Rosman, N.C.	78.561	77.056	5997.0	5996.6	5995

Figure 2 shows a time comparison between the clocks at GSFC and North Dakota during September and October 1973. The vertical scale is the time difference in microseconds. The solid curves represent the data obtained during the daytime transmission at 1200 eastern daylight time (EDT). These data have an excursion of 1 μs. The circles and crosses are data obtained during the nighttime at 2200 and 0400 EDT, respectively. Since the propagation path is increased at night, an additional propagation delay correction of 22 μs was supplied. These data have an excursion of less than ± 10 μs. Solid circles in this figure indicate those days during which sudden ionospheric disturbances (SID) were observed during the day.

Based on these results, we have made tentative conclusions that the Omega system can provide a worldwide time transmission signal to a precision of ± 1 μs if daytime transmissions are used, and less than ± 10 μs if nighttime transmissions are used. Although the data were obtained for distances of less than 2500 km, that is, 1300 miles, we have good reason to believe that this precision can be maintained for extended distances beyond 10,000 km or 6,000 miles. We have made plans to obtain additional data in the near future to substantiate this claim and to determine the number of Omega stations needed to provide worldwide coverage for time synchronization.



*(●) MARKS OBSERVATION OF SID: 9/24–1530;9/25–1015, 1445, 1645;9/27–1130; 9/28–1100; 10/4–1200, 1430; 10/5–1500.

Figure 2. Stability of VLF signals for time comparison of GSFC clock relative to North Dakota Omega station time transmission received at 0400, 1200, and 2200 EDT in Greenbelt, Maryland.

ADVANCED STUDIES IN ADAPTIVE CONTROL AS APPLIED TO SATELLITE TRACKING SYSTEMS

Andrew Rolinski

The subject of my talk is advanced studies in adaptive control as applied to satellite tracking systems. A few years ago Goddard initiated work in adaptive control because it was believed that the technique was very appropriate for precise control of large tracking systems where the system parameters change, not wildly but substantially, over an extended period of time.

Adaptive control can be conceived basically as a model, a plant, and adjustments on the plant parameters in a feedback configuration, as seen in figure 1. One looks at the reference input which goes to the model as well as to the plant. The difference between the output of the model and that of the plant is then fed into a parameter adjusting mechanism (PAM), which is simply a logic device that develops certain laws to control the parameters of the plant. PAM also adjusts the input to the plant so that the system eventually will behave the same way as the model. Therefore, we adapt the plant to the model. I might add that the plant, in our case, is the power amplifier, the antenna itself, the tower structure, tachometers, and whatever else can be included as the fixed part of our ground antenna system. Ordinarily, to reconstruct the states of the system, and thus to adapt to system changes, the input and the output and its derivatives must be known. This is rather a severe limitation on the adaptive scheme.

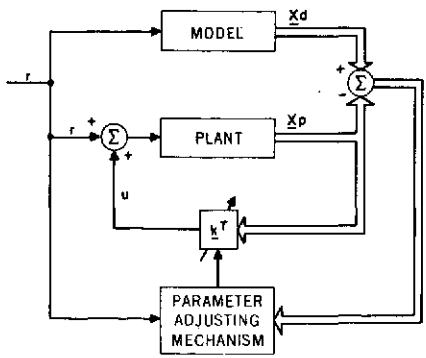


Figure 1. Adaptive control system.

The University of Massachusetts, under a NASA grant, developed an adaptive control system presently being used on a 60-cm (24-in) optical tracking mount located at the Goddard Optical Facility. The adaptation was made on the forward loop gain only. Using this simple adaptive scheme, the errors tracking the laser beam were reduced from 3.6 arc-seconds to 0.7 arc-seconds peak-to-peak.

The major part of the research at the present time is directed toward removing the basic limitation on the adaptive design, namely, the total knowledge of the states of the system. Thus, the concept of the state variable filters was introduced in the study to make a design more amenable to an engineering solution.

In figure 2, notice where the state variable filters (SVF) are introduced in the system. Through the state variable filters, we can reconstruct the various unknown states, or the unobserved states, of the system.

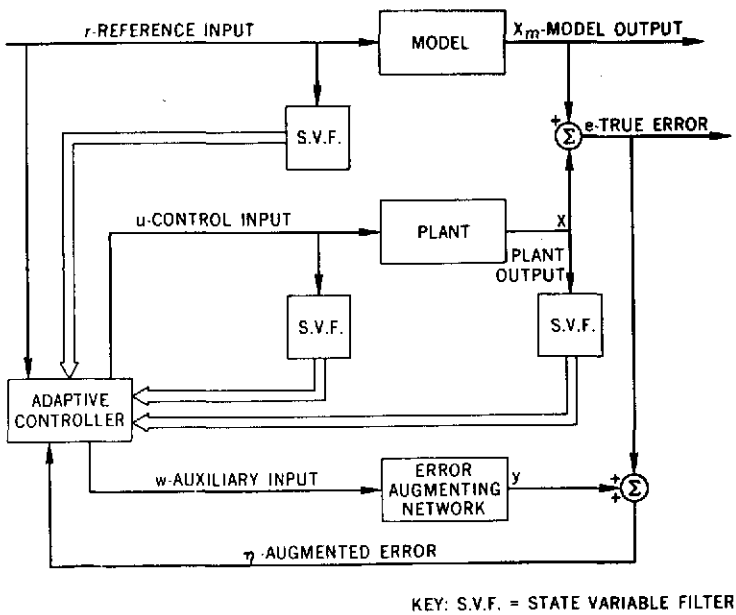


Figure 2. General system layout.

The input, Y, which is generated in the error augmenting network, is compared to the true error, e. The true error, e, is developed from a comparison of the output of the model and that of the plant. The adaptive

concept is predicated on the idea that by introducing a certain error function into the system differential equation, one can, in steady state, drive the augmented error, n, to zero. By implication, Y and e go to zero. Since e is the system state which we want to drive to zero, the whole system error converges to zero and the desired result is obtained.

The state variable filters are used to determine the unobserved states of the system; through them, a differential equation describing the auxiliary input, w, is obtained. At this time, we have performed a simulation of a third-order plant on a DEC PDP-11 machine. Adaptive adjustment was made on a single pole.

In figure 3 we see the error curve for a simulation of a third-order system. The error frequency is about 3 cycles and the period is about 5 seconds. The error curve collapses or converges to zero at about two and a half times the predominant time constant of the system.

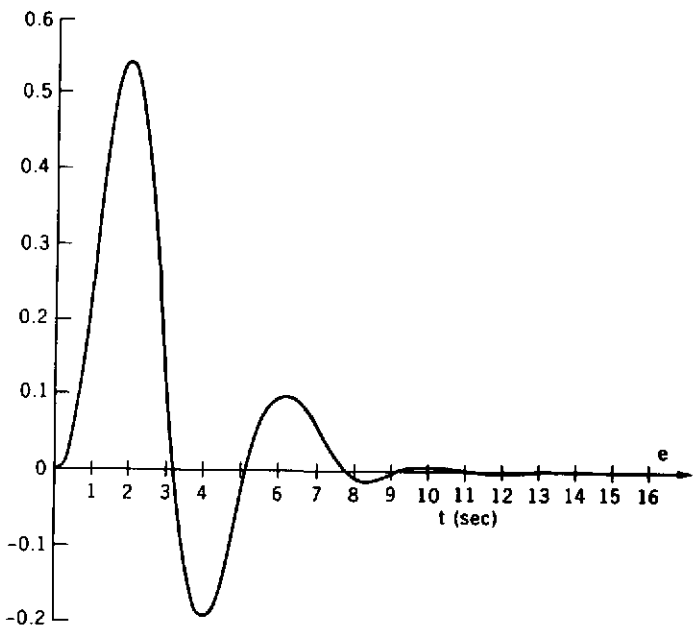


Figure 3. Error curve of the simulation of a third-order system.

Future work in this area would be directed toward the study of adaptation of more than one feedback parameter and more than one plant parameter. Study should also include adaptation of system response to noisy measurements and systems with stochastic disturbances.

Application of the concept so far has been to laser tracking systems. In the future, we would like to include radio tracking systems such as an 18-m (60-ft) Ku-band antenna, where the antenna beamwidth is 0.07°. The tracking requirements for such a system would necessitate a control system design which would reduce system errors below 30 seconds of arc.

SKYLAB TELEMETRY DATA COMPRESSION

Paul J. Pashby

I am going to talk about the telemetry data compression program which we put together for the Johnson Spacecraft Center for use on the current and past Skylab missions. This compression system, I think, is the only operational, computer-program-based data compression system in existence. It is a real-time system as well.

The typical Skylab orbital pass would see approximately 200,000 bps transmitted to the ground station. This does not include dumped data. It was not economically feasible to provide communication bandwidth to throughput this data to the control center, so an alternative had to be chosen. Numerous studies were conducted, both here at Goddard and at other NASA centers, as well as with a number of NASA contractors. They indicated that, based upon Apollo data, sufficient redundancy existed in the data streams which, if removed, could reduce the effective data rate that needed to be transmitted to the control center. It turned out that approximately a 20:1 compression ratio appeared to be feasible based upon these studies.

Figure 1 illustrates both the problem and the solution. The problem is indicated by the large amount of data shown on the left-hand side of the figure as compared to the limited output capability on the right. The solution is represented by the large block labeled data compressor, which would remove the redundancy from the data stream and thus provide a manageable output rate. You will see that I put down 200,000 bps on the left-hand side of the figure and indicated that we could achieve a 20:1 compression ratio. That is 10,000 bps, from 200,000 down to 10,000, to achieve a 20:1 compression ratio. The figure shows 20,000 bps, however, going out. The extra 10,000 is required for error information and the data identification. So the actual reduction occurred from 200,000 down to 10,000; the addition of 10,000 bps for error control was our final output bandwidth.

Various data compression techniques were studied and the one found to be most suitable is called the zero order predictor. This compression algorithm is shown graphically in figure 2 and it is a very straightforward type of algorithm. This is extremely important because it was implemented in the digital computer and it would be required to be executed for each and

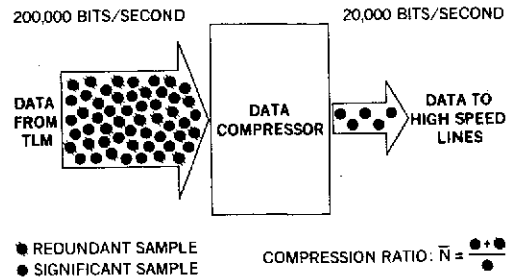


Figure 1. Model of necessary data compression.

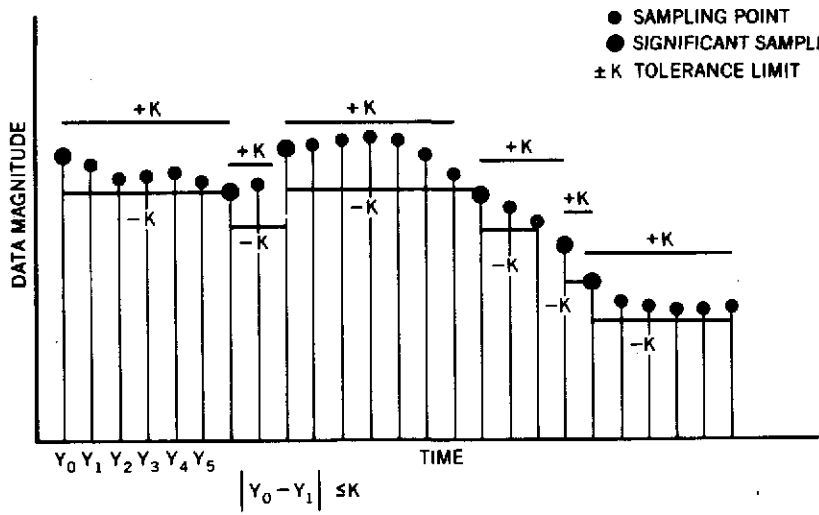


Figure 2. Zero order predictor.

every sample that was placed into the system. You can see that each successive sample of a parameter is compared to a previous nonredundant sample of that same parameter. If the absolute difference between them is greater than an amount defined by K, the parameter value is considered to be significant and it is transmitted to the user. If the difference is less than K, the sample is considered redundant and it is not transmitted to the user. There are 25 samples of this particular parameter shown in figure 2. Using the particular compression technique shown here, only six of those samples would be necessary to be transmitted to the user. The compression ratio, therefore, would be 25:6 or approximately 4.1:1.

Some of the significant design features which had to be provided in the system are as follows. First, error detection and correction was necessary

because a loss of data in a data compression system is far more significant than a loss of data in a conventional fixed-format system. A hamming code with diagonal parity was developed and it virtually eliminated the possibility of an error going undetected in the system. Additionally, it also possesses some limited error correction capability.

A second feature, automatic contingency level adjustments, was necessary to prevent the system from becoming overloaded when the compressibility of the input data stream fell below the average level expected. In other words, the system can never stop or come to a halt in a real-time situation.

Another item, sensitivity to reduced output bandwidth, is also related to the second item, in that the output capability could be reduced because of the loss of a communication line. The system was required to handle this case and to adapt to that situation as well.

A fourth feature relates to the value of K assigned to each parameter in the system. Normally, the K value would be assigned prior to a mission. However, unexpected behavior of a parameter made it necessary to provide the user with the capability to change the K factor in real time. So by reducing the K value, a user could see more samples of a parameter; by increasing it, he could effectively shut off the transmission of all samples of that parameter, such as in the case of a noisy transducer.

Finally, it was required to store on magnetic tape all values of all parameters for subsequent transmission to the control center on a post-pass or non-real-time basis. Figure 1 could have shown a bucket under the data compressor containing all of those redundant samples. We saved those and transmitted them to the control center later on.

In summary, all the design goals that we set out to achieve were achieved successfully. The compression ratios which we experienced did, in fact, average about 20:1 in those areas that we sampled, and the system has successfully supported all Skylab missions.

☆ U. S. GOVERNMENT PRINTING OFFICE: 1975—636-489 REGION NO. 3-11.