X-100-65-407

NASA TM X-55356

FINAL REPORT OF THE GODDARD SUMMER WORKSHOP PROGRAM IN ANALYSIS OF SPACE DATA AND MEASUREMENT OF SPACE ENVIRONMENTS

COMPILED BY
ELIAS KLEIN

JUNE 15 TO SEPTEMBER 15, 1965

NASA — GODDARD SPACE FLIGHT CENTER
GREENBELT, MARYLAND

X-100-65-407

FINAL REPORT OF THE GODDARD SUMMER WORKSHOP PROGRAM IN ANALYSIS OF SPACE DATA AND MEASUREMENT OF SPACE ENVIRONMENTS

June 15 to September 15, 1965

". . . Men in small groups, collaborating, can solve problems in a few weeks or months or even days that one man working alone and in isolation could never have solved had he lived a thousand years."

Pierre Teilhard de Chardin

Goddard Space Flight Center
Greenbelt, Maryland

FOREWORD

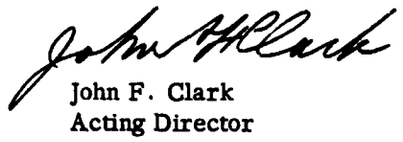
The National Aeronautics and Space Administration has been assigned the prodigious job of leading the expansion of human knowledge of the atmosphere and space. To a large extent, progress on this assignment depends on how well we use the Nation's scientific and engineering manpower. A most important reservoir of scientific talent is in our universities and colleges. To utilize this reservoir there must be good communications and mutual understanding between people on the campus and NASA technologists. This "working together" can be most effective in reaching the desired goals.

The Goddard Summer Workshop, after four years of experience, has emerged as an effective and unique instrument for generating new areas of cooperation between university people and the Goddard scientific and technical staffs. It not only produces specific solutions to some of our technical problems but, perhaps more importantly, it serves as a catalyst in bringing forth new areas of cooperation between Goddard and university circles.

An example of ongoing relationships that grow out of the Summer Workshop is the consultantships that continue after the close of each summer's session. In sessions prior to 1965 a total of some five consultants from the Summer Workshop continued to provide services to Goddard after the close of the formal session. Following the 1965 session, I am told, it is likely that there will be ten participants who will continue as active consultants for various Goddard organizational elements.

Another measure of the value of the Summer Workshop in terms of scientific accomplishment is the twenty or so papers and presentations which have been published since 1962 in scientific or technical journals, technical notes, external reports, in minutes of scientific meetings and international congresses.

We have come to regard the Goddard Summer Workshop as another effective tool for advancing the NASA-university cooperative enterprise.



John F. Clark
Acting Director

PREFACE

This document contains the substantive results of the fourth Goddard Summer Workshop. Fruitful and impressive as these results are, they reveal only the immediate needs of Goddard and represent just part of our principal objective. The intangible benefits which accrue to Goddard and NASA from this Workshop are of much greater significance to the long range space program than the outcome of current problem-solving in a 12-week effort. To be sure, this effort is most important because it provides the opportunities for Goddard working engineers and scientists to renew and update their knowledge of present-day technology. Some form of training for existing manpower is urgent in order to meet the technological needs of the space age, and, this Workshop is one approach to the rejuvination of the working scientist and engineer at Goddard.

It is gratifying to report that the educational growth of the Center staff participants in the 1965 Workshop showed marked progress. This was shown by the much larger number of participants than in any previous Workshop and by their sustained interest in the problems under study. If this trend continues, the Goddard Summer Workshop will not only increase the quality and quantity of its tangible results, it will also make sure that the technical workers at Goddard are capable of renewal and of refurbishing their professional skills. In addition, the university experts and specialists who participated in the 1965 Workshop have been very productive individually as well as in their team efforts. These diverse activities are contributing to the eventual success of this Workshop enterprise. Moreover, it seems that we are about to realize the hope expressed by Dr. Goett at the opening session of the first (1962) Summer Workshop, when he said: ". . . . I consider this Summer Workshop a very interesting experiment . . . . (but) the main product and value of this experiment would hardly appear during the twelve weeks you (university people) will be working here . . . . More important benefits should come later . . . . "

From the very beginning of the Goddard Workshop, our aim has been to achieve a three-fold objective through the in-house problem-solving efforts of the Center-university teams. By collaborating and working toward the solution of various problems, each participant is stimulated to contribute of his background and experience. Thus, there is "group interstimulation" which according to Dr. P. H. Abelson, Editor of Science (AAAS), is a major factor in creativity. And, after four years of Summer Workshop activity, Goddard can feel a sense of accomplishment in its wholesome university relationships as well as in its ability to keep its technical personnel abreast of the swift advances in space technology.

This year, again, we are grateful and proud to have had in each team one of the National Science Foundation High Ability Secondary School students. These young people contributed to the group effort with enthusiasm and understanding, despite their years. On the other hand, the active participation of these high school students amidst a group of mature scientists gave each young person an opportunity to size up the meaning of "collaborative research." Also, these group activities tend to foster professional competence in the potential scientists of the future.

Elias Klein
Summer Workshop Director

UNIVERSITY WORKSHOP PARTICIPANTS

SENIOR FACULTY MEMBERS

CHEN, Yu, Dr.—Associate Professor of Engineering Mechanics, Rutgers, The State University, New Jersey. (Summer Workshop participant 1964.)

CHENG, David H. S., Dr.—Associate Professor of Electrical Engineering, University of Missouri.

DALRYMPLE, Jean M., Dr.—Associate Professor of Mechanics, Michigan Technological University.

JAVID, Mansour, Dr.—Associate Professor, Department of Electrical Engineering, City College of New York.

JONES, Herbert W., Dr.—Professor of Physics, Florida Agricultural and Mechanical University.

MOHR, Eugene I., Dr.—Chairman, Physics Department, Columbia Union College. (Summer Workshop Participant 1963 and 1964.)

PLYLER, Earle K., Dr.—Head, Physics Department, Florida State University.

STABLER, Edward P., Dr.—Associate Professor of Electrical Engineering, Syracuse University.

WAIDELICH, Donald L., Dr.—Professor of Electrical Engineering, University of Missouri. (Summer Workshop participant 1962, 1963, 1964.)

WOGRIN, Conrad A., Dr.—Associate Professor of Engineering and Applied Science, Yale University.

JUNIOR FACULTY MEMBERS

CALLAHAN, William R., Dr.—Lecturer in Physics, Weston College.

DAVISSON, Lee D., Dr.—Assistant Professor of Electrical Engineering, Princeton University.

KINNIER, John H., Rev.—Assistant Professor of Physics, Boston College.

RAMSEY, Fred L., Dr.—Post Doctoral Fellow, Department of Statistics, Johns Hopkins University.

ROHDE, Charles A., Dr.—Assistant Professor of Biostatistics, Johns Hopkins University.

GRADUATE STUDENTS

BORIGHT, John P.—Newmann Laboratory, Cornell University.

COLEMAN, Roger D.—Department of Statistics, Johns Hopkins University.

COOPER, Melvin L.—Georgetown Observatory, Georgetown University.

DELLA VALLE, James A.—Department of Physics, University of Connecticut.

DRALLE, Anthony V.—Physics Department, Carnegie Institute of Technology. (Summer Workshop participant 1964.)

GORDON, Gary A.—Department of Electrical Engineering, The City College of New York.

JARRELL, Temple W.—Department of Physics, University of Maryland.

KORB, Charles L.—Department of Physics, Florida State University.

OIEN, Michael A.—Department of Engineering Mechanics, Cornell University.

RANCOURT, James D.—Physics Department, Carnegie Institute of Technology.

SHAPIRO, Stephen M.—Department of Physics, Johns Hopkins University.

WOODS, Robert O.—School of Engineering and Applied Science, Princeton University.

PART-TIME CONSULTANTS

GARDNER, Irvine C., Dr.—Former Chief, Division of Optics and Metrology, National Bureau of Standards. (Summer Workshop participant 1964.)

SCHWARTZ, Leonard S.—Professor of Electrical Engineering, New York University.

SUMMER WORKSHOP 1965

Program Outline and Team Participants

PROJECT A: Space Data Information Systems

Study Topics

A-1 Telemetry computer studies
A-2 Data compression and its application to video signals
A-3 An algorithmic code for monotonic sources
A-4 Information content of a q-level m-sequence with the restriction that the sum be n
A-5 The effect of waveform of correlator reference voltage on the performance of PFM systems
A-6 Confidence level of decision based on the output of correlators of a PFM system
A-7 The distribution of the outputs from a filter bank signal processing system
A-8 A minimum confidence level for signal detection by a correlator bank
A-9 Quality control in the processing of telemetric data
A-10 Some probabilistic models and statistical methods in the analysis of telemetry data
A-11 Degradation of binary coded data due to bit errors
A-12 Recursive estimation in analysis of variance models
A-13 Reconstruction of sampled input of a linear system from its sampled output

TEAM A

Academic Personnel

Dr. C. A. Wogrin
Principal Investigator
Dr. L. D. Davisson
Dr. M. Javid
Dr. F. L. Ramsey
Dr. C. A. Rohde
Dr. E. P. Stabler
Mr. R. D. Coleman
Mr. G. A. Gordon

Goddard Personnel

Thomas J. Lynch
Staff Advisor
Cyrus J. Creveling
Associate Staff Advisor
Dr. R. W. Rochelle
Associate Staff Advisor
John T. Campbell
Rodger A. Cliff
Alan M. Demmerle
Edmund J. Habib
Gerald J. Hogg
Richard L. Kutz
James C. Morakis
Larry R. Muenz
Morton Pasternack
William B. Poland
Joseph Purcell
Thomas V. Saliga
David H. Schaefer
Joseph A. Sciulli
J. W. Snively
John Y. Sos
Curtis M. Stout
James P. Strong
Harry A. Taylor
John J. Yagelowich

NSF Summer Science Student:
Carol VanWie

Part-Time Consultant: Leonard S. Schwartz

BLANK PAGE

PROJECT A: SPACE DATA INFORMATION SYSTEMS

CONTENTS

ILLUSTRATIONS

TABLES

BLANK PAGE

INTRODUCTORY REMARKS ON PROJECT A SPACE DATA INFORMATION SYSTEMS

T. J. Lynch, C. J. Creveling, R. W. Rochelle

The following thirteen (13) papers are arranged, by title, in an order approximating the data flow in a typical GSFC scientific satellite project. The first four papers deal with spacecraft telemetry problems. The next four papers are in the general area of telemetry signal reception and detection. Finally, the last five papers are in the area of data processing, data reconstruction and data analysis.

One of the outstanding problems of space Exploration by unmanned satellites (the main scientific mission of the Goddard Space Flight Center) is in the extraction of information from the masses of data points accumulated by our ubiquitous satellites. Inevitably, these must be reduced into equations, charts, graphs, pictures or models which portray their intrinsic content in such a way that the physical processes they represent can be understood. Space research being a new human endeavor, only the beginning efforts have been made to apply the very considerable experience available in other fields to handle large masses of space data. Project A of the Summer Workshop is one of these efforts to associate university researchers of several disciplines and introduce them to our problems in the hope that their several but related efforts will result in useful solutions to this problem. The following papers present the results of the first Summer Workshop in this area.

In measuring the success of the Goddard Summer Workshop one would look primarily to see if two goals were attained. The first and primary goal is the production of a set of meaningful answers posed by the problems set before the group. The second is whether the University group was sufficiently stimulated to carry back related problems to their schools for further thought and possible action. These reports certainly speak well for the first goal. On the average the reports contain material which is directly applicable to the work here at Goddard and can serve as bench marks for further work in the subjects. Several of the reports which were on more specific problems give definitive answers which stand alone and need no further work. It was not expected that the Workshop would yield these comprehensive reports in the short time which was available for the work. As for the second goal it is hoped that the University group was sufficiently stimulated to carry back some of the related problems; time will yield the full story.

TELEMETRY COMPUTER STUDIES

E. P. Stabler

INTRODUCTION

This report describes the results of a two month study of the feasibility of using special purpose computers to process telemetry data and to control the experiments in small telemetry satellites. The feasibility study comprised three main parts.

1. Investigation of typical computational tasks on small telemetry satellites.
2. Preliminary decisions on the system design of a central telemetry computer regarding memory organization, control of data channels, etc.
3. The use of a formal programming language to describe the computational tasks and the computer operation.

A fourth part of the work involved simulation of the telemetry computer operation in a simulated real time environment. This work was aimed at determining the adequacy of the computer system over a range of real time operating conditions, and determining system parameters such as computational load, waiting time, etc. This portion of the work could not be completed in the time available.

Background

The use of a centralized telemetry computer (CTM) grows more attractive as the space experiments grow in sophistication and in number.

At present, the data processing required for conversion of sensed data to transmitted data, and for control of the experiment is carried out in the experimenters package. Since there are many such packages on a given satellite the possibility of unnecessary duplication of equipment exists. This has not created any major problems in the past because so little on-board data processing is performed. However, the trend appears to be toward increased use of on-board data processing. A central data processor, time-shared by the experiments, may be desirable in future telemetry satellites.

The data processing system design for a typical telemetry satellite should consider the following factors.

1. Space and weight should be used economically. This requires that the amount of hardware and energy required to do the computation should be minimized.
2. The reliability of the system should be high. In the past this has been achieved through testing and by providing some hardware redundancy. With the use of redundant hardware, some failure modes were made less than catastrophic.
3. The data processing program is fixed for the life of the computer. That is, the computer will only run one program.
4. The input data occur and output data must be provided in real time in accordance with certain timing constraints.
5. The average computational load on the computer is low. The computational equipment will probably be idle most of the time.

6. The input-output data format is not uniform. Each input source and each experimental output signal may have a different format.

7. The total amount of equipment and energy devoted to data processing is low by any standards. As a result the central telemetry computer must be small. Sophisticated hardware suitable for larger general purpose computers is not warranted for this application. The data processing equipment on the IMP-F Satellite probably weighs less than 2 pounds, and consumes less than 2 watts.

The contrast between these requirements and those of more conventional computers is dramatic. The system designed and evaluated during this project has the following main attributes related to the previously listed system requirements.

1. Space and weight – The proposed system conserves space and weight by using a single data processor for the experiments on a time-shared basis. Since the average computation rate is low, this is entirely practical. The power consumption is reduced by the use of a passive memory instead of active registers for data. However the system requires a memory to store program information and this memory contributes some weight and occupies some space. Some comparative figures are developed.

2. Reliability – Centralization of the data processing control will introduce new modes of failure which are catastrophic. This change will be off-set by a reduction in the total equipment required, by hardware redundancy in the central control and possibly by the use of redundant encoding of information.

3. Fixed Program – The system includes a high reliability, low power, read-only memory for program storage.

4. On-line Operation – The system includes programmable interrupts which are checked frequently and automatically.

5. Low Computational Load – Serial logic is used everywhere except for transfers into and out of memory. This provides a favorable trade-off of time for hardware. Most of the equipment is turned off when no computation is required.

6. Variable Format – A major attribute of the described system is the extensive use of microprograms to do the computation. The computer has no fixed set of operations carried out on fixed format data. Instead, microprograms are used to interconnect the hardware for a certain period of time. The next set of interconnections are dictated by the next microprogram command. As a result, the computation on different format data can be performed efficiently. This mode of operation is practical because the program does not change during the life of the computer.

7. Small Size – Time sharing of common equipment will tend to reduce the weight and space devoted to data processing. This space and weight figure will increase very slowly as additional tasks are assigned to the computer because the computational load is low.

8. Universality – The central telemetry computer system described can be used for a wide variety of telemetry systems. The principal change required to adapt the computer to the new system requirements is a change in the program stored in memory.

CONCLUSIONS

The principal conclusion of the study is that a CTM will not substantially reduce the weight and energy for a telemetry satellite with few computational tasks such as the

IMP-F. A more positive result is that a CTM could easily be justified for a satellite of IMP-F size if a small amount of on-board data processing is provided for each of the experiments. This conclusion is based on the assumption that a simple conversion from integer to floating point representation of the transmitted data may be acceptable for many of the experimental numbers. Such a conversion provides about a 2:1 reduction in the number of bits needed to represent the number. If this assumption is a good one then the savings in transmitted power (4 watts for IMP-F) would easily justify the use of a CTM. One suspects that much more on-board data processing could be done to save transmitter power, but no study was made of this possibility. The computer described in this report is idle most of the time so it could easily handle more data reduction without an increase in weight.

Some rules-of-thumb which are useful in guiding system decisions for small telemetry satellite computers are described below.

1. On-board Data Processing – Suppose that a given data process is to be carried out on sensed data. A decision must be made regarding whether the process should be done on the satellite or on the ground. If the data process provides a substantial data reduction, then on-board data processing becomes desirable because of the savings in transmitter power. For example, in IMP-F it costs 40 millijoules to transmit one bit to earth. The energy required to convert a 20 bit integer to a 10 bit floating point mode is about 0.2 millijoules using the CTM system described. The energy cost of the data reduction is negligible compared to the savings in transmitter energy. The energy consumption of the computer is very low compared to that of the transmitter. As a result it is easy to justify substantial computation to achieve a small data compression.

2. Serial Logic vs. Parallel Logic – There seems to be little difference energy per unit calculation between the two modes. For a CTM, serial logic seems more desirable because less hardware is needed and because serial logic is easily adapted to the variable length data inputs from different experiments.

3. Passive Storage vs. Active Storage for Data – The question is whether or not it is desirable to provide passive storage, such as core memory, for data. It takes a great deal of energy to read information into and out of a core memory but none to save it. It takes negligible additional energy to read or write a word into a register but it takes energy to save it. Using typical figures for the energy requirements, it seems that the break-even point occurs at about 1 ms. assuming that the data word is about as large as the core memory word. Hence, storing data in a register for 100 ms. or longer is definitely wasteful in terms of energy. As a point of interest, there is not much simple register storage in the IMP-F satellite. This situation is probably typical of small telemetry satellites. The data storage requirements are so small, that efficient storage is not a strong requirement. The considerations described above do not apply to other memory applications such as buffer memories.

4. Wired Logic vs. Stored Program – It is conceivable that the program of a special purpose computer be set by the wiring of the control circuits. This is what is done on the IMP-F satellite. An alternative approach is to provide a program in a read-only memory and a switching network to interconnect hardware elements. The memory output controls the switching network so that the various hardware devices are connected as required by the program. Even for small satellites the choice does not strongly affect the energy or weight requirements. The additional weight and energy required for the memory is a small fraction of the other equipment. The stored program offers major advantages in flexibility during the design of the system because it is easy to

modify the program. If a semi-universal CTM is to be designed for use in several satellites, the stored program provides easy reprogramming, and savings in engineering design effort.

5. Distributed Logic vs. Central Logic – The question is whether a CTM can be justified because it eliminates duplication of logic hardware in the various experimental packages. For small satellites with little data processing, such as the IMP-F, there is not much duplication of equipment. As a result, duplication does not provide strong justification for a CTM unless there is substantial on-board data processing for the few experiments on the small satellite, or unless a larger satellite system is involved.

One important result of the study is the impression that there is a need for an adequate notation or language for describing the logic processes performed in preparing the experimental data for telemetry. At present there are two ways of describing these processes. One is in natural language with reference to figures or block diagrams and the other is by means of logic block diagrams. The first method suffers from the usual drawbacks of incompleteness and ambiguity while the second is difficult to write and difficult to read.

The question is, can a notation be developed which avoids the drawbacks mentioned above. In the following material a language is sketched which is promising in this regard. In this report a special notation or language is used to describe computing tasks, the data collection and the computer activity.

The description of the computer in a formal language permits simulation of the computing system behavior in a simulated real time environment at a very early point in the design. This is possible because the linguistic description is particularly convenient for programming. The simulation results provide valuable information on system characteristics such as percentage of time the computer is on, power consumption, response to real time inputs, and the effect of modified input environments. The simulation may be of major assistance in debugging the design.

ON-BOARD DATA PROCESSING

In this section some typical on-board data processes are described. In order to be specific the IMP-F satellite was used to provide the examples. The data processes are described in a language which is a modification of Iverson's (11) language. The language statements and operations will be introduced as needed in the text. The main format of the language is a table of conditions and their effects. The effects occur some time period after the corresponding conditions arise. In a more conventional programming language the effects are ordered sequentially so that the completion of one effect is all that can be accomplished in a given time period. This is a severe constraint. If the language used to describe the on-board data processing conformed to this constraint, the natural way of describing many processes would not be allowed and great complexity arises. The major differences between conventional programming languages and the language used here are:

1. Several actions can be carried out simultaneously. The notation for this is to put simultaneous effects on the same line, or to have several conditions satisfied.

2. Groups of effects can be formed for which it is understood that only the total effect of the group has any significance as far as machine behavior is concerned. The notation is a bracket around the group.
3. The conditions for initiating an action include external conditions. Hence the various start signals and the programmable interrupt become conditions which can cause effects inside the computer.

Magnetic Field Auto-correlation Experiment

This experiment calls for a major processing of experimental data to obtain auto-correlation figures. The data compression ratio is approximately 15. The CTM will spend more time on this computation than on all the other tasks combined. As a result the nature of this computation strongly affects the CTM design. The CTM must do this calculation efficiently if good overall efficiency is to be obtained. The experimental data consists of 238 8-bit numbers (7 bits and sign in 2's complement form) collected at a rate of one number every 40 ms. at 238 consecutive times during a telemetry sequence. The data collection is halted for the remaining 18 intervals making up a complete telemetry sequence. The output consists of ten 12-bit numbers, each segmented into 3 groups of 4 bits. Let the experimental numbers be

$$E_I, \quad I = 1,238$$

The 10 output numbers are:

$$R_{av} = \left(\sum_{I=1}^{230} E_I \right) \div 16 \quad \text{(11 bits + sign)}$$

$$R_J = F \left(\sum_{I=1}^{230} E_{I+J} \times E_I \right) \quad \text{(7 bits + sign + 4 bits exponent) } J = 0,8$$

where the function F is the conversion from integer representation to a form of floating point which will be described below. The CTM output to the telemetry unit is provided as 30 4-bit numbers in 30 consecutive 20 ms. time intervals, at the completion of the data collection period. We can now state more precisely the relationship between the output and input data in terms of register transfers.

E	– 8 bit binary vector	(experimental no.)
R	– 24 bit binary vector	(sum of experimental nos.)
A_J	– 24 bit binary vector	(related to auto-correlation data)
0	– 4 bit binary vector	(output data)
B_J	– 12 bit binary vector	(floating point version of A_J)
TD_I	– 1 bit binary	(condition for initiating a data read)
T_0	– 1 bit binary	(condition for initiating reset of experiment)
TE_I	– 1 bit binary	(condition for generating a new output)
D_J	– 24 bit binary vector	(stored experimental nos.)

The values of E, TD_I, T_0, and TE_I are all controlled by equipment external to the CTM.

Condition	Effect
$T_0 = 0$	$A \leftarrow 0$; $D \leftarrow 0$

This is interpreted as follows: When condition ($T_0 = 1$) is satisfied the effect is that A is set to zero. A is a matrix whose rows are the 24 bit vectors A_J. Setting the A matrix to zero means setting all the elements to zero. D is set to zero also.

$$TD_I \Downarrow D;\ D^1 \leftarrow (\epsilon(16) \times E_8) \circledcirc E$$

Before completing this program, some explanation of the meaning is needed. Iverson uses a symbol $\downarrow$ to signify circular rotation of the rows of a matrix. Circular rotation in this case means move each row of the matrix down one row and make the bottom row the top row of the matrix. In this report a new symbol, $\Downarrow$, is used to denote shifting down of one row. The difference between shifting and rotating is that the bottom row is not moved up in a shift. The top row remains unchanged unless a simultaneous operation modifies it. D^1 denotes the top row of D. $\circledcirc$ is a concatenation operation between two vectors. $A \circledcirc B$ denotes the concatenation of vectors A and B with vector A on the left. $\epsilon(16)$ is a constant vector with 16 elements all equal to 1. Therefore $E_8 \times \epsilon(16)$ denotes a vector with 16 elements all equal to E_8 where E_8 is the 8th element of E. Combining all these operations shows that the second part of the command calls for setting the top row of D equal to a vector formed by concatenating E with a 16 element vector all of whose elements are equal to E_8. The above notation is almost identical to that of Iverson which is fully explained in the referenced book. The use of simultaneous operations is new.

D^1 becomes a 24-bit 2's complement number equal to E, which is an 8-bit number.

Example:

$$E = (1\ 0\ 1\ 1\ 0\ 1\ 1\ 0)$$
$$\epsilon(16) = (1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1)$$
$$E_8 \times \epsilon(16) = (1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1)$$
$$(E_8 \times \epsilon(16)) \circledcirc E = (111111111111111110110110)$$

The D matrix and the A matrix are portions of the data memory whose words are 24 bits long. The complete description of the processing of a piece of experimental data is shown below.

$$DP \equiv DMIO \circledcirc A0;\ \ AC2 \equiv \alpha^9(24)/ACC;\ \ R \equiv \omega^9(24)/REG2$$

TD_I $\quad \Downarrow D;\ D^1 \leftarrow (E_8 \times \epsilon(16)) \circledcirc E$

$$T \leftarrow 24,\ CA \leftarrow 0$$

$$\rightarrow \downarrow RAV;\ \tau D^1;\ T \leftarrow T - 1;\ CA \leftarrow M(D_1^1,\ CA,\ RAV_1),\ RAV_{24} \leftarrow \bullet(D_1^1,\ CA,\ RAV_1)$$
$$\geq T:0$$

```
                R ← E_8 ⊙ E;  J ← 0
u               DMIO ← D^(1+J);  A0, ACC ← 0;  T2 ← 9
x   ┌──────►    A0: DMIO_1  (=, >, <) → (SHIFT, ADD, SUBT)
ADD │           T ← 9;  CA ← 0
    │  ┌──►     ↓R;  TAC2;  AC2_9 ← ⊕(AC2_1, CA, R_1);  CA ← M(AC2_1, R_1, CA);  T = T - 1
    │  └──≥     T: 0
y   │           T2: 1  = ──────► z
    └─────      T2 ← T2 - 1;  TACC;  ↓DP
SUBT            T ← 9;  CA ← 1
    ┌──►        TAC2;  ↓R;  T ← T - 1;  AC2_9 ← ⊕(CA, AC2_1, R_1);  CA ← M(AC2_1, R_1, CA)
    └──≥        T: 0 = ──────► y
z               ACC ← 8  ↓ACC
                STORE SUM (A^(1+J));  J ← J + 1
                J: 9 (=, <) → (EXIT, u)
```

In the above

$$DP \equiv DMIO \odot A0$$

asserts that DP is the name of the vector formed by concatenating DMIO with A0. $\oplus$ is the exclusive or operator and M is the majority operator. STORESUM is a subroutine listed below. The flow of control is the same as that described by Iverson and is normal for programming languages. To convert this to the condition-effect form it is only necessary to establish a condition each time which will cause execution of either the next lower command or the command which the program should branch. $\omega^9(24)$ is a constant vector with only the rightmost nine elements equal to 1. $\omega^9(24)/REG2$ is a masking operation on REG2, selecting the leftmost nine elements of REG2. $\alpha^9(24)$ is a 24 element constant with only the leftmost nine elements equal to 1.

SUBROUTINE: STORESUM (B)

```
    DMIO ← B;  CA ← 0;  T ← 24
┌──►  ↓ACC;  TDMIO;  T ← T - 1;  DMIO_24 ← ⊕(CA, DMIO_1, ACC_1);  CA ← M(CA, DMIO_1, ACC_1)
└──≥  T: 0
    B ← DMIO ──────► RETURN
```

This is the complete calculation except for setting various mode registers to indicate the calculations are complete and the interrogation of the START register. The important things to notice about this notation are that it permits description of the actual

register interactions, description of both simultaneous and sequential processes, and of equivalent processes. For example, the multiply which takes place is described completely. This is important because the method of multiplying is a novel one which avoids the difficulties introduced by negative multipliers.

The description is sufficiently precise for the logic design to proceed directly from this description. In fact, a description of this type might be used eventually as the input to a computer program which would produce the logic design.

Compare this description with a word description of the same processes. The word description becomes complex and ambiguous very quickly. Compare this description with the logic diagram of the equipment to carry out the processes. The description given is far easier to read and to write, and requires less description of peripheral events such as timing waveforms.

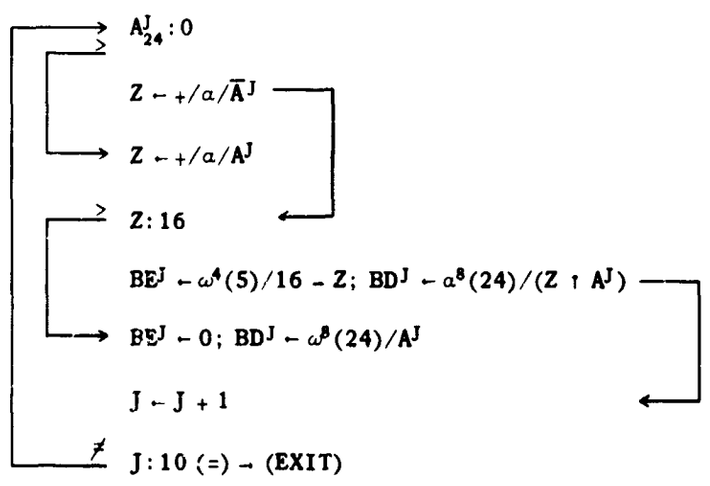
The setting of mode registers and the interrogation of the START register to check priority can be obtained from the symbolic program given in a later section.

The conversion of A_J from a 24 bit signed integer to a 12 bit floating point number is given below.

COMMENT: This program converts the auto-correlation numbers to prepare them for transmission. RAV is divided by 16 and the A^J vectors are converted to a 12 bit floating point mode. Z is a 5 element boolean vector, interpreted as a binary number where necessary. Z is the number of leading zeroes or ones in A^J.

$$BD^J \equiv \omega^8(12)/B^J;\ BE^J \equiv \alpha^4(12)/B^J$$

TE_c $\quad RAV \leftarrow 4 \top RAV;\ J \leftarrow 1$



α is the prefix vector. α/A denotes a vector of same dimension as A, and only the string of consecutive 1-elements starting on the left of A are equal to 1. All other elements are 0. The value of $+/A$ is the number of 1's in A. Hence $+/\alpha/A$ is the number of consecutive 1's in A starting from the left.

Example: $A = (1110101101)$

$\alpha/A = (11100000000)$

$+/\alpha/A = 3$

The next program describes the preparation of the telemetry signals.

COMMENT: This program generates a 4 bit output word each time condition TE = 1 is satisfied, indicating a call for a telemetry word.

$APRIME^1 \equiv RAV$; $APRIME^{1+J} \equiv A^J$

$J \leftarrow 0$; $K \leftarrow 1$

TE → $J:1$

$J \leftarrow J + 1$; $OUT \leftarrow \omega^4(24)/APRIME^K$, $4 \downarrow APRIME^K \rightarrow EXIT$

$K \leftarrow K + 1$; $J = 0$

$K:11\ (=, \neq) \rightarrow (EXIT, TE)$

This completes the description of the data processes required by the magnetic auto-correlation experiment.

Plasma Experiment

This is the only other experiment data process which will be described in detail in this section. The remaining data processing tasks are very simple and a description of them can be obtained from the symbolic program of the computer.

The plasma experiment processing has a data compression ratio of approximately 19. For each of 5 consecutive spins a plasma count C_I^J is taken during each 1/16 spin interval. C_{10}^2 would be the count during the 10th interval of the second spin. A total count per spin CT_J, the sum of squares of the individual counts ST_J, and a 4-bit indication of the spin interval during which the count was largest IM_J, are computed. The CT_J figures are converted to an 8-bit floating point notation prior to transmission. The exponent of CT_J is used to select a 4-bits of ST_J for transmission. The process quantities are

E –	19 bit binary vector	the experimental number
I –	4 bit binary vector	experimental interval
J –	a number between 1 and 5	spin count
CT^J –	24 bit binary vector	sums of spin counts
ST^J –	26 bit binary vector	sums of squares
TSP16 –	1 binary bit	indicates end of 1/16 spin
MAX –	16 bit binary vector	maximum count for present spin
SQ^J –	4 bit binary vector	indicates value of sum of squares
EXP^J –	4 bit binary vector	indicates exponent of spin count
D^J –	4 bit binary vector	indicates value of spin count

COMMENT: This program describes the data process initiated by the completion of 1/16 of a spin.

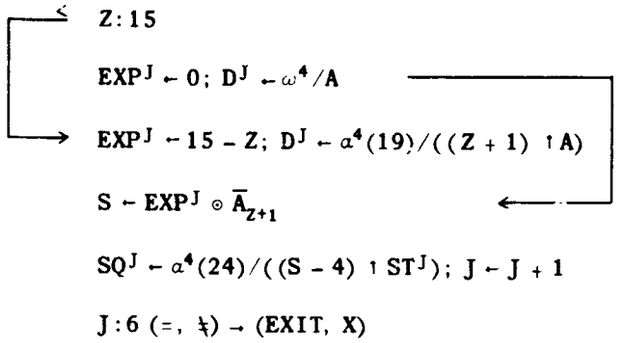
$PART\,1^J \equiv \omega^{24}/CT^J$; $PART\,2^J \equiv \epsilon^1(24)/\alpha^{25}/CT^J$;

$DP \equiv DMIO \circ A0$; $AC3 \equiv \omega^{12}(24)/ACC$; $AC4 \equiv \alpha^{12}(24)/ACC$

```
                I, J ← 1
TSP16           ACC ← 0 ⊕ α^17(19)/E
                STORESUM (CT^J)
                ≥ MAX: ACC
                → REG2, DMIO ← 5 ↑ACC; A0, ACC ← 0; T2 ← 12
X               → A0: DMIO_1 (=, >, <) → (SHIFT, ADD, SUBT)
                → IM_J ← I; MAX ← ACC
ADD             T ← 12, CA ← 0
                → . REG2; ↑AC4; T ← T - 1; CA ← M(AC4_1, REG2_1, CA); AC4_12 ← ⊕(AC4_1, REG2_1, CA)
                ≥ T: 0
Y               T2: 1 (=) → (Z)
SHIFT           ↑ACC, DP; T2 ← T2 - 1; → X
SUBT            T ← 12; CA ← 1
                → ↓REG2; ↑AC4; T ← T - 1; CA ← M(AC4_1, REG2_1', CA), AC4_12 ← ⊕(AC4_1, REG2_4', CA)
                ≥ T: 0 (=) → Y
Z               ↑ACC; DMIO ← PART1^J; T ← 24; CA ← 0
                → ↑ACC, DMIO; T ← T - 1; DMIO_24 ← ⊕(DMIO_1, ACC_1, CA); CA ← M(CA, ACC_1, DMIO_1)
                ≥ T: 0
                PART1^J ← DMIO; PART2^J ← CA; I ← 16 |_1 (I + 1)
                I: 1 (=) → (EXIT)
                J ← 5 |_1 (J + 1)
                J: 1 (=, ≠) → (HALT, EXIT)
```

COMMENT: The following program describes the processing required to prepare the plasma experiment data for transmission.

```
T_c     J ← 1
X       A ← ω^19(24)/CT^J
        Z ← +/α/Ā
```



Other Data Processes

The only other experiment computations carried out by the computer processes data from the attitude sensor on the angle of the sun and on the time of sun pulses and earth pulses. There is virtually no data processing on these data prior to transmission. However, the times of occurrence of the successive sun pulses are used to compute the length of time required for 1/16 of a spin. The TSP16 spin pulses are generated by coincidence between a 12 bit settable register (EVENT) and a portion of the clock. When such a coincidence occurs EVENT is reset by adding the value of the 1/16 spin interval time which is stored in memory. Typical condition-effect pairs are given below.

The master clock is a 40 bit counter CL counting down from 200 kc.

$$\text{CLOCK} \equiv \omega^{12}(24)/(7 \downarrow \text{CL})$$

Condition	Effect
EVENT = CLOCK	TSP16 ← 1

All of the other timing pulses occur at fixed points in the telemetry sequence. The start register (SR) is modified by a TSP16 pulse and by various timing pulses.

For example,

Condition	Effect	Remarks
$64\vert_1$ (CLOCK) = 5	SR(1) ← 1	Typical call for telemetry

Computational Load

In this section the computational loads of the major computing tasks are determined approximately. These figures depend on the time to carry out the individual operations, and the frequency with which the operations are performed.

For the magnetic auto-correlation experiment a standard operation is performed 238 times per frame. The time required for this is about 850 clock intervals. Hence a total of 202,000 time units per sequence are required for updating the auto-correlation numbers during each sequence. Assuming a 200 kc logic rate about 5% of the machine time is spent on this calculation. In a similar way, the computational load of the other data processes can be calculated.

Task	Time/Sequence	% of time/Sequence
Auto-correlation data input	202,000	5.050
Auto-correlation Figure Conversion	750	0.019
Auto-correlation Telemetry	6,000	0.150
Auto-correlation Reset	100	.003
Plasma Data Input	27,280	.682
Plasma Figure Conversion	750	.019
Plasma Telemetry	14,000	.340
Attitude Data Sun	350	.009
Attitude Data Earth	200	.005
Totals	~ 251,000	~ 6.3

All of these figures are approximate but it is safe to conclude that the magnetic auto-correlation experiment provides about 80% of the load and that the computer will be off about 90% of the time. As a result the energy consumed by the computer will consist principally of standby power which can be very low. Additional computational tasks may be assigned to the computer without increasing the equipment requirements.

GENERAL

System Description

The computer described is a microprogrammed computer. A microprogrammed computer is distinguished from more conventional computers by the method used to define and perform macro-operations. In conventional computers a few standard data formats are adopted and a set of universal arithmetic and control operations are defined. The operations are carried out by the control system of the computer using wired logic. The justification for adopting a fixed data format is that the computer will be used for a processing variety of kinds of data and the formats are not known at the time of the computer design. Once a few fixed data formats are adopted, the adoption of standard arithmetic operations is natural and efficient. For a telemetry computer the format of the data to be processed, and the data processes are known at the time of the design. The computer will normally be used only on one mission, so in a sense it runs only one program. For this reason, it is preferable to adopt efficient data processes and formats for the specific task at hand. This can be done efficiently and economically using microprogramming. The computer is thought of as a collection of storage elements, logic gates and a switching network for interconnecting the various pieces of hardware. A micro-operation consists of a particular interconnection. A macro-operation is defined as a sequence of micro-operations where the sequence is stored in memory as successive memory words. In this type of computer, a macro-operation is similar to a subroutine with efficient call and return. The microprogramming concept permits the telemetry computer to have both an 8 bit multiply and a 12 bit multiply with negligible increase in size and weight. For the task assigned to the computer it is convenient to have both and it is known that there will never be a need for some unanticipated process such as a divide. Each macro-operation, such as an 8 bit multiply, is defined by a microprogram which is stored in memory. The micro program steps are read in order when an 8 bit multiply is required. It is clear that adding or modifying a macro-operation now requires principally a change in the contents of a memory. This feature makes it feasible to design a universal telemetry computer for use in a wide variety of telemetry satellites. The universal computer would be given microprograms appropriate to the particular satellite and experiments for which it is to be used.

The computer is required to operate in real time, responding to demands for computation within certain time constraints and according to some priority rules. This is accomplished by breaking up all computations into segments which require less than 1 ms. When a segment is complete the demand is checked to see whether a higher priority demand has been made. If a higher priority demand has been made, the computation is interrupted, the register contents are saved and the higher priority computation is begun. Later the computer returns to complete the interrupted task. Because of the low computing load, a very simple interrupt system is sufficient for this computer. In a more sophisticated system it is necessary to allow interruption of tasks which were themselves initiated through an interrupt. Such systems do not require much additional hardware but the control becomes more complex.

The computer design calls for a 64 word data memory of 24 bits per word and a 512 word program memory with 16 bits per word. The program memory is not modified by the computations so a read-only memory is adequate. There are a number of active registers for storing and shifting data during computation. The registers are listed on a separate page. The transfers into and out of memory are in parallel but all register transfers are serial. As a result the switching network which interconnects the registers and the logic is small. The switching network must set up a wide variety of connections but serial logic reduces the number of gates which must be controlled.

The program memory and data memory both have an access time of about 5 μsec. The average number of memory operations is reduced by a large factor by the use of serial logic. The program memory will be read an average of about 20,000 times per second and the data memory used approximately 4,000 times per second while the computer is operating.

Figure A-1 shows in a block diagram some of the computer hardware and the switching network for interconnecting the hardware. Figure A-2 shows how the computer could be fitted into an IMP-F type system.

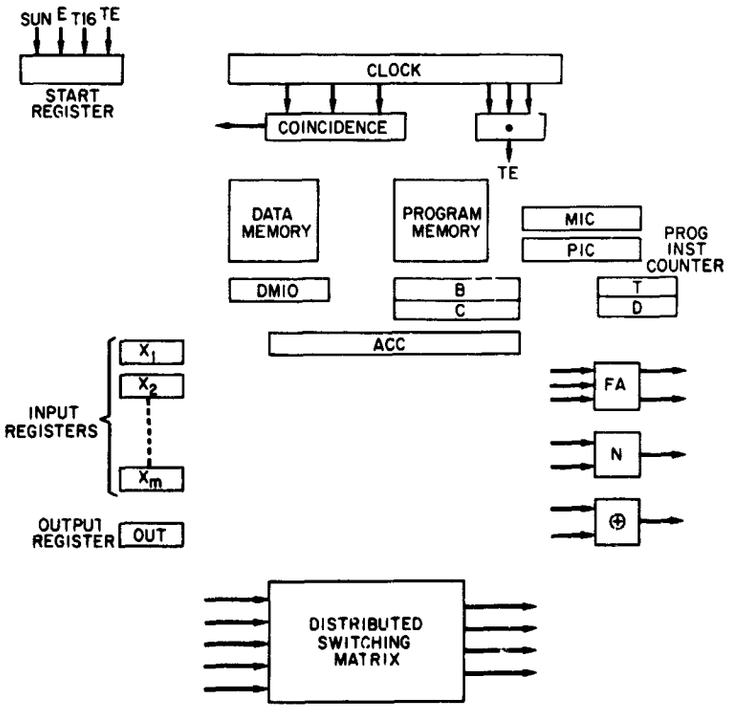


Figure A-1–Microprogrammed Computer for Telemetry Satellite

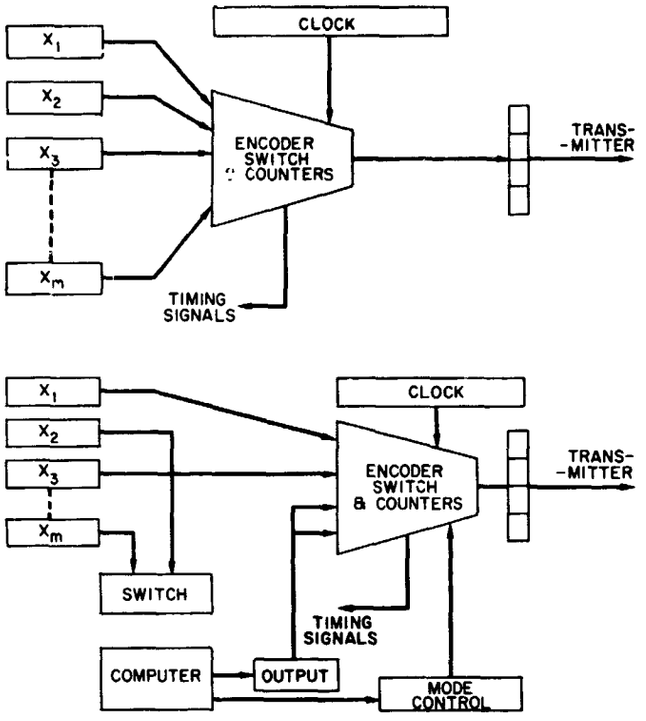


Figure A-2–Fitting the CTM into the Satellite Data Processor

SYSTEM COMPARISON

One of the purposes of this study was to compare the weight and power requirements of a satellite data processing system using a central computer with a system which has distributed data processing equipment. The most pertinent point of the comparison is that there is little difference between the systems unless substantial data processing is performed on the satellite. There is not enough data processing on IMP-F to make a major difference between the systems. This is not an entirely negative result. It implies that a central computer system is practical even for very small satellites. Since the computer on IMP-F would be idle 90% of the time, the satellite system has available an unused stock of data processing capability. New data processing tasks can be done on-board with minor changes in weight and power.

Accurate estimates of the weight and power required by the central computer system will depend on the hardware chosen to implement the computer. Most of the memory is a read-only type and many novel hardware implementations have been suggested for this application. Less than 64 words of conventional core memory is needed. A core memory of 512 words, 12 bits per word is being built for NASA which will consume approximately 200 mw. running at an average rate of 16 kc. A memory of this type could be used for both program and data storage in a system of the type described. The memory would be idle, consuming 5 mw, most of the time.

Approximately 175 register stages are needed by the central computer. Compensation for this is obtained by a reduction in the number of register stages in the plasma, attitude and auto-correlation experiment boxes. Approximately 130 register stages are eliminated from the plasma experiment. Two small magnetic memories with associated drivers and sense amplifiers are eliminated from the auto-correlation equipment in addition to approximately 80 register stages. Approximately 30 register stages are eliminated from the attitude sensing experiment.

Although these comparative figures are incomplete, they are sufficient to indicate that no major differences in weight and power should exist between the two possible implementations of the IMP-F system. In future systems it will be worthwhile to compare other system parameters such as reliability and flexibility before making a choice. In addition, increased data processing load will tend to make the central computer system more attractive.

Operation Example

Some feeling for the method of operation can be obtained by an example of the initiation of a calculation. Assume the computer is in a quiescent state, having done all the computations required. SR, the start register, is at 0, the clock is going and is being monitored by various coincidence gates which can start the computer when preset conditions occur. All the other registers in the computer are shut off to conserve energy. Now, assume the clock reaches a point where it is required to perform an auto-correlation calculation. SR_1 is set to 1. SR is used as an address to provide a branch to the appropriate point in the program memory. The value of SR determines where the program shall begin and therefore the calculation performed. The first instruction is read from the program memory and is executed. If the instruction calls for a microprogram, as most instructions do, control is transferred to the appropriate point in the program memory for execution of the microprogram. The computer will not return to the macroprogram until it completes the microprogram. Completion of one step of a microprogram initiates the reading of the next microprogram step. This procedure continues until the computation is completed or is interrupted. At intervals of 1 ms or less the value of SR is compared to the priority of the calculations being performed. If a high priority demand exists, then the calculation is interrupted. Assume that an SP16 pulse occurs while the auto-correlation calculation is being carried out. When SR is checked the high priority demand is noted. The point of interruption and the register contents are saved before branching to the high priority program. Completion of the high priority program causes a return to the interrupted calculation, with the registers restored to the correct values.

INSTRUCTION DECODING AND CONTROL

There are two formats for an instruction word, one for a microprogram instruction and one for a macroprogram instruction. The parts of a macroprogram instruction are:

K – 1 bit – indicating whether a microprogram is to be executed

B – 5 bits – indicating the operation to be performed

C – 9 bits – indicating an address

The two parts of a microprogram instruction are:

D – 10 bits – indicating the set of interconnections required

T – 6 bits – indicating the time for which the connections should be held

When K = 1 there is an automatic branch to the microprogram indicated by the contents of B. When D = T = 0 there is an automatic return from microprogram to macroprogram. When T = 0, D > 0 there is a read of the next microprogram instruction.

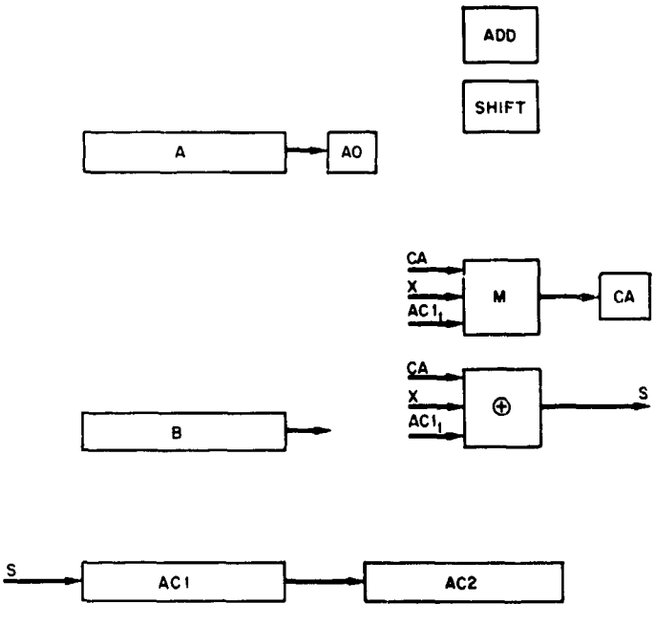
All of these processes are more precisely defined in the table of conditions and effects describing the computer operation.

The data memory can only be read into the DMIO register or written from this register. The program memory will be read into the B, C registers or the T, D registers depending on whether a microprogram is being executed or not.

ARITHMETIC UNIT

The method of doing arithmetic in this computer is different from that normally adopted. The method handles negative multipliers in a particularly convenient way.

The arithmetic is serial so a very small amount of hardware is needed. The numbers are represented in 2's complement form. In Figure A-3 the hardware used for multiplying is shown. Register A holds the multiplier, B, the multiplicand, ACC1 the partial product still being formed and, ACC2 the completed partial product. There are also 4 registers; CA, AO, ADD, and SHIFT. AO acts as an extra stage on A and is set originally to 0. CA is the carry, ADD equals 1 when an addition is called for and SHIFT equals 1 when a shift is required.



$X = ADD \wedge B_1 \vee \overline{ADD} \wedge B_1$
SET ADD IF $AO \wedge \overline{A}_1 = 1$
RESET ADD IF $\overline{AO} \wedge A_1 = 1$
SET SHIFT IF $\overline{AO \oplus A_1} = 1$
SET CA IF $M(X, CA, AC1_1) = 1$
RESET CA IF $M(X, CA, AC1_1) = 0$

Figure A-3–Arithmetic Unit

The principle of the multiply is to successively add B to ACC1, subtract B from ACC1 or to shift. The action taken depends on AO and A_1, the lowest order stage of A.

A_1	AO	Action
0	0	Shift
0	1	Add
1	0	Subtract
1	1	Shift

Subtraction is carried out using the same hardware as addition but CA = 1 initially and B is inverted before entering the full adder.

For a 9 bit multiply A, B, AC2, and AC1 are 9 bit registers.

ACC ≡ AC1 ⊙ AC2

AP ≡ A ⊙ AO

↓ is a shift without rotation. The right most bit is lost and the left most bit remains unchanged.

M (a, b, c) is the majority function

Now the complete multiply operation can be specified.

CONDITION	EFFECT
Initial	T2 ← 9; AO ← 0; SHIFT, ADD ← 0; T1 ← 9
T2.GT.O ∧ T1.EQ.9	ADD ← AO ∧ $\overline{A1}$; SHIFT ← $\overline{AO \oplus A1}$; CA ← $\overline{AO}$ ∧ A1
SHIFT. EQ. 1	↓ACC, AP; T2 ← T2-1; SHIFT ← 0; T1 ← 9
SHIFT.EQ.O ∧ ADD.EQ.1 ∧ T1.GT.O	↓AC1; ↓B; $AC1_9 \leftarrow \oplus$ ($AC1_1$, CA, $\overline{B1}$); CA ← M($AC1_1$, CA, $\overline{B_1}$); T1 ← T1-1
SHIFT.EQ.O ∧ ADD.EQ.O ∧ T1.GT.O	↓AC1; ↓B; $AC1_9 \leftarrow \oplus$ ($AC1_1$, CA, B_1); CA ← M($AC1_1$, CA, B_1); T1 ← T1-1
SHIFT.EQ.O ∧ T1.EQ.O	SHIFT ← 1

The operation of the multiplier can be illustrated by a simple example. Assume a 4 bit multiply of -4 × 3.

Let B = 0011 = 3

A = 1100 = -4

The first two operations are simple shifts because A_1 = AO = 0. The next operation calls for a subtraction of register B from the accumulator. The resulting configuration of AC1, AC2 after shift is

1	1	1	0

AC1

1	0	0	0

AC2

The next and last operation is a simple shift. The resulting accumulator content is 1 1 1 1 0 1 0 0 or -12 in 2's complement.

Simple addition and subtraction of numbers makes use of the same hardware as is used for multiply.

No provision for dividing has been included.

REGISTER LIST

DMIO	24 Bits	Normally used for data memory IO
ACC	24 Bits	Accumulator
REG 2	24 Bits	Normally used for data input
REG 3	12 Bits	Used for storing index value
REG 4	12 Bits	Used to store return address
CNTR	5 Bits	A counting register for general use
T	5 Bits	A counter used to time microprograms
D	11 Bits	Used to control microprogram conns.
B	6 Bits	Operator of instruction
C	9 Bits	Address of instruction
K	1 Bit	For branch to microprogram
SR	4 Bits	Start register
PRIORITY	4 Bits	Indicates Priority of present calculation
EVENT	12 Bits	This register is set to produce an SP16 pulse
PIC	9 Bits	Program instruction counter
MIC	9 Bits	Microprogram instruction counter

START REGISTER (SR)

This register acts as a start button for the computer

SR(1) is set to 1 when an auto-correlation sample is ready
SR(4) is set to 1 when 1/16 of a spin is complete.
SR(3) is set to 1 when a sun pulse occurs
SR(2) is set to 1 when an earth pulse occurs

SR(4) has the highest priority

EXTERNAL REGISTERS

CL	35 Bits	Countdown register from basic clock freq.
CLOCK	12 Bits	Not a separate register, Bits 8 thru 19 of CL
MAC	8 Bits	7 bits + sign for auto-correlation input
PLASMA	17 Bits	Plasma count in 1/16 spin
OUT	4 Bits	Output to telemetry

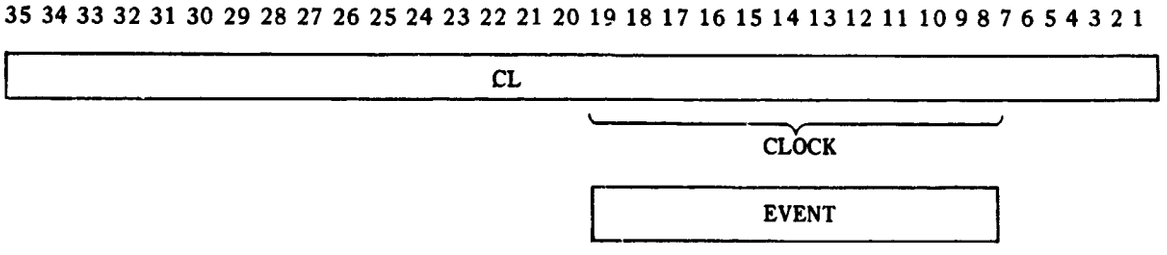
SPECIAL ONE BIT REGISTERS

MODE	Interrupt indicator – Mode = 1 if interrupt has occurred
CA	Carry – for add and subtract
ADD	Add = 1 for Add, Add = 0 for subtract
PLASACT	Activity of plasma experiment

TIMING CONSIDERATIONS

The primary clock runs at approximately 200 kc, giving a 5 μs basic computer interval. In the simulation experiments a 35 bit word CL keeps track of the number of computer intervals which have passed. The CLOCK referred to in the machine description consists of the 8th through 19th bit positions of CL. The EVENT register which is used to generate the 1/16 spin pulses monitors these bit locations.

$$\text{CLOCK} \equiv \omega^{12}(35)/(7 \downarrow \text{CL})$$



The 23rd bit has a 20.48 second period, one sequence. The 14th bit has a 40 ms period, one burst or blank. Many computer tasks can be initiated at fixed times in the sequence interval. Simple gates with CL inputs can be used to start the computer on these tasks.

PRIORITY CONTROL SYSTEM

The start register indicates the demand for machine computation. A register, PRIORITY, indicates the priority of the computation presently being carried out. At frequent intervals the computation is interrupted to compare the demand with the priority of the computation being carried out. If a new demand with greater priority has been made, the interrupt point is recorded, the register contents are recorded, and the computer begins on the more important program. When the computation is complete, the computer returns to the less important task.

It is not difficult to provide a large number of interrupt levels. This permits interruption of programs which were themselves initiated by interrupts. However, for the computing tasks assigned to the computer under study, one level is sufficient.

In order to ensure proper priority control the macroprograms must include the following.

1. When entering a program from the halt mode, store the priority value of the computation in priority.

2. When entering a program through an interrupt, store the value of the PIC and REG 2 and set the priority mode to prevent further interrupts.

3. When exiting from an interrupting program restore the registers to their former values.

4. To interrupt in order to check priority use "BL MONITOR" instruction at least once every millisecond.

MACRO OPERATIONS

In this section a table of macro operations is given along with a description of the effect of the operations. In addition estimates of the times required to execute the various macros are given.

Table A-1
Macro-operations

Symbol	Name	Description	Micro.	No.
HB	Halt and Branch	PIC · SR	N	0
SC	Subtract Constant	AC2, D2 ← D2 - C	Y	1
ZEROD	Zero	DMIO ← 0	N	13
SET TE		SR_1 · C	N	2
INPUT MAC		REG2 ← \MAC$_8$ × ϵ(16), ω^8(24), MAC\	Y	3
R DC	Read Constant	D2 ← C	Y	14
ADD2		ACC, DMIO ← REG2 + DM^C	Y	4
STRACC	Store Acc	DM^C · ACC	Y	15
STRD	Store DMIO	DM^C ← DMIO	N	5
BL	Branch and Load	REG4 ← PIC, PIC ← C, MODE ← 1		
RD2	Read into REG2	REG2 ← DM^C	Y	6
RD	Read Data	DMIO ← DM^C		
M	Multiply	ACC ← DM^C × REG2	Y	7
SHIFT DM		DMIO ← C . DMIO		
STR2	Store REG2	DM^C ← REG2	Y	8
RD ACC	Read into ACC	ACC ← DM^C	Y	10
SD	Subtract data	AC2 ← AC2 - ω^{12}(24)/DM^C	Y	11
BZ	Branch on Zero	IF ACC.EQ.O, PIC ← C	N	63
B	Branch	PIC ← C	N	62
BO	Branch on odd	IF(ACC$_1$.EQ.1), PIC · C	N	61
BP	Branch on Pos.	IF(ACC$_{24}$ .EQ.O), PIC ← C	N	60
OUT	Output	Out ω^4(24)/DM^C, DM^C ← 4 ↓ DM^C	Y	12
FLOAT 1		EXP ← + (2/(DM$^C_{24}$ × ϵ(24)) ⊕ DM^C) ω^4/(8 ↓ DM^C) ← EXP ω^8/(DM^C ← EXP · DM^C	Y	17
SQSP	Square 1	DM^{C+X} · [$2^{24}|_0$ ((5 ↓ REG2) × (5 ↓ REG2))] + DM^{C+X} DM^{C+X+1} ← DM^{C+X+1} + ([DM^{C+X} + (5 ↓ REG2) × (5 ↓ REG2)] ÷ 2^{24})	Y	18
SQ	Square	DM^C · DM^C + REG2 × REG2	Y	21
SADDX		DM^{C+X} ← REG2 + DM^{C+X}	Y	24
SD2X	Subtract data	ACC ← REG2 - DM^{C+X}	Y	25
RESET IO		Attitude Activity ← 0	N	26
FLT	FLOAT 2	EXP ← + ($\bar{\alpha}$/(ω^{19}(24)/DM^C)) IF (EXP.LT.15), ω^4/(4 ↓ DM^C) · 15-EXP ω^4/DM^C ← α^4/((5 + EXP) ↓ DM^C) SH ← \EXP, α^4(5), DM$^C_{18-EXP}$\	Y	27
PRSQ	PRSQ	A ← \DM^{C+X+1} , ω^{24}(28), DM^{C+X}\ DM^C ← α^4(28)/(SH ↑ A)	Y	28
SET EV	SET EVENT	EVENT ← ω^{12}(24)/ACC	Y	29
SET SP16		SR_4 · C	N	30

Table A-1 (Continued)

Symbol	Name	Description	Micro.	No.
SET PLASMA	Plasma Activity	Plasma Activity C	Y	34
ADD1		DMIO ← DM^C + ACC	Y	35
INP PLAS	Input Plasma	REG2 ← \0, $\omega^{17}(24)$, Plasma Count\	Y	36
BM	Branch on Mode	IF (MODE.EQ.O) PIC ← C	N	56
SSR	Subtract SR	IF (PRIORITY.GT.SR), ACC_{24} ← O IF (PRIORITY.LTE.SR), ACC_{24} ← 1	Y	37
IBP	Indirect Branch	IF (ACC_{24} .EQ.O), PIC←REG4	N	55
SETMODE		MODE←C_1	N	38
STR PR	Store Priority	DM^C←PRIORITY	Y	39
RD3	Read into REG3	X←DM^C	Y	42
RDPR	Read into PR	PRIORITY←DM^C	Y	40
SETPR	Set Priority	PRIORITY←C	N	41
OUTX	Output	out←$\omega^4(24)/DM^{C+X}$, DM^{C+X}←4↓DM^{C+X}	Y	43
STRDX	Store Data	DM^{C+X}←DMIO	Y	44
RDCL	Read Clock	REG2←CLOCK	N	45
RDANGLE		DMIO←ATTITUDE ANGLE	Y	46
SETSUN		SR(3)←C	N	47

AC2 ≡ $\omega^{12}(24)$/ACC
D2 ≡ $\omega^{12}(24)$/DMIO
X ≡ REG3

Table A-2
Times for Macro-operation Execution
(Machine Units)

Operation	Time	Operation	Time	Operation	Time
SC	26	OUT	8	SET SP16	1
SETTE	1	ZERO D	1	SET PLASMA	1
INPUT MAC	26	STRACC	27	ADD1	30
ADD2	28	FLOAT 1	80	INPUT PLASMA	26
STRD	1	SQSP	90	BM	1
RD2	27	BL	1	SSR	7
M	50	SQ	55	IBP	1
RDC	14	RD	1	SETMODE	1
STR2	27	SHIFT DM	6	STPR	7
RDACC	27	SADDX	42	RDPR	7
SD	16	SD2X	41	SETPR	1
BZ	1	RESET10	1	RD3	15
B	1	FLT	65	OUTX	22
BO	1	PRSQ	65	STRDX	16
BP	1	SETEV	14	RDCLOCK	1
				RD ANGLE	26
				SETSUN	1

SYMBOLIC PROGRAM

In this section the symbolic program is given which makes up the macroprogram stored in the program memory. It is far easier to program in this symbolic manner without reference to specific memory locations. However, each instruction shown in the symbolic program does occupy one memory location. The meaning of the symbolic macro operations can be obtained from the table of macro operations. A symbolic data memory array is shown also.

This symbolic program makes explicit the priority control system, the data process details and the flow of program control between the various computing tasks.

Table A-3
Symbolic Program

Label	OP	Address	Comments
0	HB	-	
1	B	TE	
2	B	EARTH	
3	B	EARTH	PRIORITY
4	B	SUN	BRANCHING
5	B	SUN	
6	B	SUN	
7	B	SUN	
8	B	SP16	
9	B	SP16	
10	B	SP16	
11	B	SP16	
12	B	SP16	
13	B	SP16	
14	B	SP16	
15	B	SP16	
16 TE	SETTE	0	RESET SR(1)
	SETPR	1	
	RD	TEN	TEN is a count of TE pulses MOD 512
	SC	475	
20	BP	BZ	
	BO	MAC	
22 RESETE	RD	TEN	INCREMENT
	SC	-1	TEN
	STRO	TEN	
	B	TESP	
26 MAC	BL	MONITOR	Begin Auto-correl.
	INPUT	MACWORD	Calculation
	ADD2	RAV	
	STRD	RAV	
30	SQ	RO	
	BL	MONITOR	
	M	C8	
	ADD1	R8	
	BL	MONITOR	
	M	C7	
	STRD	C8	

Table A-3 (Continued)

Label	OP	Address	Comments
	ADD1	R7	
	BL	MONITOR	
	M	C6	
40	STRO	C7	
	ADD1	R6	
	BL	MONITOR	
	M	C5	
	STRD	C6	
	ADD1	R5	
	BL	MONITOR	
	M	C4	
	STRD	C5	
	ADD1	R4	
	BL	MONITOR	
	M	C3	
	STRD	C4	
	ADD1	R4	
	BL	MONITOR	
	M	C2	
	STRD	C3	
	ADD1	R2	
	BL	MONITOR	
	M	C1	
60	STRD	C2	
	ADD1	R1	
	STR2	C1	
	B	RESETE	
64 BZ	BZ	PREP	If TEN = 475 Prepare MAC Data
65 TELEM	RD	J	
	SC	11	
	BP	RES	If J = 11, RESET all MAC Data
	RDACC	MACOUT	
	SD	MAX	
70	BP	NEXT	
	RD	MACOUT	
	SC	-1	
	STRD	MACOUT	
	RD3	J	
	OUT X	RAV	
	B	RESETE	
77 RES	BZ	*+5	
8	RD	J	
9	SC	-1	
80	STRD	J	
1	B	RESETE	
2	BL	MONITOR	
3	ZERO		
4	STRD	C1	
5	STRD	C2	
6	STRD	C3	
7	STRD	C4	

Table A-3 (Continued)

Label	OP	Address	Comments
8	STRD	C5	
9	STRD	C6	
90	STRD	C7	
1	STRD	C8	
2	STRD	RAV	
3	STRD	R0	
4	STRD	R1	
5	STRD	R2	
6	STRD	R3	
7	STRD	R4	
8	STRD	R5	
99	STRD	R6	
100	STRD	R7	
1	STRD	R8	
2	B	RESETE	
3 NEXT	BL	MONITOR	
4	RDC	1	
5	STRD	MACOUT	
6	RDC	2	
7	STRD	MAX	
8	RD	J	
9	SC	-1	
110	STRD	J	
111	B	TELEM	
112 PREP	SHIFT DM	RAV	
3	FLOAT 1	R0	
4	FLOAT 1	R1	
5	FLOAT 1	R2	
6	FLOAT 1	R3	
7	BL	MONITOR	
8	FLOAT 1	R4	
9	FLOAT 1	R5	
120	FLOAT 1	R6	
1	FLOAT 1	R7	
2	FLOAT 1	R8	
3	B	RESETE	
4	SET SP16	0	
5	SET PR	8	
6	INPUT	PLASMA	
7	RD3 X	SPINCNT	
8	SADDX	NPLAS	
9	SD2 X	NMAX	
130	BP	SET	
131 COMP	SQSP	NSQ	
2	RD	SP16	COUNT OF SP16 PULSES
3	SC	15	
4	BZ	UPSPIN	
5	RD	SP16	
6	SC	-1	
7	STRD	SP16	

Table A-3 (Continued)

	Label	OP	Addr[illegible]s	Comments
8	SETUP	RD	CLOCK	PUTS CLOCK IN REG2
9		ADD2	PERIOD	
140		SETEV		PUTS DMIO IN EVENT
1		R		
2	UPSPIN	ZERO		
3		STRD	NMAX	
4		STRD	SP16	
5		RD	SPINCNT	
6		SC	5	
147		BZ	PREP2	
148		RD	SPINCNT	
149		SC	-1	
150		STRD	SPINCNT	
151		R		
152	PREP 2	SET PLASMA	0	
153		FLT	NPLAS	
154		PRSQ	NSQ+1	
155		STRD	NSQ	
156		FLT	NPLAS + 1	
157		PRSQ	NSQ+3	
158		STRD	NSQ+1	
159		BL	MONITOR	
100		FLT	NPLAS+2	
1		PRSQ	NSQ+5	
2		STRD	NSQ+2	
3		FLT	NPLAS+3	
4		PRSQ	NSQ+7	
5		STRD	NSQ+3	
6		FLT	NPLAS+4	
7		PRSQ	NSQ+9	
8		STRD	NSQ+4	
9		ZERO		
70		STRD	SPINCNT	
71		R		
172	TESP	SC	20	IF TEN < 15 PREPARE
3		BP	TE1	PLASMA OUTPUT
4		RD3	TEN	
5		OUTX	INTERVAL	
6		R	-	
7		BZ	*+2	
8		R		
9		ZEROD		
180		STRD	T1	
1		STRD	T2	
2		STRD	T3	
3		STRD	T4	
4		RESETIO	1	
5		R	-	
186	SET	RD	SPINCNT	
		STRDX	INTERVAL	

Table A-3 (Continued)

Label	OP	Address	Comments
	STRDX	INTERVAL	
	STR2	NMAX	
	B	COMP	
190 SUN 2	RDACC	T1	SUN PULSE OCCURS
1	BZ	T4S	
2	RDACC	T4	
3	BZ	*+2	
4	R		
5	RD	CLOCK	CLK IN 2
6	STR 2	T4	PUT CLOCK READING IN ACC
7	SD 2	T1	
8	SHIFT	4	
9	STRACC	PERIOD	
200	R		
01 T1S	RD	CLOCK	
02	STR2	T1	
03	RD ANGLE	ATTITUDE	
04	STRD	ANGLE	
05	R		
06 MONITOR	SSR	-	
07	IBP	REG4	
08	SETMODE	1	
09	STR2	SAVE	
10	STRPR	SAVE PR	
11 HB	HB		
12 R	BM	HB	IF MODE IS 0 GO TO HB
13	RD2	SAVE	
14	SETMODE	0	
15	RDPR	SAVE PR	
16	IBP	REG4	
217 SUN	SETSUN	0	
18	SETPR	4	
19	B	SUN2	
220 EARTH	RDACC	T2	
1	BZ	A	
2	RDACC	T3	
3	BZ	*+2	
4	B	R	
5	RD	CLOCK	
6	STR2	T3	
7	B	R	
8 A	RD	CLOCK	
9	STR2	T2	
230	B	R	

DATA MEMORY ARRAY

0	INTERVAL	27	C8
1	↓	28	RAV
2	↓	29	R0
3	↓	30	R1
4	↓	31	R2
5	NPLAS	32	R3
6	↓	33	R4
7	↓	34	R5
8	↓	35	R6
9	↓	36	R7
10	NSQ	37	R8
11	↓	38	TEN
12	↓	39	J
13	↓	40	MACOUT
14	↓	41	MAX
15	↓	42	SPINCNT
16	↓	43	SP16
17	↓	44	PERIOD
18	↓	45	NMAX
19	↓	46	T1
20	C1	47	T2
21	C2	48	T3
22	C3	49	T4
23	C4	50	SAVE
24	C5	51	SAVE PR
25	C6	52	ANGLE
26	C7		

MICRO-OPERATIONS

A table of micro-operations is given below. This table in conjunction with the microprograms shown later constitutes a detailed description of the computer operation. The computer design was not debugged so errors probably exist in the table. However the table is useful as an illustration of compact and precise descriptions of micro operations. In addition the table indicates the typical complexity of a micro-operation and the types of operations performed. The large number of transfer operations is probably typical. Finally, the table indicates approximately the number of different micro-operations needed for a telemetry computation.

Table A-4
Micro-operations

Label	Action	No.	Remarks
XFER C, 2	$R2_{12} \leftarrow C_1$, $C \leftarrow \downarrow C$, $R2 \leftarrow \tau R2$, $T \leftarrow T-1$	1	$R2 \equiv \omega^{12}(24)/REG2$
SUBT	$AC_{12} \leftarrow \oplus(D2_1, CA, \overline{R2}_1)$, $CA \leftarrow M(D2_1, CA, \overline{R2}_1)$	2	$AC \equiv \omega^{12}(24)/ACC$
	$AC \leftarrow \tau AC$, $R2 \leftarrow \downarrow R2$, $D2 \leftarrow \downarrow D2$, $T \leftarrow T-1$		$D2 \equiv \omega^{12}(24)/DMIO$
MAC INP	$REG2_{24} \leftarrow MACWORD_1$, $REG2 \leftarrow \tau REG2$, $T \leftarrow T-1$	3	
	$MACWORD_8 \leftarrow 0$ $MACWORD \leftarrow \downarrow MACWORD$		
SHIFT 2	$REG2 \leftarrow \tau REG2$ $T \leftarrow T-1$	4	
READ DATA	$DMIO \leftarrow DM_C$, $T \leftarrow T-1$	5	
SET ADD	$ADD \leftarrow 1$, $CA \leftarrow 0$, $T \leftarrow T-1$	6	
ADD	$T \leftarrow T-1$, $CA \leftarrow M(CA, REG2_1, DMIO_1)$,	7	
	$DMIO_1 \leftarrow \tau DMIO$		
	$ACC \leftarrow \downarrow ACC$, ACC_{24}, $DMIO_{24} \leftarrow \oplus(CA, REG2_1,$		
	$DMIO_1)$ $REG2 \leftarrow \downarrow REG2$		
XFER D,2	$DMIO \leftarrow \tau DMIO$, $REG2 \cdot \cdot EG2$,		
	$REG2_{24} \leftarrow DMIO_1$, $T \leftarrow T-1$	8	
XFER C, D	$T \leftarrow T-1$, $C \leftarrow \tau C$, $DMIO \leftarrow \tau DMIO$, $DMIO \cdot \leftarrow 0$,		
	$DMIO_{12} \leftarrow C_1$	9	
XFER 2, D	$T \leftarrow T-1$, $R \leftarrow \downarrow R$, $DMIO_9 \leftarrow R_1$, $DMIO_{18}$,		
	$DMIO_{24} \leftarrow 0$, $T2 \leftarrow 9$ AO, $ACC \leftarrow 0$	11	$R \equiv \omega^9(24)/REG2$
XPLY	$T \leftarrow T-1$, $AC2_9 \leftarrow \oplus(AC2_1, CA, REG2_1)$,		
	$CA \leftarrow M(AC2_1, CA, REG2_1)$	12	ADD.EQ.1
	$AC2 \leftarrow \tau AC2$, $R \leftarrow \downarrow R$		$AC2 \equiv \alpha^9(24)/ACC$
	$T \leftarrow T-1$, $AC2_9 \leftarrow \oplus(AC2_1, CA, \overline{REG2}_1)$,		
	$CA \leftarrow M(AC2_1, CA, \overline{REG2}_1)$	12	SUBT.EQ.1
	$AC2 \leftarrow \tau AC2$, $R \leftarrow \downarrow R$		
	Needed only for simulation	13	
SHIFT	$ACC \leftarrow \tau ACC$, $T \leftarrow T-1$	14	
SETUP	$ADD \leftarrow 1$, $CA \leftarrow 0$, $DMIO \leftarrow DM_C$, $T \leftarrow T-1$	15	
ADD 24	$T \leftarrow T-1$, $DMIO \leftarrow \tau DMIO$, $ACC \leftarrow \tau ACC$,	16	
	$DMIO_{24} \leftarrow \oplus(DMIO_1, ACC_1, CA)$,		
	$CA \leftarrow M(DMIO_1, ACC_1, CA)$		
WRITE	$T \leftarrow T-1$, $DM_C \leftarrow DMIO$	17	
MIC BRNCH	$MIC \leftarrow T$, $T \leftarrow 0$	18	
XFER 2, D	$T \leftarrow T-1$, $REG2 \leftarrow \tau REG2$, $DMIO \leftarrow \tau DMIO$,	20	
	$DMIO_{24} \leftarrow REG2_1$		
XFER D, ACC	$T \leftarrow T-1$, $ACC \leftarrow \tau ACC$, $DMIO \leftarrow \tau DMIO$,		
	$ACC_{24} \leftarrow DMIO_1$	21	
SUBT	$T \leftarrow T-1$, $ACC \leftarrow \tau ACC$, $DMIO \leftarrow \tau DMIO$,	22	
	$ACC_{24} \leftarrow \oplus(\overline{DMIO}_1, CA, ACC_1)$,		
	$CA \leftarrow M(\overline{DMIO}_1, CA, ACC_1)$		
XFER D 3	$T \leftarrow T-1$, $DMIO \leftarrow \tau DMIO$, $REG3 \leftarrow \tau REG3$,	23	
	$REG3_{12} \leftarrow DMIO_{12}$		
ADD INDEX	$T \leftarrow T-1$, $C \leftarrow \tau C$, $REG3 \leftarrow \tau REG3$	24	
	$C_{12} \leftarrow \oplus(CA, REG3_1, C_1)$,		
	$CA \leftarrow M(CA, REG3_1, C_1)$		
XFER OUT	$DMIO \leftarrow \tau DMIO$, $OUT \leftarrow \tau OUT$, $OUT_4 \leftarrow DMIO_1$,	25	
	$T \leftarrow T-1$		
XFER C,D	$T \leftarrow T-1$, $DMIO \leftarrow \tau DMIO$, $C \leftarrow \tau C$, $DMIO_{12} \leftarrow C_1$,	19	
	$DMIO_{24} \leftarrow 0$		
XFER ACC, D	$T \leftarrow T-1$, $ACC \leftarrow \tau ACC$, $DMIO \leftarrow \tau DMIO$,	10	
	$DMIO_{24} \leftarrow ACC_1$		
FP	$REG2 \leftarrow \epsilon^1(24)$, IF(ACC.EQ.$\epsilon(24)$ V$\overline{\epsilon}(24)$,	26	
	$MIC \leftarrow MIC+2$		

Table A-4 (Continued)

Label	Action	No.	Remarks
FP	T←T-1, A ≡ \ACC, ω^{24}(48), DMIO\, A←↑A ↑REG2, $REG2_4$ ←⊕ ($REG2_1$, $REG2_2$) IF(ACC.EQ. $\overline{\epsilon}$(24) $\vee \epsilon$(24), T←0	27	
XFER PL 2	A ≡ \PLASMA, ω^{24}(41), REG2\, A←↑A, T←T-1	28	
SETUP	AO, ACC←0, T2←24	30	
COND	T←T-1, IF(CA.EQ.1), MIC←MIC+1	36	
	T←T-1, ↓REG2, ↑DMIO, ↑ACC, ACC_{24}←⊕($DMIO_1$, CA, $REG2_1$), CA←M($DMIO_1$, CA, $REG2_1$)	29	

SFTVD	SUBT←$\overline{AO}$ ∧ $DMIO_1$, ADD←AO ∧ $DMIO_1$,
	SHIFT←AO ⊕ $DMIO_1$
	CA←$\overline{AO}$ ∧ $DMIO_1$, T←T-1
	IF(T2.GT.O AND SHIFT.EQ.1), MIC←MIC+1
	IF(T2.EQ.O) MIC←MIC+2
		31	

SHIFT	T2←T2-1, T←0, AO←$DMIO_1$, ↑DMIO, ↓R ↑ACC, MIC←MIC-2	33	R ≡ ω^{12}(24)/REG2
ADD	T←T-1, ↓R, ↑A1, $A1_{12}$ ←⊕($A1_1$, CA, R_1) CA←M($A1_1$, CA, R_1)	32	A1 ≡ α^{12}(24)/ACC ADD.EQ.1
	T←T-1, ↓R, ↑A1, $A1_{12}$ ←⊕($A1_1$, CA, R_1), CA←M(CA, R_1, $A1_1$)	32	SUBT.EQ.1
BOVF	T←0, IF(CA.EQ.1), MIC←MIC+1	36	
SET C	C_1←1, T←T-1	37	
SETUP	ADD←1, ACC←O, DMIO←DM_C, T←T-1	38	
SETCNT	T←T-1, CNTR←0	39	
BZ	T←T-1, IF(ACC.EQ.O), MIC←MIC+1	40	
SHIFT COUNT	T←T-1, CNTR←CNTR+1, ↑AD, MIC←MIC-1 AD ≡ \ ACC, ω^{12}(24), DMIO\	41	
PLACE	T←T-1, ω^3/↓ACC←ω^3/(2↓CNTR)	42	
WRITE	DM ←DMIO, C_1←O, T←T-1	43	
COUNTDOWN	CNTR←CNTR+1, ↑DMIO, IF(CNTR.GT.O) T←O, ELSE, T T-1	44	
SET	T←T-1, $CNTR_1$←$\overline{AC}_4$, ω^4/↓CNTR←ω^4/ ↓$\overline{CNTR}$	45	
SUBT	T←T-1, ↑ACC, ↓PRIORITY, ↓SR, ACC_{24} ←⊕(CA, SR_1, $PRIORITY_1$)	46	
XFER AE	AE ≡ \ACC, ω^{12}(36), EVENT\, ↑AE, T←T-1	47	
XFER AP	AP ≡ \ACC, ω^4(28), PRIORITY, ↑AP, T←T-1	48	
XFERPD	PD ≡ \PRIORITY, ω^4(8), ω^4(24)/DMIO\, ↑PD, T←T-1	49	
XFER AGD	AGD ≡ \ANGLE, ω^{24}(32), DMIO\, ↑AGD, T←T-1	50	

The table of micro program arrays gives the actual micro programs for the individual macro operations. For example the macro ADD1 consists of a sequence of 5 micro operations and accomplishes the addition of a memory word and the accumulator. The first step is a read of data memory. The second is a setup for addition, etc. The last step causes a return to the macro program.

Table A-5
Micro Program Array

Label	T	D	Action
SQ	9	11	
	1	31	
	9	12	
	1	35	
	0	13	
	6	14	
	1	15	
	24	16	
	1	17	
	0	0	
ADD1	1	5	
	1	6	
	24	16	
	1	17	
	0	0	
RDACC	1	5	
	24	21	
	0	0	
SD	1	5	
	0	1	
	12	22	
	0	0	
RD3	1	5	
	12	23	
	0	0	
OUT X	1	6	
	12	24	
OUT	1	5	
	4	25	
	1	17	
	0	0	
SC	12	1	SET UP TO SUBTRACT CONST.
	12	2	SUBTRACT CONST
	0	0	
INPUT MAC	8	3	INPUT MACWORD TO REG 2
	16	4	
	0	0	
ADD2	1	5	24 BIT ADD
	1	6	OF DMIO AND REG 2
	24	7	RESULT IN ACC & DMIO
	0	0	

Table A-5 (Continued)

Label	T	D	Action
RD2	1	5	
	24	8	XFER DMIO → REG 2
	0	0	
m	1	5	
	0	11	
	0	12	
	9	12	
	0	13	
	6	14	
	0	0	
RDC	12	9	
	0	0	
STR2	24	20	
	1	17	
	0	0	
RDC	12	19	
	0	0	
STRACC	24	10	
	1	17	
	0	0	
FLOAT 1	1	5	
	24	21	
	8	10	
	0	26	
	15	27	
	16	20	
	0	0	
INPUT PLASMA	17	28	
	7	4	
	0	0	
SADDX	1	6	
	12	24	
	1	15	
	24	7	
	1	17	
	0	0	
SD2X	1	6	
	12	24	
	0	1	
	1	5	
	24	29	
	0	0	

Table A-5 (Continued)

Label	T	D	Action
SQSP	5	4	
	24	20	
	1	4	
	0	30	
	1	31	
	12	32	
	1	33	
	1	34	
	1	6	
	12	4	
	1	15	
	24	16	
	1	36	
	0	0	
OVF	1	37	
	1	38	
	24	16	
	1	17	
	0	0	
SHIFT DM	4	21	
	0	0	
FLT	1	5	
	24	21	
	4	10	
	1	39	
	1	40	
	1	41	
	15	10	
	1	17	
	1	45	
	0	0	
PRSQ	1	5	
	24	21	
	1	42	
	1	43	
	1	5	
	4	10	
	31	44	
	0	0	
SSR	0	1	
	4	46	
	0	0	
SET EV	12	47	
	0	0	

Table A-5 (Continued)

Label	T	D	Action
RDPR	1	5	
	4	48	
	0	0	
STRPR	4	49	
	1	17	
	0	0	
STRDX	1	6	
	12	24	
	1	17	
	0	0	
RD ANGLE	24	50	
	0	0	

REFERENCES

1. Wilkes, M. V. and Stringer, J. B., "Microprogramming" Proceedings of Cambridge Philos. Soc., Volume 49, Part 2, pages 230-238 (April 1953)

2. Carr, J. W., "Microprogramming" Section 2-17 of Handbook of Automation Computation and Control, Volume 2, Edited by Grabbe, et. al. (Wiley 1959)

3. Wilkes, M. V., et. al. "Design of the Control Unit of an Electronic Digital Computer" Proceedings of IEE, Volume 105, part B, pages 121-128 (1958)

4. Booth, A. D., Booth, K. H. V., Automatic Digital Calculators, Butterworths (1956)

5. Shea, J., "Computer Design Problems for the Space Environment," Spaceborne Computer Engineering Conference Proceedings, pages 1-8, 1962

6. Baron, R. C. and Waller, R. W., "Two Approaches to Spacecraft Data Handling Systems" Spaceborne Computer Engineering Conference Proceedings, pages 9-18, (1962)

7. Klieger, L. B., "Design of OGO Data Handling System" Spaceborne Computer Engineering Conference Proceedings, pages 19-26 (1962)

8. Maxwell, M., et. al., "An Adaptable Spacecraft Telemetry System," Internal NASA Report, Goddard Space Flight Center.

9. Van der Poel, W. L., "Microprogramming and Trickology," Digital Information Processors, pages 269-311, Wiley (1962)

10. Allen, M. W., et. al., "CIRRUS, An Economical Multiprogram Computer with Microprogram Control," IEEE Transactions on Electronic Computers, pages 663-670 (Dec, 1963)

11. Iverson, K. E., A Programming Language, Wiley (1962)

12. Schaefer, D. H. and Albus, J. S., "Satellite Attitude Determination, Digital Sensing and On-board Processing," NASA Report X-631-63-133 (July, 1963)

13. White, H. D., Jr., "Evolution of Satellite PFM Encoding Systems from 196-1965" NASA Report X-631-65-115 (March 1965)

14. Lewis, T. B., "Primary Processor and Data Storage Equipment for Orbiting Astronomical Observatory," IEEE Transaction on Electronic Computers, pages 677-687 (Dec. 1963)

BLANK PAGE

DATA COMPRESSION AND ITS APPLICATION TO VIDEO SIGNALS

L. D. Davisson

INTRODUCTION

A signal is generated by some arbitrary source. This signal is sampled and quantized in some fashion determined by the signal user. In the usual mode of system operation, $\log_2 L$ bits per sample are transmitted where L is the number of quantization levels. This report considers methods of reducing the transmission rate while maintaining an acceptable level of data quality. This reduction process is called data compression, compaction, or more generally, source encoding.

The studies reported herein are aimed primarily towards the reduction of transmission rates for TV pictures, although the general techniques and theoretical results apply as well to other data sources. Ten Tiros cloud cover pictures were selected as being typical of the class of TV pictures. Figure A-4 presents the two of these ten pictures which are used to illustrate the results of this study. The pictures of Figure 1 were taken directly from a crt display of the original analog signal. They may be used as a standard of picture quality for the photographs which follow.

Data compression is obtained by taking advantage of data redundancies which exist when the signal is "inactive" or contains some "regularity," e.g. where large areas are of the same shade. One such method is based upon the "predictability" of each succeeding frame, line or element as the case may be. In this study frame-to-frame effects were not considered since the meteorological pictures chosen are too far apart in time to have sufficient similarity. However, element and line effects have been considered and will be dealt with in this report in that order.

ELEMENT-TO-ELEMENT DATA COMPRESSION

If a sample can be predicted based on the previous samples so that the prediction error (or possibly the sample r.m.s. error or some other error criterion with "memory") does not exceed some threshold, it need not be transmitted resulting in a reduction in bit rate. The indication to predict is encoded so that the receiver knows to insert a predicted value in place of the sample. Suppose p is the probability of predicting accurately and each of the L levels is equally likely when an accurate prediction is not possible. Suppose in addition that the ability to predict is sample-to-sample independent. Then theory states that in the noiseless case a compression ratio of

$$\frac{\log_2 L}{-p \log_2 p - (1-p) \log_2 \left(\frac{1-p}{L}\right)}$$

is approachable where the compression ratio is defined as the ratio of the number of bits in the uncompressed message set to the number in the compressed message set.

It should be pointed out at this time that the concept of compression ratio can be used in a somewhat misleading manner. For example, a reduction from 6 bits to 2 bits is not the

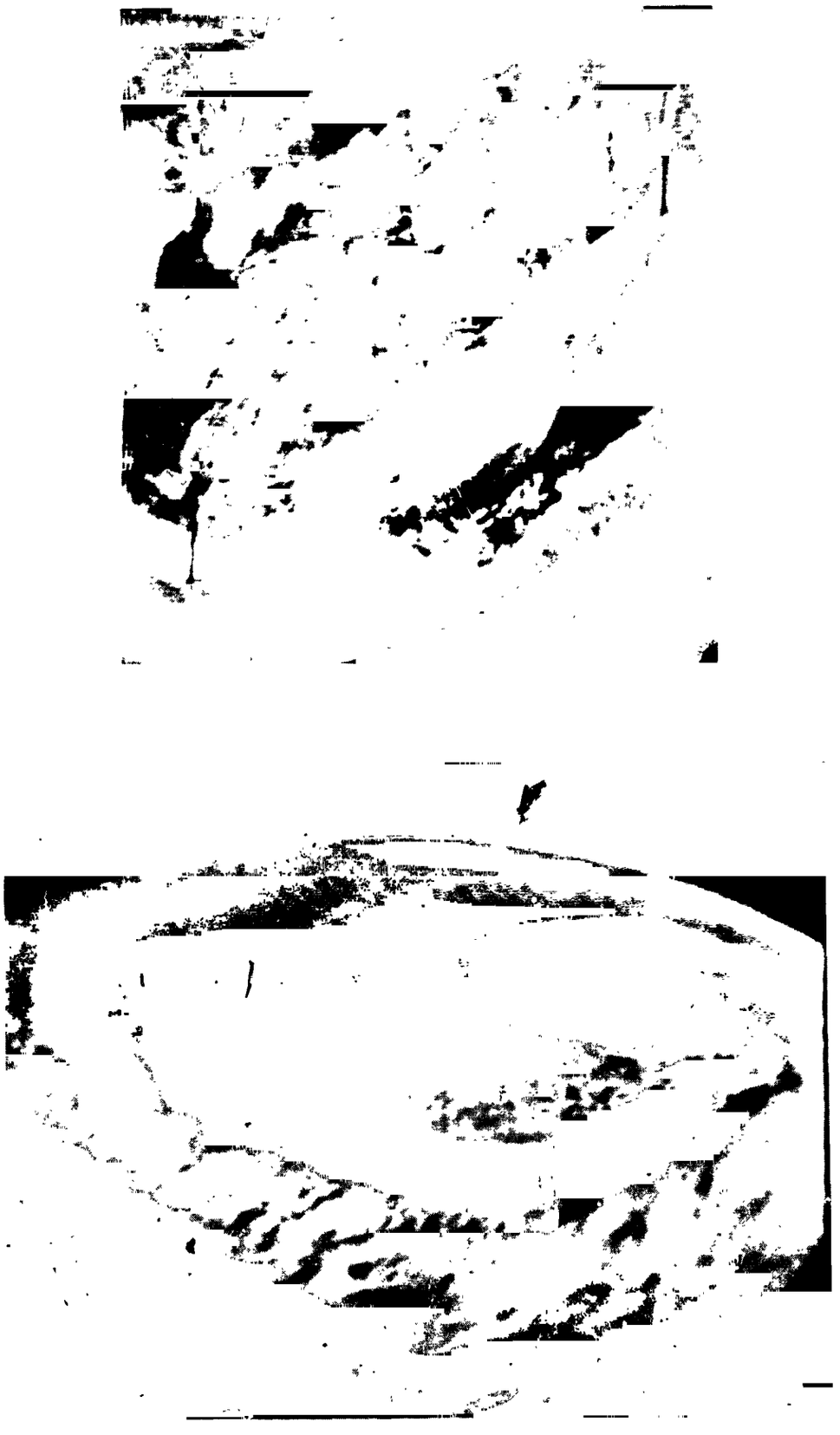


Figure A-4–Original analog pictures

same as a reduction from 3 bits to 1 bit per sample. To avoid this confusion, the message entropy will be used under the independence assumption. It must also be mentioned that the number of bits required in the noisy case for a given probability of error per sample increases the coded transmission rate. However, since the coding problem in noise has not been solved, it will be convenient to use the above definition with the understanding that it represents a lower bound on the transmission rate. Therefore, to sum up, the transmission rate per sample is stated as:

$$R(L) = -p \log p - (1 - p) \log (1 - p) + (1 - p) \log_2 L$$

$$= H(p) + (1 - p) \log_2 L.$$

R is stated as a function of L to emphasize the importance of quantizing properly. It is a simple matter to establish that $R(L + i) \geq R(L)$. This is certainly what one would expect since p can not decrease with a reduction in the number of levels.

Therefore, one should minimize the number of levels consistent with the desired data quality. For a given number of levels, L, the choice of where to place the various levels is also of considerable importance since the number of levels can be reduced the farthest if they are chosen in a careful manner which takes into consideration the information which is to be conveyed to the user. In the case of TV data the effect of quantizing too coarsely is the so called "contouring" effect. That is, there is no smooth transition of grey shades giving the picture a coarse appearance. Therefore one should set the lowest and highest levels as close together as possible so that small fluctuations in intermediate values result in the largest possible grey level changes. To put it in other terms, the contrast should be set at its desired value a priori rather than doing this a posteriori resulting in a loss in the number of effective levels.

Figures A-5 to A-8 illustrate this point. Figures A-5 and A-6 are 16 and 8 level pictures where the top and bottom levels are set at around the first and 99th percentiles of the signal. Thus most of the picture information is contained in the center of the range and a gray, coarse picture results. In Figures A-7 and A-8, however, the top and bottom levels are set at around the 15th and 85th percentiles. Thus some limiting occurs. The data of interest fluctuates through the entire range of values resulting in a higher contrast picture. In fact, the 8 level picture of Figure A-8 has about the same amplitude difference between levels as in the 16 level picture of Figure A-5 in the center of its range. Thus contouring is reduced resulting in a clearer picture. It must be further stated that if the signal is too severely limited, valuable information is lost.

Because the 8 level pictures of Figure A-8 seem to be of acceptable quality, the compression results for the 8 level data are given on the following pages.

To achieve data compression, some prediction technique must be specified. Obviously, if it is required that the prediction error be less than some peak value at all times, the optimum predictor minimizes the probability of exceeding the error threshold, or stated more simply, minimizes the probability of error conditioned on the data. That is, if the sample s_o is to be predicted, one chooses the estimate $s_o = k^{th}$ level if

$$p(k/s_{-1}, \ldots) = \max_i \sum_{j=-t}^{t} p(i + j/s_{-1}, \ldots)$$

Figure A-5–Four bit low contrast

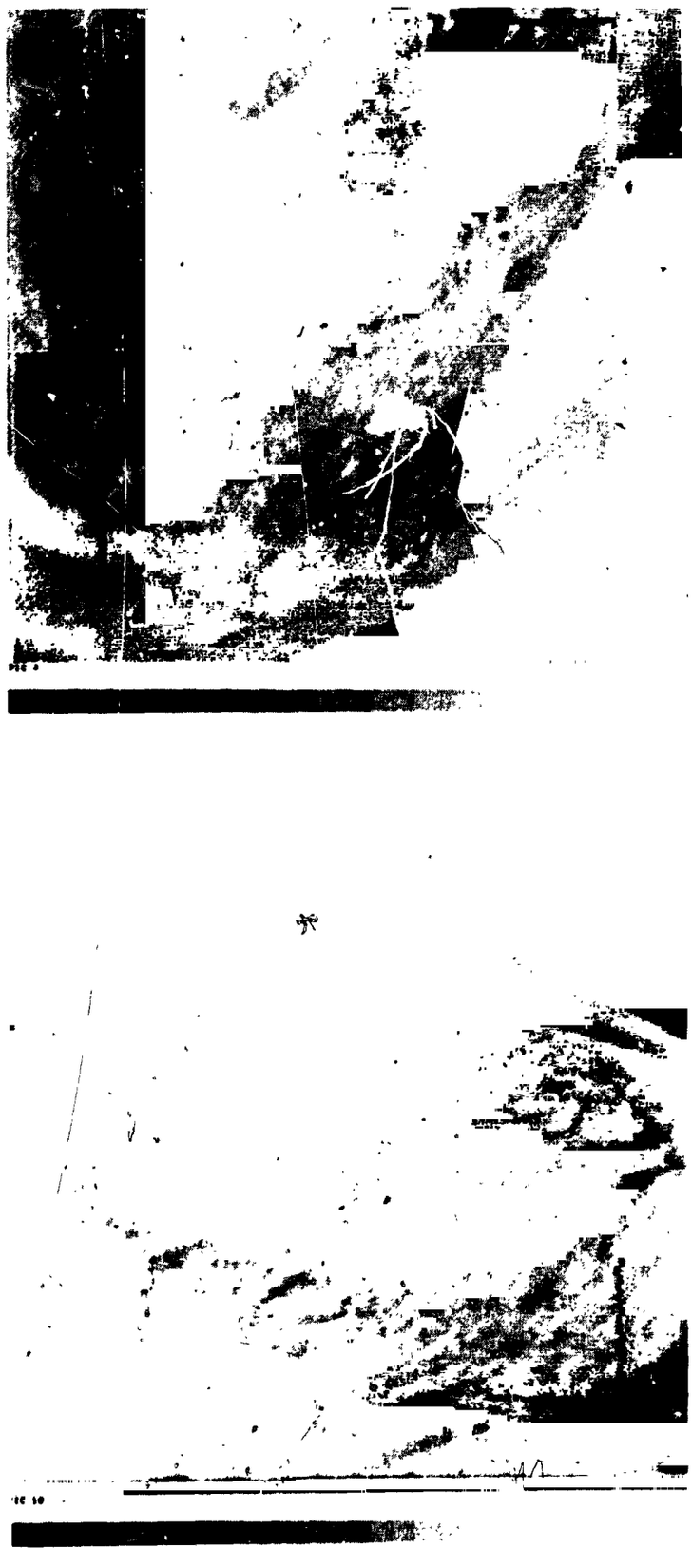


Figure A-6--Three bit low contrast

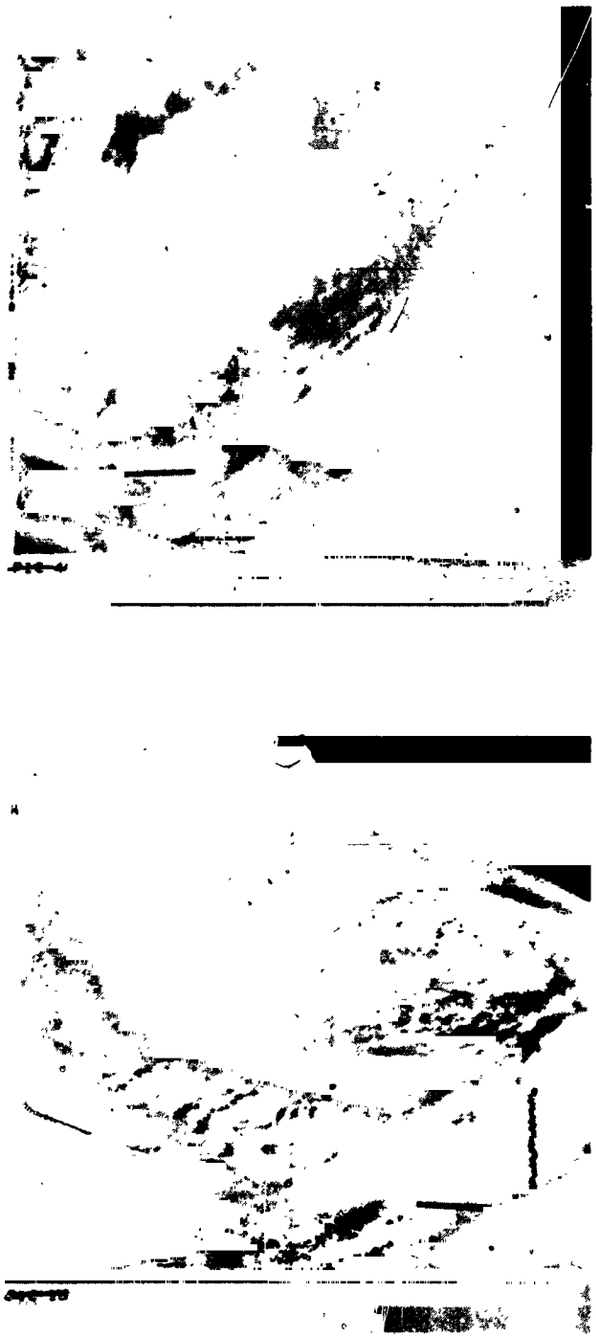


Figure A-7–Four bit high contrast

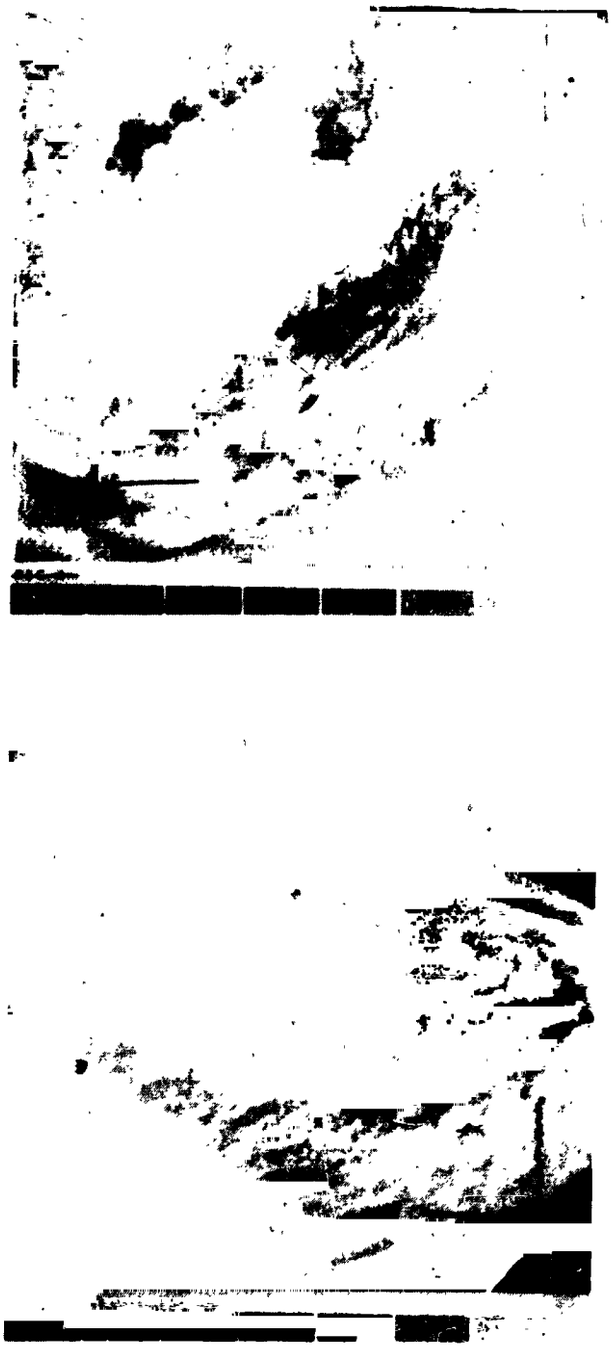


Figure A-8–Three bit high contrast

where $p(\cdot / \cdot)$ is the conditional (or transition) probability and $\pm t$ is the allowable error. Unfortunately these transition probabilities are not known and must therefore be estimated from the data. The most reasonable way to do this is to simply construct a frequency plot or histogram for each of the conditional probabilities. Of course the conditioning must be on a finite number of samples. That is, a histogram for

$$p(k/s_{-1}, \ldots, s_{-m})$$

is to be constructed as a function of k for each of the L^m ordered se's $(s_{-1}, \cdots, s_{-m})$. It is immediately apparent that an m larger than 3 or 4 is not practical from a storage viewpoint. In addition the effect of non-stationarities on the long period of time required to make a histogram of "reasonable" size limits m.

The total storage required depends on the number of bits used in each histogram. The count in each histogram does not become unbounded since one wishes to allow the histogram to depend on only the most recent points to minimize the "sluggishness" of response to statistical nonstationarities. One way to do this, for example, is to store each point and to subtract it off after a given period of time. A second way is to use a "filter" such as an exponential weight on the histogram. A third way which eliminates the need for storage and multiplication operations is to divide by 2 everytime the mode (or maximum) of the histogram overflows. Since this is a simple shift operation it is easily implemented. Supposing n bits are allowed for each histogram element (as will be shown, $n \approx 4$ is about right). Then a total memory of $L^m \times nL = nL^{m+1}$ bits is required. This might be reduced further by a priori restrictions on the histograms (e.g. transitions from level 1 to L would be assumed to have less than the maximum probability).

Another predictor to be used in the proposed method of data compression would be the sample conditioned mean (Balakrishman's second method). Since the conditional expectation minimizes mean square error, this statistic would be applicable relative to a sample mean square error (r.m.s.) threshold criterion. Everytime the sample mean square error over the previous $N \gg 1$ points (or maybe some weighting of the past) exceeds some threshold, the most recent point is transmitted.

At first glance it might appear that the sample conditional mean requires fewer storage locations. This, in fact, may or may not be true. In storing the sample conditional mean, more than $\log_2 L$ bits is required because either the sum total plus a "counter" must be stored for each of the L^m sample means, or else $\log_2 L$ bits plus a number of fraction bits must be stored. In either case, it is not clear that a significant reduction in memory requirements is achieved. In addition, multiplication or division is always required in the second method.

A third type of predictor considered is the so called "zero order hold," or previous element predictor. Here it is predicted that the next point is the same as the present value. This predictor is, in general, not optimum with respect to any criterion except simplicity. It is by far the easiest method to implement and it may, as is shown in the following, compare favorably with the optimum techniques in prediction probability.

Other error criteria than the peak error and r.m.s. error can be given. An error requirement based on a linear weighting of the previous errors could be used. This effectively causes the allowable error to depend upon its spectral properties. Since high frequency noise is less objectionable than low frequency noise in TV pictures, it is reasonable to study this criterion. The primary disadvantage to error schemes based on memory is the possibility of large local errors. On the other hand, it is possible to adjust the error threshold "smoothly" as contrasted with the peak error case where only an integer number of levels can be specified.

Table A-6 gives typical transmission rate results for the three predictors considered - the conditional mode, the conditional expectation, and the zero order hold predictors. The conditional predictors are based on the previous element and the element directly above on the previous line:

Table A-6

Predictor \ Allowable Error	0	±1 Peak Error	Weighted Error Less Than .3
Conditional Mode	2.195	.6789	1.066
Conditional Mean	2.204	.6829	1.050
Zero Order Hold	2.195	.8780	1.120

*Transmission rates in bits/sample.

In the cases where no errors or moderate errors are allowed there is no essential difference between the predictor performances. Only in the case where a large error is allowed is there a significant difference between the predictors. Even there it would be debatable whether the difference would be worth the increased complexity. It is not obvious why the spread between the transmission rates increases with increasing error, however, it apparently has to do with the error feedback mechanism. Unfortunately the large gains obtainable by the more complicated methods are not of use since picture quality requirements limit the allowable error too severely. Figures A-9 and A-10 are pictures with the 1 level and the weighted errors respectively. It appears that the 1 level peak error does not result in an acceptable picture and that the one bit pictures of Figure A-9 are about as far as one can go in reducing picture transmission rates due to element effects. Therefore, it seems apparent that the more complicated systems do not hold a sufficient advantage over the simple zero order hold predictor to warrant their application.

The conclusion that the zero order hold predictor suffices for TV data (or at least for these 10 pictures) does not mean that it will always be true for all data. Therefore it is of interest to indicate some results for the conditional mode predictor. If one specifies a conditional mode predictor, one must set the number of values, N, to be used in the histogram. Appendix A gives lower bounds and asymptotic estimates for the prediction probability, p. The following table gives a comparison between the theoretical lower bound (where the maximum empirical value was assumed to be the modal probability) and the empirical values:

N	Theoretical Lower Bound	Empirical Value
3	.50	.5596
6	.4820	.5824
12	.5670	.5954
24	.5975	.6022
48	.6057	.6053
96	.6068	.6068
∞	.6068	.5992

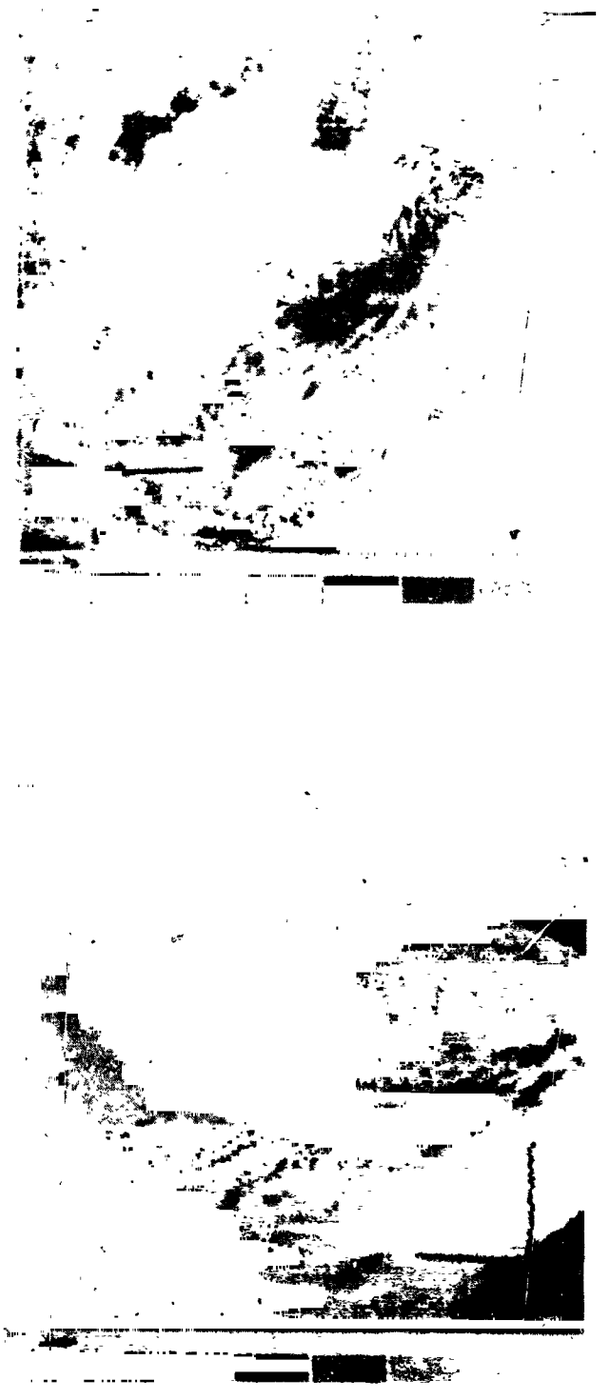
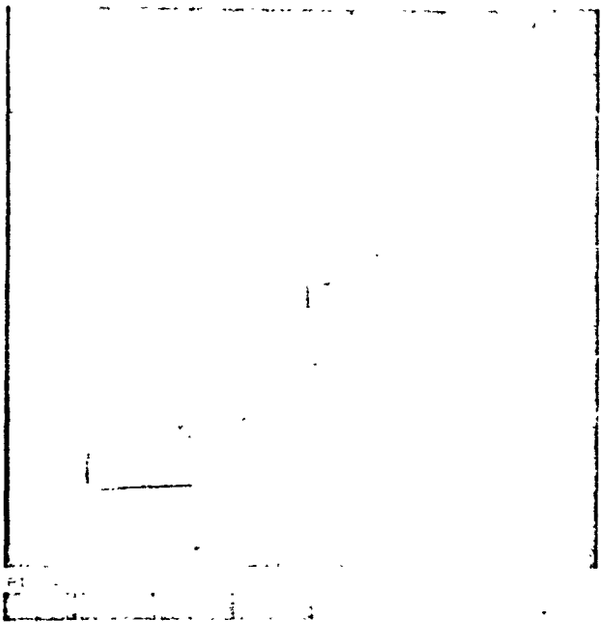


Figure A-9–Compressed to 0.6789 bits using conditional mode element predictor with ±1 level error



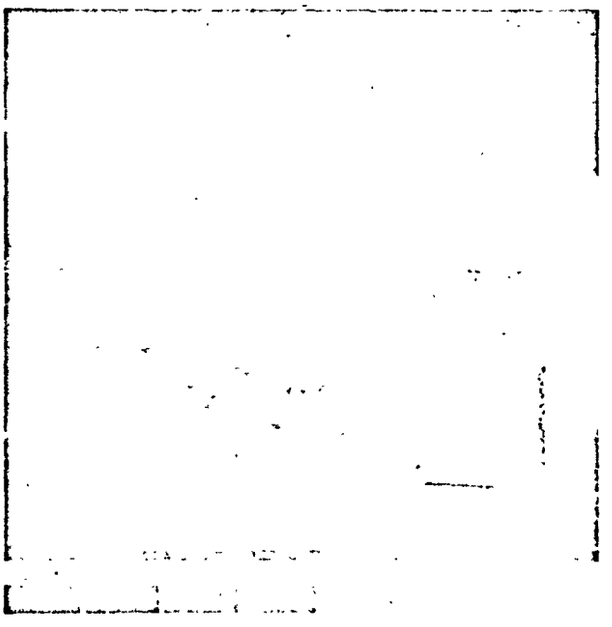


Figure A-10–Compressed to 1.05 bits using conditional mean predictor with weighted errors

Two things may be noted from this curve. First of all, although the predictor does well for relatively small values of N, it takes a surprisingly large value to "reach" the maximum. Secondly, there are non-stationary affects as evidenced by the dropping off of the prediction probability as $N \to \infty$.

An interesting thing was noted regarding the conditional histograms. It was noted that the previous element has considerably more influence than the element directly above on the preceding line. This must be due to a combination of line sync jitter plus the filtering of the signal across the line.

Line-to-line Effects

As in the element compression case, if a line can be generated with sufficient accuracy by operations on the preceding lines, it need not be transmitted. In this study only operations on the last transmitted line are considered. Let $S_m(n)$ be a function which represents the m^{th} line evaluated at the n^{th} element. If $S_m(n)$ is the last transmitted line, and it is necessary to decide whether line $m + i$ should be transmitted, one forms the estimator:

$$\hat{S}_{m+i}(n) = S_m(n + a + bn),$$

and the regression coefficients a and b are estimated. "a" represents "a" shift and "b" represents a linear expansion or contraction. This is a non-linear regression problem in the function S_m which possesses no particular smoothness properties. Therefore the "best" estimates must be found by a searching procedure over some range of values. "Best" is defined as the minimum of:

$$\sum_n f\left[S_{m+i}(n) - S_m(n + a + bn)\right]$$

where $f[\cdot]$ is some cost function. Since a search procedure is to be used, it no longer matters particularly whether or not $f[\cdot]$ is the usual squared error.*

If the minimum is less than some threshold, the parameters a and b are transmitted. Since this requires very few bits relative to the number of elements on a line (typically 500 elements), one can say that "no" bits are required.

Figure A-11 represents a 5-to-1 line compression ratio where $f(x) = x^2$. It is seen that an unacceptable amount of distortion results. In addition to a too large threshold, much of the distortion seems to result from a "blocking" effect where the minimum does not occur when sharp transition edges are lined up as one desires. To correct this, $f[x] = x^3$ was used. Figure A-12 represents a 3.5-to-1 compression ratio using this criterion. It is noted that reasonable picture quality is maintained. Table A-7 gives the relative frequency with which each of the values of a and b occur.

It is seen that $a = 0$ and $b = 0$ together occur significantly less than half of the time. Printouts of the minimum have shown it to be in fact a rather sharp minimum much of the time, making it worth while to transmit non-zero values of a and b. Future studies may

*It was found that if $f[\cdot]$ involves multiplication or exponentiation that considerable computational time can be saved by storing the L-1 values of $f[\cdot]$ as a vector.

Table A-7

a	Relative Frequency	b	Relative Frequency
-4	.0100	0	.9231
-3	.0240	-1/35	.0232
-2	.0485	+1/35	.0537
-1	.1475		
0	.3824		
1	.2293		
2	.0709		
3	.0445		
4	.0433		

find a better regression model or more satisfactory values of b. At present, computational time places a severe constraint on the range of regressors which can be tried.

Combined Line and Element Effects

It is now possible to combine line and element effects to obtain an overall transmission rate. Only those lines which cannot be reconstructed well enough from the previous line are transmitted. Every line which is transmitted is transmitted using element compression techniques. From the previous results it can be seen that less than 1/2 bit per sample should be required in the noiseless case while maintaining "reasonable" picture quality.

Figure A-13 represents pictures compressed to .37 bits which seem to be not quite good enough. Unfortunately time did not allow a more exhaustive study as to a lower bound.

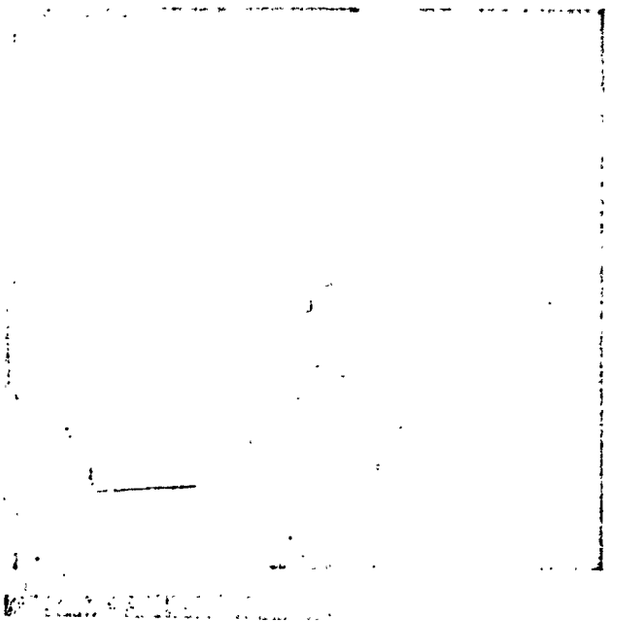
Other Possibilities Considered

In addition to the techniques mentioned above, several other ideas were considered briefly. Smoothing of the received data was considered briefly. It was found that the coarseness of the pictures was reduced at the cost of decreased sharpness.

It is felt that this problem should be explored further. A method of compression where only boundaries of areas of a given shade are transmitted was considered. This is a "paint-by-numbers" or map contour approach. Difficulties were encountered in deciding how to encode the data. Further study might be fruitful on this idea. A third idea is the use of delta modulation. Here of course one has the problems of saturation and error accumulation. However, as shown in the next section, the latter problem exists in any data compression system and perhaps can be eliminated by careful coding.

Future Research

The major outstanding data compression problem lies in the coding of the compressed message in the presence of noise. Once compression is performed, timing as well as amplitude information is generated. Since the timing information is typically sent as a change



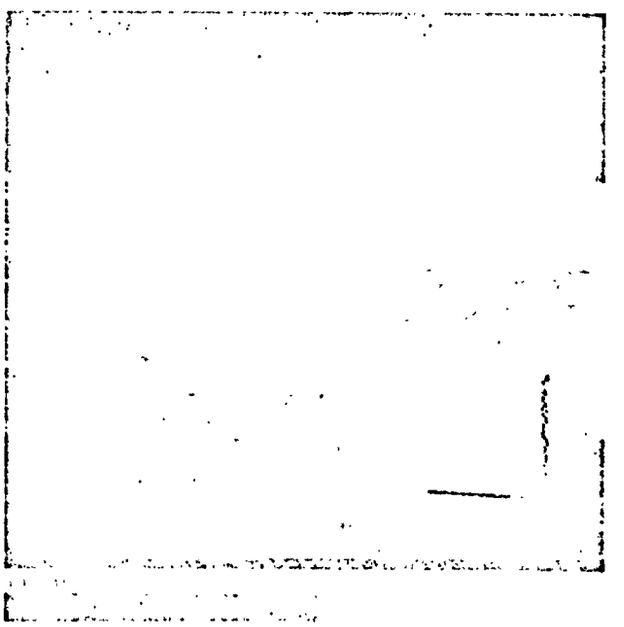


Figure A-11–5-to-1 line compression ratio

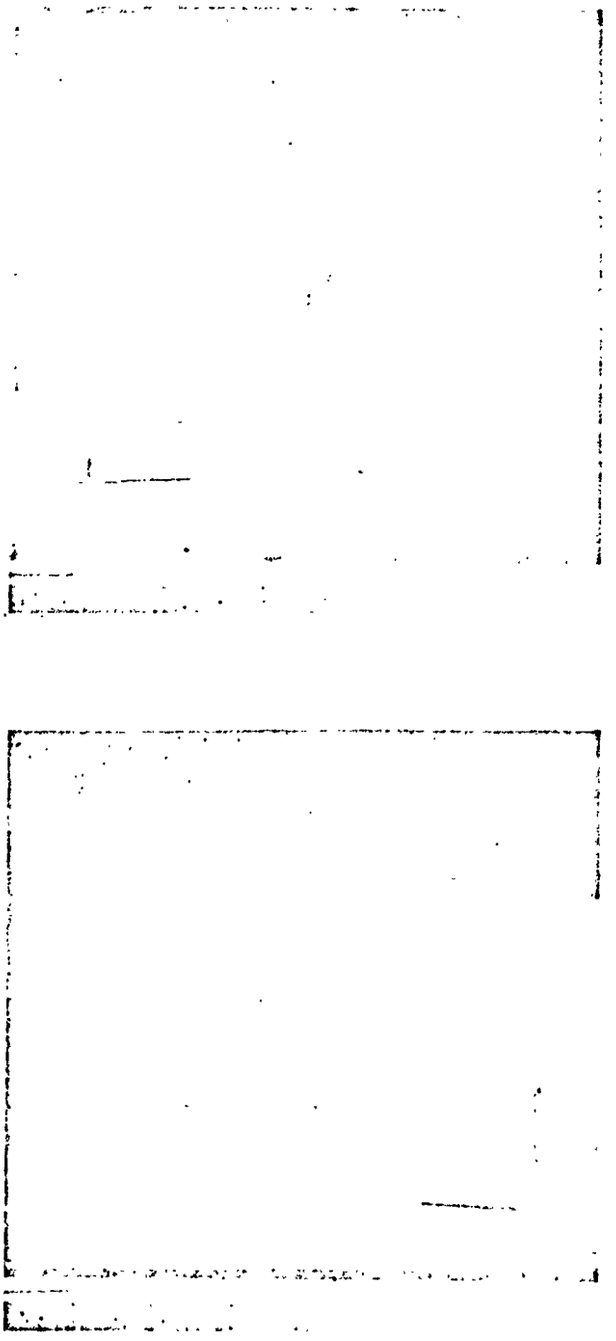


Figure A-12–3.5-to-1 line compression ratio

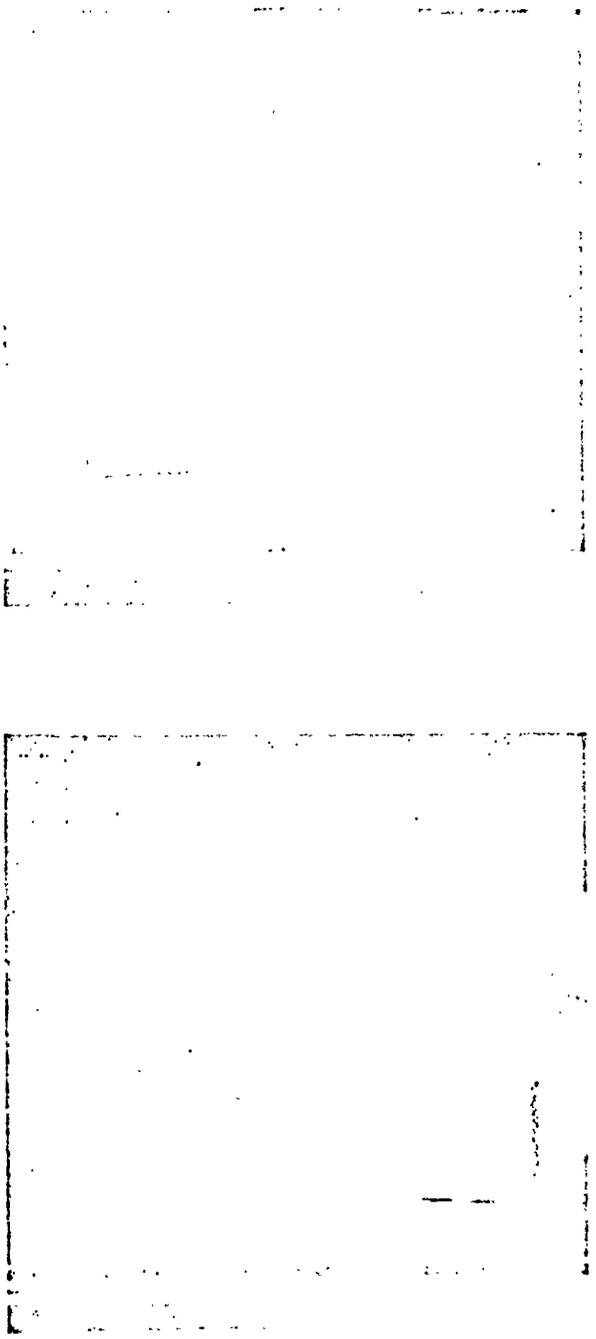


Figure A-13—Line and element compressed to 0.37 bits

from some previous value, the problem of error accumulation presents itself. How should this be handled to minimize transmission rate for a given error probability? Once this is done, it is then possible to answer the most important question. What is the transmission rate for a given compression scheme and a given probability of error?

Conclusions

In conclusion it may be stated that it is possible to reduce the message set to approximately 0.5 bits per sample by using line and element effects while still maintaining reasonable picture quality.

APPENDIX

PREDICTION PROBABILITY

In this appendix a theoretical lower bound on the prediction probability for a given modal probability ($> 1/2$) is found. In addition, an asymptotic expression is given for any given probability density. In both cases it is assumed that the values in the histogram are statistically independent of each other (as in a Markov process of the appropriate form).

Let N be the sample size and n_i be the number of values at the i^{th} level. The probability of level i is p_i. Let p_m be the probability of the modal value occurance. The probability of predicting correctly is:

$$P = \sum_{i=1}^{L} p_i \text{ Prob} \left\{ n_i = \max_j n_j \right\},$$

where L is the number of levels. For $p_m > 1/2$, the following lower bound can be used:

$$P \geq p_m \Pr \left\{ n_m = \max_j n_j \right\}$$

$$\geq p_m \Pr \left[n_m > \frac{N}{2} \right]$$

$$\geq p_m \sum_{n_m > N/2} \binom{N}{n_m} p_m^{n_m} (1 - p_m)^{N - n_m}.$$

This is the bound used in the text.

To find the asymptotic probability, let the events E_{ij} be defined as:

$$E_{ij} = \{(n_i, n_j),\ n_i \leq n_j\}$$

and let E_{ij}^c be the complement. Then

$$P = \sum_{i=1}^{L} P_i \text{ Prob} \left\{ \bigcap_{j \neq i} E_{ij}^c \right\}$$

$$= p_m \left[1 - \text{Prob} \left\{ \bigcup_{j \neq m} E_{mj} \right\} \right] + \sum_{i \neq m} P_i \text{ Prob} \left\{ E_{im}^c \right\} \text{ Prob} \left\{ \bigcap_{j \neq i,m} E_{ij}^c / E_{im}^c \right\}$$

$$= p_m \left[1 - \text{Prob} \left\{ \bigcup_{j \neq m} E_{mj} \right\} \right] + \sum_{i \neq m} P_i \text{ Prob} \left\{ E_{im}^c \right\} \left[1 - \text{Prob} \left\{ \bigcup_{j \neq i,m} E_{ij} / E_{im}^c \right\} \right]$$

where the / denotes conditioning as usual. Thus as upper and lower bounds:

$$P_m \left[1 - \sum_{j \neq m} \text{Prob}\{E_{mj}\}\right] + \sum_{i \neq m} P_i \left[\text{Prob}\{E_{im}^c\} - \sum_{j \neq i,m} \text{Prob}\{E_{ij} \wedge E_{im}^c\}\right]$$

$$\leq P \leq P_m \left[1 - \sum_{j \neq m} \text{Prob}\{E_{mj}\} + \sum_{j \neq m} \sum_{k \neq j,m} \text{Prob}\{E_{mk} \wedge E_{mj}\}\right] + \sum_{i \neq m} P_i \, \text{Prob}\{E_{im}^c\}$$

Note that both the lower and upper bounds are identical except for terms of the form

$$\text{Prob}\{E_{ij} \wedge E_{im}^c\} \leq \text{Prob}\{E_{mj} \wedge E_{im}^c\}.$$

If it can be shown that these joint probabilities go to zero faster than the other terms, then the asymptotic probability is the remaining terms on either side. Now:

$$\text{Prob}\{E_{mj}\} = \text{Prob}\{n_j > n_m\}$$

$$= \text{Prob}\{n_j - n_m > 0\}.$$

It can be shown readily that the n_i are asymptotically Gaussian with means and covariances:

$$E[n_i] = p_i N$$

$$E[n_i n_j] = -N p_i p_j \qquad i \neq j$$

$$= N p_i (1 - p_i) \qquad i = j.$$

Thus:

$$\text{Prob}\{E_{mj} \wedge E_{im}^c\}$$

$$\cong \int_0^\infty \int_0^\infty \frac{\exp -\frac{1}{2}\left[\left(\frac{y - \mu_y}{\sigma_y} - \rho \frac{x - \mu_x}{\sigma_x}\right)^2 \frac{1}{1 - \rho^2} + \frac{(x - \mu_x)^2}{\sigma_x^2}\right]}{2\pi \sigma_x \sigma_y \sqrt{1 - \rho^2}} \, dx \, dy$$

Where:

$$\mu_x = E[n_i - n_m] = N[p_i - p_m]$$

$$\mu_y = E[n_j - n_m] = N[p_j - p_m]$$

$$\sigma_x^2 = E[n_i - n_m]^2 - \mu_y^2 = N[p_i + p_m - (p_i - p_m)^2]$$

$$\sigma_y^2 = E[n_j - n_m]^2 - \mu_y^2 = N[p_j + p_m - (p_j - p_m)^2]$$

$$\rho = \frac{E[(n_i - n_m - \mu_x)(n_j - n_m - \mu_y)]}{\sigma_x \sigma_y} = -\frac{p_i p_j + p_m(1 - p_m) + p_m(p_i + p_j)}{\sigma_x \sigma_y}$$

Therefore: $\text{Prob}\{E_{mj} \wedge E_{im}^c\}$

$$\cong \int_{-\frac{\mu_x}{\sigma_x}}^{\infty} \frac{\exp -\frac{x^2}{2}}{\sqrt{2\pi}}\, dx \int_{-\frac{\mu_y}{\sigma_y}}^{\infty} \frac{\exp -\frac{1}{2}\frac{(y - \rho x)^2}{1 - \rho^2}}{\sqrt{2\pi(1 - \rho^2)}}\, dy$$

where the order of integration is such that

$$0 \le -\frac{\mu_x}{\sigma_x} \le -\frac{\mu_y}{\sigma_y}$$

Let δ be any positive number,

$$0 < \delta\sqrt{N} < -\frac{\mu_x}{\sigma_x} + \rho\frac{\mu_y}{\sigma_y}.$$

Then

$$\text{Prob}\{E_{mj} \wedge E_{im}^c\} \cong \left\{\int_{-\frac{\mu_x}{\sigma_x}}^{-\frac{\mu_y}{\sigma_y} - \delta\sqrt{N}} + \int_{-\frac{\mu_y}{\sigma_y} - \delta\sqrt{N}}^{\infty}\right\} \frac{\exp -\frac{x^2}{2}}{\sqrt{2\pi}}\, dx \int_{-\frac{\mu_y}{\sigma_y}}^{\infty} \frac{\exp -\frac{1}{2}\frac{(y - \rho x)^2}{1 - \rho^2}}{\sqrt{2\pi(1 - \rho^2)}}\, dy$$

$$\le \text{Prob}\{E_{mi}\} \left[\int_{\frac{\delta\sqrt{2}}{1-\rho^2}}^{\infty} \frac{\exp -\frac{y^2}{2}}{\sqrt{2\pi}}\, dy\right] + \int_{-\frac{\mu_y}{\sigma_y} - \delta\sqrt{N}}^{\infty} \frac{\exp -\frac{x^2}{2}}{\sqrt{2\pi}}\, dx$$

$$= \text{Prob}\{E_{mi}\} \times O(r^N),\ 0 < r < 1.$$

Thus it has been established that the probability of prediction is given asymptotically by:

$$P \cong p_m - \sum_{i \neq \mu} (p_m - p_i) \text{ Prob } \{E_{mj}\} \cong p_m - \sum_{i \neq m} (p_m - p_i) \int_{-\frac{\mu_i}{\sigma_i}}^{\infty} \frac{\exp -\frac{1}{2} x^2}{\sqrt{2\pi}} \, dx$$

where μ_i and σ_i are defined above

Note that, as one would expect, the prediction probability approaches the probability of the mode exponentially.

REFERENCE

1. E. L. Lehmann, Testing Statistical Hypotheses, Wiley (1959) page 305

AN ALGORITHMIC CODE FOR MONOTONIC SOURCES

G. A. Gordon

INTRODUCTION

Suppose the output of an information source that has an alphabet of q characters is to be coded for transmission, and it is known that the source can only emit monotonic strings of these characters, that is, if the source alphabet is denoted by

$$S = \{S_1, S_2, \cdots, S_q\}, \tag{1}$$

then the only permitted output sequences of length m are those given by

$$S_{j_1}, S_{j_2}, \cdots, S_{j_m}, \text{ where} \tag{2}$$

$$j_1 \leq j_2 \leq \cdots \leq j_m. \tag{3}$$

An obvious situation in which such a source would be considered is one where an analogue waveform is sampled and quantized, and where the waveform is known to be monotonic. Because of this natural interpretation of the source, we will often use the terms quantizing level, and sample in place of the terms source character, and character output. The case where the waveform is monotonically increasing, and the decreasing case, are equivalent. In the former case, the character S_1 corresponds to the lowest quantizing level, while in the latter case S_1 represents the highest level. In either case we will call S_1 level one, and in general, we will call S_j the j^{th} level. In terms of levels, then, the permitted sample sequences are given by j_1, $j_2 \cdots$, j_m, where Equation (3) holds.

In order to take advantage of this known restriction on the source, we would encode sequences of samples rather than coding the samples themselves. This is because, out of the q^m possible sequences of m samples, only a fraction of these represent monotonic sequences, and only these sequences need be provided with code words. The remainder of this paper is concerned with an evaluation of the benefits of coding these sequences and a method of doing so.

THE NUMBER OF MONOTONIC SEQUENCES OF LENGTH m

A source output sequence of length m can be uniquely represented by a path through a $(q \times m)$ matrix of points, where the rows correspond to the quantizing levels, and the columns correspond to the samples. The representation of several monotonic sequences is shown in Figure A-14 for $q = 4$, $m = 6$. The number of monotonic sequences of length m is clearly equal to the number of monotonically increasing paths that can be drawn in the $(q \times m)$ matrix.

The number of monotonic paths for a given q and m is calculated as a problem in combinatorial analysis. The solution is obtained by equating the problem to one that is already solved, that of calculating the number of ways in which m identical objects, say balls, can be placed in q cells. We associate a cell with each row of the matrix, and a

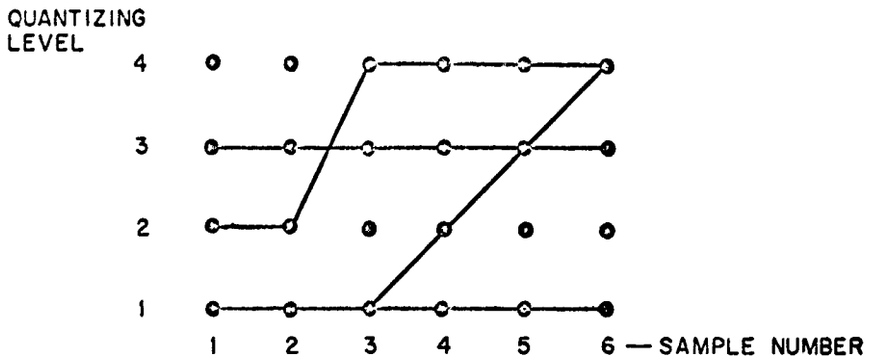


Figure A-14–The representation of several monotonic sequences.

ball with each of the m specifications involved in the designation of a particular path through the matrix.

Given any path through the matrix, we can assign a unique distribution, corresponding to that path, of m balls in the q cells by tracing the path and placing a ball in a cell whenever the trace passes through a point on that cell's row. The process is illustrated in Figure A-15.

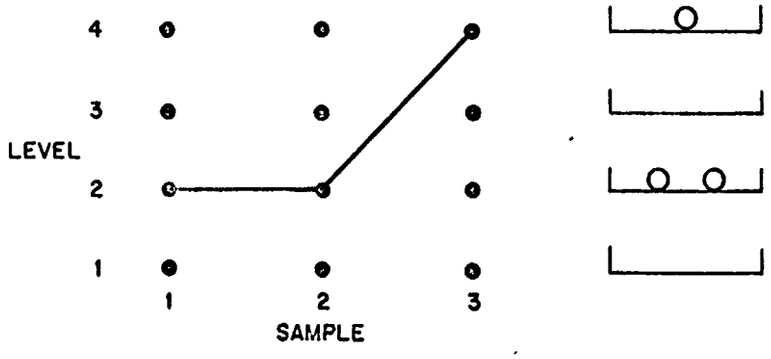


Figure A-15–The path at the left determines the distribution at the right.

Also, if we are given a distribution of m balls in the q cells, a unique monotonically increasing path can be defined as follows. We begin with the cells vertically arrayed alongside the (q × m) matrix of points, as in Figure A-16, and proceed to draw the path from left to right. We start the path at the lowest level which has an occupied cell, and continue the path on that level for as many points as there are balls in that cell in excess of one. Then the path rises to the next highest level that has an occupied cell. The process continues until all m balls have been accounted for, at which point the path is complete, as in Figure A-17.

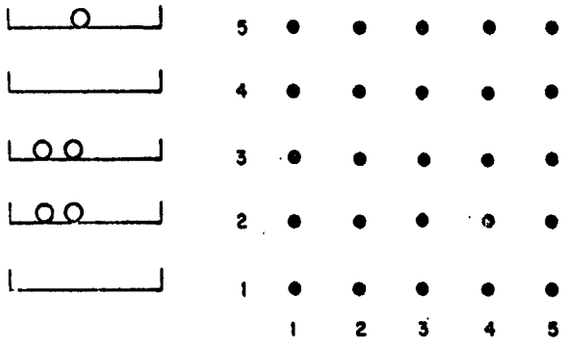


Figure A-16–Ready to draw the path.

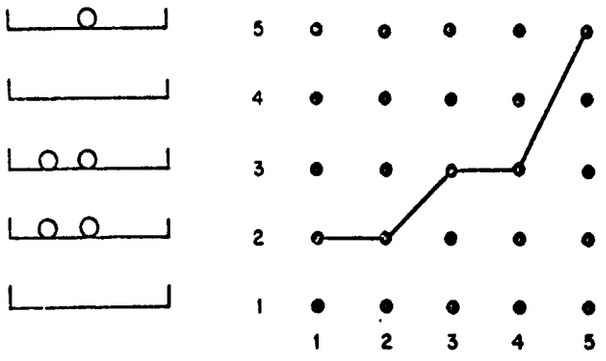


Figure A-17–The path is complete.

We have established the one to one correspondence required and we can state that the number of monotonically increasing paths of length m is given by

$$N_{q,m} = \frac{(m+q-1)!}{m!\,(q-1)!} = \binom{m+q-1}{m}. \tag{4}$$

See Reference (1).

THE BENEFITS OF SEQUENCE CODING

We see that if we code the source output in terms of sequences of m samples we require $N_{q,m}$ code words, one of which is transmitted every time m samples are emitted by the source. If we assume binary block coding, then the transmission rate will be

$$R = \frac{1}{m} \log_2 N_{q,m} \text{ bits/sample}, \tag{5}$$

while direct coding of the samples requires

$$R' = \log_2 q \text{ bits/sample}. \tag{6}$$

Since, the given m and q, the set of monotonic sequences is a subset of the set of all sequences, we have

$$N_{q,m} \leq q^m, \tag{7}$$

with equality only in the trivial cases $m = 1$, or $q = 1$. Equation (7) implies that (for m, $q \neq 1$),

$$\log N_{q,m} < \log q^m = m \log q, \text{ so that}$$

$$R < R'.$$

Thus, sequence coding always results in some saving in transmission rate.

The number of levels, q, is generally determined by considerations such as the required data precision, or the amount of tolerable quantization noise. But the choice of m is part of the coding problem. We show below that as m increases, the coding efficiency continues to improve. In particular, we prove the following

Theorem:
$$\frac{R(m)}{R'} > \frac{R(m+1)}{R'} \quad , \quad q \neq 1.$$

Proof:
$$\frac{R(m)}{R'} = \frac{\log N_{q,m}}{\log q^m} \quad ,$$

using Equations (5) and (6).

Consider the ratio

$$\frac{N_{q,m}}{q^m} = \frac{(m+q-1)!}{q^m m!\,(q-1)!} = \left(\frac{q}{q}\right)\left(\frac{q+1}{2q}\right)\left(\frac{q+2}{3q}\right)\cdots\left(\frac{q+m-1}{mq}\right).$$

Similarly,

$$\frac{N_{q,m+1}}{q^{m+1}} = \frac{(m+q)}{q^{m+1}\,(m+1)!\,(q-1)!} = \left(\frac{q}{q}\right)\left(\frac{q+1}{2q}\right)\cdots\left(\frac{q+m-1}{mq}\right)\left(\frac{q+m}{(m+1)q}\right),$$

$$= \frac{N_{q,m}}{q^m}\left[\frac{q+m}{(m+1)q}\right] = \frac{N_{q,m}}{q^m} \cdot \frac{q+m}{q+qm} \quad ,$$

$$< \frac{N_{q,m}}{q^m} \quad , \quad q \neq 1.$$

We now have that (assuming $q > 1$)

$$\frac{N_{q,m+1}}{q^{m+1}} < \frac{N_{q,m}}{q^m}$$

Taking logarithms of both sides maintains the inequality. Thus,

$$\log N_{q,m+1} - \log q^{m+1} < \log N_{q,m} - \log q^m .$$

Noting that $\log q^{m+1} > \log q^m$, we can write that

$$\frac{\log N_{q,m+1} - \log q^{m+1}}{\log q^{m+1}} < \frac{\log N_{q,m} - \log q^{m}}{\log q^{m}};$$

thus

$$\frac{\log N_{q,m+1}}{\log q^{m+1}} < \frac{\log N_{q,m}}{\log q^{m}},$$

and the proof is complete.

The theorem indicates that $R(m+1) < R(m)$; that is, each increase in m decreases the transmission rate. Of course, as m increases, our coding equipment necessarily becomes greater in size and more energy consuming, since the number of code words increases. Also, since the code word is not transmitted until all m samples are received by the encoder, increasing m results in greater delay between the production of data and its arrival at the decoder. If command decisions are to be made on the basis of the data, this factor may limit the value of m to be used.

The coding technique which will be discussed below is applicable to a source which is not strictly monotonic, but only monotonic over long sequences of outputs. In this "piece wise monotonic" case, the statistics of the monotonic run length may determine an optimum value for m.

We conclude the section with examples.

Example 1: $q = 64$, $m = 10$.

Character coding requires 6 bits per sample. Sequence coding requires 4 bits per sample. Saving = 2 bits per sample = 33%

Example 2: $q = 64$, $m = 20$.

Sequence coding requires 3.15 bits per sample. Saving = 2.85 bits per sample = 47.5%.

THE ENCODER

For fixed q and m, we can of course enumerate the possible monotonic sample sequences and assign code words to them in a tabular fashion. This coding table could then be realized in logic circuitry either as a combinatorial or sequential machine. Anyone familiar with logic circuit design will appreciate the difficulty of achieving even a suboptimal design for large q and m. Furthermore, changing the value of q or m introduces an entirely new problem with an entirely new coding machine. Finally, this approach is completely in applicable when the source is only piecewise monotonic.

Another approach is to seek an algorithmic coding procedure which could be realized with a general, or special purpose computor. By an algorithmic code we mean that the code words are formed by operating on the sample data according to a well defined mathematical program.

In order to describe the algorithmic code discovered, we represent the source output sequences by paths through a $(q \times m)$ matrix of points, as on page A-58. We associate a non-negative integer, $c_{i,j}$, with each point in the matrix and code each sequence by the

number equal to the sum of the numbers $c_{i,j}$ associated with the points (i, j) on the corresponding path. We assign the numbers to the matrix so that each monotonic path has a unique code number, and so that the highest code number used is $N_{q,m}-1$, since the code number 0 is used. In this way, we minimize the number of bits in the binary code words, and hence the transmission rate.

It is shown, in the following sections, that the matrix $(c_{i,j})$ can always be taken as an $(m \times q)$ sub-matrix of the infinite matrix shown in Figure A-18. The matrix is extended indefinately according to the rule

$$c_{i,j} = c_{i-1,j} + c_{i,j-1}, \tag{8}$$

and is called the coding matrix.

7	28	84	210	462	924 ----
6	21	56	126	252	462 ----
5	15	35	70	126	210 ----
4	10	20	35	56	84 ----
3	6	10	15	21	28 ----
2	3	4	5	6	7 ----
1	1	1	1	1	1 ----
0	0	0	0	0	0 ----

q (rows 4 through 0)

m (columns)

Figure A-18–The coding matrix.

The encoding machine is shown diagrammatically in Figure A-19. Before the first of m samples is received, the q registers are reset to the numbers in the first column of the matrix, and the accumulator is cleared. If the first sample is the number K, then the K^{th} register is addressed and its contents are added to the contents of the accumulator. Before the next sample is received, the updating circuitry generates the second column of the matrix in the registers. Since the first two rows of the matrix are constant, the lower two registers need not be updated. To update the third register, we add one to its contents. The rest of the registers are updated by adding to them the contents of the register below; that is, we add the contents of register 3 to register 4, then we add the contents of register 4 to register 5, etc. That this procedure generates the next column is evident from Equation (8). The incoming sample then addresses the proper register, whose contents are again added to the contents of the accumulator. When m samples are received, the accumulator contains the code word, which is transferred to a buffer for transmission.

The design is independent of m and q. Increasing q increases the number of registers. Increasing m increases the size of the registers. The size of the accumulator, addressing circuits, adders, and counters involved are correspondingly increased. But none

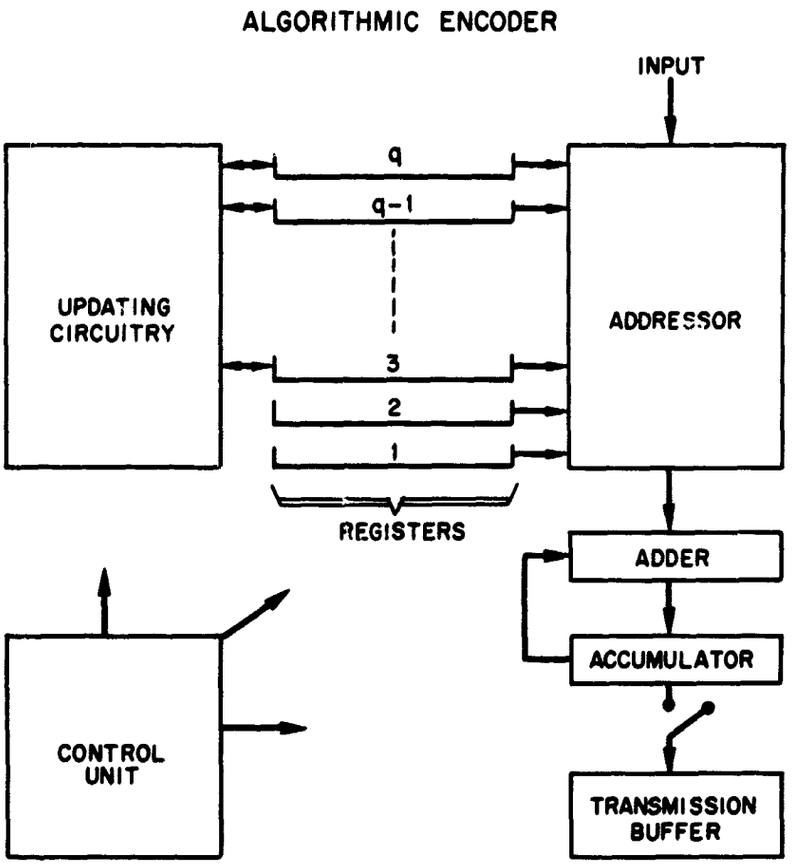


Figure A-19–The coding machine.

of these changes involve redesign of the code machine, but rather building the same machine with larger components. The components themselves are common and designs exist for a wide range of specifications.

If a machine is designed for a particular m and q, it will work, with only slight modifications for m' and q', each respectively smaller or equal to m and q. This is basically due to the fact that the $(q' \times m')$ matrix is just a submatrix of the $(q \times m)$ matrix. This also enables us to use the scheme in the piece wise monotonic case.

THE CODING MATRIX

The construction of the coding matrix is based on another matrix, called the "path count" matrix. The number associated with a point (i, j) of this second infinite matrix is equal, by definition, to the number of monotonic (increasing) paths leading to that point, and is denoted $d_{i,j}$. The matrix is shown on the following page.

The coding matrix is defined as that matrix formed by the deletion, from the path count matrix, of the first column, and the addition of a row of zeros below. Because of this relationship, we can derive many useful properties of the coding matrix by studying the path count matrix.

To begin with, we point out that every monotonic path that terminates at the point (i, j) must pass through one of the points $(i, j-1), (i-1, j-1), (i-2, j-1), - - -,$

1	6	21	56	126	252	-----
1	5	15	35	70	126	-----
1	4	10	20	35	56	-----
1	3	6	10	15	21	-----
1	2	3	4	5	6	-----
1	1	1	1	1	1	-----

Figure A-20–The path count matrix.

(2, j - 1), (1, j - 1). Also, any monotonic path which terminates at one of these points can be extended to the point (i, j) forming a monotonic path. This clearly indicates that

$$d_{i,j} = \sum_{K=1}^{i} d_{K,j-1} . \tag{9}$$

But, since every monotonic path in the (i × j - 1) submatrix terminates at one of the points (K, j - 1), K = 1, 2, - - -, i, we can also write that

$$\sum_{K=1}^{i} d_{K,j-1} = N_{i,j-1} , \tag{10}$$

where $N_{q,m}$ is as given in Equation (4). From Equations (9) and (10) we see that

$$d_{i,j} = N_{i,j-1} . \tag{11}$$

Equation (11) explains the noticeable fact that the path count matrix elements are binomial coefficients.

Finally, we rewrite Equation (9) as

$$d_{i,j} = d_{i,j-1} + \sum_{K=1}^{i-1} d_{K,j-1}$$

But, with a change of variable in Equation (9) we get,

$$d_{i-1,j} = \sum_{K=1}^{i-1} d_{K,j-1} ,$$

so that

$$d_{i,j} = d_{i,j-1} + d_{i-1,j} \tag{12}$$

It is obvious that the equivalent relation for the coding matrix, expressed in Equation (8), follows from Equation (12) and the manner in which the coding matrix is formed.

All the elements in the coding matrix appear in the path count matrix, except for the bottom row of zeros. The elements outside this row are given by

$$c_{i,j} = d_{i-1,j+1} \tag{13}$$

Applying Equation (13) to Equation (9) we get,

$$c_{i+1,j-1} = \sum_{K=1}^{i} c_{K+1,j-2},$$

and with changes of variable,

$$c_{i,j} = \sum_{K=2}^{i} c_{K,j-1}.$$

But since

$$c_{i,j-1} = 0,$$

$$c_{i,j} = \sum_{K=1}^{i} c_{K,j-1}. \tag{14}$$

Applying Equation (13) to Equation (11) we get,

$$c_{i,j} = N_{i-1,j}. \tag{15}$$

Finally, by repeated use of Equation (8), we get

$$\begin{aligned} c_{i,j} &= c_{i-1,j} + c_{i,j-1} \\ &= c_{i-1,j} + c_{i-1,j-1} + c_{i,j-2} \\ &= c_{i-1,j} + c_{i-1,j-1} + c_{i-1,j-2} + c_{i,j-3} \\ &\quad \cdots\cdots\cdots\cdots\cdots \\ &= c_{i-1,j} + c_{i-1,j-1} + \cdots + c_{i-1,2} + c_{i,1}. \end{aligned}$$

But

$$c_{i,1} = c_{i-1,1} + 1,$$

so that

$$c_{i,j} = \sum_{K=1}^{j} c_{i-1,K} + 1 \tag{16}$$

These properties of the coding matrix will be used in the following section.

THE DECODING ALGORITHM AND THE PROOF OF UNIQUE DECODABILITY

On page we stipulated that the numbers $c_{i,j}$ were to be chosen so that each monotonic path drawn in a $(q \times m)$ submatrix of the coding matrix would have a unique code number, and so that the highest code number used is given by $N_{q,m}-1$. These two stipulations indicate that the code numbers to be assigned to the monotonic paths are the numbers 0, 1, 2, ---, $N_{q,m}-1$.

We prove that the code is uniquely decodable in the following manner; we show that

A. The highest code number assigned to any monotonic path in a $(q \times m)$ submatrix is equal to $N_{q,m}-1$.

B. A monotonic path in a $(q \times m)$ submatrix can be found for each of the numbers 0, 1, 2,---, $N_{q,m}-1$ relative to some decoding scheme. That is, given any one of the above numbers, we can find a path which has that number as its code.

Statement A indicates that to every monotonic path in the $(q \times m)$ submatrix there corresponds one of the code numbers 0, 1, 2,---, $N_{m,q}-1$, while statement B assures us that for each of the above code numbers, a corresponding path with that number as its code can be found. Clearly, this one to one correspondence indicates unique decodability.

The proof of statement A is elementary. From Equation (8) we see that

$$c_{i,j} > c_{i-1,j} \quad i = 2, 3, \text{---} \tag{17}$$

so that the largest element in the j^{th} column of the $(q \times m)$ submatrix is $c_{q,j}$. Thus the highest code number, corresponding to the sequence $qq\text{---}q$, is given by

$$\sum_{K=1}^{m} c_{q,K}.$$

But from Equation (16) we see that

$$\sum_{K=1}^{m} c_{q,K} = c_{q+1,m} - 1;$$

and, using Equation (15), $= N_{q,m} - 1$. This completes the proof of A. The proof of statement B necessarily requires the definition of a decoding procedure. Decoding is performed, as is encoding, using the $(q \times m)$ submatrix of the coding matrix.

Let a_1 denote the code number to be decoded; then by hypothesis

$$0 \leq a_1 < N_{q,m}. \tag{18}$$

To begin decoding, we compare a_1 to the elements $c_{i,m}$ in the last column of the submatrix until we find the element $c_{i_m,m}$ such that

$$c_{i_m,m} \leq a_1 < c_{i_m+1,m} . \tag{19}$$

Then, by definition, the point (i_m, m) is on the path.

The first inequality in Equation (19) can always be satisfied since $0 = c_{1,m} \leq c_{i_m,m} \leq a_1$. The second can always be satisfied since $N_{q,m} = c_{q+1,m} \geq c_{i_m+1,m} > a$, using Equations (15) and (18).

The number $c_{i_m,m}$ is subtracted from a_1 to get

$$a_2 = a_1 - c_{i_m,m} . \tag{20}$$

Since (i_m, m) must be as "high" as any point on the path, the remainder of the path must be in the $(i_m \times m - 1)$ submatrix. Also, the code for this subpath must be the number a_2. That is, the matrix elements corresponding to the points on the subpath must add to a_2. Then we must have

$$a_2 \leq \sum_{K=1}^{m-1} c_{i_m,K} , \tag{21}$$

since this is the largest code number for a monotonic path in the $(i_m \times m - 1)$ submatrix.

But from Equations (20) and (21),

$$a_2 < c_{i_m+1,m} - c_{i_m,m}$$

$$a_2 < c_{i_m+1,m-1}, \text{ using (8)}$$

$$a_2 < \sum_{K=1}^{m-1} c_{i_m,K} + 1, \text{ using (16).} \tag{22}$$

Equation (21) follows, but all that this indicates so far is that we can continue to decode.

Using Equations (15) and (22) we have

$$0 \leq a_2 < N_{i_m,m-1} . \tag{23}$$

We are back at the point where we wrote Equation (18), but our path is partially defined (one point) and the submatrix has been reduced to $(i_m \times m - 1)$. Otherwise the situation is the same and we can find $c_{i_{m-1},m-1}$ such that

$$c_{i_{m-1},m-1} \leq a_2 < c_{i_{m-1}+1,m-1};$$

and, by definition, the point $(i_{m-1}, m - 1)$ is on the path.

By induction, we can continue to decode the path point by point until our submatrix is reduced to a column vector. At that time, the remainder a_m must be related to the previous point (i_2, 2) by the inequality, $0 \leq a_m < N_{i_2,2}$, by analogy with Equation (23). But $N_{i_2,2} = i_2$, so that $a_m < i_2$. The last point decoded can then be taken as the point $(a_m + 1, 1)$, since $c_{a_m+1,1} = a_m$, and since $a_m + 1 \leq i_2$, so that the point is not "higher" than level i_2.

Thus $i_1 = a_m + 1$, and the decoded sequence is $i_1 i_2 \text{---} i_m$. Repeated use of Equation (2) yields

$$
\begin{aligned}
a_1 &= c_{i_m,m} + a_2 \\
&= c_{i_m,m} + c_{i_{m-1},m-1} + a_3 \\
&\quad \text{- - - - - - - - - - - - -} \\
&= c_{i_m,m} + c_{i_{m-1},m-1} + \text{- - -} + c_{i_2,2} + a_m \\
&= \sum_{K=1}^{m} c_{i_K,K},
\end{aligned}
$$

so that a_1 is indeed the code number for the sequence defined. The unique decodability proof is now complete.

In order to decode a received number uniquely, the value of m must be known to the decoder, since this is the matrix column in which the first comparison is made. For example, if we received the code number 7 and m was not specified, then the sequence could be any of those below.

$$
\begin{aligned}
i_1 &= 8 \\
i_1, i_2 &= 2, 4 \\
i_1, i_2, i_3 &= 1, 3, 3 \\
i_1, i_2, i_3, i_4 &= 1, 2, 2, 3 \\
i_1, i_2, i_3, i_4, i_5 &= 1, 1, 1, 2, 3 \\
&\text{etc.}
\end{aligned}
$$

The reader can verify, using Figure A-17 that these sequences indeed have the code number 7.

Thus, the coder and decoder must agree beforehand on the value of m to be used. In the piecewise monotonic case the value of m agreed upon is only a "nominal" value since the value of $m' < m$ must be transmitted if the path slope changes sign in the middle of a sequence. If the monotonic run length is much greater than m, then this will not happen too often and is not a serious drawback.

CONCLUSIONS

The particular source discussed in this paper is only one of a class of sources, those q-level sources which can emit only a subset of the q^m sequences of length m. Since the appearance of this paper in preliminary form, the sequence coding technique, utilizing the concepts of the path count and coding matrices, has been applied by Mr. J. Snively to sources with bounded level jumps between successive samples, and monotonic sources with bounded level jumps. Doubtlessly, there are many other source restrictions of this type which are amenable to the coding technique.

An interesting question is whether or not a coding matrix exists for an arbitrary enumerative restriction. That is, if we enumerate a subset of sequences which are "allowed" (this can be done in 2^{q^m} different ways), can we find a coding matrix so that the allowed paths have unique code numbers, and so that the numbers used are 0, 1, 2,---, N - 1, where N is the number of paths in the subset? Perhaps this is a good topic for further study.

REFERENCE

1. Feller, W., An Introduction to Probability Theory and Its Applications, John Wiley & Sons, (1950) page 36.

INFORMATION CONTENT OF A q-LEVEL m-SEQUENCE WITH THE RESTRICTION THAT THE SUM BE n

M. Javid

A q-level m-sequence is a sequence of m samples of data quantized into q levels. Thus the sequences 1, 0, 2, 1, 6, 7 and 2, 0, 1, 3, 7, 4 are two examples of 8-level 6-sequences, where the data is quantized into 8 levels, zero to seven. In each of these the sum of the sequence is n = 17. If there are no restrictions on q-level m-sequences, then there are q^m different sequences. However, if there are specified restrictions on the sequences, their number is less than q^m, since some of the q^m sequences do not satisfy the specified restrictions. Let $N_R(q, m)$ be the number of q-level m-sequences which satisfy specified restrictions R. Then we define the information content of such sequences to be $\log_2 N_R(q, m)$ bits per sequence. This definition is motivated by the fact that for transmission of any one of the $N_R(q, m)$ q-level m-sequences one may transmit its order number ranging from 1 to $N_R(q, m)$. This requires $\log_2 N_R(q, m)$ which is always less than $m \log_2 q = \log_2 q^m$ bits per m-sequence necessary for transmission of a q-level m-sequence subject to no restrictions.

As an example there are $4^2 = 16$ 4-level 2-sequences, namely 00, 01, 02, 03, 10, 11, 12, 13, 20, 21, 22, 23, 30, 31, 32 and 33. Transmission of each of these requires $m \log_2 q = 2 \times \log_2 4 = 4$ bits per sequence. There are only 4 of these sequences which add up to 3. These are 03, 12, 21, 30. If it is known that, due to the nature of the data to be transmitted, the sum of the sequences add up to 3, then one may use $\log_2 N_3(4, 2) = \log_2 4 = 2$ bits/sequence instead of 4 bits per sequence required when there are no restrictions. The economy in bit transmission is obtained by coding the sequences 03, 12, 21 and 30 with their (arbitrarily chosen) order numbers, expressed in base 2 as 00, 01, 10 and 11, respectively.

From above discussion it is seen that analysis involved in the study of information content of q-level m-sequences consist of two distinct parts.

1. The problem of enumeration, i.e., finding an expression for $N_R(q, m)$.
2. The problem of encoding and decoding.

The latter problem may be explained in terms of the example of 4-level 2-sequences whose sum is 3, discussed above. In this example we (arbitrarily) labelled the sequences 03, 12, 21 and 30 with codes 00, 01, 10 and 11. Although in theory one may always form a correspondence table (dictionary) giving the one-to-one correspondence of the sequences and their codes, in practice this is not feasible, since with large values of m and q the number of entries in the table become very large. In such cases it is desirable to find an algorithm for coding and encoding. Thus in above example an algorithm is required to give the binary code, say, 10, given the sequence 21, and conversely, to find the sequence 21, given the binary code 10. Now we will derive the formula for $N_R(q, m)$ for restriction R, being that the sum of the sequence is n, and develop an algorithm for coding and encoding.

NUMBER OF q-LEVEL m-SEQUENCES WHOSE SUM IS n

The enumerator of this problem is (see Introduction to Combinatorial Analysis by John Riordan, John Wiley and Sons, New York 1958)

$$E(t;\, m,\, q) = (1 + t + t^2 + \cdots + t^{q-1})^m \tag{1}$$

This means that if we write

$$(1 + t + t^2 + \cdots + t^{q-1})^m = \sum_{k=0}^{m(q-1)} A_k\, t^k \tag{2}$$

Then the coefficient of the n^{th} power of t, that is, A_n, is the number of q-level m-sequences whose sum is n. The truth of this statement may be illustrated by the example used above, that is, $q = 4$, $m = 2$ and $n = 3$. If instead of $(1 + t + t^2 + \cdots + t^{q-1})^m = (1 + t + t^2 + t^3)^2$ we consider

$$(1 + xt + x^2t^2 + x^3t^3)(1 + yt + y^2t^2 + y^3t^3)$$

we note that the terms of this product in which $t^n = t^3$ occurs are obtained by taking

1. one from $(1 + xt + x^2t^2 + x^3t^3)$ and y^3t^3 from $(1 + yt + y^2t^2 + y^3t^3)$ giving y^3t^3
2. xt " " " x^2t^2 " " " xy^2t^3
3. x^2t^2 " " " yt " " " x^2yt^3
4. x^3t^3 " " " 1 " " " x^3t^3

Thus the coefficient of t^3 in the product

$$(1 + xt + x^2t^2 + x^3t^3)\,(1 + yt + y^2t^2 + y^3t^3)$$

is $(y^3 + xy^2 + x^2y + x^3)$. If we set $x = 1$ and $y = 1$ this coefficient is $A_3 = 1 + 1 + 1 + 1 = 4$. Now a term like y^3 corresponds to the sequence 03, xy^2 corresponds to 12, x^2y corresponds to 21 and x^3 corresponds to 30. Thus there are $A_3 = 4$ 4-level 2-sequences whose sum is 3. This reasoning can be generalized to give a proof of the Equation (1). The next step is to find the coefficient of t^n in Equation (1). For this purpose we note that

$$\sum_{k=0}^{m(q-1)} A_k\, t^k = (1 + t + t^2 + \cdots + t^{q-1})^m$$

$$= \left[(1 + t + t^2 + \cdots + t^{q-1} + t^q + t^{q+1} + \cdots + t^{\ell\to\infty}) - (t^q + t^{q+1} + \cdots + t^{\ell\to\infty})\right]^m$$

$$= \left[(1 + t + t^2 + \cdots + t^{\ell\to\infty}) - t^q(1 + t + t^2 + \cdots + t^{\ell\to\infty})\right]^m$$

$$= \left[(1 + t + t^2 + \cdots + t^{\ell\to\infty})(1 - t^q)\right]^m$$

$$= (1 + t + t^2 + \cdots + t^{\ell\to\infty})^m\,(1 - t^q)^m \tag{3}$$

But for $t < 1$

$$(1 + t + t^2 + \cdots + t^{\ell \to \infty}) = \frac{1}{1 - t} = (1 - t)^{-1} \tag{4}$$

$$(1 + t + t^2 + \cdots + t^{\ell \to \infty})^m = (1 - t)^{-m} = \sum_{r=0}^{\infty} \binom{-m}{r} (-t)^r \tag{5}$$

where in Equation (5) a binomial expansion is used and

$$\binom{-m}{r} = \frac{-m(-m-1)\cdots(-m-r+1)}{r!} = (-1)^r \binom{m+r-1}{r} \tag{6}$$

From above it follows that

$$(1 - t)^{-m} = \sum_{r=0}^{\infty} \binom{m+r-1}{r} t^r \tag{7}$$

Furthermore

$$(1 - t^q)^m = \sum_{p=0}^{m} (-1)^p t^{pq} \binom{m}{p} \tag{8}$$

$$\sum_{k=0}^{m(q-1)} A_k t^k = \sum_{k=0}^{\infty} \binom{m+r-1}{r} t^r \sum_{p=0}^{m} (-1)^p \binom{m}{p} t^{pq}$$

$$= \sum_{r=0}^{\infty} \sum_{p=0}^{m} (-1)^p \binom{m+r-1}{r} \binom{m}{p} t^{r+pq} \tag{9}$$

For $$n = r + pq \,, \quad r = n - pq \tag{10}$$

and the coefficient of t^n is

$$A_n = \sum_{p=0}^{m} (-1)^p \binom{m+n-pq-1}{n-pq} \binom{m}{p} \tag{11}$$

Thus the number of q-level m-sequences whose sum is n is given by

$$N_n(q, m) = \sum_{p=0}^{m} (-1)^p \binom{m+n-pq-1}{n-pq} \binom{m}{p} \tag{12}$$

and the information content of these sequences is

$$\log_2 N_n (q, m).$$

In Equation (11) the value of $\binom{a}{b}$ is assumed to be zero when $b < 0$.

The following table gives the number of 8-level 6-sequences whose sum is n = 0 to 22, and the corresponding information content (bits per sequence of 6 samples) These values were computed by Mr. J. Snively.

n =	0	2	3	4	5	6	7	8	9	10	11	12	13	14
$N_n (8, 6) =$	1	6	21	56	252	462	792	1281	1966	2877	4032	5432	7056	8856
$\log_2 N_2 (8, 6) \simeq$	.00	2.58	4.39	5.80	7.96	8.84	9.62	10.41	10.93	11.47	11.96	12.39	12.77	13.10

n =	15	16	17	18	19	20	21	22
$N_n (8,6) =$	10372	12642	14412	15946	17136	17892	18152	17892
$\log_2 N_n (8,6) \simeq$	13.37	13.61	13.80	13.95	14.05	14.11	14.13	14.11

The values for n = 22 to 42 may be found from the symmetry formula

$$N_{42-n} (8, 6) = N_n (8, 6)$$

The saving in bits may be noted by comparing $\log_2 N_n(8,6)$ with $\log_2 8^6 = 6 \times 3 = 18$.

ENCODING

Following is a description of a simple, but non-compact coding scheme. Additional steps for producing a compact code are also described. The q-level m-sequence is considered to be a number to base q. For example if m = 3, q = 8 and n = 15 then a sequence 546 is an allowed sequence and is considered to be the number to the base 8 that is

$$(546)_8 = (358)_{10}$$

From this number is subtracted the lowest of q-level m-sequences whose sum is n. For q = 18, m = 3 and n = 15 this is $(177)_8 = (127)_{10}$ $(546)_8 - (177)_8 = (358)_{10} - (127)_{10} = (231)_{10}$.

This result is then divided by q - 1 to give the code for the sequence. In the above example

$$(231)_{10} + (7)_{10} = (33)_{10}$$

is the code number for 18-level 3-sequence whose digits add up to 15.

This code is easily decipherable by going through the reverse of the procedure described above. Thus given 33 we find,

$$(33)_{10} \times (7)_{10} = (231)_{10} = (347)_8$$

$$(347)_8 + (177)_8 = (546)_8$$

where again $(177)_8$ is the smallest 3-digit number to the base 8 whose digits add up to 15.

As mentioned above this code is not compact. For example for q = 8, m = 3 and n = 2, the sequences range from 177 to 771 and from

$$N_n\,(q,\ m) = \sum_{p=0}^{m} (-1)^p \binom{m + n - pq - 1}{n - pq} \binom{m}{p}$$

There are

$$N_{15}\ (8,\ 3) = 136 - 108 = 28$$

sequences. The above encoding scheme assigns zero to 177 and

$$(771)_8 - (177)_8 + (7)_8 = (572)_8 + (7)_8 = (54)_{10}$$

Thus instead of assigning the compact codes 1 to 28 (or zero to 27) to the 8-level 3-sequences the encoding scheme assigns 0 to 54, with some of the in-between numbers not used. It is pointed out that there are $8^3 = 512$ 8-level 3-sequences, requiring $\log_2 512 = 9$ bits per sequence for their transmission. When the sum of sequences is restricted to 15, there are 28 allowed sequences requiring $\log_2 28 = 4.0$ bits, whereas with the above coding $\log_2 54 = 5.0$ an addition of one bit per sequence. There is still an economy of more than 3 bits per sequence, using the knowledge of the restriction on the sum and the described simple encoding. To make the code compact we note that the above scheme assigns the following code numbers to the case q = 8, m = 3, n = 15.

Sequence	177	267	276	357	366	375	447	456	465	474
Code	0	8	9	16	17	18	24	25	26	27
Order No.	0	1	2	3	4	5	6	7	8	9
Jump		7		6			5			
Sequence	537	546	555	564	573	627	636	645	654	663
Code	32	33	34	35	36	40	41	42	43	44
Order No.	10	11	12	13	14	15	16	17	18	19
Jump	4					3				
Sequence	672	717	726	735	744	753	762	771		
Code	45	48	49	50	51	52	53	54		
Order No.	20	21	22	23	24	25	26	27		
Jump		2								

In the above table it is seen that whenever the second digit of a sequence is 7, the code of the following sequence has a jump, this jump starting (in this example) from 7 (code of 177 is zero and code of 267 is 8 instead of one, resulting in a jump of 8-1=7) and ending at 2

In this example we can find a compact code by noting that the order number (which is compact) differs from the code by the sum of the preceding jumps and if the sequence starts with 2 jump is 7 that is the compact code is the code less 7, if the sequence is 276 with code 9, its compact code (which is the same as the order number) will be

$$\left[(276)_8 - (177)_7 \div (7)_8\right] - (7)_{10} = 2$$

Similarly if the first digit of the sequence is 3 the sum of preceding jumps is 7+6 = 13 and the compact code is the noncompact code minus 13. For a given first digit of a sequence the compact code is obtained from the noncompact code by subtracting the sum of the preceding jumps which is easily computable.

Although it is possible to formalize these steps, the above example illustrates the general procedure of coding.

BLANK PAGE

THE EFFECT OF WAVEFORM OF CORRELATOR REFERENCE VOLTAGE ON THE PERFORMANCE OF PFM SYSTEMS

M. Javid

The probability of error in a PFM System has been calculated by many authors (see References 1 to 4) using the assumption that the correlator reference voltage is sinusoidal. In some of the ground stations connected with GSFC the reference signal is a square wave. The purpose of the following study is to investigate the effect of the waveform of the reference signal on the probability of error of PFM Systems.

Let the transmitted tone be $A \sin \omega_i t$ lasting for a duration T and the received signal be

$$r(t) = \sqrt{2S} \sin \omega_i (t - \tau) + n(t), \quad 0 \leq t \leq T \tag{1}$$

where S is the signal power (mean square value), $n(t)$ is wide band white gaussian noise, and τ is delay of transmission. Consider the i^{th} correlator shown in Figure A-21.

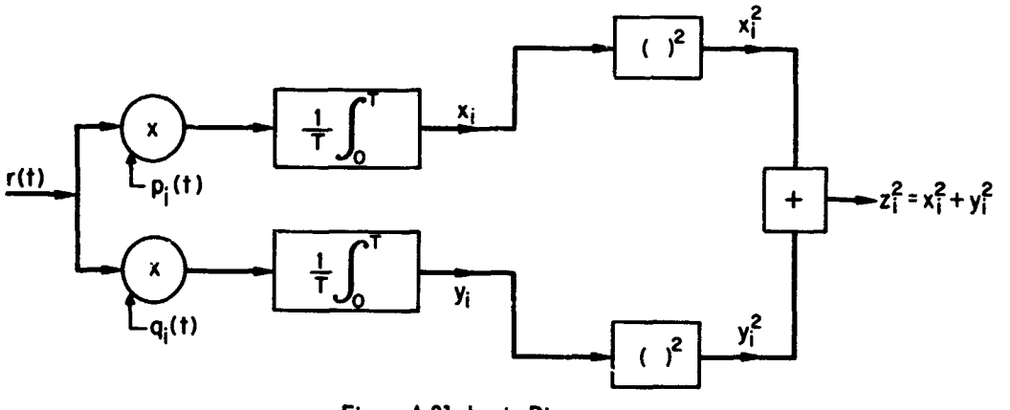


Figure A-21–Logic Diagram

For coherent detection	$p_i(t) = C \sin \omega_i t$	$q_i(t) = 0$
Non-coherent detection	$p_i(t) = C \cos \omega_i t$	$q_i(t) = C \sin \omega_i t$
Non-coherent detection with square-wave	$p_i(t)$ = square wave	$q_i(t)$ = shifted square wave

For purpose of simplification it is assumed that $\omega_i T$ is an integer multiple of 2π for all i, that is, all the tones are harmonics of a fundamental of frequency $1/T$. In the correlator shown, x_i and y_i are random variables. Assuming that ω_i has been transmitted and reference signals are sinusoidal, the expectation of x_i is:

$$\overline{x_i} = \frac{1}{T} \int_0^T \left[\sqrt{2S} \sin \omega_i (t - \tau) + n(t)\right] C \cos \omega_i t \, dt$$

$$= \frac{1}{T} \int_0^T C\sqrt{2S} \sin \omega_i (t - \tau) \cos \omega_i t \, dt = -\frac{1}{2} C\sqrt{2S}\, \delta_{ii} \sin \tau \tag{2}$$

$$\overline{y_i} = \overline{\frac{1}{T}\int_0^T \left[\sqrt{2S}\sin\omega_1(t-\tau) + n(\tau)\right] C \sin\omega_i t\, dt}$$

$$= \frac{1}{T}\int_0^T C\sqrt{2S}\sin\omega_1(t-\tau)\sin\omega_i t\, dt = \frac{1}{2}C\sqrt{2S}\,\delta_{1i}\cos\omega_1\tau \qquad (3)$$

Similarly the expectation of x_i^2 is

$$\overline{x_i^2} = \overline{\frac{1}{T^2}\int_0^T\int_0^T \left[\sqrt{2S}\sin\omega_1(t-\tau) + n(t)\right] C\cos\omega_i t \left[\sqrt{2S}\sin\omega_1(s-\tau) + n(s)\right] C\cos\omega_i s\, dt\, ds}$$

$$= \left\{\frac{1}{T}\int_0^T C\left[\sqrt{2S}\sin\omega_1(t-\tau)\right]\cos\omega_i t\, dt\right\}^2$$

$$+ \frac{1}{T^2}\int_0^T\int_0^T \overline{n(t)\,n(s)}\, C^2\cos\omega_i t\cos\omega_i s\, dt\, ds \qquad (4)$$

To simplify the computation, it is assumed that the noise $n(t)$ is white with autocorrelation function

$$\overline{n(t)\,n(s)} = \frac{N_0}{2}\delta(t-s) \qquad (5)$$

$$\overline{x_i^2} = \overline{x}_i^2 + \frac{1}{T^2}\int_0^T \frac{N_0}{2} C^2\cos^2\omega_i t\, dt = \frac{C^2}{2}\frac{N_0}{2T} + \overline{x}_i^2 \qquad (6)$$

$$\sigma_{x_i}^2 = \overline{x_i^2} - \overline{x}_i^2 = \frac{C^2}{2}\frac{N_0}{2T} \qquad (7)$$

Similarly

$$\sigma_{y_i}^2 = \frac{C^2}{2}\frac{N_0}{2T} = \sigma_{x_i}^2 = \sigma^2 \qquad (8)$$

$$\overline{x_i\, y_i} = \frac{1}{T^2} \int_0^T \int_0^T \overline{\left[\sqrt{2S} \sin \omega_1 (t - \tau) + n(t)\right] C \cos \omega_i t \cdot \left[\sqrt{2S} \sin \omega_1 (s - \tau) + n(s)\right] C \sin \omega_i s \, dt \, ds} = \overline{x}_i \, \overline{y}_i \tag{9}$$

$\therefore \sigma^2_{x_i y_i} \equiv 0$ and x_i and y_i are not correlated.

Since the noise is assumed to be gaussian, x_i and y_i are jointly normal and uncorrelated (see Reference 5). The value of their expectations takes one of the two forms.

(a) If $i = 1$ then from Equations (2) and (3)

$$\overline{x}_1 = -\frac{1}{2} C\sqrt{2S} \sin \omega_1 \tau, \quad \overline{y}_1 = \frac{1}{2} C\sqrt{2S} \cos \omega_1 \tau \tag{10}$$

(b) If $i \neq 1$ then

$$\overline{x}_i = 0, \quad \overline{y}_i = 0 \quad i \neq 1 \tag{11}$$

Therefore (see Probability & Stochastic Processes by Papoulis, pages 145-197) the probability density function of $z_i = \sqrt{x_i^2 + y_i^2}$ for normally distributed uncorrelated x_i and y_i is, for z_1

$$f_{z_1}(z) = \frac{z}{\sigma^2} e^{-(z^2 + \overline{x}_1^2 + \overline{y}_1^2)/2\sigma^2} I_0 \left(\frac{z\sqrt{\overline{x}_1^2 + \overline{y}_1^2}}{\sigma^2} \right), \quad z > 0 \tag{12}$$

where $I_0(x)$ is the modified Bessel function of the augment x and

$$\sigma^2 = \sigma^2_{x_1} = \sigma^2_{y_1} = \frac{C^2}{2} \frac{N_0}{2T}$$

$$f_{z_1}(z) = \frac{z}{\sigma^2} e^{-\left(z^2 + \frac{C^2 S}{2}\right)/2\sigma^2} I_0 \left(\frac{zC\sqrt{S/2}}{\sigma^2} \right) \tag{13}$$

For $i \neq 1$

$$f_{z_i}(z) = \frac{z}{\sigma^2} e^{-\frac{z^2}{2\sigma^2}}, \quad i \neq 1, \quad z \geq 0 \tag{14}$$

Consider $\zeta = z^2$ then

$$f_\zeta(\zeta) = f_z(\sqrt{\zeta}\,)\frac{dz}{d\zeta} = \frac{1}{2\sqrt{\zeta}}\, f_z(\sqrt{\zeta}) \tag{15}$$

To emphasize the conditional nature of the probability density in (15) we use the notation $|_{s_1}$ meaning on the condition that signal s_1 is transmitted. Thus

$$f_{\zeta_1}\Big|_{s_1}(\zeta) = \frac{1}{2\sigma^2}\, e^{-\left(\zeta + \frac{C^2S}{2}\right)\Big/2\sigma^2}\; I_0\left(\frac{C\sqrt{\zeta S/2}}{\sigma^2}\right) \tag{16}$$

$$f_{\zeta_i}\Big|_{s_1}(\zeta) = \frac{1}{2\sigma^2}\, e^{-\zeta/2\sigma^2}; \quad i \neq 1,\ \zeta > 0 \tag{17}$$

The joint p.d.f. of ζ_1 to ζ_N on condition s_1 is

$$f_{\zeta_1 \ldots \zeta_N}\Big|_{s_1} = \frac{1}{(2\sigma^2)^N}\, e^{-\frac{C^2S}{4\sigma^2}}\, e^{-\sum_{i=1}^{N}\zeta_i/2\sigma^2}\; I_0\left(\frac{C\sqrt{\zeta_1 S/2}}{\sigma^2}\right) \tag{18}$$

where N is the number of tones (codes). The probability of correct reception is

$$P_c(N) = \int_{\zeta_1=0}^{\infty}\int_{\zeta_2=0}^{\zeta_1}\cdots\int_{\zeta_N=0}^{\zeta_1} f_{\zeta_1 \ldots \zeta_N}\Big|_{s_1}\, d\zeta_1 \cdots d\zeta_N \tag{19}$$

In Equation (18)

$$\frac{C^2S}{4\sigma^2} = \frac{C^2S}{C^2N_0/T} = \frac{ST}{N_0} \triangleq r \tag{20}$$

$$\frac{C\sqrt{\zeta_1 S/2}}{\sigma^2} = \frac{C\sqrt{\zeta_1 S/2}}{\frac{C^2}{2}\frac{N_0}{2T}} = \frac{4T\sqrt{\zeta_1 S/2}}{C N_0} \tag{21}$$

$$f_{\zeta_1 \cdots \zeta_N}\Big|_{s_1} = \frac{1}{(2\sigma^2)^N}\, e^{-r}\, e^{-\sum_{i=1}^{N} \frac{\zeta_i}{2\sigma^2}}\, I_0\left(\frac{4T\sqrt{\zeta_1 S/2}}{CN_0}\right) \tag{22}$$

$$P_c(N) = \int_{\zeta_1=0}^{\infty} \int_{\zeta_2=0}^{\zeta_1} \cdots \int_{\zeta_N=0}^{\zeta_1} \frac{e^{-r}}{(2\sigma^2)^N}\, I_0\left(\frac{4T\sqrt{\zeta_1 S/2}}{CN_0}\right) e^{-\sum_{i=1}^{N} \frac{\zeta_i}{2\sigma^2}}\, d\zeta_1 \cdots d\zeta_N$$

$$u_i = \frac{\zeta_i}{2\sigma^2}, \quad \frac{4T\sqrt{\zeta_1 S/2}}{CN_0} = \frac{4T\sqrt{\sigma^2 u_1 S}}{CN_0} = \frac{4T\sqrt{\frac{C^2 N_0}{4T}\, u_1 S}}{CN_0} = 2\sqrt{\frac{TSu_1}{N_0}} \tag{23}$$

$$P_c(N) = \int_{u_1=0}^{\infty} e^{-r}\, I_0\left(2\sqrt{\frac{TSu_1}{N_0}}\right) e^{-u_1} \int_{u_2=0}^{u_1} \int_{u_N=0}^{u_1} e^{-\sum_{i=2}^{N} u_i}\, du_2 \cdots du_N$$

$$= \int_{u_1=0}^{\infty} e^{-r}\, I_0\left(2\sqrt{ru_1}\right) e^{-u_1} \{1 - e^{-u_1}\}^{N-1}\, du_1$$

$$P_c(N) = \int_{x=0}^{\infty} e^{-r}\, I_0(2\sqrt{rx})\, e^{-x}\, (1 - e^{-x})^{N-1}\, dx; \quad r = \frac{ST}{N_0} \tag{24}$$

$$P_c\left(n, \frac{ST_B}{N_0}\right) = \int_{x=0}^{\infty} e^{-n\frac{ST_B}{N_0}}\, I_0\left(2\sqrt{n\frac{ST_B}{N_0}x}\right) e^{-x}\, (1 - e^{-x})^{2^n - 1}\, dx; \quad T_B = \frac{T}{n}\, n \tag{25}$$

where "n" is the number of information bits per tone burst and T_B is the effective time per bit.

This is the probability of correct decision (no error) based on the output of the correlators.

EFFECT OF THE REFERENCE WAVEFORM

When correlation is carried out with square waves then in Equations (2), (3), (4) and (9) the terms $C \cos \omega_i t$ and $C \sin \omega_i t$ must be replaced by the Fourier series

$$p_i(t) = \sum_n C_n \cos n\omega_i t \quad \text{and} \quad q_i(t) = \sum_n C_n \sin n\omega_i t$$

and

If all of the harmonics of i^{th} square wave are non-coherent with the j^{th} subcarrier frequency ($i \neq j$) then the effect of use of square wave is as follows:

(1) The expressions for the expectations x_i and y_i remain the same, with the understanding that the term C in $x_i = -1/2\, C\sqrt{2S} \sin \omega_1 t$ and $y_i = 1/2\, C\sqrt{2S} \cos \omega_1 t$, of Equations (2) and (3), is the amplitude of the fundamental of the square wave $p_1(t)$, i.e., $C = C_1$.

(2) The expression for variance of x_i and y_i which is

$$\sigma^2 = \frac{N_0}{2} \frac{1}{T^2} \int_0^T [C \cos \omega_i t]^2 \, dt = \frac{C^2}{2} \frac{N_0}{2T}$$

will change into

$$\sigma^2 = \frac{N_0}{2} \frac{1}{T^2} \int_0^T [p_i(t)]^2 \, dt.$$

If the amplitude of the square wave is denoted by $F/\sqrt{2}$ then

$$\sigma^2 = \frac{F^2}{2} \frac{N_0}{2T}$$

With these interpretations of C and σ^2 Equations (10) through (19) remain valid for the case of square wave reference signal.

(3) In (20) the ratio

$$\frac{C^2 S}{4\sigma^2} = \frac{C^2}{F^2} \frac{ST}{N_0} \triangleq r_e$$

is to replace

$$\frac{C^2 S}{4\sigma^2} = \frac{ST}{N_0} \triangleq r.$$

Since the ratio of amplitude of the fundamental of a square wave to its amplitude is $4/\pi$, it follows that $C/F = 4/\sqrt{2}\,\pi$ and

$$r_e = \left(\frac{4}{\sqrt{2}\,\pi}\right)^2 r = \frac{8}{\pi^2} r$$

(4) In (21) and (23) the ratio

$$\frac{C\sqrt{\zeta_1 S/2}}{\sigma^2} = \frac{4TC\sqrt{\zeta_1 S/2}}{F^2 N_0} = 2\frac{C}{F}\sqrt{\frac{TSU_1}{N_0}} = 2\sqrt{r_e U_1}$$

PFM systems are ordinarily designed so that the third harmonic of the lowest frequency data tone is higher in frequency than the highest data tone. So there is no problem with the third or higher harmonics of the local square wave signal correlating with a higher data tone. Furthermore, if the local square waves have perfect quarter wave symmetry, it will contain no even harmonics which can cross correlate with double frequency data tones. Thus, with perfectly symmetric reference waveforms and the above frequency restrictions, the error probabilities are given by the formula (24) where instead of $r = N_0/ST$, the factor

$$r_e = \frac{8}{\pi^2}\frac{ST}{N_0}$$

must be used.

Therefore the effect on error probability, when using a square wave for a correlation reference instead of a sinusoid, is equivalent to a degradation of signal to noise ratio by a factor of $8/\pi^2$ or about 0.9 decibel.

REFERENCES

1. Woodward, P. M., Probability and Information Theory with Application to Radar. New York: Pergamon: (1960).

2. Turin, G. L., "Communication through noisy random-multipath channels," IRE Conv. Rec. (1956).

3. Vitebri, A. J., "On coded Phase-coherent Communication" IRE Trans. on Space Electronics and Telemetry, Vol. SET-7 (March 1961).

4. Lindsey, W. C., "Error Probabilities for Rician Fading Multipath Channel Reception of Binary and N-ary Signals," IEEE Trans. on Information Theory, Vol. IT-10, (Oct. 1964).

5. Middleton, D., Introduction to Statistical Communication Theory, McGraw Hill, New York (1960) Sections 8.1-2 and 3.

CONFIDENCE LEVEL OF DECISION BASED ON THE OUTPUT OF CORRELATORS OF A PFM SYSTEM

M. Javid

In a PFM system using correlators for signal detection, the correlator having the largest output is taken to indicate the transmitted frequency. For any signal to noise ratio, the event (that the correlator with highest output corresponds to the transmitted frequency) has the highest probability. However, if the signal to noise ratio is low the difference between the probabilities of the most likely and least likely candidates can be small. For example if there are only 4 possible transmitted frequencies and the value of signal to noise ratio is low, it is possible that the correlator with highest output indicates the transmitted frequency with the probability 0.26 and the correlator with lowest output, with probability 0.24, the other two correlators having probabilities of 0.25 each. Under these conditions the decision; that the correlator with highest output corresponds to the transmitted frequency, is not very reliable. In this discussion the measure of reliability of the decision is taken to be the probability that the correlator with the highest output corresponds to the transmitted frequency. Alternatively one may take the measure to be the ratio of the probabilities of the correlators with the two highest outputs, or in general any function of the probabilities of the individual correlators.

The probability that a given tone $s_k(t)$ is transmitted on the condition that the output of the i^{th} correlator is

$$\zeta_i = x_i^2 + y_i^2 \quad i = 1, \cdots N$$

is found by noting that, as shown in Equation (18) of report on "Effect of reference waveform on performance of PFM Systems," the joint probability density of $\zeta_1 \text{ --- } \zeta_N$ on condition that $s_k(t)$ is transmitted is

$$f_{\zeta|s_k} \triangleq f_{\zeta_1 \cdots \zeta_N|s_k} = \frac{1}{(2\sigma^2)^N} \; e^{-\frac{C^2 S}{4\sigma^2}} \; e^{-\sum_{i=1}^{N} \zeta_i/2\sigma^2} \; I_0\left(\frac{C\sqrt{\zeta_k S/2}}{\sigma^2}\right) \tag{1}$$

Removing the condition, the joint probability density of (ζ, s_k) is

$$f_{\zeta s_k} \triangleq f_{\zeta_1 \cdots \zeta_N s_k} = P_r(s_k)\, f_{\zeta_1 \cdots \zeta_N|s_k} = \frac{1}{N} f_{\zeta_1 \cdots \zeta_N|s_k} \tag{2}$$

$$P_r(s_k|\zeta) \triangleq P_r\{s_k|\zeta_1 \cdots \zeta_N\} = \frac{f_{\zeta_1 \cdots \zeta_N, s_k}}{f_{\zeta_1 \cdots \zeta_N}} = \frac{\frac{1}{N} f_{\zeta_1 \cdots \zeta_N|s_k}}{\sum_{j=1}^{N} \frac{1}{N} f_{\zeta_1 \cdots \zeta_N|s_j}} \tag{3}$$

$$P_r(s_k|\zeta) = \frac{\frac{1}{N} f_{\zeta|s_k}}{\sum_{j=1}^{N} \frac{1}{N} f_{\zeta|s_j}} = \frac{f_{\zeta|s_k}}{\sum_{j=1}^{N} f_{\zeta|s_j}} \tag{4}$$

Using Equation (1) in (4)

$$P_r(s_k|\zeta) = \frac{I_0\left(\frac{C\sqrt{\zeta_k S/2}}{\sigma^2}\right)}{\sum_{j=1}^{N} I_0\left(\frac{C\sqrt{\zeta_j S/2}}{\sigma^2}\right)} \tag{5}$$

Normalizing, as in Equation (23) of previously mentioned report

$$u_i = \frac{\zeta_i}{2\sigma^2} \qquad \sigma^2 = \frac{C^2 N_0}{4T} \tag{6}$$

we have for $u = \frac{\zeta}{2\sigma^2}$

$$r = \frac{ST}{N_0}$$

$$P_r(s_k|u) = \frac{I_0(2\sqrt{r u_k})}{\sum_{j=1}^{N} I_0(2\sqrt{r u_j})} \tag{7}$$

where

$$I_0(x) = \frac{1}{2\pi} \int_0^{2\pi} e^{x\cos\theta}\, d\theta = \sum_{n=0}^{\infty} \frac{x^{2n}}{2^{2n}(n!)^2} \tag{8}$$

$$u_i = \frac{\zeta_i}{2\sigma^2} = \frac{T(\zeta_i/C^2)}{N_0/2} \tag{9}$$

In the above expression, ζ_i/C^2 is the output of correlator normalized with respect to power in the correlating signal $C \sin \omega_i t$. Also

$$2\sqrt{r u_k} = \sqrt{4\frac{ST}{N_0}\frac{T(\zeta_i/C^2)}{N_0/2}} = \frac{2T}{N_0}\sqrt{S(\zeta_i/(C/\sqrt{2})^2)} \tag{10}$$

If we consider

$$\frac{\zeta_i}{(C/\sqrt{2})^2} = \eta_i \tag{11}$$

the correlator output, normalized with respect to mean square of correlating sinusoid $C \sin \omega_i t$. Then

$$x_k = 2\sqrt{r u_k} = \frac{2T\sqrt{S\eta_k}}{N_0} \tag{12}$$

Whereas the probability of error is a function of ST/N_0, the confidence level $P_r \{s_k | \zeta\}$ is a function of the product

$$\sqrt{\eta_k}\ \frac{T\sqrt{S}}{N_0}$$

To evaluate the confidence level given by Equation (7) it is necessary to obtain an estimate of the signal S and noise N_0. The maximum likelihood estimate of S and N_0 are the values which maximizes the probability density

$$f_\zeta = f_{\zeta_1 \zeta_2 \cdots \zeta_N} = \sum_{j=1}^{N} \frac{1}{N}\ f_{\zeta_1 \zeta_2 \cdots \zeta_N | S_j} \tag{13}$$

when the observed value of the correlator output are substituted in the expression. From Equations (1) and (13)

$$f_\zeta = \sum_{j=1}^{N} \frac{1}{N}\ \frac{1}{(2\sigma^2)^N}\ e^{-\frac{C^2 S}{4\sigma^2}}\ e^{-\sum_{i=1}^{N} \zeta_i / 2\sigma^2}\ I_0\left(\frac{C\sqrt{\zeta_j S/2}}{\sigma^2}\right)$$

$$= \frac{1}{N}\ \frac{1}{(2\sigma^2)^N}\ e^{-\frac{C^2 S}{4\sigma^2}}\ e^{-\sum_{i=1}^{N} \zeta_i / 2\sigma^2} \sum_{j=1}^{N} I_0\left(\frac{C\sqrt{\zeta_j S/2}}{\sigma^2}\right) \tag{14}$$

Substituting

$$\sigma^2 = \frac{C^2 N_0}{4T}$$

in Equation (14) we obtain an expression in which ζ_i are given by the output of correlators, and N_0 and S are noise and signal power. Differentiating Equation (14) with respect to N_0 and S and setting the result to zero, we have two equations.

$$\frac{\partial f_\zeta}{\partial N_0} = 0 \quad \text{and} \quad \frac{\partial f_\zeta}{\partial S} = 0$$

The solution of these equations gives the maximum likelihood estimate of the signal and noise in terms of the output of the correlators. Substitution of these estimates in Equation (5) gives the probability that the signal s_k has been transmitted.

BLANK PAGE

THE DISTRIBUTION OF THE OUTPUTS FROM A FILTER BAND SIGNAL PROCESSING SYSTEM

F. L. Ramsey

PHYSICAL DESCRIPTION

One word of a possible N is transmitted in a time interval T. The signals are

$$s_k(t) = A \sin \omega_k t; \quad k = 1, \ldots, N$$

where

$$\omega_k = \frac{2\pi i(k)}{T}$$

where: $i(k)$ is the appropriate integer for the k^{th} tone. When received, the message is altered by the presence of noise and by an unknown phase shift. If the k^{th} signal is transmitted, the received signal is

$$r(t) = \sqrt{2P} \sin(\omega_k t + \varphi) + n(t),$$

where φ is the phase shift, P is the signal power, and $n(t)$ is noise.

To convert the message from analogue to digital form, it is processed by a bank of N filters. In the j^{th} filter, the received tone is correlated with the sine and cosine waves of frequency ω_j.

$$S_j = \frac{1}{T} \int_0^T r(t) \sin \omega_j t \, dt$$

$$C_j = \frac{1}{T} \int_0^T r(t) \cos \omega_j t \, dt \qquad (1)$$

These quantities are squared and added, so the output of the j^{th} filter is the random variable

$$X_j = S_j^2 + C_j^2, \qquad (2)$$

which does not depend on the unknown phase φ.

ASSUMPTIONS

It is assumed that the start (and hence the finish) of each message is identifiable. That is to say, the appropriate time, $t = 0$, is known so that no overlapping will occur in the correlators when a sequence of messages is sent.

The noise process is assumed to be white Gaussian. That is, if $t_1, \cdots, t_q$ is any set of distinct time points, the joint probability density function of the random variables $n(t_1), \cdots, n(t_q)$ is

$$f_{n(t_1),\ldots,n(t_q)}(u_1, \ldots, u_q) = \prod_{i=1}^{q} f_{n(t_1)}(u_i) \tag{3}$$

$$= (2\pi\sigma^2)^{-q/2} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{q} u_i^2}; \quad -\infty < u_1, \ldots, u_q < \infty. \tag{4}$$

This holds for any number q of distinct time points. Here Equation (3) indicates that the random variables are independent, and Equation (4) indicates each has zero mean and variance

$$\mathcal{E}\, n^2(t) = \sigma^2. \tag{5}$$

Under these assumptions, the probability density function (p.d.f.) of an output from the filter $j \neq k$ (when message k is transmitted) is

$$f(x_j) = \frac{T}{2\sigma^2} e^{-\frac{T}{2\sigma^2} x_j}, \quad x_j > 0. \tag{6}$$

The output of the k^{th} filter has the p.d.f.

$$g(x_k) = \frac{T}{2\sigma^2} e^{-\frac{T}{2\sigma^2}(x_k + 2P)} \mathcal{I}_0\left(\frac{PT^2 x_k}{\sigma^4}\right), \quad x_k > 0, \tag{7}$$

where

$$\mathcal{I}_0(u) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{\sqrt{u}\cos\lambda}\, d\lambda. \tag{8}$$

STATEMENT OF THE PROBLEM

Using the output of the filters, it is required (a) that decision be made on which signal was transmitted, and (b) that a measure of confidence be put on the decision.

It is shown by M. Javid that the proper decision is to select the message corresponding to the filter with the largest output. The probability of making an error, he indicates, is then

$$P_e(\theta) = e^{-\theta} \int_{-\infty}^{\infty} \{1 - e^{-u}\}^{N-1} e^{-u}\, \mathcal{I}_0(2u\theta)\, du, \tag{9}$$

where

$$\theta = PT/\sigma^2 \tag{10}$$

is the signal-to-noise ratio.

In practice, θ is unknown so the determination of this probability depends on our ability to obtain a reasonable estimate of θ. This problem is the one which is presently under investigation.

PLAN OF INVESTIGATION

The estimation procedure proposed here is the Maximum Likelihood procedure. If $f(x_1, \cdots, x_N; P, \sigma^2)$ is the joint p.d.f. of the filter outputs, the maximum likelihood estimates (MLE) of P and σ^2 are obtained by solving the equations

$$\frac{\partial f}{\partial P} = 0$$

$$\frac{\partial f}{\partial \sigma^2} = 0. \tag{11}$$

If $\hat{P}$ and $\hat{\sigma}^2$ are the solutions, then the MLE of θ is simply

$$\hat{\theta} = \hat{P}T/\hat{\sigma}^2. \tag{12}$$

THE PROBLEM OF THE P.D.F.

The basic question now is to find an "appropriate" form for the joint p.d.f. of the outputs when it is not known which signal was transmitted. By an "appropriate" form we mean a function $f(x_1, \cdots, x_N; P, \sigma^2, k)$ such that

(a) for any fixed value of the triplet (P, σ^2, k), f is the joint p.d.f. of the output variables; and

(b) condition (a) holds for every triplet (P, σ^2, k) where $P > 0$, $\sigma^2 > 0$ and $k = 1, 2, \cdots, N$.

Condition (b) expresses the fundamental fact that we do not know what P, σ^2, and k are. If we know what value k has, then we could write down the joint p.d.f. of any set of outputs $x_1, \cdots, x_N$ from Equation (6) and (7) as

$$f(x_1, \ldots, x_N; P, \sigma^2, k) = f(x_1) \ldots f(x_{k-1})\, g(x_k)\, f(x_{k+1}) \ldots f(x_N). \tag{13}$$

But, alas, we do not know k, so we do not know which output value to plug into the $g(\cdot)$ density function. The problem is one of indexing. In the form presented, the indices of the variables cannot be identified in a form such as in Equation (13).

Two roads lead us to safety. The first involves finding an indexing system for the outputs which can be associated with an appropriate p.d.f. The second is the Bayesian solution proposed by Javid. On the following pages these two solutions are presented and their applications are discussed.

AN ORDERING APPROACH

The lack of knowledge of k, as we have seen, creates an indexing problem. This problem is solved, however, if, instead of trying to describe the outputs as indexed by their filter index, we index them according to their rank. The ordered outputs have a joint p.d.f. which is "appropriate."

Let $y_1 \geq y_2 \geq \cdots \geq y_N$ be the ordered outputs. The probability that the j^{th} largest output comes from the k^{th} filter is

$$P_j(\theta) = \binom{N-1}{j-1} e^{-\theta} \int_0^\infty \{1 - e^{-u}\}^{N-j} e^{-ju} \mathcal{J}_0(2u\theta)\, du. \tag{14}$$

The conditional joint p.d.f. of the ordered outputs, given that the k^{th} filter gives the largest output, is

$$P(y_1, \ldots, y_N | y_j = x_k) = (N-1)! \prod_{i=1}^{N} \left[f^{-1}(y_j)\, g(y_j)\right] \tag{15}$$

$$= (N-1)! \left(\frac{T}{2\sigma^2}\right)^N e^{-\frac{T}{2\sigma^2}\left[2P + \sum_{i=1}^{N} y_i\right]} \mathcal{J}_0\left(\frac{PT^2 y_j}{\sigma^4}\right)$$

for

$$0 \leq y_N \leq y_{N-1} \leq \cdots \leq y_1 < \infty.$$

Therefore, the unconditional joint p.d.f. of the ordered outputs is

$$f(y_1, \ldots, y_N | P, \sigma^2, k) = \sum_{j=1}^{N} P(y_1, \ldots, y_N | y_j = x_k)\, P_j(\theta) \tag{16}$$

$$= \left\{(N-1)!\, e^{-2\theta} \left(\frac{T}{2\sigma^2}\right)^N e^{-\frac{T}{2\sigma^2}\sum_{i=1}^{N} y_i}\right\}.$$

$$\left\{\sum_{j=1}^{N} \binom{N-1}{j-1} \mathcal{J}_0\left(\frac{\theta T y_j}{\sigma^2}\right) \int_0^\infty [1 - e^{-u}]^{N-j} e^{-ju} \mathcal{J}_0(2u\theta)\, du\right\},$$

for

$$0 \leq y_N \leq y_{N-1} \leq \cdots y_1 < \infty.$$

This form can be used to find the ML estimates in the manner described. However, as one can see, the solutions to Equations (11) will be most difficult to obtain.

BAYES APPROACH

Let the a priori densities of P, σ^2, and k be respectively $\rho(s)$, $\tau(\sigma^2)$, and $\eta(k)$. The likelihood of the sample $\mathbf{x} = (x_1, \cdots, x_N)$ given a fixed triplet (S, σ^2, k) is, from Equation (13),

$$p(\mathbf{x}|P, \sigma^2, k) = f(x_1) \cdots f(x_N) \{f(x_k)\}^{-1} g(x_k)$$

$$= \left(\frac{T}{2\sigma^2}\right)^N e^{-\frac{T}{2\sigma^2}\left[2P + \sum_{j=1}^{N} x_j\right]} \mathcal{I}_0\left(\frac{PT^2 x_k}{\sigma^4}\right) \tag{17}$$

For the a priori densities ρ, τ, and η, the joint a posteriori density of (P, σ^2, k) given the sample $\mathbf{x}$ is

$$p(P, \sigma^2, k|\mathbf{x}) = h(\mathbf{x})\, p(\mathbf{x}|P, \sigma^2, k)\, \rho(P)\, \tau(\sigma^2)\, \eta(k), \tag{18}$$

where

$$1/h(\mathbf{x}) = \sum_{k=1}^{N} \int_0^\infty \int_0^\infty p(\mathbf{x}|P, \sigma^2, k)\, \rho(P)\, \tau(\sigma^2)\, \eta(k)\, dP\, d\sigma^2. \tag{19}$$

Equation (18) is an "appropriate" form for the joint a posteriori p.d.f. of (P, σ^2, k). Observe that our lack of knowledge about k (and P and σ^2) is expressed by treating the parameters as if they were "random" variables with known a priori density functions. The marginal a posteriori density of k may be obtained by integration.

$$p(k|\mathbf{x}) = \int_0^\infty \int_0^\infty p(P, \sigma^2, k|\mathbf{x})\, dP\, d\sigma^2. \tag{20}$$

Then for a loss function $W(\hat{k}, k) \geq 0$, which expresses the loss, using the estimator $\hat{k}$ of k, when k is the actual value, an estimator for k can be obtained by minimizing

$$\mathcal{E}_p[W(\hat{k}, k)] = \sum_{k=1}^{N} W(\hat{k}, k)\, p(k, \mathbf{x}). \tag{21}$$

(Here $\mathcal{E}_p$ denotes expectation with respect to an a posteriori density.) As examples, the median of the density (20) minimizes $\mathcal{E}_p[|\hat{k} - k|]$; the mean of the density minimizes the loss function $(\hat{k} - k)^2$; etc.

The method used by Javid can be viewed as follows. Let

$$d\rho(P) = dP; \qquad P > 0$$

$$d\tau(\sigma^2) = d\sigma^2; \qquad \sigma^2 > 0 \tag{22}$$

$$\eta(k) = 1/N; \qquad k = 1, 2, \ldots, N$$

The mode of the marginal density Equation (14) is

$$\hat{k} = \max_{1 \le i \le N} \{x_i\}.$$

The probability of error using this estimator is

$$P_r\{\hat{k} \neq k\} = 1 - p_1(\theta), \tag{23}$$

from Equation (10). An estimate of this error probability is given as $1 - p(\hat{\theta})$, where $\hat{\theta} = \hat{P}T/\hat{\sigma}^2$, and $(\hat{P}, \hat{\sigma}^2)$ is the mode of the marginal a posteriori density

$$p(P, \sigma^2 | x) = \sum_{k=1}^{N} p(\hat{P}, \sigma^2, k | x). \tag{24}$$

The mode of the density is the estimator which minimizes the "simple" loss function

$$W(\hat{k}, k) = 0, \quad \hat{k} = k \tag{25}$$
$$= 1, \quad \hat{k} = k.$$

DISCUSSION

Looking at each signal separately, one may be unwilling to assume that the values of k, σ^2, and P are merely "observations" from some a priori distributions of values. This would lead to the use of the non-Bayesian model.

If the values of the parameters are more easily considered as being values from some a priori distributions, then the Bayesian approach is more appropriate. This assumption seems to us to be more reasonable in processing telemetry data. Looking at a tape as a whole we would expect that the signal power and noise variance would vary over the tape in a manner resembling random variables. Likewise, the index of the transmitted signal can be viewed in this fashion.

One question raised in Bayesian analysis is the choice of the a priori densities $\rho(P)$, $\tau(\sigma^2)$, and $\eta(k)$. In particular, the choice in Equation (22) of $\eta(k)$ may be questioned because of observed redundancy in the data. Techniques (known as "empirical Bayes" procedures) are available to cope with this difficulty. These techniques assume that true a priori densities exist, but that they are unknown. Using some initial guess for them, the accumulation of data provides the means for updating the original guesses by estimating the true a priori densities.

REFERENCES

1. Javid, M. "Confidence Level of Decision Based on the Output of Correlators of a PFM System." Goddard Summer Workshop Final Report, (1965).

For discussions of the theory and application of Bayesian inference, the reader is referred to the following texts.

2. Lindley, D. V. Probability and Statistics. Cambridge University Press, (1965).

3. Raiffa, H. and Schaefer, M. Applied Statistical Decision Theory, Harvard University Press, (1961).

4. Savage, L. J. Foundations of Statistics, John Wiley and Sons, New York, (1954).

For discussion of the empirical Bayes techniques, see:

5. Robbins, H. E. "The Empirical Bayes Approach to Statistical Decision Problems." Annals of Mathematical Statistics (March, 1964) pages 1-20.

6. Tainiter, M. "Adaptive Procedures for Pulse Detection and Extraction." The Johns Hopkins University Carlyle Barton Laboratory Internal Memorandum, CBL/64/1MA-11. (1964).

A MINIMUM CONFIDENCE LEVEL FOR SIGNAL DETECTION BY A CORRELATOR BANK

F. L. Ramsey

Estimating the probability of signal detection error from the outputs of a bank of correlators involves cumbersome and lengthy calculations when maximum likelihood or Bayesian techniques are used. These manipulations involve zero order Bessel functions their sums and ratios.

The purpose of this report is to show that a slight alteration of the underlying assumptions facilitates the use of a very simple technique which, under the worst conditions, gives an upper bound to the probability of error. The theory of order statistics provides the basis for this technique. The basic results needed for the development are presented in the appendix.

Proceeding heuristically for the moment, suppose a set of N outputs gives the following plot (Figure A-22)

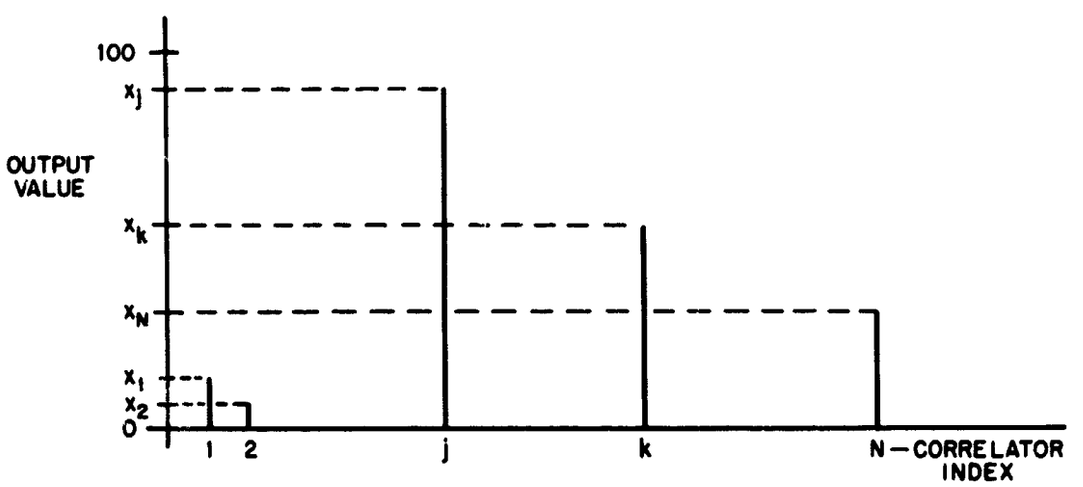


Figure A-22–Plot of N outputs.

Here x_j and x_k are respectively the largest and second largest outputs. The maximum output, x_j, is taken to be an indication that the signal transmitted was the j^{th}. Just how certain this decision is depends on how big x_j is in relation to the rest of the outputs.

A meaningful comparison should include two basic properties of the "rest" of the sample. The first question is how well x_j beats its nearest competitor. This suggests the measure $x_j - x_k$ as an indication of confidence. This measure is easily normalized so that it always lies between zero and 1. Let

$$r = 1 - \frac{x_j - x_k}{x_j} = x_k / x_j . \qquad (1)$$

The smaller r is, the more likely it is that the j^{th} signal was transmitted.

The number of correlators must also be taken into consideration. For suppose $x_j = 8$ and $x_k = 6$. If $N = 4$ and the other two outputs are 1 and 3, one would not feel very secure about judging $x_j = 8$ as having been caused by a signal. However, if $N = 100$ and the remaining 98 output values lie between 1 and 6, one would be reasonably confident that it did.

A simple way to incorporate this additional factor is to use the statistic r raised to the power $(N - 1)$.

Again on a heuristic note, let us examine situations which would indicate a lack of confidence in the decision, and factors giving rise to such situations.

Two situations are discernible. The first occurs when the array of output values is indistinguishable from a "random" array. That is, if the array might have been produced by the N correlators when no signal was transmitted, the decision should be suspect. The second situation arises when, though as a whole the array does not look random, there is one (or a few) strong competitor for the maximum value. Here the same situation prevails, because there appears to be no more than random differences between the leading candidates.

According to these considerations, confidence in the decision is based on how much r^{N-1} differs from the values one expects when r^{N-1} is calculated from a random array. Since we expect x_j to be larger when it contains a signal, the confidence level depends upon how small r^{N-1} is.

Now we will formalize the previous assumptions. Let T be the time length of the message, P be the signal power, and σ^2 be the noise variance. The signal-to-noise ratio is the (non-negative) quantity

$$\theta = PT/\sigma^2. \tag{2}$$

The output of a correlator measuring only noise is a random variable having a probability density function (p.d.f.)

$$f(u) = \left(\frac{T}{2\sigma^2}\right) e^{-\frac{T}{2\sigma^2}u}, \quad U > 0. \tag{3}$$

The p.d.f. of the output measuring the signal plus noise is

$$g(u) = e^{-\theta} f(u)\, I_0\left(\frac{\theta T}{\sigma^2}u\right), \tag{4}$$

where

$$I_0(x) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{\sqrt{x}\cos\lambda}\, d\lambda. \tag{5}$$

The two situations of interest are (i) when X_j has p.d.f. given by (4) and (ii) the random situation. Situation (ii) is equivalent to $\theta = 0$ while (i) is equivalent to $\theta > 0$. Thus we make the assumptions:

(a) the random situation ($\theta = 0$) may occur; and

(b) if $\theta > 0$, the largest output measures the signal.

In short, the assumptions state that, when the largest output does not measure the signal, the array of outputs is random.

RESULTS

Let $x_1, \cdots, x_n$ be the outputs of the N correlators. Let

$$t_i = F(x_i) = \int_0^{x_i} f(u)dv = 1 - e^{-\frac{Tx_i}{2\sigma^2}}, \tag{6}$$

($i = 1, 2, \cdots, N$). The $\{t_i\}$ are random variables and functions of the unknown noise variance σ^2. t_i represents the probability that a random ($\theta = 0$) output is less than or equal to x_i. Finally, let $0 \leq U_1 \leq U_2 \leq \cdots \leq U_N \leq 1$ be the ordered set of t_i's. That is, let

$$U_i = i^{th} \text{ smallest } t\text{-value}. \tag{7}$$

(Observe that U_i corresponds to the i^{th} smallest x-value, since F(x) is a monotone increasing function.)

Theorem: In the random ($\theta = 0$) situation, the probability that $(U_{N-1}/U_N)^{N-1}$ is less than or equal to α ($0 \leq \alpha \leq 1$) is

$$P_r \left\{(U_{N-1}/U_N)^{N-1} \leq \alpha \mid \theta = 0\right\} = \alpha. \tag{8}$$

The proof of this statement appears in the appendix.

From this theorem we can state the result.

Corollary: If r is defined as in Equation (1), then

$$P_r \left\{r^{N-1} \leq \alpha \mid \theta = 0\right\} \leq \alpha. \tag{9}$$

Proof: If

$$x_j = \max_{1 \leq i \leq N} \{x_i\}$$

and

$$x_k = \max_{i \neq j} \{x_i\},$$

then

$$r = x_k/x_j,$$

and

$$U_{N-1}/U_N = \left\{1 - e^{-\frac{Tx_k}{2\sigma^2}}\right\} \Big/ \left\{1 - e^{-\frac{Tx_j}{2\sigma^2}}\right\}$$

$$= \left\{1 - e^{-rv}\right\} / \left\{1 - e^{-v}\right\}, \tag{10}$$

where $v = tx_j/2\sigma^2$. Let A be the event $(U_{N-1}/U_N)^{N-1} \leq \alpha$ and let B be the event $r^{N-1} \leq \alpha$. The ratio (U_{N-1}/U_N) is a monotone decreasing function of v. Thus it is less than or equal to its value when $v = 0$. But $(U_{N-1}/U_N) = r$ when $v = 0$. This implies

$$(U_{N-1}/U_N)^{N-1} \leq r^{N-1} ,$$

so that B implies A. Thus

$$P_r\{A\} \leq P_r\{B\} = \alpha ,$$

as required.

CONCLUSION

Given the outputs $X_1, \cdots, X_N$, the statistic r^{N-1} is calculated. The corollary states that the probability of finding a value of the statistic less than or equal to r^{N-1} from a random array is less than or equal to r^{N-1}. Thus we are at least $(1 - r^{N-1})\%$ confident that this value was not produced by purely random effects.

DISCUSSION

What we have is a minimum confidence bound. As shown above, this minimum is achieved when $v = Tx_j/2\sigma^2 = 0$. This, of course, will never hold. However it indicates that the approximation is best when the noise variance is large; i.e. when θ is small. Since one is most concerned about situations when θ is small, this feature is desirable. (The more complicated procedure of estimating θ has the inverse effect of being best when θ is large.)

Finally, something must be said about the power of this procedure. The power of the procedure is a measure of how well it discriminates against adverse circumstances and how well it detects favorable conditions. If the correlator with the signal does not have the largest output, a "good" procedure should give a low level of confidence. But if it does have the largest output, a "good" procedure should give a high level of confidence.

On the average one expects that the confidence level based on r^{N-1} is "good" in this sense. Just how good it is has not been investigated thus far. Further studies should be made of these details. These would entail working out the distribution of r^{N-1} under non-random conditions, which is feasible but arduous.

The virtue of the procedure described is its simplicity. The basic property of Equation (9) comes from basic probability theory. This is in contrast with maximum likelihood procedures which are very complicated and whose properties are equally difficult to derive.

APPENDIX

Let $X_1, \cdots, X_N$ be independent observations on a random variable whose probability density function, $f(x)$, is continuous. Let

$$F(x) = \int_{-\infty}^{x} f(x)\,dx \tag{1}$$

denote the cumulative distribution function. Let $Z_1, \cdots, Z_N$ be the ordered observations; i.e.

$$Z_k = k^{th} \text{ smallest of the } X_j\text{'s}. \tag{2}$$

Furthermore let

$$T_k = F(Z_k), \quad k = 1, \ldots, N. \tag{3}$$

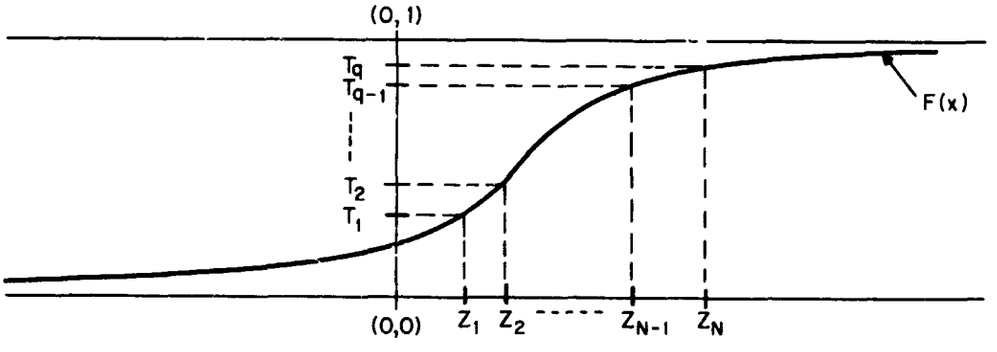


Figure A-23–Probability density function.

Lemma 1: If the random variable X has a continuous density (as above), then the random variable

$$U = F(X)$$

has the density function

$$h(u) = 1 \quad , \quad 0 \leq u \leq 1$$

$$= 0 \quad , \quad \text{elsewhere}.$$

Proof: Because $f(x)$ is continuous, the following operations are valid. The cumulative distribution function (C.D.F.) of U is

$$H(u) = \text{Prob}\{U \le u\} = P_r\{F(x) \le u\} = P_r\{X \le F^{-1}(u)\}$$

$$= F[F^{-1}(u)] = u \quad \text{for} \quad 0 \le u \le 1 .$$

Differentiation gives

$$H'(u) = h(u) = \text{desired result} .$$

Lemma 2: The joint density function of the order statistics defined in Equation (3) is

$$f(t_1, \ldots, t_N) = N!, \quad 0 \le t_1 \le t_2 \le \ldots \le t_N \le 1 . \tag{4}$$

Proof: From lemma 1, it follows that the joint density of the $U_j = F(X_j)$ is

$$h(t_1, \cdots, t_N) = 1, \quad 0 \le t_1, \cdots, t_N \le 1$$

$$= 0, \quad \text{elsewhere} .$$

This is also the joint density of the order statistics $\{T_k\}$ conditional on $T_1 = U_1$, $T_2 = U_2, \cdots, T_N = U_N$, (and conditional on any other ordering). Hence the unconditional density of the order statistics is

$$f(t_1, \cdots, t_N) = \sum_{(i_1, \ldots, i_N)} h(t_{i_1}, \ldots, t_{i_N}) ,$$

where the sum is over the N permutations $(i_1, i_2, \cdots, I_N)$ of $(1, 2, \ldots N)$. Equation (4) is correct.

In the following lemmas, the conditional notation is used. Here

$$f(t_{i_1}, \ldots, t_{i_r} \mid t_{j_1}, \ldots, t_{j_s}) = \text{conditional density of}$$

$$\{T_{i_1}, \ldots, T_{i_r}\} \text{ given that } T_{j_1} = t_{j_1}, \ldots, t_{j_s} .$$

Lemma 3: For any k,

$$f(t_1, t_2, \ldots, t_{k-1} \mid t_k, \ldots, t_N) = \frac{(k-1)!}{t_k^{k-1}}, \quad 0 \le t_1 \le \ldots \le t_{k-1} \le t_k \tag{5}$$

$$= 0, \quad \text{elsewhere} .$$

Proof: When $T_k, \cdots, T_N$ are fixed, the joint density of $T_1, \cdots, T_{k-1}$ is the same as the density of the order statistics in a sample of size $(k - 1)$ from a uniform distribution on the interval $(0, t_k)$. Thus the proof follows Lemma 2.

Lemma 4: For any k,

$$f(t_{k+1}, \cdots, t_N \mid t_1, \cdots, t_k) = (N-k)! \,/\, (1-t_k)^{N-k} . \tag{6}$$

The proof is essentially the same.

Lemma 5: For $i < k$, the unconditional density of $T_{i+1}, \cdots, T_k$ is

$$f(t_{i+1}, \cdots, t_k) = \frac{N!}{i!\,(N-k)!}\, t_{i+1}^{i}\, (1-t_k)^{N-k} \tag{7}$$

for

$$0 \le t_{i+1} \le t_{i+2} \le \cdots \le t_k \le 1 .$$

Proof: Using Lemmas 3 and 4 and the Markov character, we have

$$f(t_1, \cdots, t_i, t_{k+1}, \cdots, t_N \mid t_{i+1}, \cdots, t_k) = f(t_i, \cdots, t_i \mid t_{i+1})\, f(t_{k+1}, \cdots, t_N \mid t_k) .$$

Also, this is

$$f(t_1, \cdots, t_i, t_{k+1}, \cdots, t_N \mid t_{i+1}, \cdots, t_k) = f(t_1, \cdots, t_N) / f(t_{i+1}, \cdots, t_k) .$$

Equating right hand sides gives

$$f(t_{i+1}, \cdots, t_k) = f(t_1, \cdots, t_N) / f(t_1, \cdots, t_i \mid t_{i+1})\, f(t_{k+1}, \cdots, t_N \mid t_k)$$

$$= \frac{N!}{i!\,(N-k)!}\, t_{i+1}^{i}\, (1-t_k)^{N-k} .$$

Corollary 1: The marginal densities are

$$f(t_k) = \frac{N!}{(k-1)!\,(N-k)!}\, t_k^{k-1} (1-t_k)^{N-k}, \quad 0 \le t_k \le 1 . \tag{8}$$

Proof is obtained by putting $i + 1 = k$ in lemma 5.

Theorem 1: The random variables $Y_k = (T_k / T_{k+1})^k$, $(k = 1, 2, \cdots, N-1)$, and $Y_N = T_N^N$ are independent and uniformly distributed over $(0,1)$.

Proof: Under the condition that $T_{k+1} = T_{k+1}, \cdots, T_N = t_N$, the density of T_k is that of the k^{th} order statistic in a sample of size k from a uniform distribution on the interval $(0, t_{k+1})$.

Therefore the density of T_k / T_{k+1}, conditional on $T_{k+1} = t_{k+1}$ is that of the k^{th} order statistic in a sample of size k from a uniform distribution on the interval (0, 1). But since this is independent of t_{k+1}, it follows that this is the unconditional density. We have

$$f(t_k / t_{k+1} \mid t_{k+1}) = k\left(\frac{t_k}{t_{k+1}}\right)^{k-1} \qquad 0 \leq \left(\frac{t_k}{t_{k+1}}\right) \leq 1 .$$

Then

$$P_r \left\{T_k / T_{k+1} \leq x\right\} = x^k .$$

So

$$P_r \left\{Y_k \leq x\right\} = P_r \left\{T_k / T_{k+1} \leq x^{1/k}\right\} = \left\{x^{1/k}\right\}^k = x .$$

Noting then that the joint density of the U_j "factors" (in the sense that $1 = 1.1 \cdots 1$) into the product of the densities of the Y_k, independence of the Y_k is obvious.

REFERENCES

1. Malmquist, Sten, Skandinavisk Aktuarietidskrift. Vol. 33 (1950), page 214.

2. Sarhan, A. E. and Greenberg, B. G., Contributions to Order Statistics, John Wiley and Sons, New York, (1962).

BLANK PAGE

QUALITY CONTROL IN THE PROCESSING OF TELEMETRY DATA

L. S. Schwartz

ABSTRACT

This report is concerned with the application of quality control procedures to the processing of telemetric data. The view is presented that 100 percent data testing can be replaced by sampling methods, subject to any assignable risk and any assignable reliability. The problem of recognizing loss of synchronization in data processing is examined in detail and a systematic technique for achieving its recognition is presented. Then a method is offered for relating specified tolerances or errors in experimental data to the corresponding allowable error rate in synch words, and thereby demonstrating how a sequential sampling plan resulting in the use of a control chart for rationally accepting or rejecting data can be devised. As a supplementary effort examples are given to show how the probabilities involved may be computed by flow diagram techniques.

A quality control approach to failure in time decoding for satellite tracking systems is also briefly discussed.

Finally, the consequences of processing the data in such a way that tests for synchronization and timing may be combined in a single operation are discussed. The possibility of a significant gain in maintaining statistical control and in meeting tolerance limits on experimental data is indicated. This problem also is investigated with the aid of flow diagram techniques.

INTRODUCTION

In the processing of telemetry data from spacecraft, quality control procedures - in the sense in which they have been applied in industry - have not as yet been tried. The prevailing philosophy tends to be that of 100 percent testing. There are at least two reasons for this. One is that until now 100 percent testing could be afforded, because the data processing equipment was equal to the flow of data, and the other is the insistence on the part of the experimenters of rigid standards of quality in the data received from the processing people. Because of this insistence and because of the processing engineer's lack of familiarity with statistical methods, there appears to be a reluctance to employ the sampling techniques of quality control. Thus, the second reason for the tendency not to use quality control is, in part, psychological, a tendency that necessarily will be overcome if 100 percent data testing can no longer be afforded. That this will shortly be the case[1] may be seen from the exponential rise in the amount of data received or to be received as projected through the year 1967.

Accordingly, this report attempts to show that 100 percent data testing can be replaced by sampling, subject to any assignable risk and any assignable data reliability. As a supplementary effort examples are given to show how the probabilities involved can be arrived at by means of flow diagram techniques, the purpose being to eliminate for the engineer some of the mystery in probability analysis by showing him that the same methods for computing these probabilities are applicable for analyzing his circuits, and thus, hopefully, lower the psychological barrier.

The following material is primarily concerned with the application of quality control procedures in coping with the problem of synchronization failure in the processing of PCM data from OGO on the STARS Phase I lines. Also, the problem of failure in timing in the time decoding for satellite tracking systems is introduced. The effect on experimental precision of synchronization failures is examined; but the related problem for timing failures, although analogous, is not examined.

As further indication of the flexibility and power of flow diagram techniques in the analysis of system failure, some considerations in the theory of repair as applied to the problem of maintaining control are discussed.

FAILURE OF SYNCHRONIZATION

In the processing of PCM data from OGO on the STARS Phase I lines, the procedure now being followed by the processing people is to accept or reject data by studying error patterns in sync words. The processors agreed to call the occurrence of more than a specified number of bit errors in a sync word a sync-word error and to call the successive occurrence of a specified number or more of sync-word errors evidence of loss of sync. If "evidence" of loss of sync occurs, the processing machinery goes from a "lock" to a "search" mode until sync can again be reestablished. It will be noticed that this processing strategy bases the acceptance or rejection of data on the retention or loss of sync. Moreover, when in lock all sync words throughout the tape are examined, so that, to this extent, 100 percent testing is employed.

It is proposed to study the loss-of-sync problem by recognizing that there are two basic situations to be dealt with: (1) Loss of sync may be only apparent and (2) Loss of sync may be real and due to one or both of two constant-cause systems.

First let us consider apparent loss of sync. This condition is presumed to arise as the result of noise and to be characterized by the random occurrence of bit errors. Any frame or sync word or any succession of them may, as a result, be in error by any specified amount, assuming adequate noise. However, because of the uncorrelated nature of the phenomenon, successive frames are independently perturbed, and true loss of sync cannot be said to exist.

True loss of sync arises from one or both of two constant causes: (1) Bit slippage and (2) Equipment malfunction. Bit slippage itself may be broken down into whole and half-bit slippage. The dominant occurrence is whole bit slippage and is characterized by the introduction or subtraction of one or more bits in a frame, which in turn, delays or advances all succeeding bits. In general, the processing system cannot work itself out of this condition, except by going to the "search" mode, followed by acquisition and lock. In this sense whole bit slippage constitutes a constant cause system. Half-bit slippage, on the other band, can work itself out and thus achieves some element of randomness, so that it cannot be properly designated a constant cause system. The second constant cause system is the result of equipment breakdown and involves a highly correlated condition between frames.

It must be strongly emphasized that the methods of quality control are strictly applicable only in the case of a "bowl," i.e., random, universe, to use an expression of Shewhart.[2] For a "bowl" universe we may validly use sampling methods. In that case we say that the data are in a state of control. In other words, when constant causes such as sync slippage or equipment malfunctions occur, we cannot use sampling methods. We must first find the trouble and then remove it. On the other hand if we have been applying sampling methods on controlled data, which then goes out of control because constant

causes have arisen, our sampling techniques, if appropriately designed, will reveal the change.

In the sync problem we are dealing with a "bowl" universe when only bit errors are present. In such a universe any combination of bit errors may occur and give the appearance of being out of sync when such is actually not the case. From the data processing system standpoint the "risk" in sampling is that we may find erroneous evidence for lack of sync, and thereby reject good data. It should be mentioned, however, that even though data are rejected because of apparent "loss-of-sync," it is possible to recover them after sync is restored, but this consumes time which could be advantageously used in processing new data.

As indicated above, in applying quality control, it is important to be able to distinguish when the condition of randomness prevails and when the system is out of control (constant cause condition). Also, we need to distinguish between the two kinds of constant causes, because the measures for coping with bit slippage will be of no avail if the equipment has broken down. The only cure for the latter is to call in the maintenance people.

The procedure for coping with the "loss-of-sync," quality-control problem is now described. We assert that if there is more than some number of bit errors in a frame-sync word and if some specified number of successive frame-sync words are thus in error, there is "loss of sync." That is, we reject the hypothesis that we are in sync. In doing this, we note that we may be in error because, as mentioned above, the specified conditions for loss of sync may have arisen from the confluence of random effects which are not likely to persist over an extended period. On the other hand, we may wrongly accept the hypothesis that we are in sync, again because of random effects. Because of the special nature of the sync problem, admittedly, this second kind of error is much less likely to arise, and for that reason we shall not concern ourselves with it in the following discussion.

In Appendix A we show how to compute the mean time-between-occurrences, designated as "loss-of-sync." We call this the mean-occurrence time or the mean-time-to-occurrence and give it the symbol $\overline{T}$, measured in frames. An example would be the mean-time between occurrences of four successive sync-word errors, where a sync-word error is specified as more than three bit errors. The expression for $\overline{T}$ for this example is given in Appendix A.

If we do not have a satisfactory model for the "loss-of-sync" distribution, we should find it useful to obtain the standard deviation of the mean-occurrence time, $\sigma_{\overline{T}}$. The expression for $\sigma_{\overline{T}}$ for the above example is also worked out in Appendix A. Knowledge of $\overline{T}$ and $\sigma_{\overline{T}}$ permits us to use the Tchebycheff inequality[3] to compute the "risk" of wrong decision as to "loss-of-sync." It should be noted, however, that this inequality gives a rough estimate of the risk as compared with that obtained from knowledge of the distribution itself. Thus, the Tchebycheff inequality states that

$$P_r \left(| \overline{T} - T | \right) \geq \epsilon \leq \frac{\sigma_{\overline{T}}^2}{\epsilon^2} . \qquad (1)$$

In other words, the probability of the measured time between two occurrences of "loss-of-sync" fluctuating about the computed mean-time-to-occurrence of the population by an amount equal to or greater than a specified quantity ϵ is equal to or less than the computed variance of the population divided by the square of this same quantity ϵ. Thus, by suitable choice of ϵ (i.e., the confidence range) we can make the probability anything we desire.

Now consider the risk in relation to the Tchebycheff inequality as applied to determining loss of control. On a chance basis, if we observe the time between two occurrences of "loss-of-sync," the "bowl" from which this measurement is drawn may be one from which the actual $\bar{T}$ is significantly greater than the computed $\bar{T}$ and yet, nevertheless, find that $(\bar{T} - T) > \epsilon$. We might, therefore, erroneously assert that the system is out of control.

How frequently the data should be sampled is a function of the frequency with which out-of-control conditions arise. In this case, it would appear that the rate of slippage is the determining factor. Hence, it is suggested that the sampling rate should be at least as frequent as the average slippage rate. In fact, it would be wise to test more frequently than this. If we knew the distribution of the slippage rate, we could choose a test interval of $\bar{T}_s - t\sigma_{\bar{T}_s}$ where $\bar{T}_s$ is the mean interval between slippages, $\sigma_{\bar{T}_s}$ is the rms deviation of the slippage interval and $t > 1$. An assignment of t fixes the confidence we have that a slippage did not occur before we tested for it. If we do not know the distribution, the Tchebycheff inequality may be used to fix t for any assignable degree of confidence. It should be noted, however, that even if slippage occurs before it is observed, this does not necessarily mean that the data are irretrievably lost, because the unsynchronized data can be recorded and recovered later when synchronization is restored.

SYNCHRONIZATION AND EXPERIMENTAL ERRORS

In the previous section we have concerned ourselves exclusively with the loss of synchronization as though the ultimate goal in a major phase of data processing is to minimize the loss of synchronization. Actually, the ultimate goal is to provide data which meet the requirements on experimental error. According to this view maintenance of sync is only a means to that goal. The implication of this statement is that the emphasis on overcoming loss of sync in current data processing is misplaced. We must view loss of sync only as an indication of loss of control, and once we are in control our attention must be riveted on maintaining specified tolerances on experimental error. That is, we must set the limits on a control chart so as to maintain these tolerances, and if these tolerances are exceeded we must reject the data.

Tolerance limits fix the range through which the data may vary in accordance with the experimental specification[2]. It is insufficient, however, to center attention on the limits alone, because, although it may be desired to have the tolerance range as small as possible, it must be recognized that if it is too small the data rejection will be excessive. As has been stated elsewhere, the reasonable objective in data processing is minimum rejection consistent with maximum acceptable error probability[4]. We should consider the percentage of data that may be expected to fall within the tolerance range, i.e., of tolerance limits with an associated probability.

We come now to a consideration of how we can use the experimenter's requirements on error probability to establish a criterion for the minimum rejection of data. A derivation has been presented elsewhere[5] and is summarized in Appendix B of this report of the total rms analog error of a data word, σ_{tu}. This error includes the experimental error, and the errors caused by channel and quantizing noises. Assuming that these errors combine additively, we can use the relation, Equation (87), to compute σ_t. Thus

$$\sigma_{tu} = \left(\sigma_A^2 + \sigma_u^2 + \sigma_q^2\right)^{1/2} \tag{2}$$

where σ_A is the rms experimental error, σ_u the rms error caused by channel noise, and σ_q is the rms quantizing error. If independent observations of the same physical

quantity can be taken over parallel channels, Equation (89) can be used to achieve a smaller resultant error. Thus

$$\sigma_{mtu} = \left[\sigma_A^2 + \frac{\sigma_u^2 + \sigma_q^2}{m}\right]^{1/2}, \tag{3}$$

where m is the number of parallel channels providing independent observations of the same physical quantity.

We have now to relate the experimental error to the total rms analog error of a data word, σ_{tu}, or in the case of parallel channels, σ_{mtu}. From Equation (88) we have the average error probability in the detection of the individual pulses in the PCM train as a function of both pulse amplitude and rms noise. Hence, for a given σ_{tu}, this expression will tell us how many bits in a data word will be in error on the average, so that if we are given a permitted degradation ratio

$$k = \frac{\sigma_{tu}}{\sigma_A} \tag{4}$$

by the experimenter, we can compute the corresponding percentage of defective data, i.e., the permitted percentage defective. The risk, β, is the percentage of the time the experimenter is obliged by the sampling procedure to accept data exceeding the allowed percentage defective.

We now examine how we can compute the maximum acceptable bit-error probability knowing the error in measurement and the degradation factor permitted by the experimenter. From Equation (87) we have

$$\sigma_{tu}^2 = \sigma_A^2 + \sigma_u^2 + \sigma_q^2,$$

from which

$$k^2 = \frac{\sigma_t^2}{\sigma_A^2} = 1 + \frac{\sigma_u^2}{\sigma_A^2} + \frac{\sigma_q^2}{\sigma_A^2}. \tag{5}$$

It follows that

$$\sigma_u^2 = \sigma_A^2 \left(k^2 - 1\right) - \sigma_q^2. \tag{6}$$

From Equation (86) we have that

$$\sigma_q^2 = \frac{q^2}{12} \tag{7}$$

where q is the amplitude of the quantization interval. Also,

$$\sigma_u^2 = \sigma^2 p \frac{4^n - 1}{3} \tag{8}$$

where p is the per digit error probability and n is the number of digits in a word. It follows that

$$p = \frac{3}{4^n - 1}\left[\frac{\sigma_A^2\ k^2 - 1}{q^2} - \frac{1}{12}\right]. \tag{9}$$

This "p" we may now label $p_{e_{max}}$ because it represents the maximum acceptable bit-error probability, subject to the rms error in measurement σ_A and the degradation factor k.

It should be noted that it may not be possible to achieve the specified degradation factor k with data from a single ground station (tape). In that case it might be desirable to acquire independent data from more than one station, in which case, Equation (89) could be used to bring σ_{mtu} to the value required by the assigned k. That is, we would have

$$k = \frac{\sigma_{mtu}^2}{\sigma_A}. \tag{10}$$

We shall now say a few words about the problem of setting tolerance limits in data processing. When the experimenter asserts that the rms error in his measurements is σ_A he means, or should mean, that x percent of the samples he takes include the true value of the characteristic being measured (assuming systematic errors have been removed) within the range between $-\sigma_A$ and $+\sigma_A$, centered on the sample mean[2]. The larger $|\sigma_A|$ the larger the range and the less precisely the experimenter knows the true value. If he takes a range based on $|t\sigma_A|$, $t > 1$, more than x percent of the samples will contain the true value in the range between $-t\sigma_A$ and $+t\sigma_A$, centered on the sample mean; i.e., he locates the true value in his confidence range more of the time but knows its value less precisely. If we ask the experimenter to accept a range based on $|\sigma_{tu}| > |\sigma_A|$, we are asking him, in effect, to be less confident that before that the true value is in the range between $-\sigma_A$ and σ_A, centered on the sample mean. It is up to him to say how much reduction in confidence he will accept. If the channel and quantization noises are such that this reduction exceeds the acceptable limit, the data should be rejected by the data processing system.

Having determined the $p_{e_{max}}$ from the specified σ_{tu} or σ_{mtu}, we are then in a position to construct a control chart for sampling on the basis of the maximum allowed error probability, determined from the specified tolerance on experimental error. We note that in sampling from a population of data whose per-bit error probability exceeds a prescribed value, there is a calculable probability that change effects will cause the sample per-bit error probability to be less than the prescribed value, so that the data are accepted; or if the population value is below the prescribed value there is a calculable probability that the sample per bit error probability will be greater than the prescribed value, so that the data are rejected. Either way there is a risk involved, for the experimenter in the first instance and for the data processor in the second.

We are now in a position to see the distinction between the quality control test procedure based upon experimental tolerances and that based upon loss-of-sync. In the latter the purpose of the test is to recognize loss of control. In the former the purpose of the test is not to recognize loss of control but to insure that the controlled data meet specified experimental tolerances, assuming control is maintained. Hence, the purpose of the test for meeting tolerances is served by two quality control procedures: (1) A test for control ("loss-of-sync" test) and (2) A test to insure that the data comply with specified tolerances.

We shall find that the purpose of the second test is contained in a sequential sampling plan which is applied in the form of a control chart, such as that shown in Figure A-24.

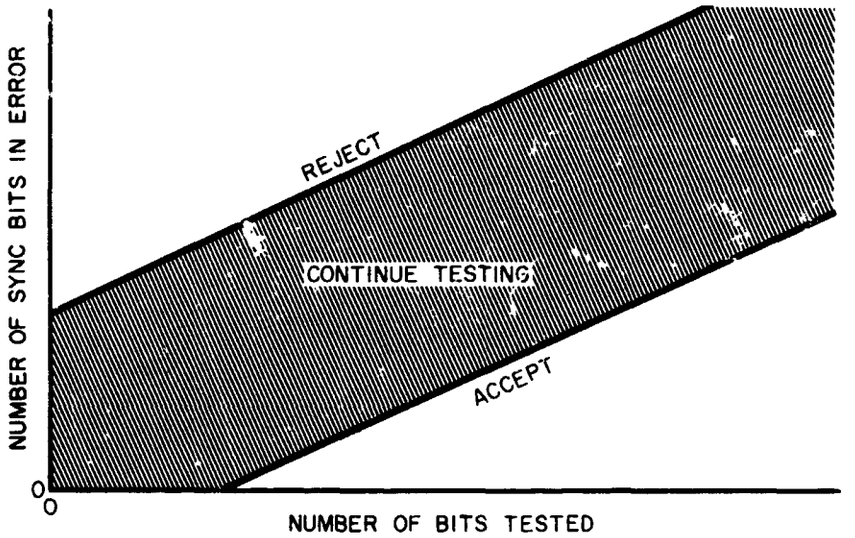


Figure A-24–Control chart for sequential test based on a quality level of $P_{e_{max}}$ for synchronization errors.

To understand how the "reject" and "accept" lines are plotted, we must introduce some terms. We begin by specifying who is the "producer," who is the "consumer" and what is produced in the contest of data processing. We define the producer to be the data processing system, the consumer to be the experimenter and the product to be the data. The goal of the producer is to maximize the output of data while meeting the "consumer's" requirements on the reliability of the data to a specified degree of confidence. He should achieve this with a minimum number of tests because his tests destroy data and also because the flow of data will soon exceed his capacity for 100 percent testing. The goal of the consumer is to receive maximum data, subject to his requirements on reliability. Accordingly, the producer sets himself a confidence limit, with a high enough expectation to meet the experimenter's requirements but not so high that his probability of rejecting a good lot of data becomes excessive. Conversely, the consumer (experimenter) must set himself a confidence limit which he will permit a certain fraction of the data to be bad. In each of these cases there is a risk, but the risk must be faced to bring the processing cost within acceptable limits. The risk for the producer is the probability of rejecting a good lot of data. The risk for the consumer is the probability of accepting a bad lot of data.

We shall now introduce the following terms.*

R_0 = reliability associated with producer's risk

R_1 = reliability associated with consumer's risk

α = probability of rejecting a good lot (producer)

β = probability of accepting a defective lot (consumer)

*Details on how to compute α, β, R_σ, R_1 and the "accept" and "reject" lines are given in pp. 468-491 of (3).

Accept Line =

$$\frac{a_1}{b_1 - b_2} + \frac{(-b_2)\, t}{b_1 - b_2} \tag{11}$$

and Reject Line =

$$\frac{a_2}{b_1 - b_2} + \frac{(-b_2)t}{b_1 - b_2} \tag{12}$$

where t = observation time and

$$a_1 = \ln \frac{\beta}{1 - \alpha} \tag{13}$$

$$a_2 = \ln \frac{1 - \beta}{\alpha} \tag{14}$$

$$b_1 = \ln \frac{1 - R_1}{1 - R_0} \tag{15}$$

$$b_2 = \ln \frac{R_1}{R_0}. \tag{16}$$

The probability of rejecting good data, which we call the producer's risk, α, and the probability of accepting bad data, which we call the consumer's risk, β, can both be computed from the percentage of bit errors in a sync word. We can estimate the percentage of bit errors by simply counting and averaging the number of bit errors in a succession of frames, say for example, ten. Suppose this is done, and the percentage of errors is found to be p. There is then a probability, r, that the percentage, p, or any other larger percentage, taken as the threshold above which we announce a sync-word error, will be exceeded in a given sample. Hence, the probability, r, constitutes the risk that we will reject data. On the other hand the average p may lie above the threshold of rejection, but nevertheless, there is a non-zero probability that in a sample the percentage of errors will fall below the rejection threshold, so that there is the corresponding risk of accepting bad data.

The producer's risk, α, and the consumer's risk, β, are computed from a quality level based on the maximum acceptable probability of error, $P_{e_{max}}$. That is, we count the errors in sync words, and when the percentage of errors exceeds $P_{e_{max}}$ we reject the data in a lot as being outside tolerance limits. When the percentage is equal to or less than $P_{e_{max}}$ we accept the data. Hence, the control chart has for abscissa the number of sync bits tested and for ordinate the number of sync bits in error, as shown in Figure A-24.

If a sample falls in the null region between the "accept" and "reject" lines, then testing must continue until the result of test moves into the "accept" or "reject" region; if into the "accept" region, testing may then stop. It should be noted that we may find ourselves in the "continued testing" region for an impractically long period. As this is precisely the condition we are trying to avoid, it would then become necessary to adopt a truncation rule, so that, effectively, the two control lines are brought together after a specified number of observations. Standard methods for doing this are given in the quality control literature, e.g., pp 521-522 of (6).

OGO tapes are generally of two types, good or bad, and if they begin good, they remain good and vice versa. This behavior would suggest the need for two categories of control charts. However, the dichotomy into two quality levels for tapes while true for OGO, is not true for all vehicles. Thus, in the Scout vehicles for example, the signal-to-noise ratio varies widely with the position of the vehicle in its orbit relative to the earth. Hence, the quality of the data on a tape varies accordingly, so that a single control chart for the tape in that case would not be suitable.

Let us explore the implications of a variation in quality throughout a tape. The universe percentage of errors, p, clearly is a function of signal-to-noise ratio, as seen by the fact that we expect nonerror occurrence in a noise-free channel and error occurrence in a noisy channel. Now if we were to hold the rejection threshold at a given value, it is clear that the α and β would vary with signal-to-noise ratio, since p does. For example, if the signal-to-noise ratio is large, p is small, the risk of exceeding the threshold is small, and the data are, of course, good. On the other hand if the signal-to-noise ratio is small, p is large, the risk of exceeding the rejection threshold is large, and the data are bad. From this it should be clear that the assignment of a fixed rejection threshold in a fluctuating signal-to-noise ratio environment is likely to result in unsatisfactory or at least inefficient data processing. Hence, a better plan would seem to be to adjust the rejection threshold in accord with the variation in signal-to-noise ratio in order to hold α and β fixed. This objective can be achieved by monitoring the number of sync word errors to form a running estimate of p. From this running estimate we can compute "accept" and "reject" lines for the control chart, the position of the lines changing with the value of the estimate. It would be seen then that the width of the "continuous testing" region would automatically widen with small signal-to-noise ratios and narrow with large signal-to-noise ratios.

It is noted that the two tests - one for lack of control and the other for meeting tolerances on experimental data - may be conducted simultaneously. Thus, if control is lost the results of both tests are negative, whereas if control is maintained, the tolerance test may be either in the "accept" or the "reject" region. If in the "reject" region the data are too noisy to be useful to the experimenter. In that event there is only one thing to do and that is to continue taking samples at frequent intervals until the quality of the data improves to the extent of carrying the test into the "accept" region, at which point the data are again useful. It is emphasized that if control is maintained but the sample is outside tolerance limits we reject the data and continue to sample until the data are again acceptable, but he does not go into the "search" mode to look for sync or stop the test to look for mechanical trouble.

FAILURE OF TIMING

Time is an important independent variable in correlating data with satellite position. The time code for taped data indicates the data reception time in Universal time, serially recorded on a channel of the tape adjacent to the incoming recorded data.

To promote the reliability of time readings, an error detecting and correcting time decoder operating on the following principle is used[7,8]. A Binary Coded Decimal (BCD) detector reads the BCD time code and a Serial Decimal (SD) code detector the SD time code. The time decoded in either the BDC or the SD decoders is automatically set into the accumulator register at the beginning of each processing run or when the mode of operation changes. This register is updated by the 1 kc output from an internal oscillator. A comparison is then made between the next BCD time reading, the SD time reading and the updated time in the accumulator, all of which usually agree. If, however, they disagree, the circuits automatically check themselves for malfunction. If there is no circuit malfunction, a check is made to see whether or not the BCD circuit is in synchronism with the incoming BCD time code. If it is in sync, a new BCD time code is set in the accumulating register and updated, so that when there is a break in the data acquisition period, the time decoder automatically picks up the new time at the start of the new period. If there is a BCD circuit malfunction or a signal loss, the BCD circuit is inhibited, and the SD circuits provide the time reading. The converse holds if the SD system fails.

Thus, the decoder makes use of the redundancy of having both the BCD and SD time signals available for processing. If only one of these two signals is present, only that one is considered.

The first step in the quality control of timing errors is to be able to distinguish between the "bowl" universe and constant causes. It will be noted from what has just been said that constant causes are failures on the part of the BCD or the SD circuits or failure on the part of these circuits to be in sync with their time codes. In addition to these broad categories of constant causes, there are others for which there is no present means for recognition. Thus, we have (1) A BCD bias error and an SD bias error where the two bias errors are equal; (2) A BCD bias error and no SD code anywhere on the tape; (3) No BCD code anywhere on the tape, an SD bias error; (4) No BCD code anywhere on the tape, an SD bias error, and the linearizing frequency missing for very short periods of time. It would seem that some means of error detection should be incorporated against these, because, otherwise, these constant-cause systems may, if they occur with sufficient frequency, negate all efforts at quality control.

Factors contributing to the creation of a "bowl" universe are time fade-outs, noise and bias errors. These can create all the conditions for a failure in timing on a chance basis, and we may expect that any combination of timing errors may occur giving the appearance of actually being out of timing when such is actually not the case. Here, as in the case of the sync problem, there are two risks: (1) From the standpoint of the data processor, according to which evidence is found for timing errors outside tolerance limits when such is actually not the case, resulting in the rejection of a good data and (2) From the standpoint of the experimenter, according to which data actually outside tolerance limits are accepted because sampling has shown the data to be good.

There is a provision in the timing decoder for adjusting the number, N, of consecutive unequal comparisons of BCD holding and accumulating registers or for adjusting the number, M, of consecutive unequal comparisons of SD holding and accumulating registers. An increase in N or M decreases the chance that the code will be erroneously reported as "out-of-sync" with its time code, causing it to be checked for a circuit malfunction that may not actually exist, a condition which would, of course, interrupt the flow of data and result in a needless loss of time. On the other hand, if N or M is increased too much, there is a risk that too much erroneous data will be accepted before the decision is made that there is loss of timing.

We see then a situation entirely parallel with that of loss of synchronization: (1) We must be able to recognize the state of lack of control; and (2) we must know when we are

outside the tolerance limits on error as specified by the experimenter. In timing, also, there are two constant cause systems leading to lack of control. One is the introduction of bias errors which can be eliminated by updating in the accumulator, and the other is equipment failure. The time decoder is capable of working itself out of the first kind of constant cause system but not of the second, so that the first is only a temporary constant cause system. The foregoing suggests the need for two quality control procedures, (as in loss of synchronization) one for indicating loss of control and the other for indicating operation outside the tolerance limits. Loss of control is indicated when the number of consecutive unequal comparisons between the BCD holding and accumulating registers equals a specified number, N, or when the number of consecutive unequal comparisons between the SD holding and accumulating register equals a specified number, M. An unequal comparison is defined as any difference exceeding a time discrepancy greater than $|y|$%. The number N is that number for which the probability is less than some number p on the assumption of a "bo/l" universe and similarly for M. To assign the numbers N and M, it will be necessary to know the probability of a single unequal comparison. This probability can be estimated by recording the number of unequal comparisons in a test sample. Knowing the probability of a single unequal comparison, we can compute the probability of its occurring a given number of times. We choose the minimum number of times, N or M as the case may be, to make this probability less than p. The assumption is that N or M or more successive events is so unlikely on a chance basis alone that its occurrence may be regarded as resulting from a constant cause system.

Let us agree to call the compound event; the successive occurrence of N or M or more events, evidence of "loss-of-timing." Then, as in the case of "loss-of-synchronization," we can compute the mean-time-to-occurrence, $\overline{T}_t$, and variance, $\sigma_{\overline{T}_t}$, of this compound event. As in the case of testing for loss of synchronization, knowledge of $\overline{T}_t$ and $\sigma_{\overline{T}}$ permits us to compute the "risk" of wrong decision as to "loss-of-timing" by means of the Tchebycheff inequality or preferably from the distribution of "loss-of-timing," if a satisfactory model of that distribution is available.

As we have seen, however, indication of the state of control on timing is insufficient. We must then test for compliance with the tolerance limits on experimental error. Specifically, knowing the experimenter's rms error in timing, σ_{A_t}, the error in timing caused by quantizing, σ_{q_t}, and the permitted amount of degradation in timing,

$$k_t = \frac{\sigma_{(tu)t}}{\sigma_{A_t}}, \tag{17}$$

we can solve for the permitted rms error in timing caused by the channel noise. From this we can compute the maximum acceptable error probability, $p_{e_{max}}$, by the method described in the previous section on failure of synchronization. Then $p_{e_{max}}$ can be used as the basis for computing the "producer" and consumer "risks," α and β respectively, and from these we obtain the "accept" and "reject" lines by the methods already described. Thus, we have the sequential control chart for timing tolerances, as shown in Figure A-25.

Precisely the same approach is applicable for both the BCD and SD time decoders, each of which must have its own control chart for tolerance limits. Tests for control and tolerance limits may be made simultaneously as explained in the previous section.

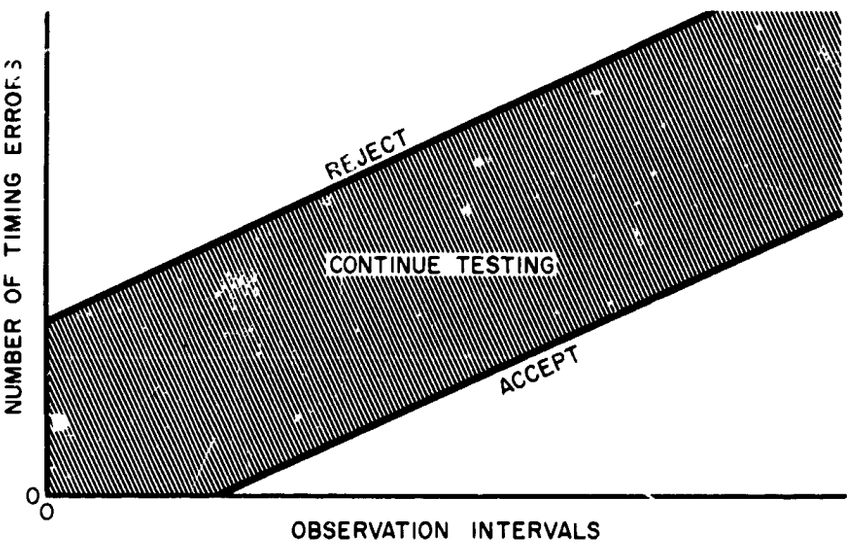


Figure A-25–Control chart for sequential test based on a quality level of a $P_{e_{max}}$ for timing errors.

RE-ESTABLISHMENT OF SYNCHRONIZATION AND TIMING AS PROBLEMS OF REPAIR

Having described the factors involved in testing for a state of control and for meeting tolerance limits on experimental data, we are now in a position to take up the re-establishment of sync and/or of timing, not as isolated but rather as interrelated problems. Data processing fails if the synchronizer and/or the time decoding links fail. Hence, we may look upon the two steps in processing as analogous to a tandem connection of two systems, as in Figure A-26. We see that two data to be processed the two systems

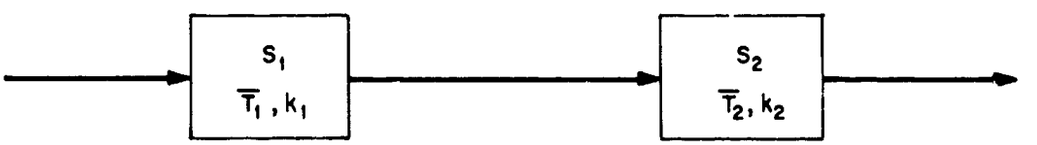


Figure A-26–Two links in the chain of data processing, with a mean-time-to-occurrence of failure of $\overline{T}_1$ and $\overline{T}_2$ and of error probability p_1 and p_2, respectively.

must jointly exhibit a state of control; moreover, they must jointly meet their respective tolerance limits.

Since "search" and "acquisition" are steps in the re-establishment of control for S_1 and "updating" is a step in the re-establishment of control for S_2, these measures may be regarded as phases of the repair problem, which, of course includes maintenance as well. In the following we shall not be concerned with maintenance as such. What, then, is the total rejection rate and the total error probability of tandem systems with prescribed values of each? We would expect the resultant mean-time-to-occurrence to be decreased and the resultant error probability to be increased. Let us examine these two effects.

Instead of working specifically with mean-time-to-occurrence, it will be convenient to introduce the notion of the probability of system availability[9,10]. This probability is related to being in or out of control, since clearly the system when out of control is not available for processing but is, in fact, in a failed state. Hence, we shall treat the probability of availability as the probability that the system is in a state of control. We have then to obtain two probabilities of availability, one for each of the sub-systems of the data processor. In this connection it should be noted that controlled operation in the timing decoder depends on controlled operation in the preceding synchronization stage.

Appendix A indicates how flow diagram techniques can be used to obtain the probability of availability of a single system with repair or of two systems in cascade with repair in each.

We define an availability function $A(t)$ as the probability that at any time t the system is in an operating state[9,10]. If we are interested in the long-term availability of the system, we can let $t \to \infty$ and

$$A(\infty) = \frac{\mu}{\lambda + \mu}, \tag{18}$$

where λ is the probability per unit time of a processing stage going out of control and μ is the probability per unit time that the process will be restored to control by "search" and "acquisition" or by "updating," as the case may be. For example, $A(\infty)$ may be regarded as the percentage of time that the processor is in synchronism with the tape.

In the case of the STARS Phase I data-processing system, tests for synchronization and timing are conducted simultaneously, but the system may be synchronized and not in timing. It cannot, however, be in timing and not synchronized. Now suppose that we define the availability of the processor relative to synchronization as

$$A_1(\infty) = \frac{\mu_1}{\lambda_1 + \mu_1} \tag{19}$$

and the availability of the processor for timing tests as

$$A_{21}(\infty) = \frac{\mu_{21}}{\lambda_{21} + \mu_{21}}, \tag{20}$$

where the subscript "21" indicates that this probability is conditional on the first stage being in a state of control.

The probability of the first and second stages jointly being in a state of control is

$$A_{121}(\infty) = A_1(\infty)\, A_{21}(\infty) = \left(\frac{\mu_1}{\lambda_1 + \mu_1}\right) \cdot \left(\frac{\mu_{21}}{\lambda_{21} + \mu_{21}}\right). \tag{21}$$

Thus, cascading operations results in reduced availability or, in other words, reduced control, as might be expected.

We wish to examine next the effect of cascading processing operations on meeting tolerance limits. Suppose that for each processing stage we compute the maximum

acceptable error probabilities, p_1, and p_2. Then the resultant error probability is

$$p_r = p_1 + p_2 - p_1 p_2 \tag{22}$$

assuming independence in the occurrence of errors in the two stages. Since p_r is larger than p_1 or p_2, it is evident that cascading has increased the percentage of the time our data will lie outside tolerance limits.

We may look at the error problem in another way. If we assume that errors propagate indpendently throughout the data processing system and that for the first stage of processing the rms error is $\sigma_{(tu)_1}$ and for the second $\sigma_{(tu)_2}$, the resultant rms error is

$$\sigma_r = \left(\sigma_{(tu)_1}^2 + \sigma_{(tu)_2}^2\right)^{1/2}. \tag{23}$$

Thus, the effect of increasing the total rms error is either to decrease the confidence of the experimenter in the data he receives and/or to reject more data in processing.

Is it possible to rationalize the data processing system, so that tests for control and for meeting tolerance limits, now done in two steps, could be combined into a single step? In the present procedure telemetric data received at the ground stations are recorded on so-called analog tapes which are then sent to the data processing center where they are converted to digital tapes for use in the digital computers. There seems to be no inherent reason why the telemetric data, timing signals included, cannot be recorded directly in the form of digital tape data. If this step were practicable, it would be unnecessary to conduct synchronization and time decoding tests independently. A single test would suffice to check for synchronization and timing, and this step would be essentially a test for synchronization. Let us see how this would affect the probability of availability and the overall error. For the single test

$$A(\infty) = \frac{\mu}{\lambda + \mu}, \tag{24}$$

which probably would approximate $A_1(\infty)$. In the case of error probability the single step value would approximate p_1, or, alternatively, the resultant rms error would approximate σ_{u1}, values which, in each case, would be less than those for the cascaded operation.

CONCLUSION

Within the framework of data processing the role of quality control is to insure a given level of quality in experimental data without at the same time being wasteful of data and to accomplish this with minimum testing. The minimum testing requirement may seem unnecessary at a time, such as the present, when the processing facilities are equal to the flow of data; but with the anticipated exponential increase in the data volume, the time may soon come when the flow will exceed the capacity of the processing system. In that case the only alternative to a totally unreasonable mushrooming of processing equipment costs will be a more efficient utilization of the existing facilities. Costs, however, are not the only factor. One must also consider the possibility of freeing the processing system for conducting a greater variety of tests, with the hope that such tests might reveal unexpected phenomena in the data. Specifically, looking ahead to the introduction of the STARS Phase II System, we note that the computer will check data quality as well as perform data decommutation, to determine whether the data quality has deteriorated in

the satellite or ground telemetry links or in the STARS Phase I operation. The fewer checks of a given kind the computer is required to make the more available it is to check for hidden periodicities or non-stationarities in the data, conditions which may profoundly affect the interpretation of experimental results.

We have seen how quality control procedures can be used to indicate whether the processing system is operating satisfactorily, i.e., is in a state of control, whether there is bit slippage, loss of timing, or equipment failure. Moreover, this same approach can be used to determine whether the data are being processed with sufficient reliability to meet the experimenter's requirements for precision, and this can be done concurrently with testing for control. If the tolerance limits specified by the experimenter cannot be met with a particular lot of data, then that lot must be discarded. An indication has been given of a method for relating tolerance limits and error probability and of how this can provide a rational basis for accepting or rejecting data.

The consequences of processing the data in such a way that tests for synchronization and timing could be combined in a single operation were discussed, and it appears that a significant gain in maintaining control and in meeting tolerance limits on experimental data would result.

Finally, it is emphasized that the views expressed in this report constitute only an introduction to the subject of quality control in the processing of telemetric data from space vehicles. Before the full implications of the use of quality control can be understood, careful study must be made of the advantages and limitations of various testing plans and of the possibilities that these afford for widening the interpretation of data. There should also be a careful examination of the role of quality control in the STARS Phase II system of processing and of other possibly practical systems for rationalizing the processing operation. In addition, there should be a thorough examination of how the precision of experimental data is affected by processing, particularly in view of the enormous redundancy in the data. It will be noted that this question of data redundancy in its relation to the dependence of precision of measurement on channel errors was not raised in this report.

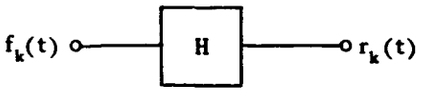
APPENDIX A

ERROR ANALYSIS IN DATA PROCESSING BY FLOW GRAPH METHODS

INTRODUCTION

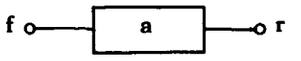
In some applications flow graphs are intended to provide a pictorial description of the systems as a guide in mathematical analysis.[10, 11, 12] Moreover, flow graphs appear to be evolving into a new type of mathematical notation that provides a concise, easily visualized description of system structure, capable of being manipulated and "solved" just as the conventional set of equations describing the system may be manipulated and solved.

Crucial to the use of flow graphs is the notion of representation of an operation by its transmittance. Thus, if $f(t)$ is the input to an operator H and $r(t)$ is the response, we have the representation



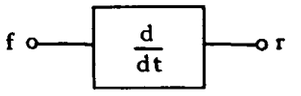
where $H_k = r_k(t)/f_k(t)$ is the transmittance of the operator H to the characteristic signal $f_k(t)$. Some important linear stationary operators are:

Scalar:



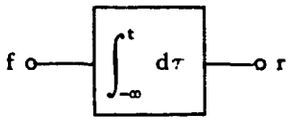
$af(t) = r(t)$

Differentiator:



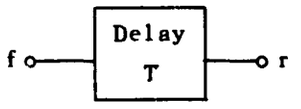
$\frac{df}{dt} = r(t)$

Integrator:



$\int_{-\infty}^{t} f(\tau)\, d\tau = r(t)$

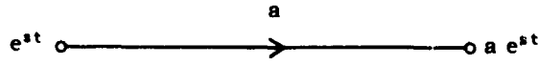
Delayor:



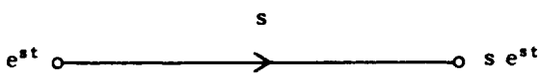
$f(t - T) = r(t)$

The corresponding flow graph representations are:

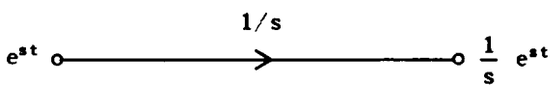
Scalor:



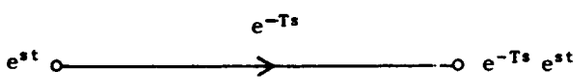
Differentiator:



Integrator:



Delayor:



where s is the complex frequency of the exponential signal $A\,e^{st}$.

Flow-graph methods may be used not only to represent deterministic systems as above but also non-deterministic systems, as we shall now see. In illustrating the application of flow graphs to non-deterministic systems, we shall take as example the case of apparent loss of sync. We shall assume that we are concerned with the OGO experiment in which the data are transmitted via PCM over a binary-symmetric channel (BSC).

ERROR ANALYSIS FOR LOSS OF SYNC*

We may represent the binary symmetric channel by the transition diagram (Figure A-27).

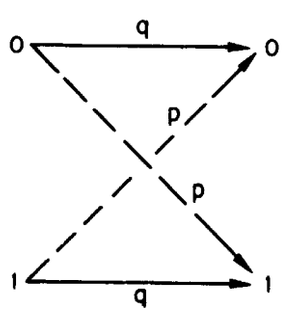


Figure A-27–Binary-symmetric channel (BSC),
q ≡probability of error-free bit,
p ≡probability of erroneous bit.

Let us designate the correct occurrence of a bit by y and the incorrect occurrence by z. Then the probability generator for a sequence of y's and z's is shown in Figure A-28. To start the generator a unit impulse of probability, $\delta(t)$, is delivered to the "start" or y node at $t = 0$. If $x = e^{-s}$ represents the transmittance of a unit delayor, at time $t = 1$ an impulse of strength p will appear at the z node, signifiying that there is a probability p that the first bit will be z. Also, at $t = 1$ there is an impulse of strength q at the y node, signifying that there is a probability q that the first bit will be y.

The probability generator for the first occurrence of z is characterized by the graph transmittance[10]

*See report of C. A. Rohde for alternative approach.

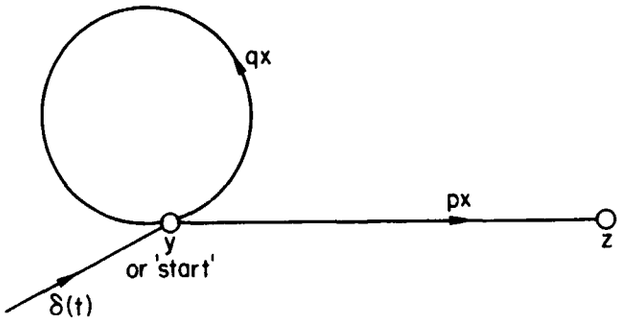


Figure A-28–Flow-graph for single event.

$$F_z(x) = \frac{px}{1 - qx}$$

$$= px + pqx^2 + pq^2 x^3 + \cdots \tag{25}$$

Thus, the probabilities of the system entering the state z, after it is started with unit probability at $t = 0$, are at $t = 1, 2, 3$ etc. p, pq, pq^2, etc. These are the probabilities for the event consisting of a single bit error.

Suppose we are interested in the compound event consisting of the successive occurrence of a given event. An example might be a decision in the processing of data from OGO in which we ask for the probability of the first occurrence of a run of four successive Z's, which we designate by the event ϵ. In this example we shall specify that each Z stands for the occurrence of a sync word with more than three errors, i.e., a sync-word error. Hence, we shall designate the occurrence of a sync word with three or fewer errors by the letter Y, i.e., sync-word not in error. The probability of Z we shall call P and that of Y we shall call Q, so that $P + Q = 1$.

The flow graph and the probability generating function for ϵ are as follows. The flow-graph is shown in Figure A-29 and the probability generating function for ϵ is

$$F_\epsilon(x) = \frac{(Px)^4}{1 - Qx - PQx^2 - P^2Qx^3 - P^3Qx^4}$$

$$= P^4x^4 + P^4Qx^5 + (P^5Q + P^4Q^2)\,x^6 + \cdots \tag{26}$$

Thus the probability of a run of four successive Z's occurring for the first time in six frames is

$$(P^5Q + P^4Q^2).$$

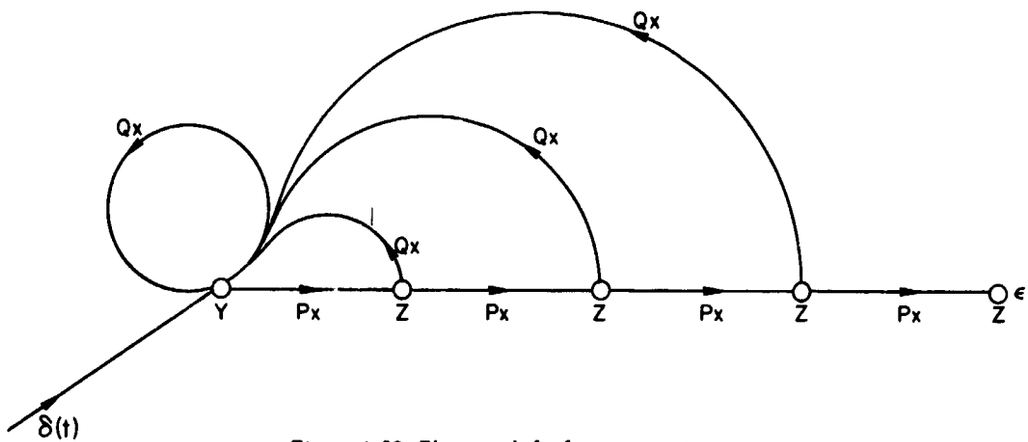


Figure A-29–Flow-graph for four successive events.

The mean-occurrence time until the first occurrence is*

$$\bar{T} = \frac{d}{dx} \ln F_{\epsilon}(x)\Big|_{x=1} = \frac{4 - 3Q - 2PQ - P^2Q}{1 - Q - PQ - P^2Q - P^3Q}, \tag{27}$$

where in the present application $\bar{T}$ is measured in frames. $\bar{T}$ is the mean-time-to-occurrence between the events called "loss-of-sync." We have

$$P = \text{Prob.}\,(r > 3)$$

where r is the number of bits in error. Hence

$$P = \sum_{r=4}^{27} \binom{27}{r} p^r q^{27-r}, \tag{28}$$

since in the case of OGO there are 27 bits in a sync word. Also

$$P = 1 - \sum_{r=0}^{3} \binom{27}{r} p^r q^{27-r}. \tag{29}$$

In addition to the mean-time-to-occurrence of the event which is the criterion for losing sync, we are interested in the variance of this time-to-occurrence, because by means of it we can establish the limit for any specified degree of confidence regarding the state of control, as explained in the text. The variance in the mean-occurrence time is**

$$\sigma_{\bar{T}}^2 = \left[\frac{d}{dx} \ln F(x) + \frac{d^2}{dx^2} \ln F(x)\right]_{x=1} \tag{30}$$

*See p. 761 of Reference (10).
**See p. 672 of Reference (10).

$$\sigma_{\bar{T}}^2 = \frac{4 - 3Q - 2PQ - P^2Q}{1 - Q - PQ - P^2Q - P^3Q}$$

$$+ \frac{8Q - 3Q^3 + 10PQ - 6PQ^2 - 8P^2Q - 10P^2Q^2 + 20P^3Q + 24P^3Q^2 + 5P^4Q^2 - 2P^5Q^2 - 4}{1 - 2Q + Q^2 + 2PQ + 2PQ^2 + 2P^2Q + 3P^2Q^2 - 2P^3Q + 4P^3Q^2 + 3P^4Q^2 + 2P^5Q^2 + P^6Q^2} \quad (31)$$

For purposes of illustration suppose that the bit-error probability is 0.10. Then $P = P(r > 3) = 0.282$ from Equation (28) or (29) and $Q = 1 - P = 0.718$. Hence, from Equation (27) $\bar{T} \doteq 178$, and $\sigma_{\bar{T}} \doteq 13.4$ from Equation (31). By the Tchebycheff inequality the probability that $|T - \bar{T}| \geq 3\sigma_{\bar{T}} \leq \sigma_{\bar{T}}^2/9\sigma_{\bar{T}}^2 = 1/9$, so that we are roughly 90% confident that the deviation about $\bar{T} = 178$ will not exceed 40.

LOSS OF CONTROL AND REPAIR IN SINGLE AND CASCADED SYSTEMS

Let us define a quantity λdt as the probability of a system going out of control in a small time interval dt and μdt as the probability of the system being restored to control in the same interval dt. We shall assume that these probabilities are stationary and independent of the state of the system prior to the present instant. The flow-graph for system failure is shown in Figure A-30. The nodes p_0 and p_1 are the probabilities of the system being operative or inoperative, respectively. The nodes $\dot{p}_0$ and $\dot{p}_1$ are the rates: $\dot{p}_0 = dp_0/dt$ and $\dot{p}_r = dp_1/dt$. These rates are interpreted as the expectations of the system entering (or leaving) the states "0" or "1" in any small interval of time dt. The condition $p_0 + p_1 = 1$ must be satisfied, so that $\dot{p}_0 + \dot{p}_1 = 0$, and an increase in p_1 must be balanced by a decrease in p_0. These transitions are indicated by the branches labeled λ and $-\lambda$ in Figure A-30.

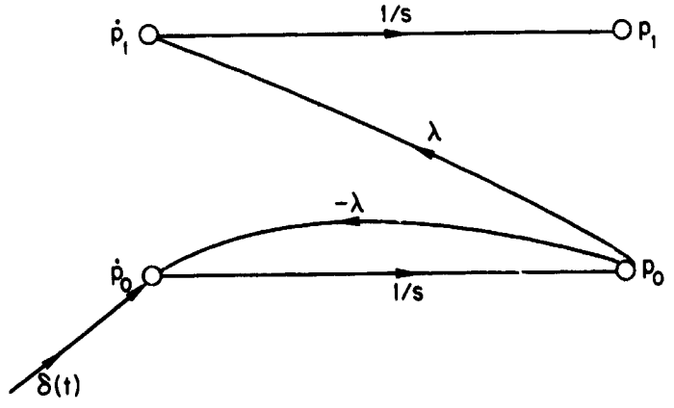


Figure A-30–Flow-graph for system failure.

We assume that at time $t = 0$ an impulse of probability is applied to the node $\dot{p}_0$ and immediately following the system is in the state "0" with unit probability. The transmittance of the probability generator is

$$P_0(s) = \frac{\frac{1}{s}}{1 - \frac{-\lambda}{s}} = \frac{1}{s + \lambda} \quad (32)$$

and thus

$$p_0(t) = e^{-\lambda t} \tag{33}$$

from which

$$p_1(t) = 1 - e^{-\lambda t}. \tag{34}$$

The relations (33) and (34), respectively, are the probabilities that the system is in the "operating" or the "failed" state at time t .

We next introduce the provision for recovering control, assuming that it has been lost, a problem analogous to the repair of equipment in failure analysis. Control is recovered by introducing a "renewal" path from state "1" back to state "0", as shown in Figure 31.

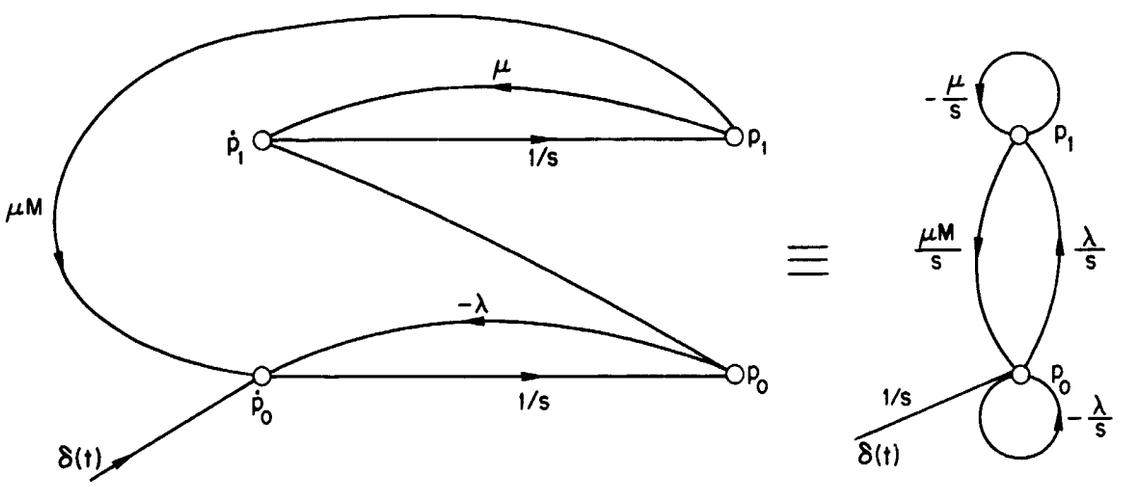


Figure A-31–Flow-graph and its reduced equivalent for failure system with renewal.

Restoration of control or "renewal" is to take place at an average rate μ, determined by the branch labeled μM between nodes p_1 and p_0, and to conserve probability a corresponding amount must be subtracted from the p_1 node. "M" is a tagging variable for keeping track of the number of renewals.

The probabilities of the system being in the operative and inoperative states are, from the grapn transmittances, respectively,

$$u_0(s, M) = \frac{s + \mu}{s^2 + (\mu + \lambda)\, s + (1 - M)\, \mu\lambda}, \tag{35}$$

$$u_1(s, M) = \frac{\lambda}{s^2 + (\mu + \lambda)\, s + (1 - M)\, \mu\lambda}, \tag{36}$$

from which it can be shown[10] that the steady-state probability of operation, called by some writers the probability of availability, is[9]

$$A(\infty) = \frac{\mu}{\lambda + \mu} . \tag{37}$$

This probability may also be regarded as the probability that the system is in a state of control.

If n independent, identical systems are cascaded, the joint probability of overall-system availability is

$$A_r(\infty) = \left(\frac{\mu}{\lambda + \mu}\right)^n . \tag{38}$$

ACKNOWLEDGMENT

The author is grateful for much helpful information and many valuable suggestions received in discussions with the following Goddard personnel: Messrs. C. J. Creveling, T. J. Karras, R. Lee, T. J. Lynch, M. Pasternack, W. P Poland, J. Quann and C. M. Stout and from the following academic members of the Summer Workshop 1965; Drs. C. A. Wogrin, M. Javid, F. L. Ramsey, C. A. Rohde, and Messrs. R. D. Coleman and G. A. Gordon.

APPENDIX B

MEASUREMENT AND COMMUNICATION ERRORS

MEASUREMENT ERRORS (ERROR INDICES)

In the treatment of experimental data the average of fluctuations rather than individual fluctuations in the observations are of interest.[2] Two well-known measures of error in observation are the standard deviation and the probable error. These are called error indices. By definition the probable error is strictly applicable only in the case of normal, that is Gaussian, distributions since it is defined as that value of the deviation from the mean of a normal distribution which is just as likely to be exceeded as not. Thus, a probable error range has a probability of 0.5. The standard deviation, on the other hand, is defined as the root-mean-square deviation from the average of observations. This definition is applicable for all distributions, so that it is a general measure of average fluctuation or error. In the case of a normal distribution, there is a fixed relation between the probable error and the standard deviation. The standard deviation range includes 68.3 percent of all observations. In the case of a discrete observable X_i the expression for the standard deviation σ is:

$$\sigma = \left[\sum_{i=1}^{N} \frac{(X_i - \bar{X})^2}{N} \right]^{1/2}, \tag{39}$$

where X_i is the i^{th} observation and $\bar{X}$ is the average of N observations, i.e.,

$$\bar{X} = \sum_{i=1}^{N} \frac{X_i}{N}. \tag{40}$$

In the case of a continuous observable x,

$$\sigma = \left[\int_{-\infty}^{\infty} (x - \bar{x})^2 \, p(x) \, dx \right]^{1/2}, \tag{41}$$

where

$$\bar{x} = \int_{-\infty}^{\infty} x p(x) dx, \tag{42}$$

and $p(x)$ is called the probability density function. For a normal distribution

$$p(x) = \frac{1}{\sqrt{2}\pi\sigma} e^{-x^2/2\sigma^2}. \tag{43}$$

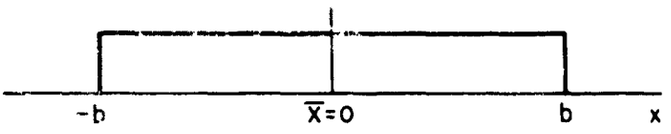
Knowledge of the standard deviation does not automatically give the probability, because it is necessary to know in addition the probability density function.

In many cases involving experimental data, observations are normally distributed, but not in all cases. In those cases where they are, it is possible to find the probability for any range from tabulated data.

One case where the data are clearly not normally distributed is that where the deviations from an average lie in some range -b to +b. A useful approach in that case is to assume that deviations from the mean are uniformly distributed, i.e., according to a rectangular distribution. One can then be sure that the computed standard deviation will be pessimistic and that any error estimates therefrom will be conservative. For example, consider the distribution in Figure A-31. This distribution is what is meant by a uniform distribution. Its standard deviation is

$$\sigma^2 = \int_{-b}^{b} x^2 \left(\frac{1}{2b}\right) dx = \frac{1}{3} b^2 \tag{44}$$

$$\sigma = \frac{b}{\sqrt{3}}$$



This distribution weights extreme deviations equally with lesser deviations. Although in practice large deviations tend to occur less frequently than small deviations, it may be safe to assume a uniform distribution in the absence of knowledge concerning the actual distribution of the data.

On the other hand, if errors from independent distributions are combined additively, the sum distribution quickly converges to a normal distribution according to the Central Limit Theorem. In fact, Shewhart has shown that this convergence takes place, within engineering accuracy, when errors from four or more distributions are added.[13]

PROPAGATION OF ERROR INDICES UNDER MATHEMATICAL TRANSFORMATION

We shall consider the laws for the propagation of error indices first for the cases of product or quotient, and finally for the general case.[14]

Sum or Difference

Let

$$\bar{U} = \bar{X} + \bar{Y}, \tag{45}$$

where $\bar{X}$ and $\bar{Y}$ are the means for two separate independent sets of observations with standard deviations $\sigma_{\bar{x}}$ and $\sigma_{\bar{y}}$ respectively.

The individual X and Y may be paired to give (X_1, Y_1), (X_2, Y_2) etc. Now if u_1, x_1, y_1, etc. are deviations from the means $\bar{U}$, $\bar{X}$, and $\bar{Y}$,

$$U_1 = \bar{U} + u_1 = X_1 + Y_1 = \bar{X} + x_1 + Y + y_1,$$

$$U_2 = \bar{U} + u_2 = X_2 + Y_2 = \bar{X} + x_2 + \bar{Y} + y_2, \tag{16}$$

and so forth. It follows that

$$u_1 = x_1 + y_1$$

$$u_2 = x_2 + y_2 \quad . \tag{47}$$

Moreover,

$$\sigma_{\bar{u}}^2 = \sum_{i=1}^{n} \frac{u_i^2}{n}, \tag{48}$$

where

$$\sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} x_i^2 + 2 \sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} y_i^2 .$$

Since, by hypothesis, errors x_i and y_i are independent,

$$2 \sum_{i=1}^{n} x_i y_i = 0 \tag{49}$$

in the limit as $n \to \infty$ with probability one. Therefore,

$$\sigma_{\bar{U}}^2 = \sigma_{\bar{X}}^2 + \sigma_{\bar{Y}}^2 . \tag{50}$$

If

$$\bar{U} = \bar{X} - \bar{Y}, \tag{51}$$

we have

$$\sum_{i=1}^{n} u_i^2 = \sum_{i=1}^{n} x_i^2 - 2 \sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{n} y_i^2 \tag{52}$$

and it is evident that Equation (50) is still applicable. Thus, Equation (50) is the law for sum or differences.

Product or Quotient

Let $\bar{U} = \bar{X}\ \bar{Y}$. (53)

Then, $U_1 = X_1\ Y_1$, etc., (54)

and $\bar{U} + u_1 = (\bar{X} + x_1)(\bar{Y} + y_1) = \overline{XY} + x_1 \bar{Y} + y_1 \bar{X} + x_1 y_1$, etc.. (55)

where the product $x_1 y_1$ is of the second order compared with the other terms, provided $\bar{x}$ and $\bar{y}$ are not small. Usually, x and y are independent of the amplitude. Hence,

$$u_1 \doteq x_1 \bar{Y} + \bar{y}_1 \bar{X}, \text{ etc.} \tag{56}$$

Now

$$\sum_{i=1}^{n} u_i^2 = \bar{Y}^2 \sum_{i=1}^{n} x_i^2 + 2\bar{X}\bar{Y} \sum_{i=1}^{n} x_i y_i + \bar{X}^2 \sum_{i=1}^{n} y_1^2. \tag{57}$$

As above

$$\sum_{i=1}^{n} x_i y_i = 0,$$

in the limit as $n \to \infty$ with probability one, so that

$$\sigma_{\bar{U}}^2 = \sum_{i=1}^{n} \frac{u_i^2}{n} = \bar{Y}^2 \sum_{i=1}^{n} \frac{x_i^2}{n} + \bar{X}^2 \sum_{i=1}^{n} \frac{y_i^2}{n} = \bar{Y}^2 \sigma_{\bar{X}}^2 + \bar{X}^2 \sigma_{\bar{Y}}^2 , \tag{58}$$

and

$$\frac{\sigma_{\bar{U}}^2}{\bar{U}^2} = \frac{\sigma_{\bar{X}}^2}{\bar{X}^2} + \frac{\sigma_{\bar{Y}}^2}{\bar{Y}^2} . \tag{59}$$

Thus Equations (58) and (59) is the law for the propagation of error indices in the case of products. It is also the law in the case of quotients, but before showing this we shall obtain the result for the general case.

General Case

Let

$$\bar{U} = f(\bar{X}, \bar{Y}) \tag{60}$$

Expanding Equation (60) in Taylor Series, we have

$$\bar{U} + u_1 = f(\bar{X} + x_1, \bar{Y} + y_1) = f(\bar{X}, \bar{Y}) + \left(\frac{\partial U}{\partial X}\right)_{U=\bar{U}} x_1 + \left(\frac{\partial U}{\partial Y}\right)_{U=\bar{U}} y_1 + \cdots \tag{61}$$

and

$$u_1 = \left(\frac{\partial U}{\partial X}\right)_{U=\bar{U}} x_1 + \left(\frac{\partial U}{\partial Y}\right)_{U=\bar{U}} y_1 + \cdots \tag{62}$$

$$\sum_{i=1}^{n} U_i^2 = \left(\frac{\partial U}{\partial X}\right)_{U=\bar{U}} \sum_{i=1}^{n} x_i^2 + 2 \left(\frac{\partial U}{\partial X}\right)_{U=\bar{U}} \cdot \left(\frac{\partial U}{\partial Y}\right)_{U=\bar{U}} \sum_{i=1}^{n} x_i y_i + \left(\frac{\partial U}{\partial Y}\right)_{U=\bar{U}}^2 \sum_{i=1}^{n} y_i^2 + \cdots \tag{63}$$

Therefore,

$$\sigma_U^2 = \left(\frac{\partial U}{\partial X}\right)^2_{U=\bar{U}} \sigma_{\bar{X}}^2 + \left(\frac{\partial U}{\partial Y}\right)^2_{U=\bar{U}} \sigma_{\bar{Y}}^2 \qquad (64)$$

Thus, Equation (64) or an extension of it is the law for the propagation of error indices in the general case.

From Equation (64) we can readily see that Equation (59) is valid for quotients as well as products. Given that

$$\bar{U} = \frac{\bar{X}}{\bar{Y}}$$

we find $\frac{\sigma^2}{\bar{U}}$ by means of Equation (64):

$$\left(\frac{\partial U}{\partial X}\right)_{U=\bar{U}} = \frac{1}{\bar{Y}}$$

and

$$\left(\frac{\partial U}{\partial Y}\right)_{U=\bar{U}} = \frac{-\bar{X}}{\bar{Y}^2} ,$$

so that

$$\sigma_{\bar{U}}^2 = \frac{\sigma_{\bar{X}}^2}{\bar{Y}^2} + \frac{\bar{X}^2}{\bar{Y}^4} \sigma_{\bar{Y}}^2 .$$

Therefore,

$$\frac{\sigma_{\bar{U}}^2}{\bar{U}^2} = \frac{\sigma_{\bar{X}}^2}{\bar{X}^2} + \frac{\sigma_{\bar{Y}}^2}{\bar{Y}^2} ,$$

which checks.

The above results may be readily extended to more than two variates.

Precision Index of the Arithmetic Mean

If communication channels are operated in parallel and if the statistical fluctuations in the data caused by the noise in each channel are independent, a reduction in the precision index for each channel may be achieved by averaging the output from the channels. The result is the well known standard deviation in the mean

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}} \qquad (65)$$

where $\bar{X}$ is the arithmetic mean and σ_x is the standard deviation for each of n similar channels.

PRECISION INDEX OF A NETWORK

The preceding discussion has suggested a method for determining the precision index at the output of a single link communication system, knowing the precision of the measurement data going into the link. In discussing the propagation of errors through a network, it is important to note that the techniques for network analysis are not limited by the manner in which the link precision is found or by its value. Furthermore, it should be emphasized at this point that the concept of precision has two facets and that one cannot be considered without the other; namely, a specified deviation range and the probability that an observation will lie in that range. Hence, when taking up the question of the overall precision for a network, we must consider the output deviation range and the probability of the output lying within that range.

If the links are connected in series, the precision index for the chain is found by whichever law of combination is appropriate. For example, if each succeeding link linearly processes the data from the preceding link, the rule indicated by Equation (50) is used. That is, the precision indices are combined according to the sum of the squares, so that if all links in a chain of m links have the same precision index σ, the resultant precision index is $\sigma\sqrt{m}$.

If links are connected in parallel and process the same data, fed to a common sink, Equation (65) applies, according to which the resultant precision index for m links in parallel is $\sigma/\sqrt{m}$.

If we know the probability density function for the output error, we can find the probability that a datum will lie in a range that is some multiple of the output precision index. It must be emphasized that knowledge of the output probability density function is required before the probability and thus the precision can be found. However, the properties of the Gaussian process can be of assistance here; namely, that if the errors in individual links are Gaussianly distributed, the errors in the output of series or parallel-connected networks are also Gaussianly distributed. As noted previously, even if the errors may not be Gaussianly distributed, we can assume that they are in the knowledge that our resulting estimates of precision will thereby be conservative.

We shall now apply the above principles of error propagation to two cases of practical significance.

Propagation of Computer Errors

Suppose that inputs to a computer may each introduce an error ϵ_1 having one of three discrete values: 0, + 0.0005, and + 0.001. Suppose also that each of the inputs introduces these errors with equal likelihood, so that each error will have a probability weighting of 1/3. This will make the expectation of error, i.e., the average value, equal to

$$\bar{x} = \frac{1}{3}(0) + \frac{1}{3}(0.0005) + \frac{1}{3}(0.001) = 0.0005,$$

which is, as it were, a center of gravity for the discrete input errors. It is, however, the fluctuation from the average that determines the uncompensatable portion of the error, so that we need to find the index which is a measure of this fluctuation, namely, the standard deviation. It is

$$\sigma_i = \left[\sum_{i=1}^{3} \frac{(x_i - \bar{X})^2}{3}\right]^{1/2} = \left[\frac{(0 - 0.0005)^2 + (0.0005 - 0.0005)^2 + (0.001 - 0.0005)^2}{3}\right]^{1/2} = 4.1 \times 10^{-4}.$$

This σ_i is now taken as the index of each input error.

If internal errors were not also introduced by less than perfect operation of computer components, we could use a general relation such as Equation (64) to get the resultant computer error, but in this case multipliers and amplifiers both introduce errors, so that we must proceed in a step-by-step manner instead. We note that each multiplier introduces an error ϵ_2 which varies randomly between +0.001 and -0.001. We shall be conservative and take this to mean that the multiplier error has a uniform distribution between these limits, so that we can use relation, Equation (44), to get the index for it, i.e.,

$$\sigma_m = \frac{b_m}{\sqrt{3}} = \frac{0.001}{\sqrt{3}} = 5.77 \times 10^{-4}.$$

Additionally, the operational amplifiers may each introduce an error which varies randomly between +0.0005 and -0.0005. Again we assume a uniform distribution, so that the standard deviation for each amplifier is

$$\sigma_a = \frac{b_a}{\sqrt{3}} = \frac{0.0005}{\sqrt{3}} = 2.88 \times 10^{-4}.$$

Thus, the three precision indices are:

$$\sigma_i = 4.1 \times 10^{-4},$$

$$\sigma_m = 5.8 \times 10^{-4},$$

$$\sigma_a = 2.9 \times 10^{-4}.$$

Computation of Resultant Computer Error

We now take the relation for the computer output and proceed with the calculation of its output error. The output relation for the computer is

$$\Delta d = \Delta r \frac{[Y_1 Y_2 - X_1 X_2 + Z_1 Z_2]X_1}{[X_1 Y_2 + X_2 Y_1] Z_1} , \quad (66)$$

where

$$Z = \sqrt{X_1^2 + Y_1^2} \quad \text{and} \quad Z_2 = \sqrt{X_2^2 + Y_2^2} .$$

Set $\Delta W = \Delta d / \Delta r$, since Δr is exact, so that

$$\Delta W = \frac{[Y_1 Y_2 - X_1 X_2 + Z_1 Z_2] X_1}{[X_1 Y_2 + X_2 Y_1] Z_1} .$$

If each multiplier were perfect, the output standard deviation in terms of the input standard deviation would be, by Equation (58), simply

$$\begin{aligned}
\sigma^2_{\bar{Y}_1 \bar{Y}_2} &= \bar{Y}_2\ \sigma_i^2 + \bar{Y}_1^2\ \sigma_i^2 = \sigma_i^2\ (\bar{Y}_1^2 + \bar{Y}_2^2), \\
\sigma^2_{\bar{X}_1 \bar{X}_2} &= \sigma_i^2\ (\bar{X}_1^2 + \bar{X}_2^2), \\
\sigma^2_{\bar{X}_1 \bar{Y}_2} &= \sigma_i^2\ (\bar{X}_1^2 + \bar{Y}_2^2), \\
\sigma^2_{\bar{X}_2 \bar{Y}_1} &= \sigma_i^2\ (\bar{X}_2^2 + \bar{Y}_1^2),
\end{aligned} \tag{67}$$

where $\bar{X}_1^2$, $\bar{X}_2^2$, $\bar{Y}_1^2$, and $\bar{Y}_2^2$ are the average values of the inputs squared. The median values of their ranges would be appropriate to use for the average. Since the multipliers are not perfect, however, they introduce an error in operation, and this error in each case has the index σ_m, as discussed above. In general, the operating error and the propagated errors are independent, so that the resultant standard deviation for the two of them combined is the square root of the sum of their standard deviations squared. That is

$$\sigma^2_{r\bar{Y}_1 \bar{Y}_2} = \sigma^2_{\bar{Y}_1 \bar{Y}_2} + \sigma_m^2, \tag{68}$$

and likewise for the other multipliers.

Moreover, since the amplifiers introduce errors of their own as well as propagating those errors which are introduced into them, we must, when adding or subtracting, account for this error also. We shall assume it to be independent of the propagated error. Hence, the index for the sum $Y_1 Y_2 - X_1 X_2$ becomes

$$\sigma^2_{sum} = \sigma^2_{\bar{Y}_1 \bar{Y}_2} + \sigma^2_{\bar{X}_1 \bar{X}_2} + 2\sigma_m^2 + \sigma_a^2. \tag{69}$$

Similarly, for the bracket in the denominator, we have

$$\frac{\sigma^2}{[\ ]\ denom.} = \sigma^2_{\bar{X}_1 \bar{Y}_2} + \sigma^2_{\bar{X}_2 \bar{Y}_1} + 2\sigma_m^2 + \sigma_a^2. \tag{70}$$

We now address ourselves to obtaining the indices for Z_1, and Z_2 where,

$$Z_1 = (X_1^2 + Y_1^2)^{1/2}$$

and

$$Z_2 = (X_2^2 + Y_2^2)^{1/2}.$$

To obtain the error indices for these, we use the general Equation (64) and proceed as follows:

$$\sigma^2_{Z_1} = \left(\frac{\partial Z_1}{\partial X_1}\right)^2_{Z_1=\bar{Z}_1} \sigma^2_{\bar{X}_1} + \left(\frac{\partial Z_2}{\partial Y_1}\right)^2_{Z_1=\bar{Z}_1} \sigma^2_{\bar{Y}_1},$$

where

$$\left(\frac{\partial Z_1}{\partial X_1}\right) = \frac{X_1}{(X_1^2 + Y_1^2)^{1/2}}$$

and

$$\left(\frac{\partial Z_1}{\partial Y_1}\right) = \frac{Y_1}{(X_1^2 + Y_1^2)^{1/2}} .$$

Hence,

$$\sigma^2_{Z_1} = \frac{\bar{X}_1^2 \sigma^2_{\bar{X}_1} + \bar{Y}_1^2 \sigma^2_{\bar{Y}_1}}{(\bar{X}_1^2 + \bar{Y}_1^2)}, \tag{71}$$

where

$$\sigma^2_{\bar{X}_1} = \sigma^2_{\bar{Y}_1} = \sigma_i^2 .$$

Similarly,

$$\sigma^2_{Z_2} = \frac{\bar{X}_2^2 \sigma^2_{\bar{X}_2} + \bar{Y}_2^2 \sigma^2_{\bar{Y}_2}}{(\bar{X}_2^2 + \bar{Y}_2^2)} . \tag{72}$$

The above two results, $\sigma^2_{Z_1}$ and $\sigma^2_{Z_2}$, do not, however, take account of errors in the amplifiers used to form the sum $(\bar{X}_1^2 + \bar{Y}_1^2)$ and $(\bar{X}_2^2 + \bar{Y}_2^2)$. Hence, the complete error index in each case is

$$\sigma^2_{Z_1} + \sigma_a^2$$

and

$$\sigma^2_{Z_2} + \sigma_a^2 .$$

The error index for the produce Z_1 Z_2 is

$$\sigma^2_{\bar{Z}_1\bar{Z}_2} = \bar{Z}_2^2 \sigma^2_{\bar{Z}_1} + \bar{Z}_1^2 \sigma^2_{\bar{Z}_2} . \tag{73}$$

To this must be added the error index for the multiplier, so that

$$\sigma^2_{r\bar{Z}_1\bar{Z}_2} = \sigma^2_{\bar{Z}_1\bar{Z}_2} + \sigma_m^2 .$$

We are now in a position to combine and consolidate all the previous results to obtain the resultant computer error index. The error index for the square bracket in the numerator,

$$[\ ] = [Y_1 Y_2 - X_1 X_2 + Z_1 Z_2],$$

is

$$\frac{\sigma^2}{[\ ]} = \sigma^2_{\bar{Y}_1 \bar{Y}_2} + \sigma^2_{\bar{X}_1 \bar{X}_2} + \sigma^2_{\bar{Z}_1 \bar{Z}_2} + 3\sigma^2_m + 2\sigma^2_a . \tag{74}$$

For X_1 the error index is σ_1^2. Hence for the entire numerator

$$N = [Y_1 Y_2 - X_1 X_2 + Z_1 Z_2] X_1 ,$$

we have the error index

$$\sigma_N^2 = \bar{X}_1^2 \frac{\sigma^2}{[\ ]} + \overline{[\ ]}^2 \sigma_i^2 . \tag{75}$$

For the error index in the denominator

$$D = [X_1 Y_2 + Y_2 Y_1] Z_1 ,$$

we have

$$\sigma^2_{\bar{D}} = \bar{Z}_1^2 \frac{\sigma^2}{[\ ]\ \text{denom.}} + \frac{2}{[\ ]\ \text{denom.}} \sigma^2_{\bar{Z}_1} . \tag{76}$$

The total resultant computer error index is thus the error index for the numerator over the denominator, i.e.,

$$\Delta W = \frac{N}{D} ,$$

or

$$\sigma^2_{\frac{\Delta W}{}} = \bar{D}^2 \sigma^2_{\bar{N}} + \bar{N}^2 \sigma^2_{\bar{D}} . \tag{77}$$

Thus Equation (77) the terms of which can be found from the previous steps, is the relation to use to obtain the resultant computer error.

The final result cannot be given in numerical form until the average or median values of the inputs X, Y, X and Y are specified. Once they are given, however, computation of the numerical result is straightforward, following the detailed steps given above.

It is emphasized that use of a probable error index instead of the standard deviation is justifiable only in the event that the error data are known to be normally distributed to a good approximation. If they are, then it is valid to substitute probable error indices in place of the standard deviation indices throughout. In that event one can assertthat the probability is 50 percent that the resultant computer error will not lie outside the probable error range.

PRECISION INDEX FOR PCM SYSTEM

Our next illustration is concerned with finding the precision index for a PCM communication system.[5] By sampling, quantization and coding, an analog waveform may be represented as a set of discrete pulses which may be transmitted over a noisy communication channel as PCM. The pulses normally have equal amplitudes and widths, and the

individual digit error probability, after decoding at the receiver, is the same for all pulse positions in the binary sequence. The positions of the pulses in the sequence, however, have varying degrees of importance with respect to the fidelity of the reconstructed analog waveform. That is, for binary numbers the value of the i^{th} pulse in a sequence is 2^{i-1}, which is seen to increase by powers of two as the order i increases. The error resulting from the insertion or deletion of a pulse in the i^{th} position grows correspondingly. Therefore, from a fidelity viewpoint, it is desirable to modify the PCM process such that account is taken of pulse position.

In conducting the error analysis, we shall assume a binary symmetric PCM channel, perturbed by additive Gaussian noise with coherent detection on a bit basis. The average error probability in the detection of the individual pulses in the PCM train is a function of both pulse amplitude and rms noise and is determined by

$$p_i = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{a_i}^{\infty} e^{-n^2/2\sigma^2}\, dn, \tag{78}$$

where p_i is the error probability for the i^{th}-position pulse in a group, n is the noise voltage, and σ is the rms noise voltage. In ordinary, unweighted PCM, each pulse is transmitted at the same amplitude, and, therefore, all the p_i's are equal, $i = 1, 2, \cdots n$, in a code word.

The probability of a single error in the i^{th} position of an n-digit word is

$$p_i(1, n) = p_i \prod_{\substack{j=1 \\ j \neq 1}}^{n} (1 - p_j) \approx p_i \tag{79}$$

for small digit error probabilities. For unweighted PCM, $p_i = p$ for all i and thus

$$p_i(1, n) = p. \tag{80}$$

The effect of these digit errors on the original analog signal can be expressed in terms of the rms error σ_u.

The mean-squared analog error is

$$\sigma_u^2 = p \sum_{i=1}^{n} e_i^2, \tag{81}$$

where

$$e_i = \pm q^{2^{i-1}}$$

is the analog error of the i^{th} order pulse, q being the amplitude of the quantization interval, as is shown in (5). Hence,

$$\sigma_u^2 = q^2 p \sum_{i=1}^{n} 4^{i-1} \tag{83}$$

$$= q^2 p \frac{4^n - 1}{3}. \tag{84}$$

The error given by Equation (84) is that fraction of the total error due to channel noise. The total error includes the original measurement error, the channel noise error and the error in the quantization of the original analog signal. If e_q is the error voltage in the quantized equivalent of the actual instantaneous signal, the mean-squared analog error due to quantization is

$$\sigma_q^2 = \frac{1}{q} \int_{-q/2}^{q/2} e_q^2 \, de_q \tag{85}$$

and the rms quantization error is

$$\sigma_q = q/2\sqrt{3}. \tag{86}$$

Hence, the total rms analog error of a pulse group is

$$\sigma_{tu} = (\sigma_A^2 + \sigma_u^2 + \sigma_q^2)^{1/2}, \tag{87}$$

assuming that the original measurement error, the channel noise error, and the quantization error combine in a linearly additive manner. This assumption is valid for channels with additive Gaussian noise.

By way of further illustration, in a series-connected network for m-similar links the resultant precision index for an unweighted PCM system is found from Equation (50). The final index is

$$\sigma_{mtu} = \left[\sigma_A^2 + m(\sigma_u^2 + \sigma_q^2)\right]^{1/2}. \tag{88}$$

Notice that σ_A^2 occurs only once since it is a measure of the original precision of measurement, transmitted through relays to a final destination.

For a parallel-connected network of m-similar, statistically independent links the resultant precision for an unweighted PCM system is from Equation (65):

$$\sigma_{mtu} = \left[\left(\sigma_A^2 + \frac{\sigma_u^2 + \sigma_q^2}{m}\right)\right]^{1/2}. \tag{89}$$

REFERENCES

1. Habib, E. J., "Data Processing Portion of the Tracking and Data Acquisition Program Review," (February 29, 1964).

2. Shewhart, W. A., "Statistical Method From the Viewpoint of Quality Control," Ed. by W. Edwards Deming, The Graduate School of The Department of Agriculture, Washington, (1939).

3. Hoel, P. G., Introduction to Mathematical Statistics, 3rd Ed., John Wiley and Sons, Inc., New York, (1962).

4. Schwartz, L. S., "Discussion of Telemetry Performance Standards," Tech. Memo. No. 40 on New York University Contract NAS 5-3508, (June 1964).

5. Hauptschein, A., "Weighted PCM," Tech. Memo No. 16 on New York University Contract NAS 5-408, October 15, (1961).

6. Von Alven, W. H., Ed., Reliability Engineering, Prentice-Hall, Inc., (1964), pages 468-491.

7. Demmerle, A., McCeney, P. J. and Karras, T. J., Time Decoding for Satellite Tracking Systems.

8. Coates, R., Creveling, C., Habib, E., Mahoney, M., and Stout, C., Telemetry Data Processing for the Eccentric Orbiting Geophysical Observatory Satellite, Goddard Space Flight Center, Greenbelt, Maryland, X-545-63-212, October 1963.

9. Sandler, G. H., Systems Reliability Engineering, Prentice-Hall, Inc. Englewood Cliffs, N. J., (1963), pages 112-127.

10. Flagle, C. D., Huggins, W. H., and Roy, R. H., Operations Research and Systems Engineering, The Johns Hopkins Press, Baltimore, (1960), pages 609-684.

11. Huggins, W. H., "Signal-Flow Graphs and Random Signals, "Proc. IRE, 45, No. 1, January, 1957.

12. Feller, W. An Introduction to Probability Theory and Its Applications, 2nd ed.; New York: John Wiley and Sons (1957) Chaps. 11, 12, 13.

13. Shewhart, W. A., The Economic Control of Quality of Manufactured Product, D. Van Nostrand Co., Inc., New York, N. Y., (1931).

14. Wortling, A. G. and Geffner, J., Treatment of Experimental Data, John Wiley and Sons, Inc., New York, (1943).

SOME PROBABILISTIC MODELS AND STATISTICAL METHODS IN THE ANALYSIS OF TELEMETRY DATA

C. A. Rohde and R. D. Coleman

INTRODUCTION

We consider the processing of a digital tape consisting of a sequence of data words and sync words, each in turn consisting of zeros and ones. Between each M data words a sync word is placed consisting of m bits. The purpose of the sync word is to detect errors in the processing line, e.g., bit slippage. The sync word pattern transmitted from the satellite consists of a pre-set sequence of zeros and ones. Due to noise, bit slippage or malfunction of the telemetry system, analogue to digital conversion unit or data processing line, errors are introduced which alter the sync word. Thus proper recognition of the sync word, which serves as a guide post to the position of the data words, is an important problem. The current procedures utilize the autocorrelation (simply a count of the number of mismatches or errors) of the pre-set sync word pattern with the transmitted sync word pattern. Clearly a high autocorrelation indicates low noise and/or proper hardware performance while a low autocorrelation indicates high noise or more probably loss of the sync word pattern due to bit slippage or other hardware malfunction.

SYNC DETERMINATION

In order to detect sync loss, whatever its cause, two strategies are currently under consideration:

Strategy 1: Examine each sync word and make the decision that the line is out of sync when k_1 or more errors occur in a sync word.

Strategy 2: Examine each sync word and make the decision that the line is out of sync when k_2 or more errors occur for r sync words in a row.

The problem of investigating the optimality of the above strategies is being investigated using simulation methods by Goddard personnel and has been theoretically investigated by Hauptschein (3).

ASSUMPTIONS

To investigate relevant properties of Strategies one and two, certain assumptions about the behavior of the bit probabilities of error will be made. These assumptions are commonly made (3,5) but are perhaps not realistic (4). It is hoped that the procedures to be developed on page A-148 are reasonably robust or conservative enough so that violation of the assumptions is not disastrous. On page A-149 some comments will be made on possible effects of violation of the assumptions.

We assume that the bit stream is such that errors occur in an independent or random fashion. In probabilistic terms such an assumption simply means that the probability of an error in both the i^{th} bit and in the j^{th} bit is the product of the probability of an error in the i^{th} bit and the probability of an error in the j^{th} bit. In addition we assume that the probability of error per bit, say p_e, is constant throughout the tape. Since the

signal to noise ratio of the transmitted signal directly influences the bit probability of error this assumption is equivalent to the assumption of a constant signal to noise ratio.

The assumptions made above imply that the bit stream is a set of Bernoulli trials (1) with probability of success (observation of an error) given by p_e. It also follows that the probability of observing x errors in the sync word is

$$P[x \text{ errors in sync word}] = \binom{m}{x} p_e^x (1-p_e)^{m-x}.$$

DISTRIBUTIONS UNDER STRATEGY ONE

Under the assumptions of the previous section and the procedure of Strategy one the sequence of sync word correlations constitute a set of Bernoulli trials with probability of success (detection of k_1, or more errors)

$$p_1 = \sum_{r=k_1}^{m} \binom{m}{r} p_e^r (1-p_e)^{m-r}.$$

A trial in this context is the observation of the number of errors in the transmitted sync word pattern correlated with the pre-set sync word pattern. Consider the distribution of the random variable T, where T is defined as the number of sync words tested until the decision that the line has lost sync is made. Since $T = t$ if and only if we observe a sequence of t failures (no sync loss) followed by a success it follows that

$$P[T=t] = P[1^{st}\ t \text{ trials result in failure, } (t+1)^{st} \text{ trial results in success}]$$

$$= P[1^{st}\ t \text{ trials result in failure}]\ P[(t+1)^{st} \text{ trial results in success}]$$

$$= (1-p)^t p$$

for $t = 0, 1, 2, \cdots$

The random variable T thus has the geometric distribution [Feller (1) page 156] with parameter p. In terms of measuring the quality of a tape it is clear that some information is contained in the length of time to sync loss. In particular a small value of such a parameter is indicative of a noisy tape. A reasonable estimate for the parameter time to sync loss is simply the average of the times to sync loss which occur during the processing of the tape. To discuss the properties of such an estimate we need to determine its probability distribution.

If we let $T_1, T_2, \cdots T_N$ be the times to sync loss we seek the distribution of the random variable

$$X = \sum_{i=1}^{N} T_i$$

where the T_i ($i = 1, 2, \cdots, N$) are independent random variable with a common geometric distribution. The probability distribution of X is most easily found using the

probability generating function (the Z-transfer of sequence of probabilities with Z replaced by Z^{-1}). The probability generating function of T_i is

$$\pi_{T_i}(z) = \sum_{t_i=0}^{\infty} z^{t_i} (1-p)^{t_i} p = p\,[1-(1-p)Z]^{-1}$$

for $|Z| < 1$. Since the probability generating function of the sum of N independent and identically distributed random variables is $[\pi(z)]^N$ (Feller (1) page 251), where $\pi(z)$ is the common probability generating function, it follows that the probability generating function of X is given by

$$\pi_x(z) = p^N\,[1-(1-p)z]^{-1}.$$

It is well known that the probability generating function uniquely determines the probability distribution so that determination of the probability generating function of X will suffice to determine the distribution of X. The probability distribution

$$f(x) = \binom{-N}{x} p^N\,[-(1-p)]^x;\ x = 0,\ 1,\ 2,\ \cdots$$

has probability generating function

$$\pi(z) = \sum_{x=0}^{\infty} \binom{-N}{x} p^N\,[-(1-p)]^x\, z^x$$

$$= p^N \sum_{x=0}^{\infty} \binom{-N}{x} [-z(1-p)]^x$$

$$= p^N\,[1-(1-p)z]^{-N}.$$

It follows that the distribution of X is (t)

$$f(x) = \binom{-N}{x} p^N\,[-(1-p)]^x;\ x = 0,\ 1,\ 2,\ \cdots$$

The distribution (t) is called the negative binomial distribution. Thus the sum of the times to sync loss has a negative binomial distribution with index N and mean $N(1-p)/p$. On page A-147 we shall indicate how one could use the above result to estimate P_e and also to construct a confidence interval for P_e.

In certain investigations the experimenter might consider using the total number of accumulated errors to estimate the bit probability of error, p_e. Let $N_1, N_2, \cdots, N_{t+1}$ be random variables representing the number of errors in the sync word. Here $t+1$ is the number of the sync word where k_1 or more errors first appear. The quantity of interest is then

$$N^* = \sum_{i=1}^{t} N_i.$$

It is important to note that because of the definitions of $N_1, N_2, \cdots, N_t$ we have

$$N_1 < k_1, N_2 < k_1, \ldots, N_t < k_1.$$

The distribution of N_i $(i = 1, 2, \ldots, t)$ is thus

$$f(n_i) = \frac{\binom{m}{n_i} p_e^{n_i} (1-p_e)^{m-n_i}}{1 - p(m)}; \quad n_i = 0, 1, \cdots, k_1 - 1.$$

where

$$p(m) = \sum_{r=k_1}^{m} \binom{m}{r} p_e^r (1-p_e)^{m-r}.$$

The distribution of N_i is a truncated binomial distribution; the truncation resulting from the strategy which stops as soon as k_1 or more errors are observed in a sync word. From the above expression it is possible to obtain the probability generating function of N_i and hence the probability generating function of N^*, the convolution of the N_i. The resulting probability generating function could be inverted (Fisz (2) page 119) to obtain the probability distribution of N^*. The computations involved would probably be complicated due to the truncation involved in the distributions of the N_i so some sort of approximation might have to be used. For many purposes, however, knowledge of the mean and variance of N^* might provide sufficient information. To illustrate the calculations consider the expected value of N^*.

By definition the expected value of N_i is

$$\mathcal{E}[N_i] = [1-p(m)]^{-1} \sum_{n_1=0}^{k_1-1} n_i \binom{m}{n_i} p_e^{n_i} (1-p_e)^{m-n_i}$$

$$= \frac{m p_e}{1-p(m)} \sum_{n_1'=0}^{k_1-2} \frac{(m-1)!}{n_1'!\,(m-1-n_i')!} p_e^{n_1'} (1-p_e)^{m-n_i'-1}$$

$$= \frac{m p_e}{1-p(m)} \left[1 - p(m-1) - \binom{m-1}{k_1-1} p_e^{k_1-1} (1-p_e)^{m-k_1}\right].$$

It follows that

$$\mathcal{E}[N^*] = t\,\mathcal{E}[N_i] = \left[\frac{t m p_e}{1-p(m)}\right] \left[1 - p(m-1) - \binom{m-1}{k_1-1} p_e^{k_1-1} (1-p_e)^{m-k_1}\right].$$

It is clear from the expression for $\mathcal{E}[N^*]$ that the "natural estimate" of p_e, i.e.

$$p_e = \frac{\text{total number of errors in sync words}}{\text{total number of bits in sync words}}$$

is a biased estimate of p_e. In particular such an estimate underestimates, on the average, the value of p_e.

DISTRIBUTIONS UNDER STRATEGY TWO

Under the previously mentioned assumptions we can consider the sequence of sync word tests as a set of Bernoulli trials with parameter p_2 where

$$p_2 = P[k_2 \text{ or more errors observed in sync word}]$$

$$= \sum_{r=k_2}^{m} \binom{m}{r} p_e^r (1-p_e)^{m-r}.$$

In this case a trial consists of observing the number of errors in the sync word and calling the result a success if k_2 or more errors are observed.

According to Strategy Two, sync loss is declared when ℓ sync words in succession have k_2 or more errors, i.e., ℓ successes in a row have occurred. Letting $\mathcal{E}$ denote this event it follows that $\mathcal{E}$ is a recurrent event (Feller (1) Chapter 13). In particular $\mathcal{E}$ is a success run of length ℓ. To determine the probability distribution of the number of trials until $\mathcal{E}$ occurs we follow the development of Feller (pages 299-301). Let u_n be the probability that $\mathcal{E}$ occurs on the n^{th} trial (n^{th} sync word test) and let f_n be the probability that $\mathcal{E}$ occurs for the first time on the n^{th} trial. The probability that the ℓ sync words numbered n, n-1, $\cdots$, n - ℓ + 1 each result in success is p_e^{ℓ}. Note that

$$P[\mathcal{E} \text{ occurs at trial } (n-k) \text{ and successes occur on trials } n-k+1, \cdots, n]$$

$$= P[\mathcal{E} \text{ occurs at trial } (n-k)]\ P[\text{successes occur on trials } n-k+1, \cdots, n]$$

$$= u_{n-k}\ p_2^k.$$

Thus we have

$$P[\text{success on trials } n, n-1, \cdots, n-\ell+1]$$

$$= P[\mathcal{E} \text{ occurs on trial } n] + P[\mathcal{E} \text{ occurs on trial } n-1, \text{ trial } n \text{ is a success}]$$

$$+ \cdots\cdots + P[\mathcal{E} \text{ occurs on trial } n-\ell+1, \text{ trials } n-\ell+2, \cdots, n \text{ are successes}]$$

$$= u_n + p_2 u_{n-1} + \cdots\cdots + p_2^{\ell-1} u_{n-\ell+1}$$

for $n \geq \ell$. It is clear that $u_1 = u_2 = \cdots = u_{\ell-1} = 0$ and for simplicity we assume $u_0 = 1$. Multiplying both sides of the equation

$$p_2^{\ell} = u_n + p_2 u_{n-1} + \cdots + p_2^{\ell-1} u_{n-\ell+1}$$

by z^n yields

$$z^n p_2^{\ell} = z^n u_n + z p_2 \left(z^{n-1} u_{n-1}\right) + \cdots + z^{\ell-1} p_2^{\ell-1} \left(z^{n-\ell+1} u_{n-\ell+1}\right)$$

Hence

$$p_2^{\ell} \sum_{n=\ell}^{\infty} z^n = \sum_{n=\ell}^{\infty} z^n u_n + zp \sum_{n=\ell}^{\infty} z^{n-1} u_{n-1} + \cdots + (zp)^{\ell-1} \sum_{n=\ell}^{\infty} z^{n-\ell+1} u_{n-\ell+1}.$$

Since $u_1 = u_2 = \cdots = u_{\ell-1} = 0$ the above equation becomes

$$p_2^{\ell} z^{\ell} (1-z)^{-1} = [u(z)-1]\left[1 + zp_2 + \cdots + (zp)^{\ell-1}\right]$$

when

$$u(z) = \sum_{n=0}^{\infty} u_n z^n .$$

Thus

$$u(z) - 1 = \frac{p_2^{\ell} z^{\ell} (1-zp)}{(1-z)\,[1-(zp)^{\ell}]}$$

and

$$u(z) = \frac{1 - z + z^{\ell+1} p^{\ell} (1-p)}{(1-z)\,[1-(zp)^{\ell}]} .$$

If

$$F(z) = \sum_{n=0}^{\infty} f_n z^n$$

is the probability generating function of the f_n sequence Feller (1) page 286 shows that

$$F(z) = \frac{u(t) - 1}{u(z)} .$$

It follows that the probability generating function of the time to sync loss under Strategy Two is

$$F(z) = \frac{(1-zp)\, p_2^{\ell} z^{\ell}}{1 - z + z^{\ell+1} p_2^{\ell} (1-p_2)}.$$

From the generating function one could find the probability distribution of the time to sync loss by inversion of the generating function or approximations using partial fraction expansions (Feller (1) page 257). The mean and variance of time to sync loss can easily be obtained from the above generating function and can be found to be (Feller (1) page 300)

$$\mu = \frac{1 - p_2^{\ell}}{(1-p_2)\, p_2^{\ell}}$$

(*)

$$\sigma^2 = \frac{1}{[(1-p_2)\, p_2^{\ell}]^2} - \frac{2\ell + 1}{(1-p_2)\, p_2^{\ell}} - \frac{p_2}{(1-p_2)^2}.$$

If we let y_M denote the number of times sync is lost in M frames it can be shown (Feller (1) page 300) that y_M is asymptotically normal with mean and variance given by

$$n\mu \text{ and } \frac{\sigma^2 n}{\mu^3}$$

where μ and σ^2 are as in (*). The asymptotic normality of y_M could be used to construct sequential procedures similar to those to be discussed on page A-148.

The distribution of the total number of sync word errors, due to the complexity of the conditioning event, as defined by the strategy, appears to be complicated and further analysis, perhaps leading to adequate approximations, is indicated.

A BRIEF SURVEY OF SEQUENTIAL TESTING

Having indicated the types of probabilistic models one can set up under the previously mentioned assumptions we shall now consider the important practical problem of using statistical methods to indicate departures from the theoretical models. The procedure we shall discuss will be sequential in nature although non-sequential problems of estimation of probability of error need further consideration. In this section we present the basic concepts of sequential analysis. A more detailed account can be found in the books by Wald (6) and Fisz (2).

One essential problem in sequential analysis is that of testing hypotheses. Thus suppose we wish to test the hypothesis that a parameter θ has the value θ_0 versus the alternative that θ has the value θ_1. We have available a sample (assumed random) $X_1, X_2, \cdots, X_n$ from a population having probability density function $f_\theta(x)$. Consider the random variable

$$S_n = \sum_{i=1}^{n} z_i$$

where

$$z_i = \log \left[\frac{f_{\theta_1}(x_i)}{f_{\theta_0}(x_i)} \right].$$

Under modest assumptions (existence of the expected value of e^{hz_i} for all real h and an identifiability condition on f_θ) it can be shown that the optimum sequential test (in the sense of leading to a minimum sample size subject to fixed probabilities of error of the first and second kind) is given by the sequential probability ratio test (SPRT). The SPRT is given by

Reject the hypothesis $\theta = \theta_0$ if $\sum_{i=1}^{n} z_i \geq \log A$

Continue sampling if $\log B < \sum_{i=1}^{n} z_i < \log A$

Accept the hypothesis $\theta = \theta_0$ if $\sum_{i=1}^{n} z_i \leq \log B$.

The numbers A and B are given (to a sufficiently close approximation) by the equations

$$A = \frac{1-\beta}{\alpha}, \quad B = \frac{\beta}{1-\alpha}$$

where α and β are the (pre-set) probabilities of the first and second kinds of error. An error of the first kind (Type I error) is said to occur if the test procedure rejects the hypothesis when it is true. An error of the second kind (Type II error) is said to occur if the test procedure accepts the hypothesis when it is false. The probability of a Type I error is often called the "producers risk" while the probability of a Type II error is called the "consumers risk" – see Schwartz (5) for an intuitive discussion of the concept and interpretation of the terms producer's and consumer's risk in the processing of telemetry data. The use of the SPRT will ensure (to an adequate approximation) that the producer's risk is α and the consumer's risk is β, α and β being pre-set by the data processing system and the experimenter respectively.

APPLICATION OF SEQUENTIAL ANALYSIS TO STRATEGY I

As an illustration of the applicability of sequential testing to digital data of the type described on page A-141 consider sequential methods applied to data gathered under Strategy I. Formally, we wish to test the hypothesis that $p = p_0$ against the alternative that $p = p_1$ where $p_0 < p_1$. Recall that

$$p_0 = \sum_{r=k_1}^{m} \binom{m}{r} p_{e_1}^r \, (1-p_{e_1})^{m-r}$$

and

$$p_1 = \sum_{r=k_1}^{m} \binom{m}{r} p_{e_2}^r \, (1-p_{e_2})^{m-r}$$

so that the hypotheses under consideration are equivalent to hypotheses about the bit probability of error p_e. We first compute

$$z_i = \log \left[\frac{(1-p_1)^{T_i} \; p_1}{(1-p_0)^{T_i} \; p_0} \right]$$

$$= \log \left(\frac{p_1}{p_0} \right) + T_i \log \left(\frac{1-p_1}{1-p_0} \right)$$

Note that $\log(1-p_1/1-p_0)$ is negative since $p_0 < p_1$. The SPRT thus rejects whenever

$$\sum_{i=1}^{N} z_i \geq \log A$$

or whenever

$$\left[\sum_{i=1}^{N} T_i \right] \log \left[\frac{1-p_1}{1-p_0} \right] \geq \log A - \log \left(\frac{p_1}{p_0} \right)$$

Using the fact that $\log[(1-p_1/1-p_0)]$ is negative the SPRT thus rejects whenever

$$\sum_{i=1}^{N} T_i \leq \log\left[\frac{1-p_1}{1-p_0}\right]\left\{\log\left(\frac{A p_0}{p_1}\right)\right\}.$$

In a similar fashion one can derive the region where one continues testing and the region where one accepts the hypothesis. The resulting procedure is

Reject the hypothesis $p = p_0$ if $\sum_{i=1}^{N} T_i \leq \log\left(\frac{1-p_1}{1-p_0}\right) \log\left[\frac{A p_0}{p_1}\right]^N$

Continue testing if $\log\left(\frac{1-p_1}{1-p_0}\right) \log\left[\frac{A p_0}{p_1}\right]^N \leq \sum_{i=1}^{N} T_i \leq \log\left(\frac{1-p_1}{1-p_0}\right) \log\left[\frac{B p_0}{p_1}\right]^N$

Accept the hypothesis $p = p_0$ if $\sum_{i=1}^{N} T_i \geq \log\left(\frac{1-p_1}{1-p_0}\right) \log\left[\frac{B p_0}{p_1}\right]^N$.

The equation delineating the regions are seen to be linear in N and hence a graph similar to that presented in Schwarz (5) can be constructed. Note that the procedure described by the SPRT is highly intuitive since if

$$\sum_{i=1}^{N} T_i$$

is small it is clear that the probability of error p_e is large (unless an event with probability less than α has occurred) so we reject.

The SPRT just developed could be used to reduce the amount of quality checking on digital data. A strategy which could be used would be to examine each sync word until the hypothesis was accepted or rejected. If rejected each sync word would continue to be tested as in the current procedures. If accepted, one could print out all data until y sync words have passed and then check for sync loss. Choice of the number y involves selection of a strategy but it would presumably be a function of the expected value of T_i. If tapes from a given station on a particular satellite indicate uniform quality such a procedure could reduce substantially the amount of quality checking.

PROBLEMS FOR FUTURE INVESTIGATION

Having indicated the possibility of applying sequential procedures to redo the amount of quality checking of telemetry data we now indicate some further problems which would appear to merit further investigation. We also indicate other uses for the data gathered by Strategies one and two.

A major problem untouched by the present investigation is that of deciding when the processing line is in sync. Clearly, strategies similar to those used in detecting loss of sync could be considered (and are in fact in use) and their effect on the distributions derived on pages A-142 to A-146 discussed. Of importance would be the distribution of the total number of accumulated errors in the sync word since this can provide information on the bit probability of error, P_e.

The distribution of the time to sync loss can also provide information on the bit probability of error in the form of a confidence interval. A companion of this interval with that constructed from the total number of accumulated errors would be of interest. Such a comparison might indicate that less "housekeeping" might provide adequate information on p_e.

The estimation of the bit probability of error, p_e, indicated in the preceding two paragraphs is, of course, of importance to both the experimenter and the telemetry personnel. Investigation of estimation of the bit probability of error should also include providing the experimenter with methods to estimate p_e on the basis of the quality card data which is routinely supplied to him. Presumably the method of maximum likelihood or minimum chi-square could provide a simple answer to this problem. Also of importance to the experimenter is some sort of check designed to indicate whatever the bit probability of error, p_e, as estimated using the sync word data is applicable to his data words. If there are periods when his experiment is transmitting highly redundant data the estimate obtained from such data could be compared with the estimate from the sync word data.

The complex problem of comparing the effect of stations, experiments, processing lines etc. on the bit probability of error needs to be considered. Knowledge of such effect could influence the design of the experiment as well as the design of the telemetry equipment. Some form of analysis of variance (with appropriate transformation) might be useful or a first step.

The importance of failure of the assumptions to hold needs to be investigated further. If, for example, p_e is not constant one might use the Lapunav form of the Central Limit Theorem to obtain an approximate sequential procedure of the same general nature as that indicated on page A-149. If the successive bits are not independent more sophisticated models need to be set up c.f. Rohde and Coleman (4). A superficial study indicates that sequential procedures become quite complicated for such models so that approximations simple enough for practical use need to be investigated.

REFERENCES

1. Feller, W., An Introduction to Probability Theory and Its Applications, John Wiley and Sons.

2. Fisz, M., Probability Theory and Mathematical Statistics, John Wiley and Sons.

3. Hauptschein, A., Report, New York University College of Engineering, Contract, Goddard Space Flight Center, NAS-5, 1964.

4. Rohde, C. and Coleman, R., Marker Chain Model for a Data Processing System. Mid-Session Progress Report of the Goddard Summer Workshop – 1965.

5. Schwartz, L. S., Quality Control in the Processing of Telemetry Data. 1965 Goddard Summer Workshop Final Report.

6. Wald, A., Sequential Analysis. John Wiley and Sons.

BLANK PAGE

DEGRADATION OF BINARY CODED DATA DUE TO BIT ERRORS

G. A. Gordon

Whenever data is transmitted over a system, some degradation of the data occurs. But for any given application, there exists a degradation level beyond which the data ceases to be usable. Also, it is not desirable to attempt to decrease the degradation indefinitely, since the cost involved (e.g., money, power supply drain, bandwidth) becomes prohibitive. It is necessary, then, to design the system to operate at a degradation level between the two extremes, and in order to do so, we require a relationship between a measure of degradation and cost.

Such a relationship is established in what follows for a particular transmission system. In this system, an analogue waveform is uniformly sampled and uniformly quantized, with subsequent binary coding of the samples. The coded samples are then transmitted over a noisy (binary symmetric) channel. We concern ourselves with the data error introduced in the channel through errors in individual bits. The errors due to sampling and quantization can be calculated separately and the effects combined to give the total degradation. We begin by forming a model of the system.

We consider a source which has as outputs the numbers 0, 1, . . . , q, where $q = 2^n - 1$, and let p_i, $i = 0, 1, \ldots, q$ denote the stationary probability that the number i is emitted by the source. The outputs are coded as n bit binary sequences, represented by the expression $x_{n-1}\, x_{n-2} \cdots x_1 . x_0$. The code sequence for each number is taken to be its binary number representation, but it is convenient to let -1 take the place of 0 in the expressions. That is, each x_j is either $+1$ or -1. The number i is related to its code sequence in an obvious manner. If we let

$$y_j = \begin{cases} +1 & \text{when } x_j = +1 \\ 0 & \text{when } x_j = -1 \end{cases} \tag{1}$$

then,

$$i = y_{n-1}\, 2^{n-1} + y_{n-2}\, 2^{n-2} + \cdots + y_1\, 2^1 + y_0\, 2^0. \tag{2}$$

The coded outputs are considered to be transmitted over the binary symmetric channel represented in Figure A-32 and then decoded using Equations (1) and (2). Bit errors introduced in the channel produce errors in the data numbers, where the amount of the error produced when a given number is transmitted depends on how many bit errors occur and where they occur in the code sequence. We seek expressions relating the bit error probability to the degradation of the transmitted data, taking as measures of this degradation the mean squared and average absolute data error. The probability of error, which depends on such things as signal power, size of antennae, etc., is a measure of the system cost.

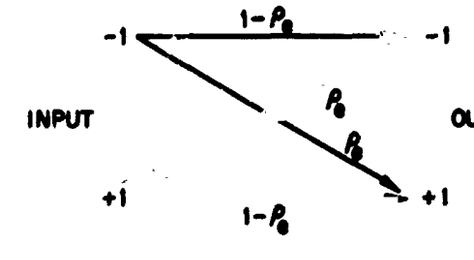


Figure A-32—The binary symmetric channel. p_e is the probability that a transmitted one or minus one is received in error.

With the code word $x_{n-1}\ x_{n-2}\ \text{---}\ x_1 x_0$ for i transmitted and with k bit errors occurring at positions $j_1, j_2, \text{---}, j_k$, where

$$\left.\begin{aligned} & j_1 < j_2 < \cdots < j_k \text{ and} \\ & 0 \le j_\ell \le n-1,\ \ell = 1, 2 \cdots, k, \end{aligned}\right\} \tag{3}$$

the data error introduced is given by

$$\epsilon_{i,k,\bar{j}} = -\left\{x_{j_1} 2^{j_1} + x_{j_2} 2^{j_2} + \cdots + x_{j_k} 2^{j_k}\right\}. \tag{4}$$

The error, ϵ, is a random variable defined on the set of possible outcomes of the experiment implied by the source and channel. These outcomes, or sample points, are the events { the number i is emitted and k bit errors occur in transmission at positions $j_1, j_2, \text{---}, j_k$ }, abbreviated $\{i, k, \bar{j}\}$. The mean squared and average absolute errors are then the expectations of ϵ^2 and $|\epsilon|$ respectively, calculated over the sample space $\mathcal{S}$ of possible outcomes. That is,

$$\left.\begin{aligned} (\epsilon^2)_{av} &= \sum_{\mathcal{S}} \epsilon_{i,k,\bar{j}}\ p\{i, k, \bar{j}\} \\ |\epsilon|_{av} &= \sum_{\mathcal{S}} |\epsilon_{i,k,\bar{j}}|\ p\{i, k, \bar{j}\}. \end{aligned}\right\} \tag{5}$$

Since our channel is symmetric, bit errors occur without regard to the original value of the bit. As a result, the events $\{k, \bar{j}\}$ and $\{i\}$ are independent, so that

$$p\{i, k, \bar{j}\} = p_i\ p\{k, \bar{j}\}.$$

The probability $p\{k, \bar{j}\}$ is the probability that a transmitted sequence x_{n-1}, x_{n-2}, ---, x_1, x_0 is received as

$$x_{n-1}, \cdots, -x_{j_k}, \cdots, x_{j_{k-1}}, \cdots, -x_{j_{k-1}}, \cdots, -x_{j_1}, \cdots, x_1, x_0,$$

for example. This minus signs indicate the errors at the k bit positions. Since bit errors occur with mutual independence this probability is given by

$$p\{k, \bar{j}\} = (1-p_e) \cdot \ldots \cdot p_e \cdot \ldots \cdot (1-p_e) \cdot \ldots \cdot p_e \cdot \ldots \cdot p_e \cdot \ldots \cdot (1-p_e)^2 = p_e^k (1-p_e)^{n-k}.$$

The result is independent of $\bar{j}$, the positions where the errors occur. We can now rewrite Equation (5) as

$$\left.\begin{aligned} (\epsilon^2)_{av} &= \sum_{\mathcal{S}} \epsilon^2_{i,k,\bar{j}}\ p_i\ p_e^k\ (1-p_e)^{n-k} \\ |\epsilon|_{av} &= \sum_{\mathcal{S}} |\epsilon_{i,k,\bar{j}}|\ p_i\ p_e^k\ (1-p_e)^{n-k} \end{aligned}\right\} \tag{6}$$

The sum over s can be expressed as an iterated summation over i, k, and $\bar{j}$. That is,

$$\sum_{s} = \sum_{i} \sum_{k} \sum_{\bar{j}} \tag{7}$$

When $k = 0$, the error ϵ is also zero, so in calculating the expectations of ϵ^2 and $|\epsilon|$ we take $k = 1, 2, \cdots, n$. For each value of k, there are $\binom{n}{k}$ arrangements of the k errors in the n positions, with corresponding vectors $\bar{j} = j_1 j_2 \cdots j_k$. The vector elements take on values in accordance with Equations (3). We can expand the operator $\sum_{\bar{j}}$ in the form

$$\sum_{\bar{j}} = \sum_{j_1=0}^{n-k} \sum_{j_2=j_1+1}^{n-k+1} \cdots \sum_{j_{k-1}=j_{k-2}+1}^{n-2} \sum_{j_k=j_{k-1}+1}^{n-1} \tag{8}$$

so that the index vector $\bar{j} = j_1 j_2 \cdots j_k$ systematically takes on the $\binom{n}{k}$ proper values. write that

$$(\epsilon_i^2)_{av} = \sum_{i=0}^{2^n-1} p_i \sum_{k=1}^{n} p_e^k (1-p_e)^{n-k} \left\{ \sum_{j_1=0}^{n-k} \cdots \sum_{j_k=j_{k-1}+1}^{n-1} \left[x_{j_1} 2^{j_1} + \cdots + x_{j_k} 2^{j_k} \right]^2 \right\}. \tag{9}$$

$$|\epsilon|_{av} = \sum_{i=0}^{2^n-1} p_i \sum_{k=1}^{n} p_e^k (1-p_e)^{n-k} \left\{ \sum_{j_1=0}^{n-k} \cdots \sum_{j_k=j_{k-1}+1}^{n-1} \left| x_{j_1} 2^{j_1} + \cdots + x_{j_k} 2^{j_k} \right| \right\}.$$

The values of $x_{j_1} x_{j_2} \cdots x_{j_k}$ can of course be determined from i using Equations (1) and (2), but it is convenient to expand the summation,

$$\sum_{i=0}^{2^n-1}$$

to make the values $x_{j_1} x_{j_2} \cdots x_{j_k}$ explicit. We do so by observing that as i varies over its range, the sequence $x_{n-1}\ x_{n-2} \cdots x_1 x_0$, where $x_j = \pm 1$, takes on all of its 2^n possible values. We write,

$$\sum_{i=0}^{2^n-1} = \sum_{x_{n-1}=\pm 1} \sum_{x_{n-2}=\pm 1} \cdots \sum_{x_1=\pm 1} \sum_{x_0=\pm 1}, \tag{11a}$$

which gives us

$$(\epsilon^2)_{av} = \sum_{x_{n-1}=\pm 1} \sum_{x_0=\pm 1} p_i \sum_{k=1}^{n} p_e^k (1-p_e)^{n-k} \left\{ \sum_{j_1=0}^{n-k} \cdots \sum_{j_k=j_{k-1}+1}^{n-1} \left[x_{j_1} 2^{j_1} + \cdots + x_{j_k} 2^{j_k} \right]^2 \right\} \tag{11a}$$

$$|\epsilon|_{av} = \sum_{x_{n-1}=\pm 1} \cdots \sum_{x_0=\pm 1} p_i \sum_{k=1}^{n} p_e^k (1-p_e)^{n-k} \left\{ \sum_{j_1=0}^{n-k} \cdots \sum_{j_k=j_{k-1}}^{n-i} \left| x_{j_1} 2^{j_1} + \cdots + x_{j_k} 2^{j_k} \right| \right\} \tag{11b}$$

Finally, to give the expressions greater generality, we divide Equation (11b) by 2^n and take the square root of Equation (11a) again normalizing with respect to 2^n. The resulting expressions are the root mean square and average absolute errors in percent of full scale and are thus independent of any units associated with the quantizing levels.

Although the expressions in Equation (11) appear cumbersome, they are very easily handled with the aid of a digital computer, and were in fact formulated with a final computer solution in mind. The logic involved in a computer realization is discussed in Appendix A, and some sample results are given in Appendix B.

With the probabilities p_i of the data known or assumed, Equations (11) yield plots of the percentage degradation as a function of p_e. A plot of root mean squared error, for example, might appear as in Figure A-33.

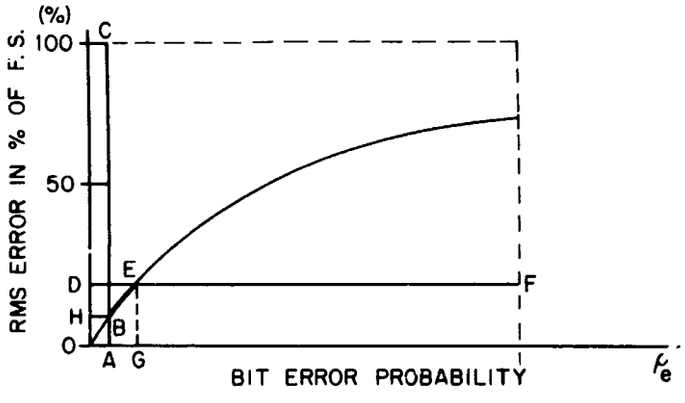


Figure A-33–A plot of r.m.s. error for some distribution of p_i.

The level of tolerable degradation is represented by the line $\overline{DF}$ with intersection E. The position of this line is determined by those who will finally apply the data. The line $\overline{AC}$ represents the lower bound on the achievable probability of bit error, which depends on such things as the available signal power, the noise power in the channel, the bandwidth of the system and the required data transmission rate. The system must be designed to operate somewhere on the interval $\overline{BE}$ which gives $\overline{AG}$ as the allowable range for the bit error probability as a system design criterion.

We thus achieve a rational basis for determining the point at which we stop trying to decrease p_e. That is, when we have decreased p_e to be point where the degradation level is well below the threshold value, say at the level

$$\frac{D + H}{2},$$

we accept the system design. Thus we strike a balance between the desire to have our data as reliable as possible and yet to keep the cost of system design and operation as low as possible.

APPENDIX A

COMPUTER ORIENTED ALGORITHMS USEFUL IN CALCULATING EQUATION (11)

In Equation (11), the multiple summation is performed, in accordance with convention, with the indices on the right of the expression varying most rapidly and the indices varying less rapidly as we proceed to the left. That is, for a fixed sequence $x_{n-1}\, x_{n-2}$ --- $x_1\, x_0$, we vary k from 1 to n, and for each value of k, we vary the index vector $\bar{j}$ over its $\binom{n}{k}$ values. This is obviously realized through the use of nested iterations.

The variation of the index sequences x_{n-1}, $x_{n-2}, \ldots, x_1, x_0$ and $j_1, j_2, \ldots, j_k$ in Equation (11) is not, however, best realized in the form of nested iterations, expecially in the latter case where the number of nexted iteration loops would be k, a variable.

Instead, the index sequences are stored in vector arrays which are systematically updated so as to eventually take on all of the values corresponding to points in the space $\mathcal{S}$. The algorithms used to update these arrays are given below. The general logical flow diagram is given in Figure A-34 for evaluation of Equation (11a).

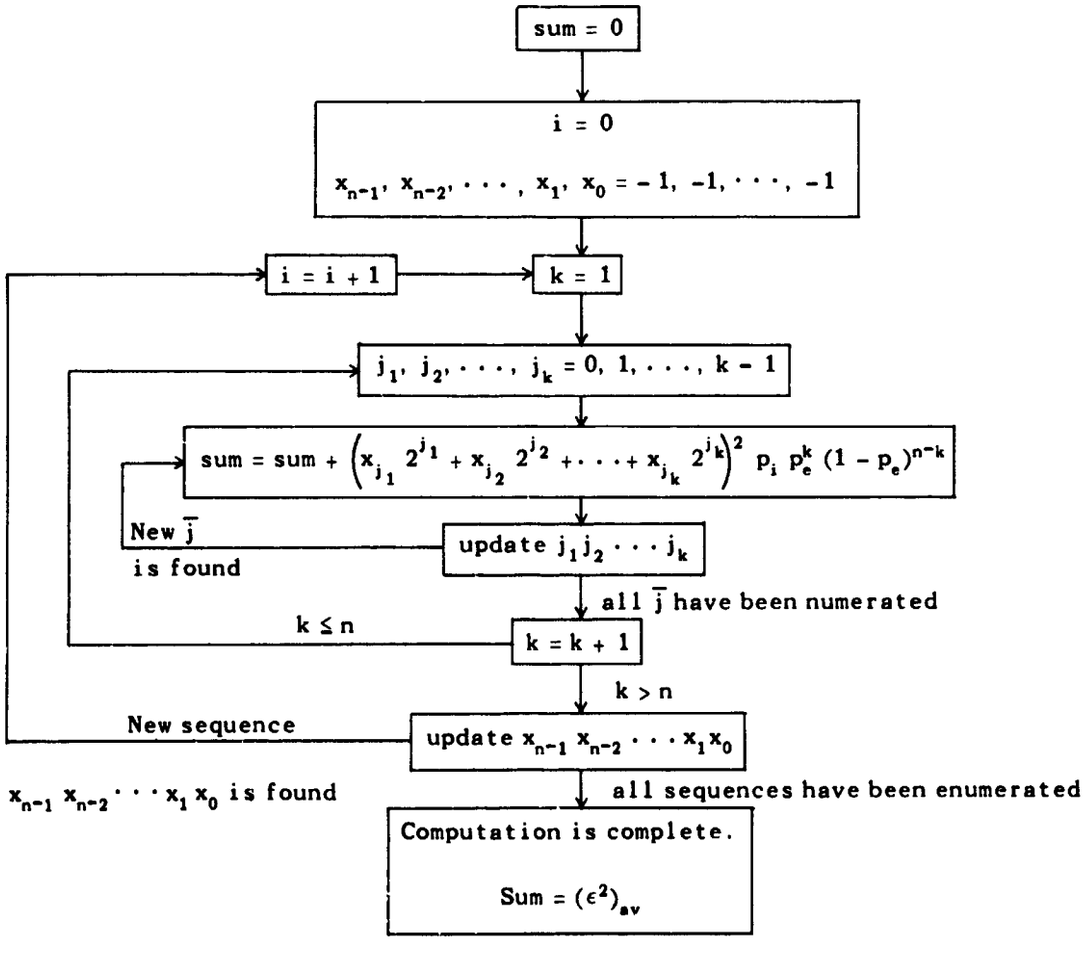


Figure A-34--Logic diagram.

The sequence $x_{n-1}, x_{n-2}, \cdots, x_1, x_0$ is initialized as

$$x_{n-1}, x_{n-2}, \cdots, x_1, x_0 = -1, -1, \cdots, -1.$$

When it becomes necessary to update the sequence, a search is made, starting from the right of the sequence, for the first element which has the value -1. When it is found, it is changed to + 1, and all elements to the right are changed to -1. This rule repeatedly applied to the above sequence, yields the list of sequences:

$$-1, -1, \cdots, -1, -1; \quad i = 0$$

$$-1, -1, \cdots, -1, +1; \quad i = 1$$

$$-1, -1, \cdots, +1, -1; \quad i = 2$$

$$-1, -1, \cdots, +1, +1; \quad i = 3$$

$$-1, -1, \cdots, +1, -1, -1; \quad i = 4, \text{ etc.}$$

It can be seen that the procedure systematically takes the sequence through its 2^n possible values. Eventually all of the sequence elements will be +1, corresponding to $i = 2^n - 1$. At this point the search for a - 1 element will fail, indicating that the computation is complete.

The sequence $j_1, j_2, \cdots, j_k$ is initialized as

$$j_1, j_2, \cdots, j_k = 0, 1, \cdots, k-1,$$

and is updated by a similar rule. As seen in Equation (11), the element j_ℓ can take on the values

$$j_\ell = \ell - 1, \ell, \ell + 1, \cdots, \ell + n - k - 1.$$

Thus each element in the sequence $j_1, j_2, \cdots, j_k$ has a maximum value which it can assume, depending on the position of the element.

To update the sequence $j_1, j_2, \cdots, j_k$, a search is made, starting from the right of the sequence, for an element which is less then its maximum value. When it is found, the element is increased by one, and all elements to the right are decreased as much as possible without violating Equation (3). With $n = 5$ and $k = 3$, the procedure yields the following results:

j_1	j_2	j_3
0	1	2
0	1	3
0	1	4
0	2	3
0	2	4
0	3	4

j_1	j_2	j_3
1	2	3
1	2	4
1	3	4
2	3	4.

We see that the $\binom{n}{k}$ sequences are systematically generated by the algorithm. When the search for a sub-maximum element fails, it indicates that all $\binom{n}{k}$ sequences have been generated, and we must increment k, as shown in Figure A-34.

APPENDIX B
SOME SAMPLE RESULTS

Results are calculated for the case of a uniform data distribution, i.e., $p_i = 2^{-n}$, all i. In this case the degradation is given by

$$\frac{[(\epsilon^2)_{av}]^{1/2}}{2^n} = \frac{1}{2^{3n/2}} \left\{ \sum_{k=1}^{n} p_e^k (1-p_e)^{n-k} \cdot \sum_{x_{n-1}=\pm 1} \cdots \sum_{x_0=\pm 1} \sum_{j_1=0}^{n-k} \cdots \sum_{j_k=j_{k-1}+1}^{n-1} \left[x_{j_1} 2^{j_1} + \cdots + x_{j_k} 2^{j_k} \right]^2 \right\}^{1/2} \quad (12a)$$

$$\frac{|\epsilon|_{av}}{2^n} = \frac{1}{4^n} \sum_{k=1}^{n} p_e^k (1-p_e)^{n-k} \cdot \sum_{x_{n-1}=\pm 1} \cdots \sum_{x_0=\pm 1} \sum_{j_1=0}^{n-k} \cdots \sum_{j_k=j_{k-1}+1}^{n-1} \left| x_{j_1} 2^{j_1} + \cdots + x_{j_k} 2^{j_k} \right| \quad (12b)$$

These are plotted as functions of p_e in the following graphs for three values of n. (Figures A-35 and A-36).

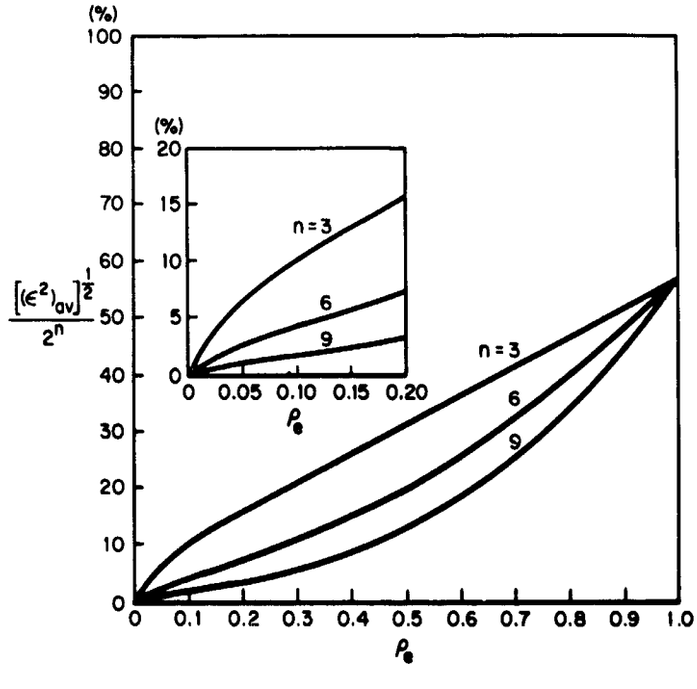


Figure A-35–Mean squared error.

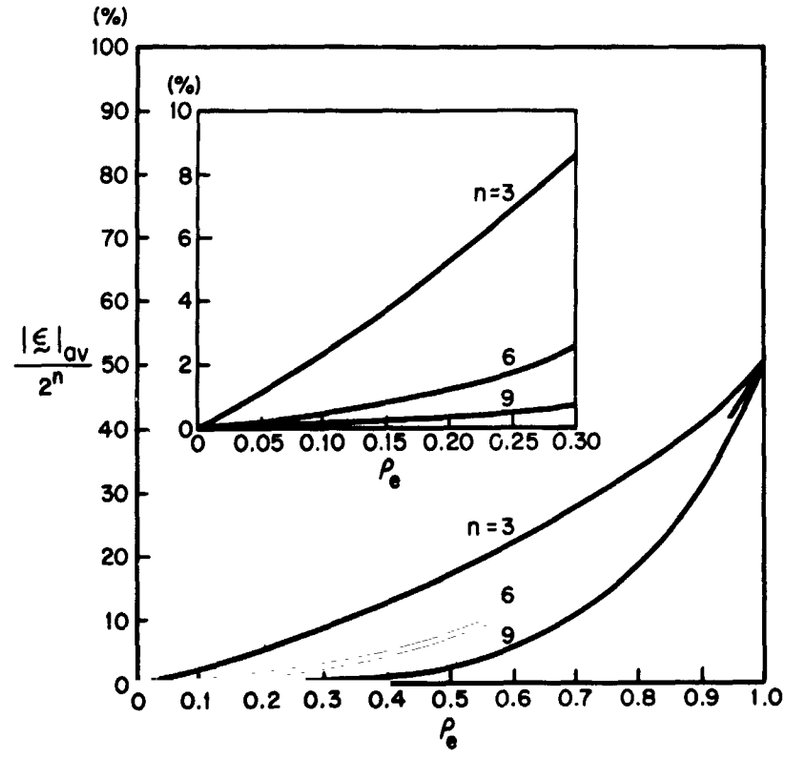


Figure A-36–Average absolute error.

RECURSIVE ESTIMATION IN ANALYSIS OF VARIANCE MODELS

C. A. Rohde

In many experimental situations an observed vector of data y can be represented by a general linear model of the form:

$$\begin{aligned} &(i) \quad E[y] = X\beta \\ &(ii) \quad \mathrm{Var}[y] = \sigma^2 I. \end{aligned} \tag{1}$$

In this model y is an $n \times 1$ vector of random variables, X is a known $n \times p$ matrix of rank $r \leq p < n$ and β is a $p \times 1$ vector of parameters. E denotes the expectation operator and I the $n \times n$ identity matrix. Alternatively one can write Equation (1) in the form

$$\begin{aligned} &(i) \quad y = X\beta + \varepsilon. \\ &(ii) \quad E[\varepsilon] = 0 \text{ and } \mathrm{Var}[\varepsilon] = \sigma^2 I \end{aligned} \tag{1'}$$

which simply expresses it in the more familiar "signal-plus-noise" form.

Interest usually focuses on the estimation of β called the vector or regression coefficients in the statistical literature, or the state vector in the engineering literature. There has recently been a great deal of work in the control theory literature on obtaining recursive estimates of β, i.e., given an estimate $\hat{\beta}(n)$ at time n and a new observation $y(n+1)$ to find the optimal (which is usually a linear estimate unbiased that minimizes squared error loss) estimate $\hat{\beta}(n+1)$ at time $n+1$. In addition formulas for the variance covariance matrix of such estimators are often of interest. In certain investigations one wants to consider part of β, i.e., β_2 as a random variable (which can be taken to have zero mean with no loss of generality). In such cases one desires information in the variance of β_2. Since variance is a quadratic quantity it is natural to estimate $\mathrm{Var}[\beta_2]$ with quadratic forms. Thus, we need to develop recursive formulae for certain quadratic forms at time $(n+1)$ in terms of the quadratic form at time n and a correction due to the new observation. As a by-product we will obtain recursive expressions for $\hat{\beta}(n+1)$ for the case where Rank $X < p$, which are simpler computationally than those previously suggested in recent literature.

Under the model Equation (1) it is well known that the optimal (in the sense of being a linear unbiased estimator which minimizes squared error loss) is given by the solution to the least squares or normal equation,

$$X'Xb = X'y. \tag{2}$$

if $(X'X)$ is non-singular. If $(X'X)$ is singular it is known that a solution to Equation (2) can be represented as

$$b = (X'X)^{c}X'y.$$

where $(X'X)^g$ is a generalized inverse of $X'X$ and satisfies the equation,

$$(X'X)\,(X'X)^g X'X = X'X$$

In this case, estimates of functions $\ell'\beta$ are such that a c exists so that $c'X = \ell'$ are uniquely determined and are given by $\ell' b$, even though b is not uniquely determined by Equation (2). The minimum value of the error sum of squares $(y - Xb)'(y - Xb)$ is given by the equation,

$$y'y - y'Xb = y'[I - X(X'X)^g X']\,y,$$

and is also uniquely determined. In order to obtain recursive formulas for the above model we rewrite $X'X$ in the following form:

$$(X'X) = A(n) = \left[\begin{array}{c|c} B(n) & C(n) \\ \hline & D(n) \end{array}\right] \tag{3}$$

where $B(n)$ is $r \times r$ of rank r and $A(n)$ is $p \times p$ of rank r. Note that the rearrangement of $A(n)$ effects a rearrangement of X and also of β. In Equation (3) it is well known that

$$D(n) = C'(n)\,[B(n)]^{-1}\,C(n). \tag{4}$$

It is easily verified that a convenient generalized inverse of $A(n)$ can be represented by

$$[A(n)]^g = \left[\begin{array}{c|c} [B(n)]^{-1} & Q \\ \hline Q & Q \end{array}\right]$$

where Q's denote null matrices of appropriate dimensions.

If a new observation $y(n+1)$ is added to the available vector y of n observations we have the following model:

$$\begin{bmatrix} y \\ y(n+1) \end{bmatrix} = \begin{bmatrix} X \\ x'(n+1) \end{bmatrix} + \begin{bmatrix} \epsilon \\ \epsilon(n+1) \end{bmatrix}$$

The least squares estimate or optimal estimate of β thus becomes $b(n+1)$ where $b(n+1)$ is any solution to the equations

$$[X'X + xx']\,b(n+1) = [X'x] \begin{bmatrix} y \\ y \end{bmatrix}$$

where we have dropped the $(n+1)$ notation on $x(n+1)$ and $y(n+1)$ for simplicity. The problem of obtaining recursive equations for statistics computed at time $(n+1)$ in terms of the statistics at time n and "correction terms" thus reduces to finding a recursive equation relating $A^g(n+1)$ to $A^g(n)$. To do this we utilize the generalized inverse given by Equation (4). Letting x^* denote $x(n+1)$ and rewritten to conform with the rewriting of $A(n)$ we have

$$A(n+1) = A(n) + x^* x^{*1}$$

or

$$A(n+1) = \left[\begin{array}{c|c} B(n) & C(n) \\ \hline C'(n) & D(n) \end{array}\right] + \left[\begin{array}{c|c} z_1 z_1 & z_1 z_2' \\ \hline z_2 z_1' & z_2 z_2' \end{array}\right]$$

where the partitioning of x^* conforms with the partitioning of $A(n)$.

We should recall that

$$[I + a\, b']^{-1} = I - a(1 + b'\, a)^{-1}\, b'.$$

Hence if $B(n+1) = B(n) + z_1 z_1'$, we have the equation

$$[B(n+1)]^{-1} = [B(n)]^{-1} - k\,[B(n)]^{-1} z_1 z_1' [B(n)]^{-1}, \tag{5}$$

where $k = 1 + z_1' [B(n)]^{-1} z_1$ (5'). Letting $C(n+1) = C(n) + z_1 z_2'$, $D(n+1) = D(n) + z_2 z_2'$, and performing some tedious though routine algebraic manipulations, we find that

$$D(n+1) - C'(n+1)\,[B(n+1)]^{-1} C(n+1) = k\,[z_2 - C'(n)\,[B(n)]^{-1} z_1]\,[z_2 - C'(n)\,[B(n)]^{-1} z_1]' \tag{6}$$

If a symmetric matrix is partitioned in the form

$$\left[\begin{array}{c|c} B & C \\ \hline C' & D \end{array}\right],$$

where B is non-singular, then a generalized inverse can be shown to be

$$\left[\begin{array}{c|c} B^{-1} + B^{-1} C Q^g C' B^{-1} & -B^{-1} C Q^g \\ \hline -Q^g C' B^{-1} & Q^g \end{array}\right], \tag{7}$$

where $Q = D - C' B^{-1} C$.

Using Equations (5), (6) and (7) we have the following computational procedure for obtaining $A^g(n+1)$:

Available at time n: $[B(n)]^{-1}$, $C(n)$, $D(n)$.

Available from time $(n+1)$: z_1, z_2.

Procedure:

(A) Compute $[B(n)]^{-1} z_1$ and $C'(n)\,[B(n)]^{-1} z_1$.

(B) Compute $[1 + z_1' [B(n)]^{-1} z_1]^{-1} = k^{-1}$

(C) Compute $R(n+1) = [B(n)]^{-1} - k\,[B(n)]^{-1} z_1 z_1' [B(n)]^{-1}$.

(D) Compute $\eta = z_2 - C'(n)\,[B(n)]^{-1} z_1$.

Case:I: $\eta = 0$

If $\eta = 0$, then

$$A^g(n+1) = \left[\begin{array}{c|c} R(n+1) & \underset{\sim}{0} \\ \hline \underset{\sim}{0} & \underset{\sim}{0} \end{array}\right]. \tag{8}$$

Case II: $\eta \neq 0$

(i) If $\eta \neq 0$ compute $\eta\eta'$ and select the first non-zero diagonal element of $\eta\eta'$, say α. Call the corresponding row and column of $C'(n+1)$ and $C(n+1)$, α' and α respectively. Rearrange $A(n+1)$ so that the first $r+1$ rows and columns are

$$\left[\begin{array}{c|cc} B(n+1) & \alpha & \underset{\sim}{0} \\ \hline \alpha' & \alpha & 0' \\ \underset{\sim}{0} & 0 & \underset{\sim}{0} \end{array}\right]$$

(ii) Compute α^{-1} and compute $[B(n)]^{-1}\alpha$, using the appropriate column of $[B(n)]^{-1}C(n)$ added to $[B(n)]^{-1} z_1 z_2'$.

(iii) Compute $\alpha^{-1}[B(n)]^{-1}\alpha\alpha'[B(n)]^{-1}$.

(iv) Compute

$$A^g(n+1) = \left[\begin{array}{c|c|c} [B(n)]^{-1} + \alpha^{-1}[B(n)]^{-1}\alpha\alpha'[B(n)]^{-1} & -\alpha^{-1}[B(n)]^{-1}\alpha & \underset{\sim}{0} \\ \hline -\alpha^{-1}\alpha'[B(n)]^{-1} & \alpha^{-1} & 0' \\ \hline \underset{\sim}{0} & 0' & \underset{\sim}{0} \end{array}\right] \tag{9}$$

Using Equations (8) and (9) it should be possible to develop recursive formulae for $b(n+1)$ in terms of $b(n)$ and a correction term as well as recursive formulae for other statistics of interest: such as quadratic forms, estimates of variance-covariance matrices etc. These formulae will be the extensions of those obtained in References (1) through (4).

REFERENCES

1. Kalman, R. E., "A New Approach to Linear Filtering & Prediction Problems," Journal ASME, (March 1960).

2. Sittler, R. W., "Recursive Estimation with Application to Trajectory Problems," Paper presented at 105th Eastern Regional Meeting of the Institute of Mathematical Statistics and American Statistical Association (April 29-May 1, 1965).

3. Schweppe, Fred C., "An Introduction to Estimation Theory for Dynamical Systems," MIT Lincoln Laboratory Report 22G-15.

4. Lee, Robert, C. K. Lee, "Optimal Estimation Identification and Control," Research Monograph No. 28, MIT Press, Cambridge, Massachusetts.

BLANK PAGE

RECONSTRUCTION OF SAMPLED INPUT OF A LINEAR SYSTEM FROM ITS SAMPLED OUTPUT

M. Javid

In determination of the density of ions in space, the shape of the waveform obtained from the spectrometer output used in the experiment is of importance (1). The data telemetered by a satellite is not the sampled value of this waveform, but corresponds to the output of the amplifier and filter system which follows the spectrometer. This data must be processed to recover the shape of the waveform input to the amplifier-filter system. At present an empirical procedure is used for this processing. The following analysis leads to an exact procedure suitable for computer processing which reconstructs the sampled input of a linear system from its sampled output. This procedure also makes it possible to compute the delay between the maximum (peak) value of input and the output of the linear system. Since the location of this maximum is indicative of the mass of the corresponding ion, the prescribed procedure should improve the accuracy of the experiment.

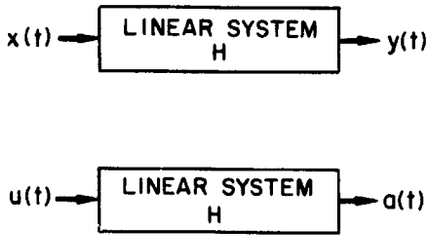


Figure A-37–Linear system flow diagram.

Let $x(t)$ and $y(t)$ be the input and output of a linear system H, described by its (indicial) response $a(t)$ to the unit step function $u(t)$, as shown in Figure A-37. It is known that

$$y(t) = x(0)\, a(t) + \int_{\tau=0}^{t} \frac{dx(\tau)}{d\tau}\, a(t-\tau)\, d\tau. \tag{1}$$

If the input $x(\tau)$ is sampled at intervals T_s and the sampled values of x and y are designated by

$$x_n = x(n\,T_s), \quad y_n = y(n\,T_s), \tag{2}$$

then, from Equation (1)

$$y_n = x_0\, a(n\,T_s) + \int_{\tau=0}^{n\,T_s} a(n\,T_s - \tau)\, dx(\tau). \tag{3}$$

The value of y_n may be approximated by

$$y_n \simeq x_0\, a(n\,T_s) + \sum_{m=0}^{n} a[(n-m)\,T_s]\, \Delta x\,(m\,T_s). \tag{4}$$

Designating the k^{th} sample of $a(t)$ by

$$a_k = a(k\,T_s), \tag{5}$$

we have

$$y_n = x_0\, a_n + \sum_{m=0}^{n} a_{n-m}\,(x_{m+1} - x_m). \tag{6}$$

The problem of reconstruction is to obtain x_j for $j = 0$ to K given y_k; $k = 0$ to K. From Equation (6) we have

$$y_0 = a_0 x_0 + a_0 (x_1 - x_0) = 0. \tag{7}$$

Since $a_0 = a(t = 0) = 0$ (no jump in $a(t)$ at $t = 0$),

$$y_1 = a_1 x_0 + a_1(x_1 - x_0) + a_0(x_2 - x_1) = a_1 x_1, \tag{8}$$

$$y_2 = a_2 x_0 + a_2 (x_1 - x_0) + a_1(x_2 - x_1) = (a_2 - a_1) x_1 + a_1 x_2, \tag{9}$$

$$y_3 = a_3 x_0 + a_3(x_1 - x_0) + a_2(x_2 - x_1) + a_1(x_3 - x_2) = (a_3 - a_2)x_1 + (a_2 - a_1) x_2 + a_1 x_3. \tag{10}$$

The trend is now clear and is given by

$$y_k = \sum_{m=1}^{k} (a_{k+1-m} - a_{k-m}) x_m, \tag{11}$$

or

$$y_k = \sum_{m=1}^{k} (a_m - a_{m-1}) x_{k+1-m}. \tag{12}$$

From Equation (11) we find x_j from y's recursively:

$$x_1 = \frac{1}{a_1} y_1, \tag{13}$$

$$x_2 = \frac{1}{a_1} [y_2 - (a_2 - a_1) x_1], \tag{14}$$

$$x_3 = \frac{1}{a_1} [y_3 - (a_2 - a_1) x_2 - (a_3 - a_2) x_1], \tag{15}$$

$$x_4 = \frac{1}{a_1} [y_4 - (a_2 - a_1) x_3 - (a_3 - a_2) x_2 - (a_4 - a_3) x_1], \tag{16}$$

$$x_j = \frac{1}{a_1} \left[y_j - \sum_{m=0}^{j-1} (a_{j-2+m} - a_{j-3+m}) x_{j-m} \right], \tag{17}$$

or

$$x_j = \frac{1}{a_1} \left[y_j - \sum_{n=2}^{j} (a_n - a_{n-1}) x_{j+1-n} \right]. \tag{18}$$

The above result is the algorithm for reconstruction of the sampled input given the sampled output of a linear system and its unit step response.

This algorithm also makes it possible to find the location of the peak of the input waveforms by using instructions in the computer program which result in storage of the current maximum value of the input and its index (location). A similar procedure facilitates determination of distance between the peaks of the input and output wave shapes.

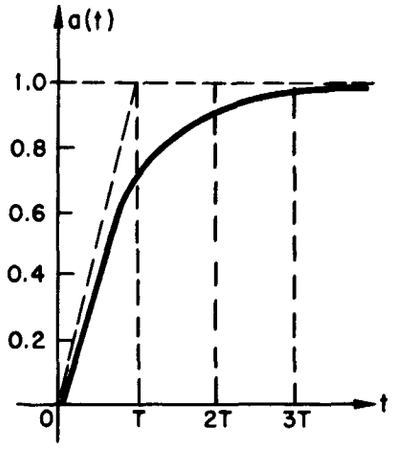


Figure A-38–Step response of the amplifiers used in the S-49 satellite.

SPECIAL APPLICATION

The step response of the amplifiers used in the S-49 satellite are shown in Figure A-38. They may be represented as

$$a(t) = 1 - e^{-t/T}, \tag{19}$$

where $1/T$ is the slope of tangent to $a(t)$ at $t = 0$. From the above diagram it can be seen that for $t > 3T$, $a(t) \simeq 1$. This indicates that, for

$$n > \frac{3T}{T_s}, \quad a_n - a_{n-1} = 0. \tag{20}$$

Substituting this value in Equation (18) results in

$$x_j = \frac{1}{a_1}\left[y_j - \sum_{n=2}^{3T/T_s} (a_n - a_{n-1})\, x_{j+1-n}\right], \quad j \geq \frac{3T}{T_s} \tag{21}$$

Equation (21) reveals that for the step response shown at most $[3T/T_s]$ samples may be used in evaluation of x_j. Thus, $3T$ may be thought of as the memory of the linear system and $3T/T_s$ is the maximum number of samples used in processing x_j, where the range of j is from 1 to T_w/T_s and T_w is the "duration" of waveform in which its value is appreciably different from zero. Substituting these in Equation (21) from Equation (19) we have

$$x_j = \frac{1}{a_1}\left[y_j - \sum_{n=2}^{K} e^{-n\,T_s/T}\,(e^{T_s/T} - 1)\, x_{j+1-n}\right], \tag{22}$$

where

$$K = \min\{j,\ 3T/T_s\}. \tag{23}$$

Let

$$e^{T_s/T} = s. \tag{24}$$

Then Equation (22) may be written as

$$x_j = \frac{1}{a_1}\left[y_j - \sum_{n=2}^{K} s^{-n}\,(s - 1)\, x_{j+1-n}\right], \tag{25}$$

which is a more convenient form for computer evaluation.

REFERENCES

1. W. H. Bennett, Journ. of App. Phys. Vol. 21, page 143, February 1950.

2. Javid and Brenner, Analysis, Transmission and Filtering of Signals, McGraw Hill Book Co., (1963) page 147.

SUMMER WORKSHOP 1965

Program Outline and Team Participants

Study Topics

B-1 Conical radiation detectors

B-2 Testing and calibration of detectors used in flight

B-3 Photometric principles governing the vacuum optical bench

B-4 The tungsten lamp as a secondary standard for total radiant flux

B-5 Standardization of 1000w quartz-iodine lamps

B-6 A prismatic cutoff filter for transmitting short wavelengths

B-7 Spectral radiant flux in the solar environment simulator

B-8 Spectral transmittance of Jarrell-Ash model 82-000 monochromator

B-9 Infrared reflectance spectra of igneous rocks, tuffs, and red sandstone from 0.5 to 22 microns

B-10 The gravity reference system

TEAM B

Academic Personnel

Dr. Earle K. Plyler,
principal investigator

Dr. Irvine C. Gardner
Dr. Eugene I. Mohr
Dr. William R. Callahan
Charles L. Korb
James D. Rancourt
Stephen M. Shapiro

Goddard Personnel

Dr. Matthew P. Thekaekara,
staff advisor

Dr. Fred W. Paul
Robert L. Appler
Gustave J. Comeyne
James L. Diggins
Hubert H. Hoehn
Dr. Warren A. Hovis
Nathan J. Miller
Andrew W. McCulloch
James T. McLean
Barbara Sparkman
Arthur R. Winker
Neil P. Zylich

NSF Summer Science Student
Sandra E. Read

BLANK PAGE

PROJECT B: SPACE OPTICS, RADIATION SENSORS, AND SOLAR SIMULATION

CONTENTS

ILLUSTRATIONS

TABLES

CONICAL RADIATION DETECTORS

C. L. Korb, J. D. Rancourt,
and M. P. Thekaekara

INTRODUCTION

Accurate measurement of the energy incident on the test floor is extremely important in the testing and evaluation of spacecraft and experimental packages. Eppley thermopiles, which are frequently used for total energy measurement, cannot conveniently be employed in vacuum chambers at liquid nitrogen (LN_2) temperatures. G. E. detectors do not yield sufficient accuracy. Solar cells have certain advantages, but they have a high spectral selectivity. The black ball spherical integrator requires a very long time for measurement. Some form of conical cavity was thought of as an alternate method combining the advantages of a high degree of accuracy, uniform response to all spectral ranges, stability of response, and suitability for vacuum at LN_2 temperatures. Experiments were performed with two types of conical cavities: thin copper shells and wire-wound cones. The results are discussed in the following pages.

EXPERIMENTAL DATA ON THIN CONICAL SHELLS IN THE SEAR CHAMBER

The conical shells for the present investigations were made by electrodepositing copper on a mandril to a thickness of about 6 mils (0.15 mm). The apex angles of the cones were 20, 30 and 40 degrees, and the diameter of the cone aperture was about 1 inch. The inner surface of all shells was spray painted with 3M velvet black paint, no. 9564. The spraying technique produces a better result than brushing which gives a glossier and less even finish. The outer surface of each shell was treated differently. One cone was spray painted black inside in the same way as the outside. The second was spray painted white with a paint made up of an aluminum oxide pigment in a potassium silicate vehicle. The rest of the cones were electrolytically gold plated. Some of these cones had a much brighter finish than others due to the thickness of the gold coating, the thicker coatings producing the brighter surface. No gold coating was produced by vacuum evaporation.

The cones were so oriented that radiation fell on the aperture and into the cavity as shown in Figure B-1. The radiant energy is absorbed and this heats the cone. The equilibrium temperature attained by the cone should be a measure of the radiant flux in the plane of the cone aperture. In order to measure this temperature two thermocouples were mounted fore and aft on the outside surface of the cone, positioned so that neither would receive direct radiation from the light source. They were mounted, after all electroplating but before painting, with a silver epoxy manufactured by Epoxy Products Inc., no. 3021. Epoxy was used to ensure a good thermal as well as mechanical bond and to add as little thermal mass as possible to the cones.

Copper and constantan wires were used to make the thermocouples. The wiring scheme for the thermocouples is shown in Figure B-2. A DPDT switch permits rapid change from one thermocouple to another so that their e.m.f. can be read alternately. The reference junction was a commercial solid-state ice-point reference system manufactured by Joseph Kaye and Co., Inc., Model PRJAI. The e.m.f. was read by a digital voltmeter, manufactured by Dana Laboratories Inc., Model 5600, which has a sensitivity of 0.2 microvolt.

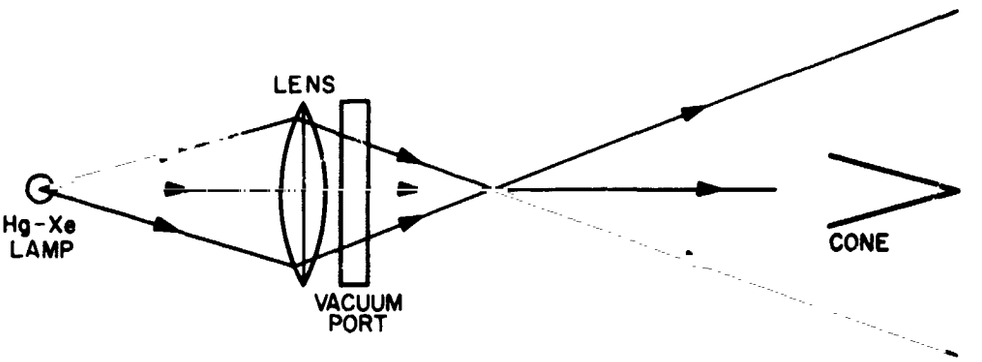


Figure B-1–Optical schematic for illumination of the cone.

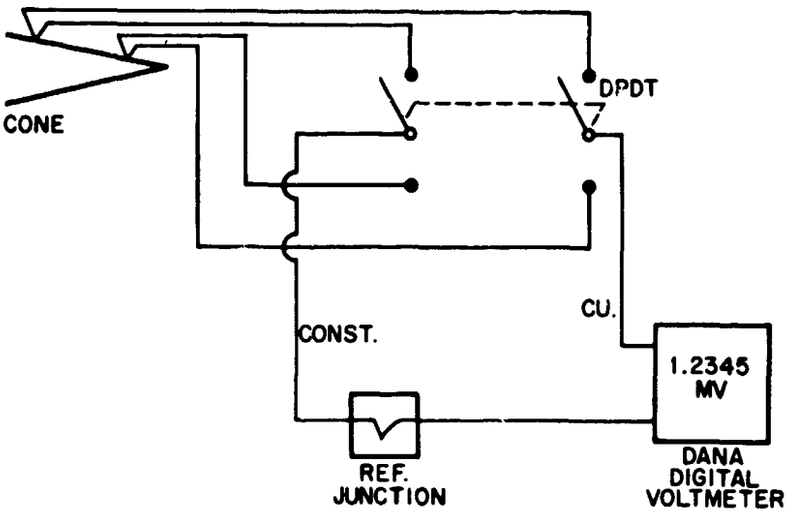


Figure B-2–Wiring schematic for measuring the temperature of the cone.

The source of radiant energy was a 2500 watt mercury-xenon lamp, operated from a Westinghouse type WS direct current arc power supply. This high power source was preferred to the more stable tungsten lamp partly to ensure a flux level of the order of one solar constant or more and partly to duplicate more closely the conditions in the Solar Environment Simulator. The power input to the lamp was continually monitored by a recording wattmeter, and readings were also taken from an ammeter and voltmeter connected to the lamp. These meters, having an accuracy of 1 percent, were more reliable than the wattmeter. An attempt was made to check the stability of the Hg - Xe arc by means of a photocell and a recording potentiometer; this did not help as much as expected, since its sensitivity to small lamp fluctuations was still limited, but it did help with the larger variations.

The lamp was calibrated at 2750 watts, using a scanning Eppley thermopile in the plane later to be occupied by the cone aperture. It was found that the flux was about 1.1 solar constant in an area 1.5 inches in diameter. The optical arrangement is shown in Figure B-1. A quartz lens of focal length about 11 cm is mounted about 22 cm in front of the lamp. The light beam focusses to a point and diverges to illumine the cone, which is located about 49 cm on the other side of the lens. The port of the vacuum chamber is also of quartz to allow ultraviolet radiation to enter the chamber.

The tests were conducted at atmospheric pressure or in a vacuum of about 10^{-7} mm Hg in the SEAR chamber. A shroud cooled with LN_2 was used when the tests were made in vacuum. This shroud was a cylindrical shell with LN_2 pipes on its entire outside surface. The outside was coated with an aluminum paint and the inside was black. Five parts were cut on the shroud and the chamber wall to allow access to the inside. The hole at the bottom was just of the size of the pipe leading to the diffusion pump. A baffle cooled with LN_2 was placed over the hole to shield the test object from the hot environment. Four of the openings in the shroud were 3.5 inches in diameter; the fifth was about 1.25 inches.

Thermocouples were placed at various points on the inside and outside surfaces of the shroud. The temperatures on the outside always registered about 80°K. The inside thermocouples gave results which were questionable, since no really satisfactory way to mount them was found. Those on the wall opposite the entrance port registered relatively high temperatures.

Figures B-3 through B-7 show the e.m.f. of the thermocouples on the cone and the temperature of the cone as a function of the time elapsed since the beginning of the test. Only the results of a few of the more significant tests are presented. A summary of the results is given in Table B-1.

The tests under atmospheric pressure – for example, the one shown in Figure B-3 – were conducted mainly as an equipment test; but they also showed the general behavior of the cone in air.

Figure B-4 shows a typical cooling curve for the black cone when placed in vacuum and surrounded by a LN_2 shroud. The latter part of the curve is on a contracted time scale. Note that the front and rear thermocouples reach the same temperature very rapidly after the shutter is closed. Initially the front thermopile is 1.4°K higher than the rear thermopile. Figure B-5 shows a comparison of the heating curves of the black cone at two different levels of incident energy (input to the lamp 2700 w and 2300 w) with that of the gold cone. It was not possible to test the gold cone at power levels above 2300 watts, since the cone temperature would have exceeded the safe limit of the epoxy used to mount the thermocouples. A cooling curve for the gold cone is shown in Figure B-6. Again note the rapid

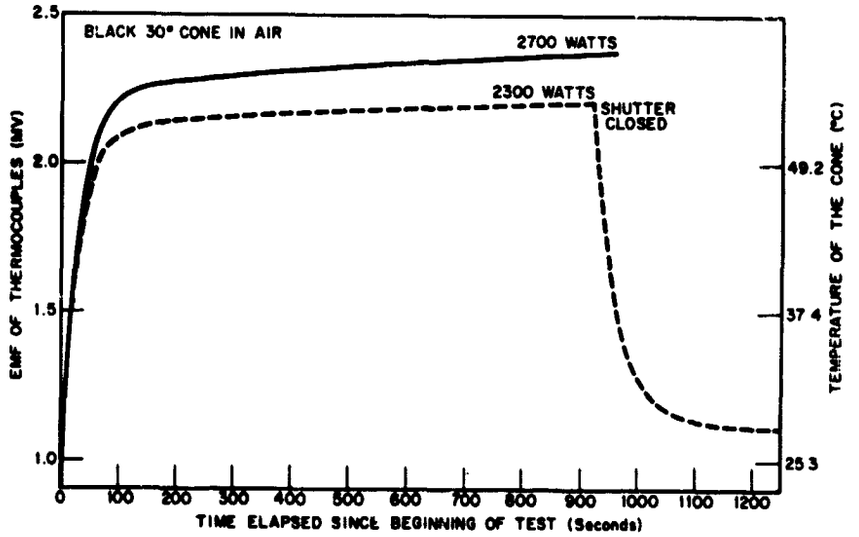


Figure B-3–Heating and cooling curves of front thermocouples of black cone, at atmospheric pressure, for input power of lamp 2700w and 2300w.

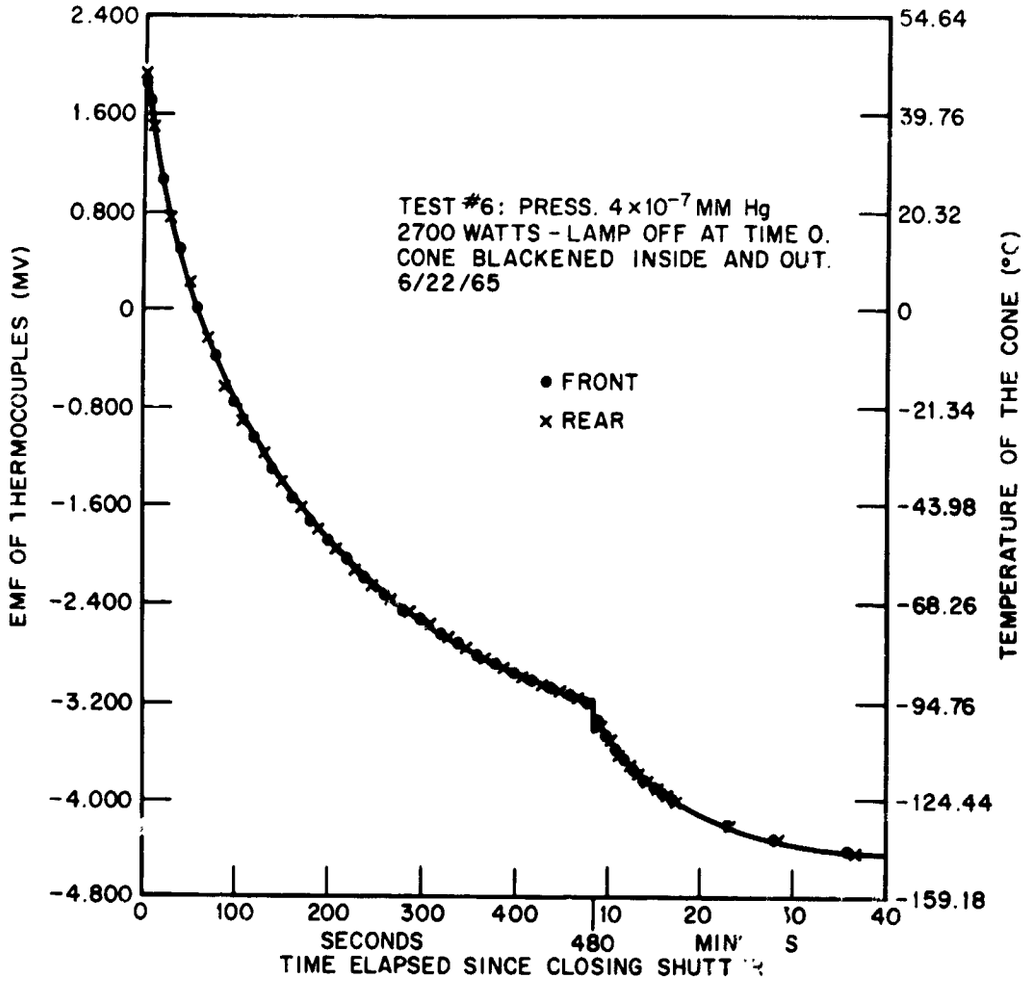


Figure B-4–Cooling curve of black 30° cone in vacuum.

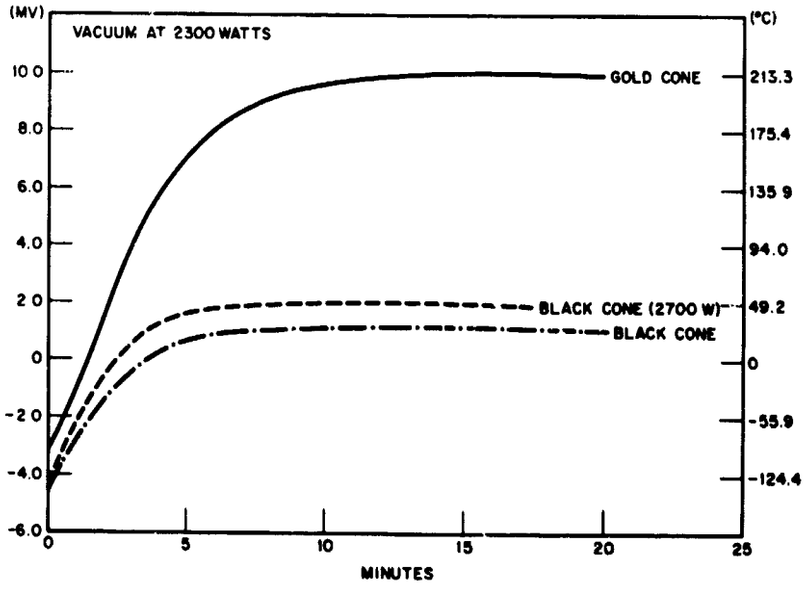


Figure B-5–Heating curves of cones in vacuum.

equalization of the two thermocouple temperatures. The temperatures at time zero when the shutter was closed were 484.7°K and 481.7°K for the front and rear thermocouples respectively. Figure B-7 shows the behavior of the temperature measured by the front thermocouple during a stepwise increase in the power input to the lamp.

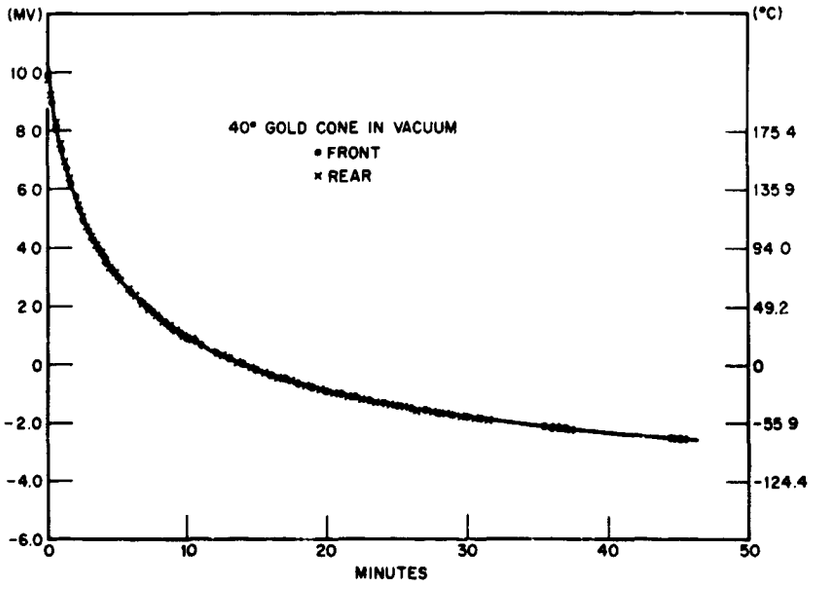


Figure B-6–Cooling curve of gold 40° cone in vacuum.

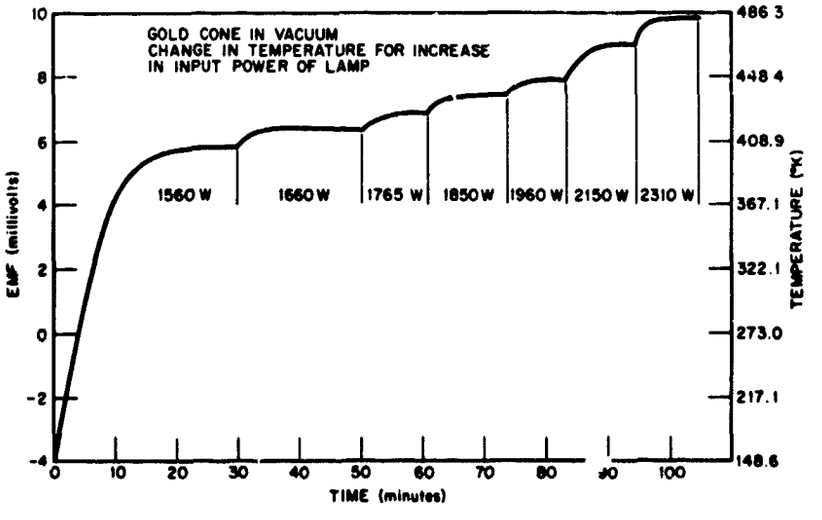


Figure B-7–Variation of cone temperature for various power inputs to lamp.

Table B-1
Summary of Results Presented in Figures B-3 through B-7.

Fig.	Lamp Input (watts)	Press.	Equil. Front (°K)	Temp. Rear (°K)	F-R Diff. (°K)	Rise To 90% (sec)	Fall To 90% (sec)	Cone Type
B-3	2300	Atm	327.6	326.9	.7	100	180	30° black
	2700	Atm	332.8	331.9	.9	160	–	30° black
B-4	Off	Vac	–	–	–	–	840	30° black
B-5	2700	Vac	320.0	318.6	1.4	220	840	30° black
	2300	Vac	302.5	300.9	1.6	270	540	30° black
	2300	Vac	485.2	481.9	3.3	840	2400	40° gold
B-6	Off	Vac	–	–	–	–	2400	40° gold
B-7	Var	Vac	–	–	–	–	–	40° gold

ANALYSIS OF THE RESULTS

The results of each test were analyzed to determine the feasibility of using the cone as a relative or absolute method for measuring the radiant flux density. The analysis showed the relative importance of the parameters which govern the final equilibrium temperature of the cone. Some of the computations which seemed more significant will be briefly discussed here.

Under certain simplifying assumptions the heat balance equation of the cone at equilibrium temperature may be written as follows:

$$P A_1 \alpha = \sigma (\epsilon_1 A_1 + \epsilon_2 A_2) (T_e^4 - \bar{T}_{v.c.}^4), \quad (1)$$

where P is the flux per unit area from the Hg - Xe lamp incident on the cone aperture; α, ϵ_1, A_1 are respectively the absorptance, emissivity and area of the cone aperture; ϵ_2, A_2 are the emissivity and area of the outside curved conical surface; σ is the Stefan-Boltzmann constant, T_e, $\bar{T}_{v.c.}$ are the equilibrium temperature of the cone and the average radiative temperature of the inner surface of the vacuum chamber. It is assumed that the vacuum chamber is sufficiently large and of absorptance equal to unity, so that the view factor of the cone need not be considered.

If the test is conducted under atmospheric pressure, an extra term expressing the loss of heat due to convection has to be added to the right-hand side of Equation (1).

The value of α may be assumed to be unity. P is known from independent thermopile measurements. The other quantities can be experimentally determined. Assuming $\bar{T}_{v.c.}$ to be that of liquid nitrogen, T_e was calculated for each of the tests. In every case the calculated temperature was seen to be lower than the observed temperature. An obvious explanation is that a considerable portion of the radiation entering the chamber falls on the back wall of the shroud, and hence the average radiative temperature $\bar{T}_{v.c.}$ is very much higher than the liquid nitrogen temperature recorded by the thermocouples on the outer walls of the shroud. The gradual increase of T_e under atmospheric pressure and no LN_2 cooling (Figure B-3), contrasted with the relative steadiness of T_e with LN_2 cooling (Figure B-5) shows the effect of a gradual heating of the shroud, and a consequent change in $\bar{T}_{v.c.}$.

Hence $\bar{T}_{v.c.}$ is an important and relatively uncertain parameter. It can be calculated from the energy balance equation during the cooling phase. Equating the heat lost by radiation to the thermal loss of the cone as shown by the rate of fall of temperature, we obtain:

$$\bar{T}_{v.c.} = \left[\frac{\sigma (\epsilon_1 A_1 + \epsilon_2 A_2) T_c^4 + m C(T_c) dT_c/dt}{\sigma (\epsilon_1 A_1 + \epsilon_2 A_2)} \right]^{1/4} . \qquad (2)$$

In the above equation, T_c is the temperature and dT_c/dt is the time rate of change of temperature of the cone at time t, m is the mass of the cone and $C(T_c)$ is the specific heat of the material of the cone at temperature T_c.

The values of T_c and dT_c/dt can be experimentally determined from the cooling curve. Using Equation (2), the values of $\bar{T}_{v.c.}$ were calculated at five different points on the cooling curve of Figure B-4. The results are presented in Figure B-8. The computations show that $\bar{T}_{v.c.}$ is considerably higher than LN_2 temperature even after 20 minutes. This is due to the ports which are not covered by the LN_2 shroud. Values of $\bar{T}_{v.c.}$ computed at shorter time intervals after closing the shutter are higher. At t = 1 minute, $\bar{T}_{v.c.}$ is 150°K. An extrapolation of this curve from t = 1 minute to t = 0 gives an approximate value of $\bar{T}_{v.c.}$ during equilibrium conditions to be 165°K.

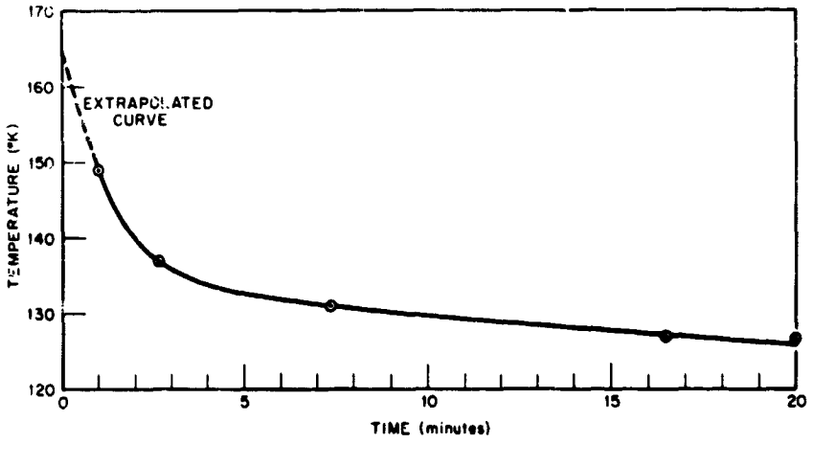


Figure B-8–Calculated curve of the average vacuum chamber temperature vs. time for the 30° blackened copper cone. Data for calculations taken from test 6.

Figure B-9 shows a similarly calculated curve of $\bar{T}_{v.c.}$ where the values of T_c and dT_c/dt were taken from test no. 18, which was run using a 40° gold cone in vacuum, of which the cooling curve is given in Figure B-6. The computations for $\bar{T}_{v.c.}$ show a value of 312°K after 3 minutes of cooling, and approaching 150°K after 100 minutes of cooling. An extrapolation of this curve from t = 3 minutes to t = 0 gives an approximate equilibrium value of $\bar{T}_{v.c.}$ of 360°K.

A partial explanation of the two different equilibrium values of $\bar{T}_{v.c.}$ which are "seen" by the blackened cone and the gold cone is that the value of $\epsilon_2 A_2$ is considerably different for the two cases, as shown in Figure B-10. The gold cone which has a low emissivity and absorptance on its slant surface is little responsive to the LN_2 shroud, and most of its heat exchange is with the hot quartz window of the chamber and the area surrounding it. The blackened cone, however, "sees" the whole vacuum chamber due to its highly emissive and absorptive slant surface. This shows that the assumption in Equation (1) that the view factor may be neglected is not fully justified. This result is

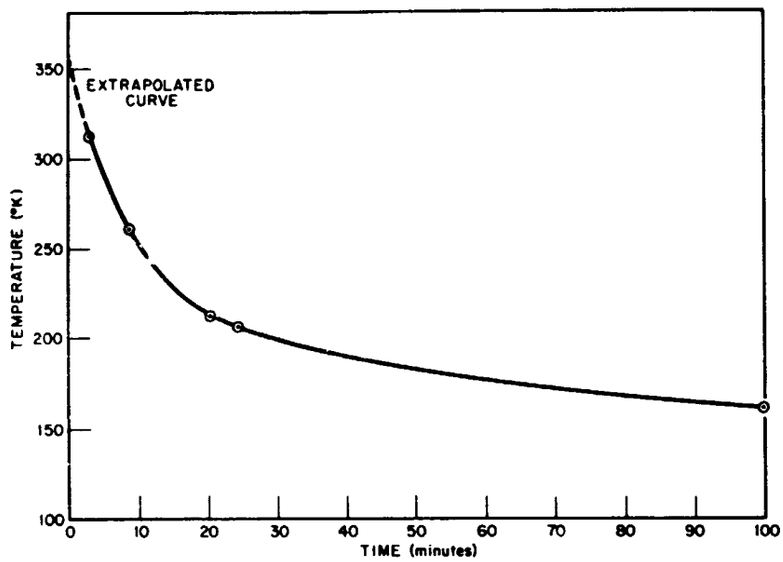


Figure B-9–Calculated curve of the average vacuum chamber temperature vs. time for the 40° gold plated copper cone. Data for calculations taken from test 18.

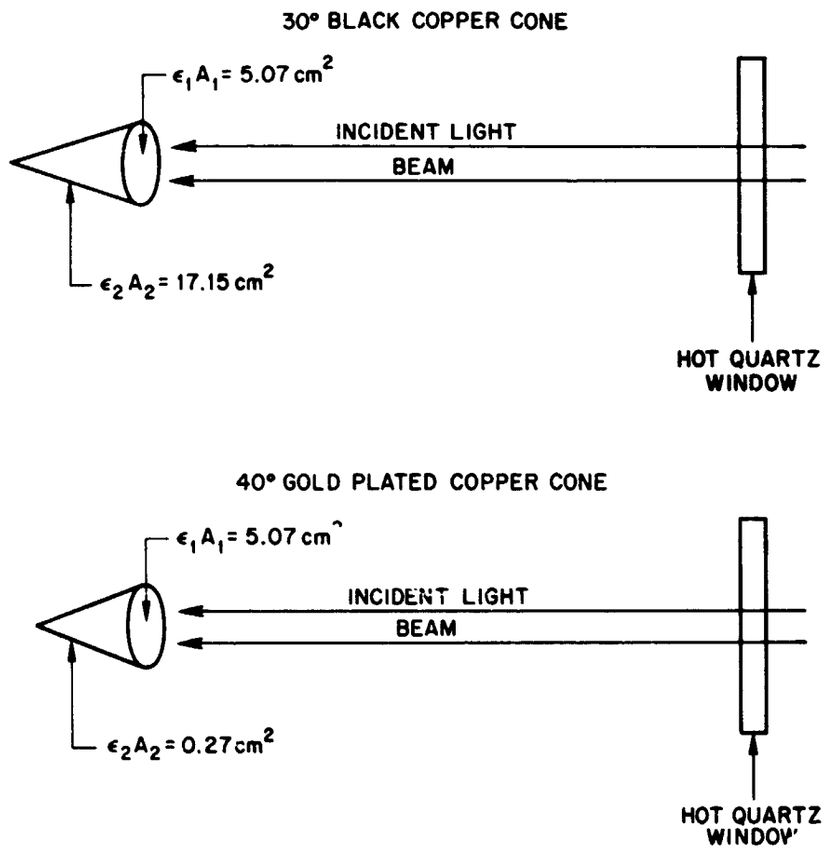


Figure B-10–Comparison of the emissive areas of the 30° black and 40° gold plated copper cones.

significant also in other tests made in a vacuum chamber, as for example, with a "black ball" spherical integrator or a black plate.

The theoretical equilibrium cone temperature, T_e, may be calculated from the following equation for the energy balance between the cone, its surroundings and the Hg-Xe lamp:

$$T_e = \left[\frac{P A_1 \alpha + \sigma (\epsilon_1 A_1 + \epsilon_2 A_2) \bar{T}^4_{v.c.} - 2 h A_1 \Delta T}{\sigma (\epsilon_1 A_1 + \epsilon_2 A_2)} \right]^{1/4}. \quad (3)$$

This equation can be obtained by adding the term $2h A_1 \Delta T$ for convection losses Reference (1) to the right hand side of Equation (1) and solving for T_e. The quantity h is an approximate convection coeffecient and ΔT is the temperature difference between the cone and the surrounding air. An approximation of the 30° cone with a horizontal cylinder of mean radius $\bar{r}$ and with a length equal to that of the cone, so that the cylinder and the cone have the same area, was used in Equation (2) for calculating h. The convection coefficient h for a horizontal cylinder of radius $\bar{r}$ is given by

$$h = 1.53 \times 10^{-4} \left(\frac{\Delta T}{2\bar{r}} \right)^{1/4}. \quad (4)$$

The factor of 2 in the convection loss term, $2h A_1 \Delta T$ of Equation (3) comes from the assumption that both the inner and outer surfaces of the cone will be losing energy to the surrounding air by convection at the same rate.

The values of T_e as given by Equation (3) were calculated for four of the tests with the 30° black cone and for three of the tests with the 40° gold cone. The results are presented respectively in Tables B-2 and B-3 in the fifth column. At low pressure the convection term was set equal to zero. The incident flux P had been experimentally determined only for lamp power of 2700 watts. The flux at the lower power levels was computed on the assumption established by independent measurements that within narrow ranges the radiant flux due to the Hg-Xe lamp is nearly proportional to the input power. The term which is the least certain is $\bar{T}_{v.c.}$, the average radiative temperature of the chamber. The values 160°K and 360°K for the black cone and gold cone respectively at low pressure were obtained by the extrapolation method discussed earlier (Figures B-8 and B-9). For the tests at atmospheric pressure, the value 298°K was assigned on a semi-empirical basis. The method consisted of measuring the temperature T at different points on the inside of the shroud and taking the fourth root of the area-weighted average of T^4. This method was also applied to the case of the black cone in vacuum and yielded a value close to 160°K. The method is not, however, applicable for the case of the gold cone in vacuum, because of the dominant effect of the chamber wall facing the cone and the entrance port.

A comparison of the theoretically calculated temperature and the corresponding experimentally measured temperature shows that they differ on an average by 3.2 percent for the tests made at atmospheric pressure and by 7.3 percent for those made in vacuum.

The difference between the theoretical and experimental values of the equilibrium temperatures can be explained by two factors: first, the actual value of the average vacuum chamber temperature is substantially higher than the values used in the calculations because the quartz window, the entrance port of the light beam, and the area surrounding it are being heated to relatively high temperatures, and secondly, some of the incident light scattered or reflected by the back wall of the vacuum chamber and absorbed

Table B-2
Theoretically Calculated and Experimentally Observed Equilibrium Cone Temperatures for the 30° Black Cone

Tank Press. (mm Hg)	Hg-Xe Lamp Power (watts)	Incident Flux P (watt/cm²)	Av. Vac. Chamber Temp. (°K)	Theo. Cone Temp. (°K)	Exp. Cone Temp. (°K)
5×10^{-7}	2700	0.150	160	287	320
5×10^{-7}	2300	0.128	160	283	300
760	2700	0.150	298	316	332
760	2300	0.128	298	313	327

Table B-3
Theoretically Calculated and Experimentally Observed Equilibrium Cone Temperature for the 40° Gold Cone

Tank Press. (mm Hg)	Hg-Xe Lamp Power (watts)	Incident Flux P (watt/cm²)	Av. Vac. Chamber Temp. (°K)	Theo. Cone Temp. (°K)	Exp. Cone Temp. (°K)
5×10^{-7}	2350	0.131	360	443	485
760	2300	0.128	298	334	342
760	2700	0.150	298	356	351.5

by the cone has not been considered in setting up the energy balance, Equation (1). These two factors were not accounted for in the initial calibration of the incident flux with the Eppley thermopile at atmospheric pressure. The effect of these factors is to increase the flux falling on the cone and thus to give relatively higher experimental cone temperatures.

The time required for the cone temperature, T_c to increase from some initial value T_{c1} to some final value T_{c2} can be calculated from the integrated form of the following energy balance equation, which holds during the time interval in which the cone temperature is increasing (for the case of vacuum):

$$P A_1 \alpha = \sigma \epsilon A (T_c^4 - \bar{T}_{v.c.}^4) + m C(T_c) \frac{dT_c}{dt} \quad . \tag{5}$$

Here for convenience sake we have defined $\epsilon A = \epsilon_1 A_1 + \epsilon_2 A_2$. From Equation (1)

$$P A_1 \alpha = \sigma \epsilon A (T_e^4 - \bar{T}_{v.c.}^4). \tag{6}$$

Subtracting Equation (6) from Equation (5), transposing terms, and integrating over t gives

$$t_2 - t_1 = -\frac{m}{\sigma \epsilon A} \int_{T_{c1}}^{T_{c2}} \frac{C(T_c)\, dT_c}{T_c^4 - T_e^4} .$$

Now, for a small interval of time, we may approximate $C(T_c)$ as $C_0 + C_1 T_c$, where C_0 and C_1 are constants. Thus,

$$t_2 - t_1 = -\frac{m}{\sigma \epsilon A} \int_{T_{c1}}^{T_{c2}} \frac{(C_0 + C_1 T_c)\, dT_c}{T_c^4 - T_e^4} .$$

Integrating the right hand side over T_c, we obtain

$$t_2 - t_1 = \frac{m C_0}{2 T_e^3 \sigma \epsilon A} \left\{ \frac{1}{2} \ln \left[\frac{(T_e + T_{c2})(T_e - T_{c1})}{(T_e - T_{c2})(T_e + T_{c1})} \right] + \tan^{-1}\left(\frac{T_{c2}}{T_e}\right) - \tan^{-1}\left(\frac{T_{c1}}{T_e}\right) \right\}$$

$$+ \frac{m C_1}{4 T_e^2 \sigma \epsilon A} \ln \left[\frac{(T_e^2 + T_{c2}^2)(T_e^2 - T_{c1}^2)}{(T_e^2 - T_{c2}^2)(T_e^2 + T_{c1}^2)} \right] . \qquad (7)$$

Two theoretical rise curves calculated from Equation (7) for the black cone are shown in Figure B-11, along with the corresponding experimental curve for the same cone. The values of the equilibrium cone temperatures, T_e, used in Equation (7) for calculating the theoretical curves were 287°K and 320°K. These temperatures are respectively the theoretical and experimental equilibrium values of cone temperatures shown in Table B-2, line 1.

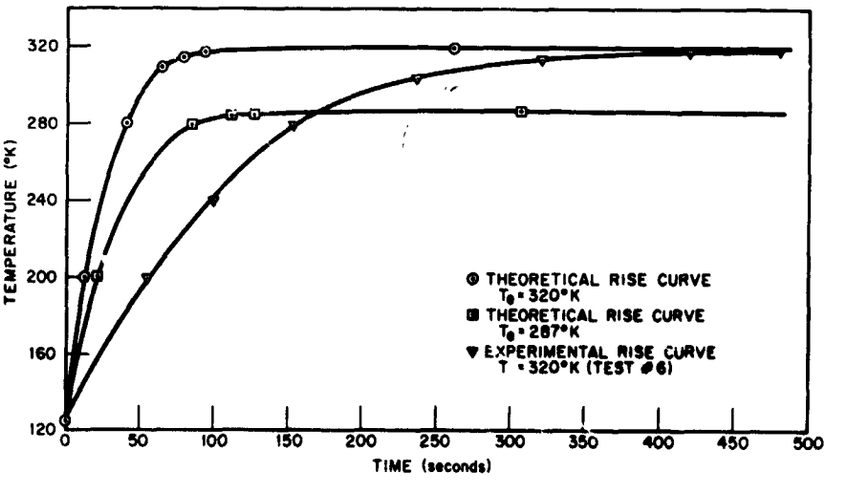


Figure B-11–Theoretical and experimental curves of cone temperature vs. time for the 30° blackened copper cone.

From Figure B-11 it is seen that the rise times for the theoretical curves with T_e = 320°K and 287°K are 95 and 112 sec respectively; the rise time for the experimental curve, however, is 420 sec. The rise time as used here is defined as the time required to reach the temperature $T_{ci} + 0.99\,(T_e - T_{ci})$, where T_{ci} is the initial cone temperature.

It is probable that a combination of three effects is causing this long experimentally observed rise time. First, since the cones are only 6 mil thick, the present method of attaching the thermocouples to the cone with a small bit of epoxy will cause the cone to have a high local thermal mass at the position of the thermocouple. Secondly, the low thermal conductivity of the epoxy relative to that of copper may act partially to insulate the thermocouple from the cone. Thirdly, if the high experimental cone temperature of 320°K (theoretical cone temperature of 287°K) is attributed to a slow increase in the vacuum chamber temperature with time, then the equilibrium value of the cone temperature, T_e, will also be increasing with time. From Equation (7) it is seen that the effect of small increases in the temperature T_e, when the difference $(T_e - T_{c2})$ is small, is greatly to increase the rise time.

TESTS IN THE WESTINGHOUSE CHAMBER

The analysis of the results of tests made in the relatively small SEAR chamber showed that several improvements are required in the experimental arrangement. A schematic of the improved facility which was used in the latter phase of the research is shown in Figure B-12. A considerably larger vacuum chamber, the Westinghouse chamber No. 245, was used instead of the SEAR chamber. The cone to be tested was mounted inside a cylindrical vessel which was covered by a circular plate with a 1 inch aperture. This formed a small shroud within the larger shroud of the vacuum chamber. It was cooled by LN_2. The aperture of the cone alone was exposed to the radiation and the inside of the small shroud was maintained at a more uniform temperature. A baffle with a 2 inch aperture was placed between the quartz window of the vacuum chamber and the lamp assembly. An air blower helped to maintain the temperature of the quartz window near the room temperature.

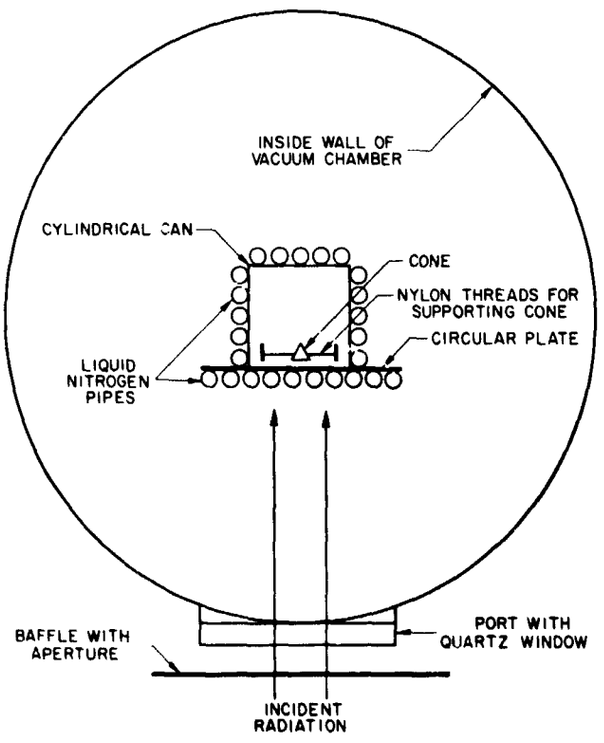


Figure B-12–Top view of new experimental arrangement.

Figures B-13 and B-14 show the circular plate and cylindrical vessel which form the inner shroud. Figure B-13 is the front view, and Figure B-14 is the rear view. The front plate is bolted to the rectangular frame which hangs from the upper lid of the vacuum chamber. The 1 inch aperture of the front plate and the LN_2 coils are clearly visible in the figure. The rear view, Figure B-14, shows the Eppley thermopile clamped to the front plate, as also the hose for water cooling to the right and the cylindrical vessel to the lower left. The cylindrical vessel was bolted to the front plate and the whole assembly was lowered into the vacuum chamber before taking the measurements.

The source of radiation was a d.c. stabilized quartz-iodine lamp. It provided a more steady radiant flux than the Hg-Xe arc.

Two types of conical receivers were used, the copper conical shell painted black both outside and inside, which was used in earlier tests, and a

Figure B-13–Front plate of the inner shroud, showing LN_2 coils and one inch aperture at the center.

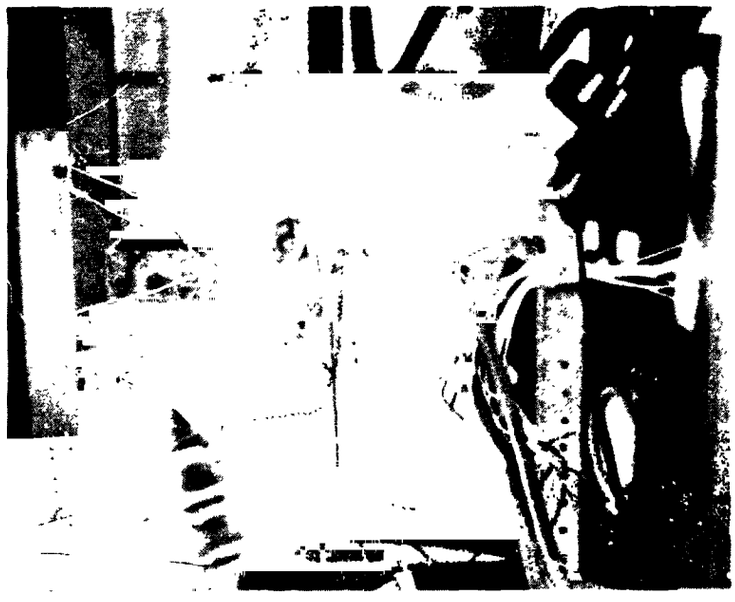


Figure B-14–Rear view of the front plate, with Eppley thermopile mounted in position. The cylindrical vessel with LN_2 coils has been detached and is shown at lower left.

wire-wound cone. The wire-wound cone was manufactured by Minco Products, Inc. of Minneapolis, Minnesota. The wire was of nickel-iron alloy, of thickness 0.0025 cm. It was wound to form a cone of apex angle 29° 45' and an aperture of diameter 0.9464 cm. The wall thickness was 0.0062 cm. The cone was suspended by Nylon threads inside a small copper cylinder open at both ends. The terminals for the resistance wire were provided on the outer surface of the copper cylinder.

The experimental procedure with the copper cone was the same as in the previous tests. The cone was exposed to radiation from the tungsten lamp. The e.m.f. of the two thermocouples was measured at equal intervals of 20 seconds until equilibrium conditions were attained. The average readings of the thermocouples for four different flux levels in the first series of measurements are shown in Figure B-15.

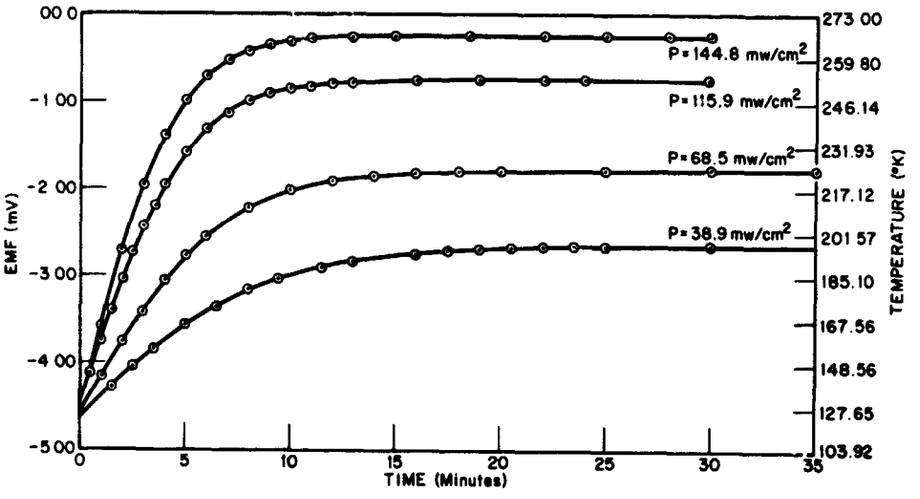


Figure B-15–Heating curves for the 30° copper cone in vacuum when exposed to radiation of intensities 144, 115.9, 68.5, and 38.9 mw/cm^2.

The flux levels were changed by changing the power input to the lamp. The value of the flux in each case was measured independently before the cone was mounted in the chamber, when the inside of the chamber was at atmospheric pressure. An Eppley thermopile, calibrated with reference to a standard lamp, was used for these measurements. Because of the size of the thermopile casing it was not possible to measure the flux at the same distance (0.68 cm) from the front plate at which later the cone was to be located. Hence measurements were made at three slightly greater distances, and the value at 0.68 cm was obtained by extrapolation. The estimated error in this method of extrapolation is ±2 percent.

A summary of the results of the tests with the copper cone is presented in Table B-4. Two series of tests were performed. The first which is presented in Figure B-15 was with an aperture diameter of 1" in the LN_2 cooled front plate of the inner shroud. For the second series of tests the aperture was slightly enlarged in order to lessen the risk of any part of the cone opening being shielded by the front plate, without, however, letting too much of the radiation enter the shroud. The equilibrium temperature was determined from the average of the readings of the two thermocouples. The average radiative temperature $\bar{T}_{v,c}$ of the inner walls of the small cylindrical shroud was assumed to be the same as that recorded initially by the thermocouples before the shutter was opened. This value ranged between 121°K and 127°K, considerably higher than that of LN_2. Even allowing the system to settle down under continual LN_2 flow for a whole day did not bring the initial temperature of the cone any lower. Columns 5 and 6 of Table B-4

Table B-4
Results of Tests with the 30° Black Copper Cone

Aperture Diam. (in.)	Equil. Temp. T_e (°K)	Initial Temp. $\overline{T}_{v.c.}$ (°K)	Lamp Power L_p (watts)	Intensity, $p\left(\frac{mw}{cm^2}\right)$			calc $\epsilon_2 = .96$ / Eppley Ratio
				Calculated		Eppley	
				$\epsilon_2 = .86$	$\epsilon_2 = .96$		
Distance from Cone to Front Plate = 0.68 cm.							
1.00	196.65	121.16	178.8	32.55	35.39	38.9	.9098
	223.97	122.02	303.0	57.30	62.30	68.5	.9095
	253.15	127.65	495.9	95.92	104.28	115.9	.9000
	266.79	125.87	613.7	120.23	130.73	144.8	.9028
Distance from Cone to Front Plate = 0.34 cm.							
1.04	194.07	123.65	163.54	29.60	32.19	35.5	.9068
	220.15	123.65	276.75	52.85	57.46	63.5	.9049
	248.89	123.65	455.2	90.05	97.91	108.2	.9049
	262.39	123.65	563.4	112.61	122.44	135.2	.9056

gives two sets of values of calculated intensity. These values were calculated on the assumption that Equation (1) is strictly applicable. An important parameter in this equation is ϵ_2, the emissivity of the 3-M velvet paint of the outside curved surface of the cone. The value of 0.86 had been accepted in all earlier calculations, but more recent measurements made by J. Triolo of the Thermal Systems Branch, Spacecraft Technology Division, seem to show that 0.96 is a more accurate value for ϵ_2. ϵ_1 and α are taken to be unity. The computed values are compared in columns 7 and 8 with the values obtained by measurements with the Eppley thermopile. The final column gives the ratio of the values in columns 6 and 7.

The measured values of flux were obtained from the input power of the lamp with the aid of the graph shown in Figure B-16. The relation between lamp power and flux is approximately linear.

We observe that the ratios for all 8 tests lie between 0.90 and 0.91. One possible source of error which was suspected after the first series of tests was that the inflow of LN_2 to the front plate causes a slight warping of the brackets which support the cone, and hence that the front plate partially shields the cone aperture. For this reason, the front plate aperture was slightly increased for the second series of tests. These tests were performed with a different lamp and after an independent Eppley measurement. The ratio of calculated to measured radiant flux has practically the same value in the second series of tests.

Another possible explanation is that the flux per unit area is non-uniform in the plane of the cone aperture, being higher at the center and less on the periphery. The sensitive area of the Eppley thermopile being considerably smaller than that of the cone, the flux per unit area as recorded by the Eppley thermopile would be higher.

Perhaps a more significant error is introduced in the assumption that the initial temperature of the cone is the average radiative temperature of the shroud, and that the

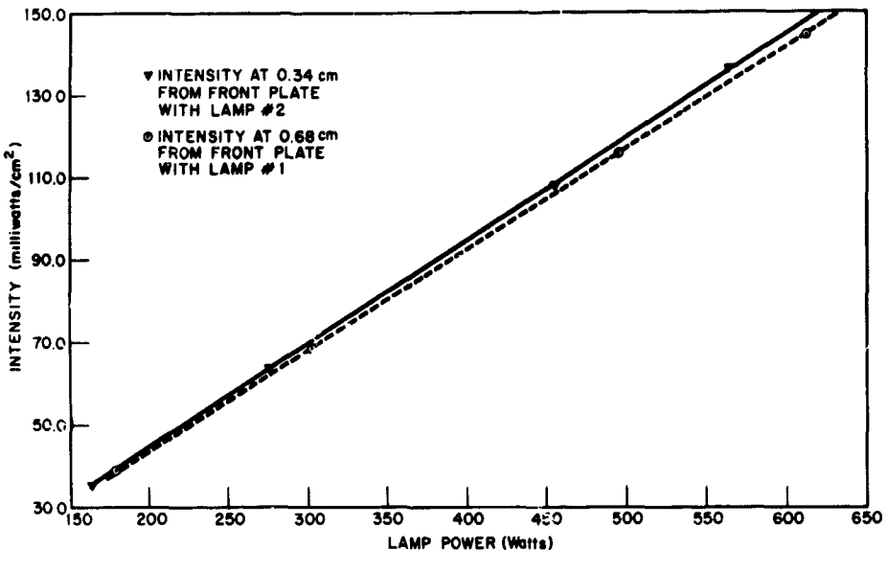


Figure B-16–Calibration of the intensity of two 600 watt quartz iodine lamps with Eppley thermopile # 5888 as a function of lamp power.

thermocouples indicate correctly the final temperature of the cone. Because of the insulating effect of the epoxy the temperature of the thermocouple is possibly less than that of the cone, and the radiative temperature of the shroud is closer to that of LN_2. Increasing T_e or decreasing $\overline{T}_{v.c.}$ would increase the calculated value of the radiant flux.

The wire-wound cones eliminate the need of measuring the temperatures of the cone and of the shroud, and hence are capable of yielding a higher degree accuracy. The cone is surrounded by a cylinder which is blackened on the inside. The outside of the cone has an enamel finish which lessens the radiant loss and raises the equilibrium temperature. The resistance of the cone under equilibrium conditions was measured by means of a Shallcross 638-R Kelvin-Wheatstone bridge with an accuracy of ±1 ohm. Measurements were made at different power levels, and at each power level the readings were repeated on different days. The readings showed a fairly high degree of consistency, with an average deviation of 1/20th of one percent. Care is taken so that the bridge current passes through the cone only for a short interval of time and that no extra heating of the wire takes place because of the process of measurement. Equilibrium conditions are reached in about 6 to 10 minutes after exposing the cone to the radiation. Measurements of resistance were made at intervals of 2 minutes for about 10 minutes after steady state was reached. The rate of change of resistance with temperature for one of the typical series of tests is shown in Figure B-17. These tests were made in the first week of September, 1965. Three other series of tests similar to the one shown in the figure were made on the following days. A random sequence of power input to the lamp was intentionally chosen to find out whether the resistance showed any influence of temperatures attained in earlier tests. No such influence was detected.

After making the measurements with the radiant flux, the shutter is closed, the cone is allowed to come to a low equilibrium temperature, and then electric current from an external source is passed through the wire. A simple electrical circuit shown in Figure B-18 was employed for heating the wire and for measuring the voltage and current supplied to the wire. At each power setting a Dana digital voltmeter is used to measure the voltage across the leads and that between the terminals of a 1 ohm precision resistance. The resistance of the leads to the cone was determined by a separate experiment by shunting the cone and was found to be 2.054 ohms. If E is the voltage, I the current, r the resistance of the leads, the resistance of the cone is $R = E/I - r$; and the electric power supplied to the cone is $P_E = I^2R$.

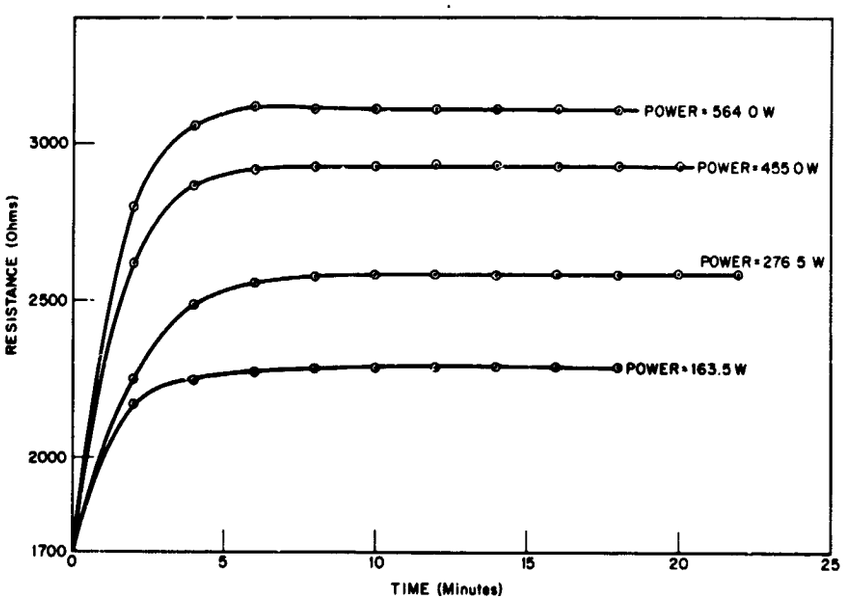


Figure B-17–Wire-wound cone: rate of change of temperature with time for four values of power input to the lamp tests made on Sept. 2, 1965.

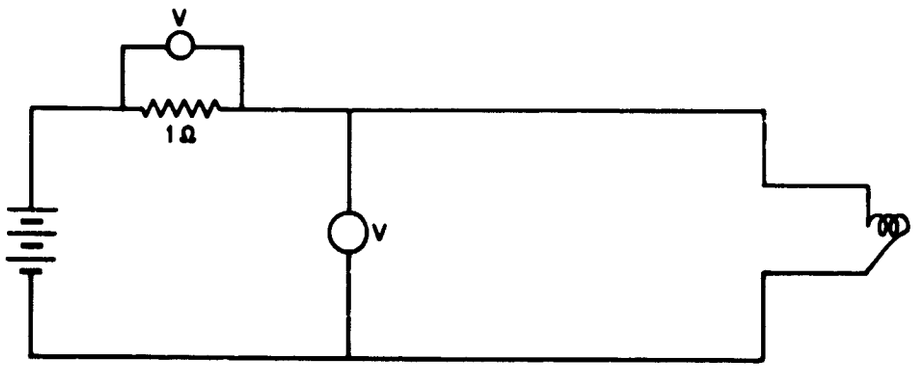


Figure B-18–Electrical circuit for heating wire-wound cone.

The electric power for any given resistance may be determined by drawing a graph of P_E versus R. The observed points lie on a smooth curve which is very nearly a quadratic. A more accurate procedure is to use a least-squares method for a curve fitting program and to prepare a table of values of P_E for R. A third power polynomial curve was fitted to 12 pairs of values of P_E and R. The maximum difference between the observed and calculated values of P_E was 0.06 mw/cm² for P_E about 100 mw/cm².

An alternate method which does not require a computer but can yield sufficiently accurate results is to draw a curve showing the departure of measured values of P_E from values P_c calculated on the assumption of a linear dependence of P on R, according to the equation $P_c = mR + C$ where m is the slope of the line relating P_E and R, and C is a constant. In the case of the set of measurements made on August 16, 1965, m = 0.08475 mw/cm² -ohm and C = - 184.37 mw/cm². Figure B-19 shows the curve of the difference between observed and calculated values of P as a function of R. P_E for any value of R obtained by exposing the cone to the source of radiation is given by $mR + C + \Delta P$. This

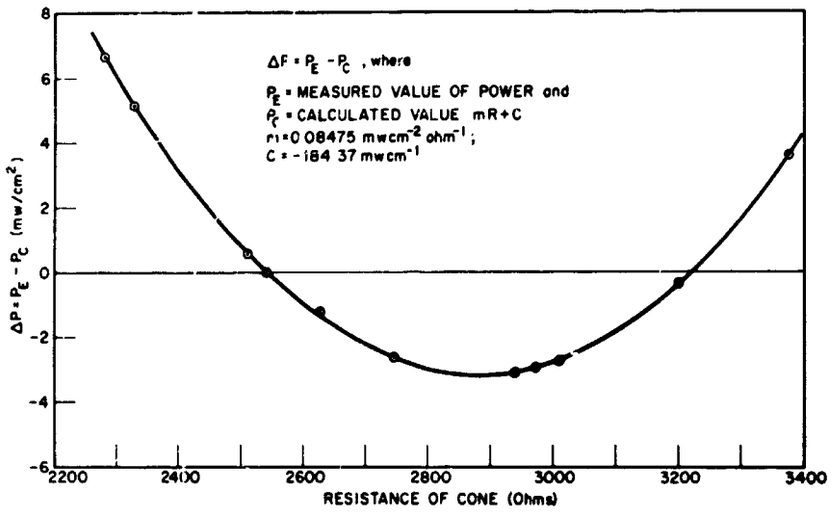


Figure B-19–Curve of Δp versus R for determining power of the cone for given resistance.

graphical method enables the power to be read correct to 0.01 mw/cm², which corresponds to an accuracy of .02 percent.

An analysis of two sets of measurements of P and R made on two days three weeks apart by two different experimenters showed that each set of values had an internal consistency of 0.03 percent, and that the two sets differed from each other by 0.3 percent. This latter difference must be attributed to a slight change in the temperature of the inner shroud.

The theory underlying the case of the wire-wound cone is as follows. During the first phase, when the cone is illumined by the external source of radiation, the heat balance equation is the same as Equation (1): $P A_1 \alpha = \sigma (\epsilon_1 A_1 + \epsilon_2 A_2)(T_e^4 - \bar{T}_{v.c.}^4)$. During the second phase, when the cone is heated by electric power, if it were possible to bring it to the same equilibrium temperature T_e, average radiative temperature of the inner shroud being maintained constant, the heat balance equation would be

$$P_E = \sigma (A_1 \epsilon_1 + A_2 \epsilon_2)(T_e^4 - \bar{T}_{v.c.}^4); \text{ hence}$$

$$P = P_E / A_1 \alpha. \qquad (8)$$

Since the inner shroud is cooled by LN_2, $\bar{T}_{v.c.}$ remains constant; and hence the value of P for different values of R can be determined once for all. Any error or approximation in the right-hand side of the equation does not affect the final value. Thus we have in the wire-wound cone an absolute method of measuring the radiant flux.

The results of the tests made on August 16 and succeeding days are presented in Table B-5. The aperture of the cone was at a distance 0.34 cm behind the front plate. The input power of the lamp was set at 6 different values ranging from 164 to 563 watts. The third column gives the resistance of the cone as measured by the Wheatstone bridge. The electrical power which would give the same resistance as given in the fourth column was determined from the second phase of the experiment. The dimensions of the cone aperture were carefully measured with a travelling microscope. The inner area of the cone is (0.7034 ± .0035) cm². The rim of the cone which is also exposed to the radiation is partially black (α = 0.9) and partially varnished with a yellow tint (α = 0.6); it has an effective area of (.0064 ± .002) cm². The total area of the cone aperture is thus A_1 = 0.7093 cm² with an uncertainty of 0.77 percent.

Table B-5
Results of Tests with the 29° 45' Wire-Wound Cone

Lamp Power (watts)	Equil. Cone Resistance (ohms)	Electrical Power to Cone (mw) P_E	Calc. Intensity (mw/cm²) P	Eppley Intensity (mw/cm²)	Calc/Eppley Ratio
Distance From Cone to Front Plate: 87 cm					
163.9	2400	24.9	35.28	35.6	.991
277.2	2716	44.7	63.34	63.8	.993
455.2	3110	76.75	108.76	108.2	1.005
457.2	3115	77.0	109.11	108.4	1.006
565.0	3306	95.7	135.61	135.5	1.001
563.4	3307	95.75	135.68	135.2	1.003

Another equally important parameter is the absorptance, α, of the cone aperture. A survey of pertinent literature shows that there are considerable differences between authors concerning this. André Gouffé (2) was one of the first to treat the problem extensively on a theoretical basis. He assumes the inner surface of the cone to be a perfect diffuser. De Vos (4) later developed a theory which does not require this assumption. A more recent discussion by Sparrow and Jonsson (4) considers the radiant interchange within the conical cavity and gives numerical results for effective emissivity which are claimed to be highly accurate. Sparrow and Jonsson state that "a comparison with the approximate analysis of Gouffé disclosed large errors in this prior work." A disturbing feature about the results of Sparrow and Jonsson is that the apparent hemispherical emissivity of the cone does not tend to unity as the cone angle tends to zero, whereas in Gouffé's theory it does tend to unity. Apparently little theoretical work has been done on the absorptance of a conical cavity. E. J. Gilham (5) of the National Physical Laboratory, Teddington, England, has discussed the use of a cavity type detector as an absolute radiometer. The cavity is not a cone but a cylinder with baffles. Using a paint of absorptance 0.98, Gilliam assumes the effective absorptance of the cone to be 0.999. We accepted on a provisional basis the theory of Gouffé and assumed that the absorptance is the same as the emissivity, and that the emissivity is given (6) by the formula $\epsilon = \epsilon_0 (1 + k)$

where

$$\epsilon_0 = \frac{\epsilon_s}{\epsilon_s\left(1 - \frac{A}{S}\right) + \frac{A}{S}},$$

$$k = (1 - \epsilon_s)\left(\frac{A}{S} - \frac{A}{S_0}\right). \qquad (9)$$

Here ϵ is the effective emissivity of the aperture of the cone, ϵ_s is the emissivity of the inner surface, A is the aperture area, S is the total area including the conical curved surface and the aperture, S_o is the area of a spherical cavity having the same length as the cone. The dimensions of the wire-wound cone used in our measurements were substituted in Equation (9), and ϵ was computed for different values of ϵ_s, .7, .8, .95, .96.

The corresponding values of ϵ were found to be respectively .9406, .969, .995 and .996. In our present case ϵ, as also α, was assumed to be 0.9945, which corresponds to a value ϵ_s = 0.945 for the absorptance and emissivity of the black paint inside the cone.

These values of A_1 and σ were substituted into Equation (8) to give the calculated values of intensity given in column 5 of Table B-5.

The next column gives the intensity values as given by the Eppley thermopile. As in the case of the copper cone, these values were obtained not by direct measurement in the position occupied by the cone aperture, but by extrapolation from measurements made at slightly greater distances.

The final column shows that the two sets of values are in close agreement. The agreement is in fact closer than might be expected given the many sources of error in both the cone method and the Eppley measurement.

Figure B-20 presents the results of the measurements made with both the wire-wound cone and the copper cone in the Westinghouse chamber, as compared with the Eppley measurements. The points corresponding to the wire wound cone lie on a line at 45° to both axes, as should be expected. The two series of measurements made with the copper cone give another straight line inclined at 42° - 10' to the x-axis.

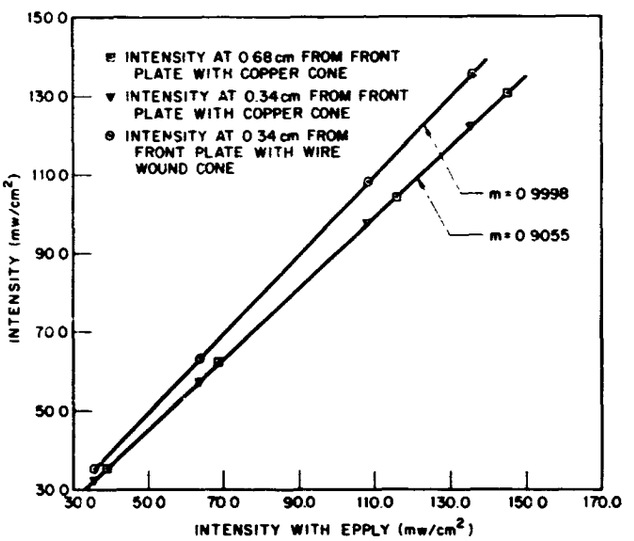


Figure B-20–Intensities as measured by copper and wire-wound cones vs. intensities measured by Eppley thermopile #5888.

CONCLUSION

From these results it may be concluded that the wire-wound cone gives a highly accurate technique for measuring radiant flux. The technique is slow and cannot be used for a rapidly changing flux. If sufficient care is taken to maintain the ambient temperature constant, results of better than 1 percent accuracy can be obtained. The method has the great advantage of being an absolute method which does not require an independent source for calibration.

The copper cone is considerably less satisfactory, partly because the temperatures of the cone and that of the surrounding cannot be accurately determined.

ACKNOWLEDGMENTS

The authors wish to acknowledge their indebtedness to A. Winker and N. Zylich whose close collaboration was most valuable in all phases of the research. A. Winker made the experimental setup in most cases and took several of the measurements.

The success of this project owes a great deal to E. K. Plyler, Head, Physics Department, Florida State University, who was available for consultation frequently during the summer. Thanks are due also to R. Kruger and J. Rogers of the Thermodynamics Branch who did the preliminary work on the use of conical shells for radiometry.

REFERENCES

1. W. H. McAdams, Heat Transmission, McGraw-Hill Book Company, Inc., New York, (1954) page 177.

2. André Gouffé, Rev. Opt. 24, Nos 1-3 (1945).

3. J. C. De Vos, Physica, 20, 669 (1954).

4. E. M. Sparrow and V. K. Jonsson, J. Opt. Soc. Am. 53,816 (1963).

5. E. J. Gilham, "Recent Investigations in Absolute Radiometry," Proc. Roy. Soc. (London) Series A 269, 249 (1962).

6. Charles S. Williams, "Discussion of the Theories of Cavity Type Sources of Radiant Energy," J. Opt. Soc. Am. 51, 564 (1961).

THE TESTING AND CALIBRATION OF DETECTORS USED IN FLIGHT

E. I. Mohr

INTRODUCTION

Some members of Team A of the 1964 Summer Workshop, in collaboration with members of the former A & M Division Staff, studied and reported on the "Evaluation of a 6-foot Spherical Integrator as a Wide-Angle Source" (1). In that report data are given for the calibration of channel 3 of the Tiros five-channel medium-resolution scanning radiometer #306 by means of the spherical integrator. This calibration was compared with a calibration made by members of the former A & M Division staff by means of an extended source composed of a sheet of diffusing white Kodak paper illuminated by a 2740-candlepower tungsten filament lamp. The data given by the spherical integrator for intensities which correspond to given output voltages of the radiometer are from 28 to 35 percent higher than the intensities as determined by the diffusing white paper. It was concluded, therefore, that the method or methods used to calibrate a radiometer by the use of an extended source required further study.

In the above-mentioned study, the calibrated Eppley water-cooled thermopile No. 4928A was used to measure the radiant flux at the windowless port of the spherical integrator. Since the accuracy of the radiometer calibration is dependent on a correct evaluation of the radiant flux at the port of the integrator, it is necessary to re-examine the use of the thermopile in determining this flux, as well as the methods used to calibrate the radiometer.

For this reason a number of investigations were proposed. These were: (A) determine the field of view of the thermopile; (B) calibrate the thermopile by means of 200 w and 1000 w standard quartz-iodine lamps; (C) obtain an independent calibration of the thermopile by means of a blackbody at several temperatures; (D) obtain a new set of measurements of the response of the thermopile to the radiant flux of the spherical integrator; (E) compare the response of the thermopile to the radiant flux from a diffusing screen (Kodak paper) when illuminated by a standard 1000 w quartz-iodine lamp with the response when illuminated by a standard 2740-candlepower lamp; (F) compare the results obtained for the sensitivity of the thermopile with and without the use of diffusing plates.

Due to failure of some of the equipment, waiting for the repair of the same, and the necessity of repeating some investigations a number of times due to this failure, it was not possible to complete all of the investigations which were planned. The present report presents the results obtained for those investigations carried out. An attempt is made to evaluate their bearing on the main problem in spite of the fact that a number of the most significant portions of the investigation could not be made.

Field of View of Thermopile 4928A

The Eppley water-cooled thermopile No. 4928A has 15 Bi-Ag junctions. The "hot" junctions are mounted on a disk placed at the small end of a hollow truncated cone which has its large base at the front surface of the thermopile. This cone is 4.8 cm deep. The large and small bases have diameters of 1.8 cm and 1.0 cm respectively. The absorbing disk and the cone wall are coated with Parson's black. According to studies made by the National Bureau of Standards (2), Parson's black has a fairly constant spectral absorptance. The response of thermopile 4928A is rather slow, with a time constant of about five seconds.

In order to evaluate the radiant flux from an extended source, one must know the true field of view of the thermopile. This was determined from the dimensions of the thermopile and gave a half angle for the field of view of 10.73°.

Since it was difficult to measure the dimensions of the truncated cone and of the absorbing disk without damage to the coating of Parson's black, there was some uncertainty in the value of the field of view as calculated by means of the dimensions. For this reason it was proposed to determine the field of view directly by using the 0.6328 μ beam from a Spectra-Physics Gas Laser No. 131. This laser has a beam with a diameter of about 2 mm.

The thermopile and laser were mounted on an optical bench and adjusted so that the laser beam coincided with the axis of the thermopile and hence hit the absorbing surface at its center. The thermopile was mounted on a carriage which has a graduated circle and may be rotated about a fixed axis through an angle of 360°. The thermopile itself was mounted on this carriage in such a way that the axis of rotation of the carriage lay in the plane of the absorbing surface and also passed through the center of this surface. In this way the laser beam always fell on the center of the absorbing surface as the thermopile was rotated on the axis of the carriage. By keeping the carriage in a fixed position and rotating the thermopile about the fixed axis, it was possible to measure the angle through which the thermopile could be rotated and still "see" the laser beam. The thermopile response versus angle of the beam to the normal is shown in Figure B-21. Since the laser beam is not a geometrical ray or line, the value of the field of view was taken at the "half-width" of the response curve. This gave a half angle of the field of view of 10.68° compared to 10.73° calculated from the dimensions. The former is within 0.47% of the latter value, so that the calculated value may be taken as being correct within the limits of experimental error.

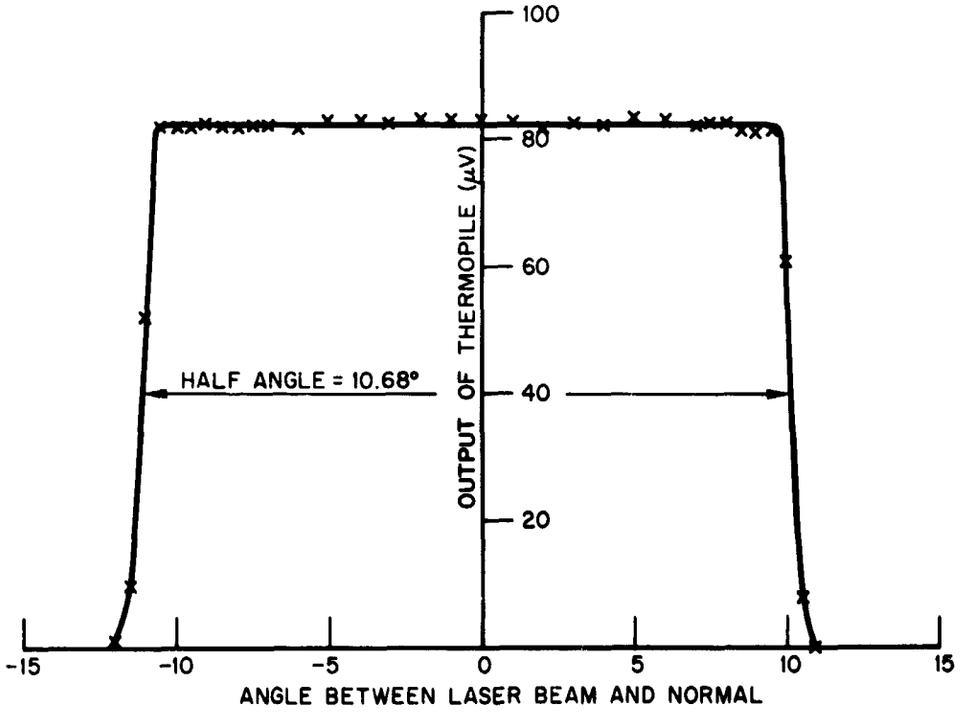


Figure B-21–Field of view of thermopile 4928A.

Calibration of Thermopile 4928A by Use of Quartz-Iodine Lamps

The calibration was made by use of 200 w and 1000 w quartz-iodine lamps which had been calibrated by the National Bureau of Standards. An unpublished Goddard report (3) by Thakaekara and Mohr discusses the extreme care with which such a calibration must be made.

The source and the thermopile were mounted on an optical bench. When using the 1000 w lamp as source, two parallel screens of masonite, rigidly held 3 cm apart by spacers, each one with an aperture of 3" × 2", were placed in front of the source. The apertures were machined to a knife edge. The screens had aluminum foil on the side toward the source and were coated with 3-M black velvet paint on the side toward the detector. They were rigidly mounted so that the apertures of the two screens were in alignment. (A different pair of parallel screens 5 cm apart with apertures of 2-1/4" × 1.5" were used with the 200 w lamp.) The optical bench was covered with a piece of black cloth. The source was shielded from the detector by a mask placed between the source and the screen aperture except when readings had to be taken. The intensities were varied by changing the distance between source and detector. The distances used were such that the inverse square law could be applied without correction for the finite size of the source.

The output of the thermopile was read on a Dymec Integrating Digital Voltmeter No. 2401 B . Two lamps, QM-95 and QL-45 were used as a source. The data are presented in Figure B-22. There is very close agreement between the responsivities of the thermopile to the two lamps. These responsivities will be compared with that for the blackbody at 1243°K later in this report. (See Figure B-27.)

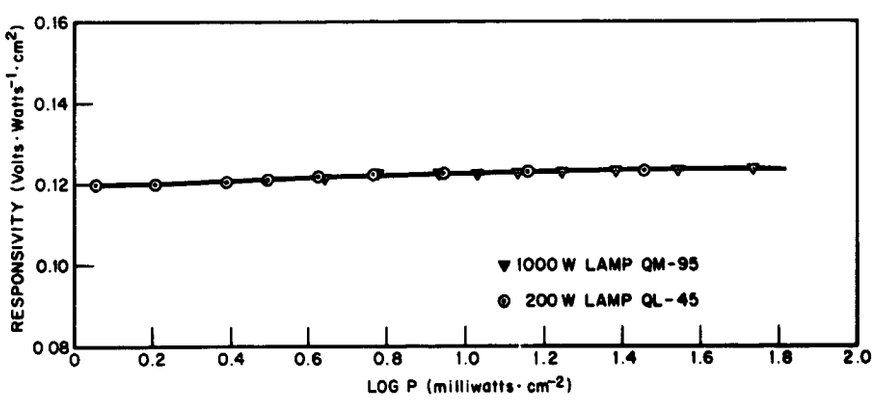


Figure B-22–Responsivity of thermopile 4928A to quartz-iodine lamps.

Independent Calibration of Thermopile 4928A by Means of a Blackbody at Several Temperatures

A Barnes Blackbody, Model RS-8B was used. It contains a central core with a conical cavity having a 14° apex angle. The conical cavity is finely grained and coated with a blackening material to give an emissivity of 0.99 ± 1% from 1 to 40 microns. It is designed to operate over a temperature range of 473°K to 1273°K. The temperature is set by means of a single turn potentiometer and held invariant by a proportional electronic controller. A blackened aluminum shield in front of the blackbody carries a selector dial which gives a choice of seven apertures ranging from 0.0081 inches to 0.2564 inches in diameter.

The thermopile was mounted on a one-meter optical bench placed in front of the blackbody so that the axis of the latter coincided with the normal to the center of the absorbing surface of the thermopile. This facilitated a change in the intensity of the radiation received by the detector by changing the distance between source and detector.

The shield or faceplate of the blackbody is not provided with a cooling system to keep it at ambient temperature. Therefore, it becomes quite hot and is a secondary source of radiation in the form of a graybody at a temperature materially above ambient temperature but well below the blackbody temperature. To prevent the latter radiation from reaching the detector, a shield was placed four centimeters from and parallel to the faceplate. The shield had an aperture 1.35 cm in diameter with its center coinciding with the axis common to blackbody and detector.

The output of the thermopile was read with the Dymec Integrating Digital Voltmeter. The zero reading was obtained by covering the shield aperture on the side toward the blackbody so that the thermopile was able to see everything in the field of view except the blackbody aperture.

In order to determine the time constant of the thermopile, the integrating voltmeter was connected to a print-out circuit that recorded the output voltage of the detector as seen by the integrating voltmeter once every second. Starting with the zero or background reading, the voltages were recorded second by second from the instant the blackbody aperture was uncovered until the voltage reached a maximum value. The detector output increased exponentially from zero to its maximum value in about 60 seconds. For this reason all readings of the detector response to blackbody radiation and to background or zero were taken after a 60 second waiting period.

The blackbody was operated at temperatures of 900°K, 1100°K and 1273°K as determined by the dial setting of the temperature controller. The radiant flux emitted per unit area into a hemisphere was determined by the Stefan-Boltzman Law $W = \sigma T^4$ in watts $\cdot$ cm^{-2}.

The radiation received by the detector was determined by the inverse square law as modified for radiation between two parallel disks of finite size. This equation, derived by H. B. Keene (4) is of the form $E = (1/4)\pi R(r_1 - r_2)^2$ where R is the net radiation per square centimeter per second, and r_1 and r_2 are the greatest and least distances between two points, one on each circle bounding the apertures. When this is rewritten and expanded one obtains, by neglecting higher powers, the form

$$E = \frac{R S_1 S_2}{\pi d^2} \left[1 - \frac{a^2 + b^2}{d^2}\right]$$

where S_1 and S_2 are the areas of blackbody and detector, with corresponding radii a and b respectively, and d is the perpendicular distance between the apertures. The correction given by this relation was a maximum of about 0.3% for the shortest distance between blackbody and detector used in the present investigation.

The responsivity of the thermopile in volts $\cdot$ watts$^{-1} \cdot$ cm^2 was determined for each combination of blackbody temperature and detector-blackbody separation used. The responsivity for each blackbody temperature was plotted against the common logarithm of the power incident on the detector. This is shown in Figure B-23.

An examination of the curves in Figure B-23 reveals a number of significant features. First the responsivity varies with blackbody temperature. For the same detector-blackbody separation, at 900°K the responsivity is 8% higher than at 1100°K, and the latter 5%

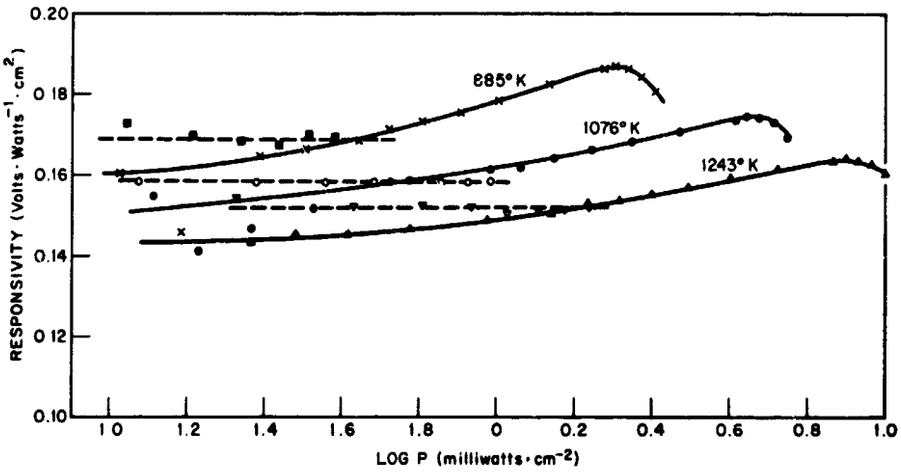


Figure B-23–Responsivity of thermopile 4928A to blackbody (shield with 1.35cm aperture diameter).

higher than at 1273°K. Since this might be due to an error in blackbody temperature, a platinum-platinum-rhodium thermocouple was inserted into the blackbody cavity to check the temperature. It was found that the temperatures were 885°K, 1076°K and 1243°K instead of 900°K, 1100°K and 1273°K respectively. These temperatures were double checked with a Leeds-Northrup Optical Pyrometer which gave values of 1075°K and 1246°K compared to 1076°K and 1243°K respectively. (The 885°K value cannot be measured with the optical pyrometer.) Using the corrected temperatures of the blackbody the responsivity was recalculated. The differences in responsivity for corresponding distances at the three temperatures changed from the former values of 8% and 5% to values of 4% and 6% respectively. This still represents a large difference in responsivity. Figure B-23 actually shows the responsivity for the corrected temperatures.

Secondly, it will be observed in Figure B-23 that the responsivity of the detector decreases for distances shorter than 14.5 cm. When the shield with an aperture of 1.35 cm was replaced by one with an aperture of 1.83 cm, the responsivity no longer decreased for the shorter distances. This is shown in Figure B-24.

This may be explained in terms of the reflection of radiation from the conical wall of the thermopile. This reflection was studied by means of the 0.6328 μ beam from the Spectra-Physics Gas Laser. The beam was directed normally at the center of the absorbing area of the thermopile to determine its response to the total laser beam energy as collimated by a very small aperture. The thermopile was then moved horizontally toward one side and the other a millimeter at a time, while keeping its axis parallel to the laser beam. In this way the laser beam was made to fall on various points along a diameter of the absorbing surface of the thermopile and also on various points along the conical wall. It was found that the energy received by the absorbing disk by reflection from the conical wall amounted to 5.4% of the energy received when it fell on the absorber directly. When this reflected energy is integrated over the whole area of the conical wall, if one assumes constant spectral reflectance for the Parson's black, it is found to amount to 12.9% of the energy received by the absorber directly. When using the shield with a 1.35 cm aperture, more and more of this conical wall would not "see" the blackbody as the distance decreased below 14.5 cm and consequently the fraction of energy received by the absorber by a single reflection would decrease continuously. Hence one would expect the responsivity to decrease for distances smaller than 14.5 cm. On the other hand, in the case of the shield with an aperture of 1.83 cm, all parts of the conical wall would see the blackbody at all distances so that no decrease would be expected for distances less than 14.5 cm.

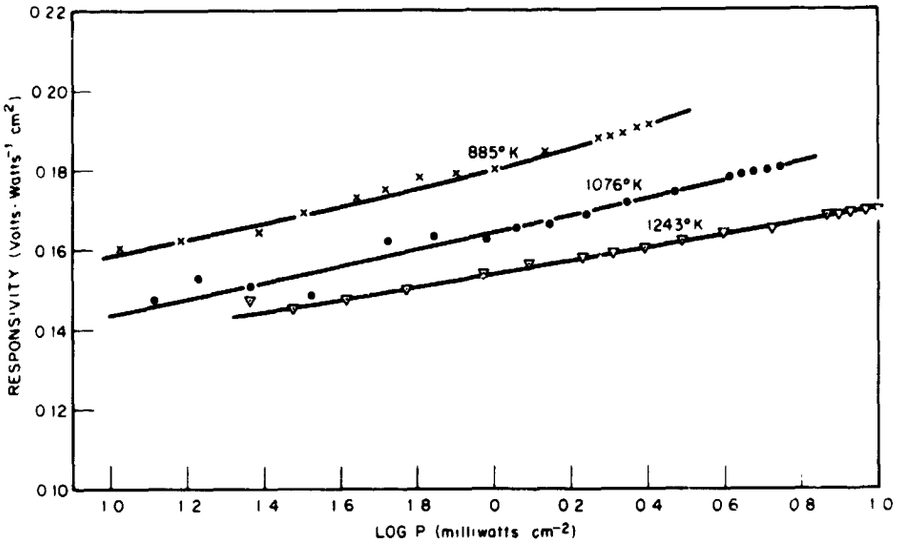


Figure B-24—Responsivity of thermopile 4928A to blackbody (shield with 1.81 cm aperture diameter).

Thirdly, it will be observed in Figure B-23 as well as Figure B-24, that for a given blackbody temperature the responsivity decreases with increasing distance between detector and blackbody. In other words, the detector receives smaller and smaller amounts of energy than has been calculated by the modified inverse-square law as the distance between detector and blackbody increases. Moreover the curves for the three temperatures do not have the same slope.

In order to check whether this might be a characteristic of the detector, a rotating sector disk was designed and used. By means of this sector disk, the intensity of the flux from the blackbody could be varied in steps of 1/8 while keeping the distance between the detector and blackbody constant. It was found that the responsivity of the detector remained constant within experimental error as the intensity varied for a given blackbody temperature. In this case the responsivity of the detector for the three temperatures differ by 6.6% and 4.5%. These data are plotted in Figure B-24 and indicated by the horizontal dashed lines. The corresponding data are given in Table B-6.

Table B-6
Responsivity of Thermopile 4928A Using Sector Disk with Blackbody

T = 885°K		T = 1076°K		T = 1243°K	
Energy	Responsivity	Energy	Responsivity	Energy	Responsivity
.439	.1686	.960	.1583	1.709	.1521
.384	.1693	.840	.1583	1.495	.1518
.329	.1702	.720	.1583	1.282	.1513
.274	.1679	.600	.1583	1.068	.1507
.220	.1682	.480	.1583	.854	.1522
.165	.1697	.360	.1583	.641	.1529
.110	.1727	.240	.1583	.427	.1522
.055	.1636	.120	.1583	.214	.1542

Note: w is given in milliwatts · cm^{-2}; R is given in volts · $watts^{-1}$ · cm^2

The sector disk was used also with the 0.6328 μ laser beam as a source. The responsivity of the detector was found to remain constant in this case also.

It seemed advisable to double check the responsivity at the several temperatures by use of more carefully designed shields and corresponding apertures. The new shields consisted of two parallel screens of masonite rigidly held two centimeters apart by spacers. Both screens had carefully aligned circular apertures 1.83 cm in diameter and machined to a knife edge. A second pair of screens were made identical to this except that the circular apertures had a diameter of 1.25 cm. In each case both screens had aluminum foil on the side toward the blackbody and 3-M black velvet paint on the side toward the thermopile.

An analysis of the data taken with these improved screens revealed some important facts: the decrease in responsivity with distance was not as large as before; in addition the change in responsivity with temperature was considerably smaller. In looking for a possible reason, a careful examination showed that the detector was seeing not only a blackbody at a given temperature but also a graybody around the blackbody in the form of a part of the hot faceplate of the blackbody. While this faceplate was at a much lower temperature than the blackbody, the total area of the graybody was considerably larger than that of the blackbody. Thus the graybody would contribute a considerable amount of additional energy to the thermopile. Thus the radiant flux emitted into a hemisphere would be given by

$$R = S_b \sigma_b T_b^4 + S_g \epsilon_g \sigma_g T_g^4$$

where the subscript "b" refers to the blackbody and "g" to the graybody. In the case of the screen with an aperture of 1.25 cm, the area of the graybody seen by the detector was 4.8 times that of the blackbody for a detector distance of 20 centimeters and decreased to a ratio of 3.2 at a distance of 115 centimeters. This large ratio of the area of graybody to blackbody would mean that the detector saw much more energy than the blackbody flux which was used in calculating the responsivity. Also, since the ratios of the areas decrease with increasing distance, the experimental responsivity would decrease with distance. The effect of this graybody radiation on the responsivity might also be different at different temperatures.

In order to check this effect, a new shield was made consisting of two parallel sheets of masonite rigidly held 7 centimeters apart by spacers. Both screens had carefully aligned circular apertures 0.7 cm in diameter and machined to a knife edge. Both screens had aluminum foil toward the blackbody and 3-M black velvet paint on the side toward the thermopile. The arrangement is shown in Figure B-25.

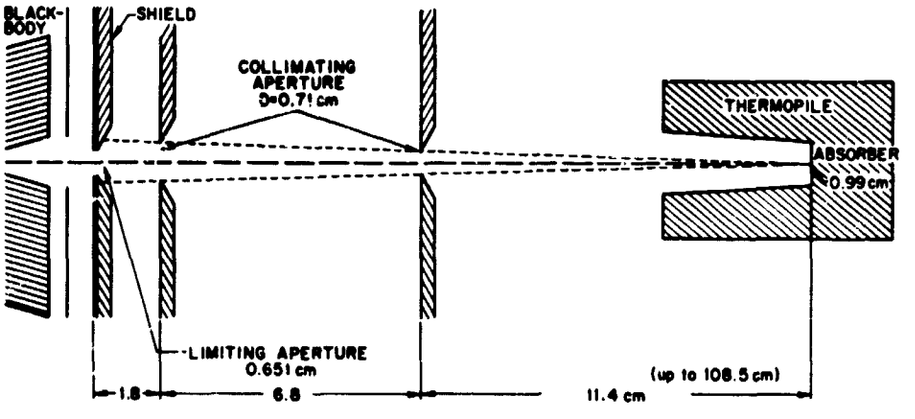


Figure B-25–Arrangement of blackbody and shield for responsivity measurements.

Using the new shield the responsivity of the thermopile to blackbody radiation was again determined at the same three temperatures. The responsivity of the thermopile is lower than with the previous shields. For blackbody temperatures of 885°K and 1076°K the responsivity at intermediate distances now differ by only 2.5% while at 1243°K it still differs by about 5%. The resulting data is plotted in Figure B-26. (Figure B-26 compares the 1243°K blackbody and quartz-iodine lamp calibrators.) It will be noted that the decrease in responsivity with increasing distance between blackbody and detector is smaller than before. Calculations show that when the detector is within 20 cm of the blackbody, it still sees a graybody area about 0.44 of the area of the blackbody rather than 4.8 times the latter area which was true with the previously used shield. The ratio of the graybody area to blackbody area decreases as the distance increases so that it is about 0.28 at 115 cm. This means that the detector sees much less graybody energy than when the larger apertures were used, so that the responsivity of the detector represents more nearly its true value as well as a smaller decrease in value with distance.

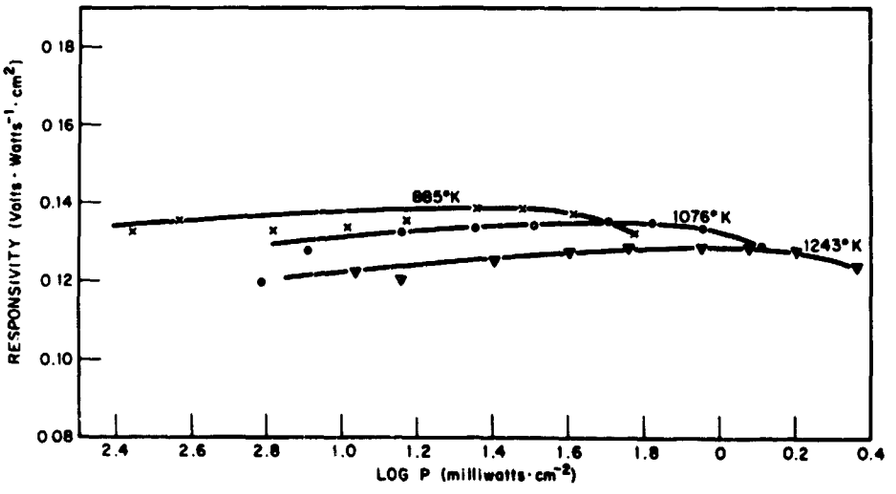


Figure B-26–Responsivity of thermopile 4928A to blackbody (shield with 9.7 cm aperture diameter).

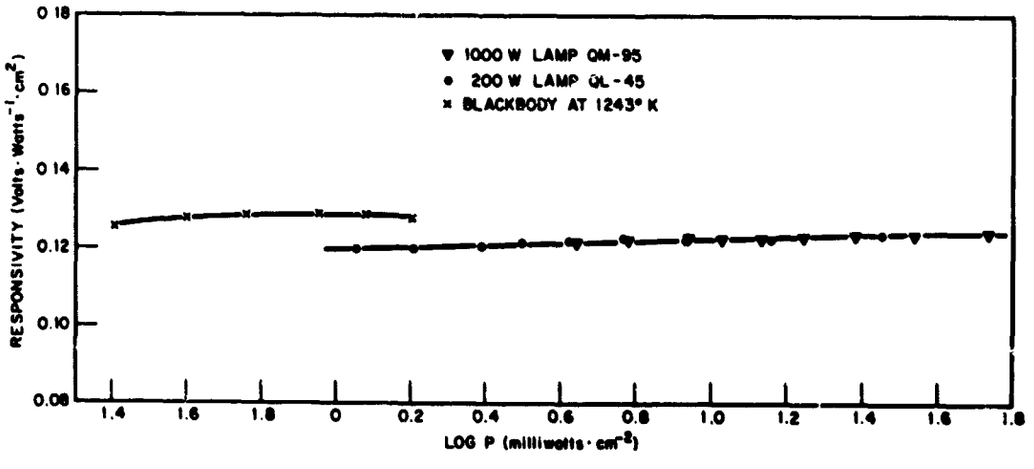


Figure B-27–Comparison of responsivity to lamp and blackbody.

It will also be obvious that when the 0.7 aperture is used for distances shorter than 30 cm the responsivity drops rapidly. This is because smaller and smaller areas of the conical wall of the detector "see" and hence reflect energy to the absorbing surface. This is similar to the case with the 1.25 cm aperture for distances shorter than 14.5 cm.

Fourthly, one observes in Figure B-26 that for a given blackbody temperature, the responsivity is about constant for intermediate distances and then drops slowly as the distances increase. As mentioned before, this may be explained at least in part by the small but decreasing graybody area still seen with the 0.7 cm aperture. Another contributing factor may be that of infrared absorption by water vapor in the air.

The blackbody radiation curves have their maxima at 3.3 μ, 2.7 μ, and 2.3 μ for the blackbody temperatures of 885°K, 1076°K and 1243°K respectively. These peaks lie near the strong water vapor absorption band at 2.7 μ as well as a fairly strong carbon dioxide band at about 2.75 μ. Hence absorption by water vapor might in part account for a decrease in responsivity with increasing distance and also have some bearing on the differences in the responsivity of the detector to radiation at different blackbody temperatures.

It had been assumed that water vapor absorption would be negligible for the paths used in these measurements. However, because of the results referred to above, it seemed worthwhile to explore the possibility of water vapor absorption. Measurements were made of the relative humidity of the laboratory at 300°K and found to be 51%. This represents 0.001 cm of precipitable water vapor in a 76 cm path. Calculations on the theoretical infrared transmittance of water vapor have been reported by Wyatt, et al. (5). The smallest amount of water vapor for which they give data is for 0.001 cm of precipitable water vapor (a 76 cm path at 300°K and 51% rel. hum.) which is very nearly that of the 75 cm path used as one of the distances in the present measurements. Additional data given by Wyatt represent paths of 760 cm, 7600 cm, etc. under the same conditions of ambient temperature and relative humidity. Since the absorption by water vapor is rather complex, it has as yet not been possible to evaluate the absorption by water vapor at the various distances used in these measurements other than at the 75 cm path.

Calculations on the infrared transmittance of carbon dioxide were made by Stull et al. (6). The smallest amount of carbon dioxide for which they give data is 0.2 atm-cm, which represents an air path of 732.6 cm at 1 atm. and 300°K. This path is about 10 times the 75 cm path considered above in the case of absorption by water vapor. Preliminary calculations based on the data for a 75 cm path indicate that the responsivity might be 7.2%, 5.8% and 5% higher at 885°K, 1076°K and 1243°K respectively if there were no absorption due to water vapor or carbon dioxide. The total probable effect of this absorption can only be determined after further study.

In the light of the investigations on the effective use of a blackbody for detector calibration which have been completed to date, it would seem that the following conditions should pertain: (A) The blackbody shield and aperture disk should be designed to keep it at ambient temperature regardless of the blackbody aperture used. (B) The smallest separation between detector and blackbody used for calibration should be such that all parts of the absorbing surface and of the conical walls will see essentially the same radiating area. (C) The relative humidity should be very low so that no corrections need be made for water vapor absorption, or else the humidity should be determined at the time of the experiment and appropriate data be obtained to make corrections for this absorption. (D) In certain applications of the detector calibration, it may be necessary to correct for the energy reflected by the conical wall of the thermopile. Finally (E) it would be desirable to have a blackbody which operates at approximately the temperatures at which the standard lamps operate. This would make comparison of the calibrations easier. Also, since the maximum radiation would come at a shorter wavelength, the fraction absorbed by any water vapor present might be a smaller fraction of the total energy emitted by the source.

Radiant Flux at Open Port of Spherical Integrator

The thermopile 4928A without its quartz window was used to measure the response of the thermopile to the radiant flux from the spherical integrator due to the flux from 1 to 12 200-w quartz-iodine lamps. The procedure used was similar to that described in the 1964 Summer Workshop Report (1) but modified by not using a quartz window on the thermopile and by using the Dymec Integrating Digital Voltmeter to measure the output of the thermopile.

Using the calibration of the thermopile as obtained by means of the quartz iodine lamps (page B-25 of this report), preliminary calculation were made on the corresponding flux at the port of the spherical integrator. This flux was used to check the calibration of radiometer #306 which was reported in the 1964 report (1). These preliminary calculations seem to indicate that for a given voltage output of the radiometer, it takes about 60% more power from the spherical integrator than from the Kodak paper, according to present calibrations, rather than approximately 30% more which was obtained in 1964.

Before accurate calculations can be made it will be necessary to determine accurately the spectral distribution of the radiant flux at the Windowless port of the spherical integrator. In addition it will be necessary to investigate the possible effect which the energy reflected by the conical wall of the thermopile has on the calibration of an extended source. In using a point source, the radiation received by the thermopile is approximately a bundle of parallel rays, except for very short distances, so that the energy reflected from the conical wall to the absorber makes an angle of approximately 85° to the normal to the surface. As stated on page B-27 of this report, this reflected energy amounts to 5.4% so that the integral value of this reflected energy amounts to 12.9% of that received by the absorber directly.

In the case of an extended source, the conical wall of the thermopile would receive and reflect to some extent rays making angles with the normal varying between 60° and 90° if one considers only those rays reaching the absorber after a single reflection. In order to see what effect this change in angle might have on the reflectivity of the Parson's black on the conical wall, use was made of the 0.632 μ laser beam.

The thermopile was mounted on the optical bench carriage previously used to determine the field of view (see page B-24). In this case, however, the thermopile was mounted so that the axis of rotation of the carriage passed through a point on the conical wall of the thermopile about half way between the two bases of the cone. The laser beam was directed at this point of rotation (point P, Figure B-28). By rotating the thermopile through known angles, it was possible to determine the reflectivity of Parson's black at various angles. The reflectivity for various angles (measured from the normal) is shown in Figure B-29. It will be observed that the reflectivity varies from about 1% for an angle of 75° to 12% for an angle of 88° and presumably approaches 100% as a limit at 90°. Work is under way to integrate the energy from a perfectly diffuse extended source which would suffer a single reflection by Parson's black and reach the absorbing surface of the thermopile. If this total reflected energy is more or less than the 12.9% reflected in the case of the point source, it will be necessary to correct the calibration for the extended source accordingly. These results will have to be reported later.

Time did not allow investigations (E) and (F), mentioned in the introduction, to be carried out. Work on these problems will be continued during the coming months. It is believed that no answers in regard to the overall problem are possible until these investigations have been completed.

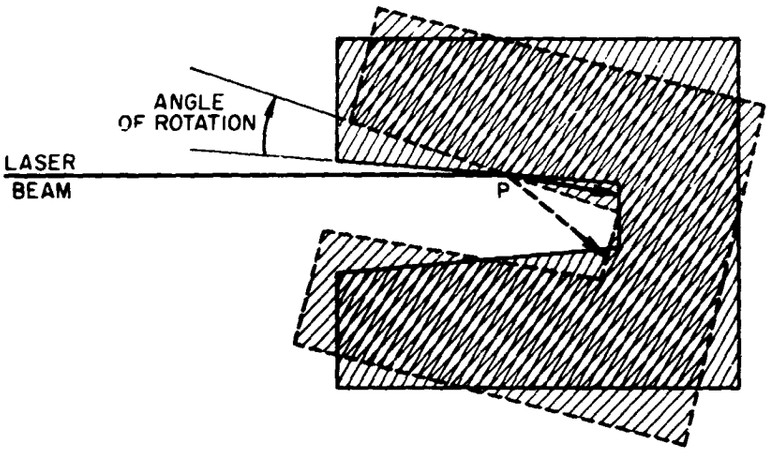


Figure B-28–Axis of rotation for measurement of reflectivity versus angle.

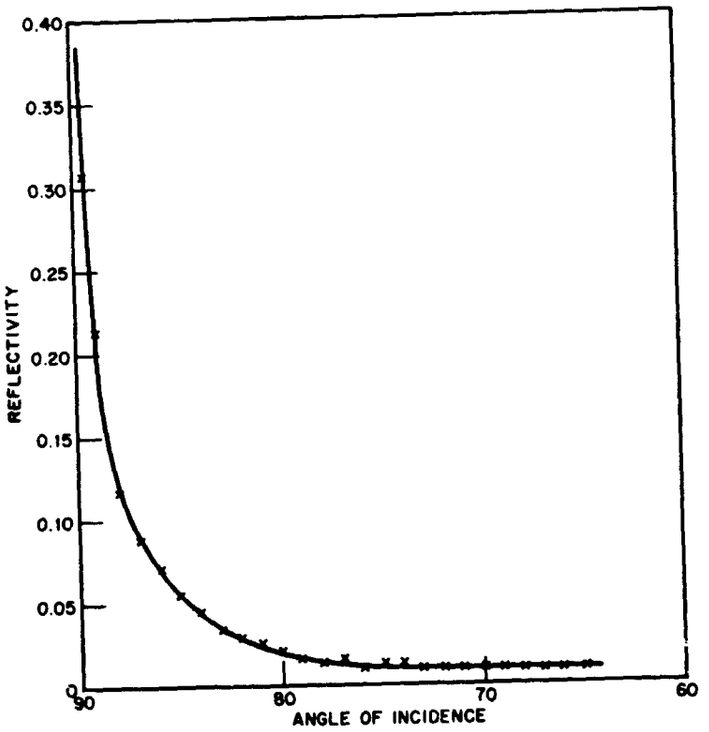


Figure B-29–Reflectivity of parson's black.

I wish to express my appreciation for the help given by Melvin L. Cooper who joined the Summer Workshop for one month, and to Barbara Sparkman and Nathan Miller of the Planetary Radiation Branch of the Laboratory for Atmospheric and Biological Science for their help in carrying out the experimental work. I also wish to express my gratitude for helpful discussions with Andrew W. McCulloch and I. L. Goldberg of the Planetary Radiation Branch and to Dr. M. P. Thekaekara, and Dr. F. W. Paul of the T & E Division and to the principal investigator Dr. E. K. Plyler.

REFERENCES

1. GSFC Final Report of the Summer Workshop Program, (1964) page A-17.

2. Ralph Stair, William E. Schneider, William R. Waters, and John K. Jackson, "Some Factors Affecting the Sensitivity and Spectral Response of Thermoelectric (Radiometric) Detectors," Applied Optics, Volume 4:703-710 (1965).

3. M. P. Thekaekara and E. I. Mohr, "Comparison of the Calibration of Total Radiation Sensors with Different Standard Lamps."

4. H. B. Keene, "A Determination of the Radiation Constant," Proc. Roy. Soc. A Vol. 88, pp. 49-60 (1913).

5. Philip J. Wyatt, V. Robert Stull and Gilbert N. Plass, "The Infrared Transmittance of Water Vapor," Applied Optics 3:229-241 (1964).

6. V. Robert Stull, Philip J. Wyatt and Gilbert N. Plass, "The Infrared Transmittance of Carbon Dioxide" Applied Optics, 3:243-254 (1964).

BLANK PAGE

PHOTOMETRIC PRINCIPLES GOVERNING THE VACUUM OPTICAL BENCH

I. C. Gardner

A fundamental equation of photometry is

$$N_\lambda \, d\lambda \cos\theta \, dA d\Omega = \text{radiant flux.} \tag{1}$$

This is the equation which gives the radiant flux or rate of flow of energy in an elementary beam of radiation. In this equation, N_λ is a characteristic of the particular light source that is employed. It is a function of λ, and $N_\lambda \, d\lambda$ is the energy transmitted in unit time within the spectral range lying between λ and $\lambda + d\lambda$, within unit solid angle through unit projected area. The angle θ is the angle between the normal to the element of area dA and the direction of incidence of the elementary beam. The product $\cos\theta \, dA$ is the area of the element dA projected into a plane normal to the direction of the ray. The elementary solid angle is $d\Omega$. The radiant flux has the dimension of power and may conveniently be measured in watts, or in photons per second.

This equation will be applied to the cross-section of the large beam of collimated light produced by the collimating system of the vacuum optical bench (VOB). Let it be assumed that the monochromator and optical bench are free from aberration and that all parts, including the source are correctly aligned. This represents an ideal condition that can be approached but never completely attained. It is well worth while to study this assumed condition because it represents the maximum performance for which one may strive. With this ideal condition, the radiant flux over any area of the cross-section is given by the equation

$$\text{Total radiant flux over area A} = \int_A \int_\Omega \int_{\lambda_1}^{\lambda_2} N_\lambda G R^6 \cos\theta \, d\lambda \, d\Omega \, dA \tag{2}$$

A grating may be considered as operating upon the incident light in two ways. The light within the spectral range λ and $\lambda + d\lambda$ is partially deviated by diffraction in a useful direction, that is, toward the exit slit of the monochromator, and another large portion of the light is diffracted into the spectra of other orders which are rejected by the monochromator. The factor G is the ratio of the usefully diffracted light to that which is rejected. Also, a grating, like any other reflecting surface, reflects only a portion of the incident light. For simplicity the reflectance of the grating is assumed to be the same quantitatively as that of the other reflecting surfaces in the VOB. Counting in this manner, in the VOB, as now in use, there are 4 reflecting surfaces, 2 in the monochromator and 2 in the collimating system. With the new monochromator, now on order, there will be a total of 6 reflecting surfaces, 1 in the illuminating system, 3 in the monochromator, and 2 in the collimating system. In Equation (2) the loss of light at the grating and at the reflecting surfaces is taken into account by the factors G and R^6. The three integrations of Equation (2) are over the spectral range, over the solid angle of the beam, and over the area. This last of the three integrations will be over unit area in order to give the radiant flux per unit area. With extended limits of integration this triple integration would be a difficult one to evaluate. However, in each instance, the limits of integration include a very small range, the various parameters may consequently be assumed constant without great error, and the three integrations may be performed independently.

According to the performance specifications for the VOB, the spectral range λ_1 to λ_2 for any setting is 1 angstrom. Within this small range, N_λ can be considered constant although its value will be different for different settings of the monochromator as well as for different sources. Consequently the result of the integration with respect to λ is simply N_λ.

If the cone of light proceeding toward any point of the cross-section of the large beam in the VOB is assumed symmetrical about its axis,

$$d\Omega = 2\pi \sin\theta \, d\theta \tag{3}$$

where θ is the half-angle at the vertex of the cone. The parallax is limited to 1 second and θ, therefore, is 0.5 second which equals 2.5×10^{-6} radians. If the value of $d\Omega$ given in Equation (3) is substituted for $d\Omega$ in Equation (2), the integration with respect to θ becomes

$$2\pi \int_0^\theta \sin\theta \cos\theta \, d\theta = \pi \sin^2\theta \tag{4}$$

For present purposes θ and $\sin\theta$ may be considered equal and the integral becomes 1.96×10^{-11}. Letting A equal 1 to obtain the flux per unit area and substituting the results of the several integrations in Equation (2) one finally obtains the result

$$\frac{\text{Radiant flux}}{\text{per unit area}} = 1.96 \, N_\lambda \, G R^6 \times 10^{-11}. \tag{5}$$

Equation (5) represents ideal performance and is a very fundamental equation applying to the operation of the VOB. It will be noted that it has no terms referrring to the optical system of the VOB. The focal lengths of the primary and secondary mirror, the focal lengths of the mirrors of the monochromator, the size of the grating, and other details can not change the result of Equation (2) and it is well to know this because it may forestall the undertaking of modifications and changes which are destined to be fruitless. It is necessary that the apertures be "filled," that is to say the source of light must be sufficiently large and so adjusted that, for any component, flux is transmitted over the entire area that is supposed to play a part in the transmission of the radiant flux. This is equivalent to saying that the design shall be functionally correct and that the artisanship involved in its construction be first class. Equation (5) has this fundamental character because the required performance is very fully specified when the bandwidth is limited to 1 angstrom and the maximum parallax to 1 second.

Although Equation (2) has great universality, its usefulness for the derivation of quantitative results is not great because of the lack of the necessary engineering data. We do not have determinations of G for standard gratings as a function of wavelength and values of N_λ for different types of sources as a function of wavelength are very limited. As the wavelength becomes smaller this lack of the necessary data becomes more pronounced.

If the specifications for a bandwidth of 1 angstrom and a maximum parallax not exceeding 1 second are retained, there are only three parameters that can be controlled to modify the radiant flux, namely N_λ, G, and R. For any increase in the value of G one will have to appeal to the manufacturers of gratings.

For λ = 1600, reflecting surfaces with the reflectance as great as 0.7 have been obtained. When $R = 0.7$, $R^6 = 0.12$. Consequently, with perfect reflectors, the radiant flux at this wavelength would be multiplied by 8. For shorter wavelengths the conditions are much less favorable. For λ = 1000, reflection coefficients as large as 0.2 have been obtained with a coating of aluminum and magnesium fluoride. With a coating of aluminum and lithium fluoride this coefficient may be increased to 0.5. For the aluminum and magnesium fluoride $R^6 = 0.00006$ and this suggests that it will be very difficult, if not impossible, to obtain adequate radiant flux with this coating. If the reflection coefficient is increased to 0.5, $R^6 = 0.016$, a value which is still discouragingly small. There are three possible solutions for this difficulty. A spectacular improvement in the light source used to increase the value of $N\lambda$ for the shorter wavelengths, a spectacular improvement in the reflecting coatings available for these shorter wavelengths, or a rearrangement of the optical bench and the monochromator so that during the testing of the experiment package it receives light direct from the source instead of after passage through the monochromator as is now the case. This would eliminate three reflections and increase the available flux by a factor of approximately 100 for λ = 1000 and the aluminum-magnesium coating. Dr. Paul has suggested one such arrangement. By such an arrangement the test will not be as direct as at present and possible sources of error are introduced but the resulting increase of flux is a tremendous advantage and the extremely short wavelengths are in some respects the most important to be observed. The possibilities of the new method and of other possible new methods will be explored.

THE TUNGSTEN LAMP AS A SECONDARY STANDARD FOR TOTAL RADIANT FLUX

M. P. Thekaekara and E. I. Mohr

INTRODUCTION

In many applications of radiometry a compact source of known radiant intensity is highly desirable. To meet the need for such a source, the National Bureau of Standards, Washington, D. C., developed originally in 1914, the carbon filament standard lamp of total radiance (1). It has found wide acceptance in many laboratories and many subsequent studies have been made on its calibration (2) (3), thus showing its wide range of usefulness. Its disadvantages are the relatively low radiance, the low temperature at which it operates, and the rather large size of the filament. For a typical lamp (3) the radiant flux at a distance of 2 meters is 84.9 microwatts per cm^2.

More compact lamps of higher temperature and greater radiance have been issued by the National Bureau of Standards as standards of spectral radiance. In its original form it was a tungsten strip enclosed in a large cylindrical pyrex envelope, with or without a neck and a quartz window. The earlier method of finding its spectral energy distribution was to determine the temperature of the ribbon by visual methods. The color temperature thus determined was used to find the actual temperature of the radiating surface and hence the spectral radiance, through known values of the blackbody radiation function, the emissivity of tungsten and the transmittance of the quartz window. This basic procedure has been used by other laboratories to set up independent standards of spectral radiance (4). A tungsten strip standard of spectral radiance calibrated directly in comparison with a blackbody source was developed by the N.B.S. in 1960 (5).

A more recent development (6) is the so-called quartz-iodine lamp. It is a coiled coil tungsten filament lamp manufactured commercially by the General Electric Company. The filament is enclosed in a quartz envelope which contains a trace of iodine along with its usual inert gas filling. Two types of these lamps have been issued, the 200 watt lamp, G. E. type T 4 Q/1 C L - 200-w and the 1000 watt lamp, G. E. type D X W - 1000-w. These lamps have several advantages which make them superior to the earlier strip lamp standard (7) (8). They do not blacken with prolonged use; the bulb remains clear during the entire life of the lamp, since an "iodine cycle" deposits the evaporated tungsten back on the lamp. The filament can be operated at a higher temperature than earlier types of lamps of comparable life-time, which means a greater radiance in the ultraviolet. The quartz envelope makes these lamps more compact for a given power rating.

This paper presents the results of studies which have been made on the applicability of these lamps as secondary standards for total radiance. Standards of total radiant flux are usually defined not in terms of a standard source of radiant flux but in terms of a standard instrument for measuring (9) (10). Since most of the work of standardization has been done in connection with measurements of solar energy, the instrument is the pyrheliometer. Two standard scales have been widely used in meteorological institutes, the Smithsonian scale and the Angstrom scale. Both have undergone periodic revisions. Intercomparisons between the two scales have shown a difference which is not constant but has a wide scatter, averaging about 3.5 percent. A new standard scale, the International Pyrheliometric Scale, was adopted in 1956 by the World Meteorological Organization; by definition, measurements on this scale are 1.5 percent greater than those on the Angstrom scale of 1905 and 2 percent less than those on the Smithsonian scale of 1913.

TOTAL RADIANCE OF A TUNGSTEN SOURCE

An alternate approach is to adapt a standard of spectral radiance as one for total radiance. The use of such a source as a standard for temperature has been discussed by Jones (11). This would give a total radiance standard of which the spectral radiance is also known accurately. No detector surface is perfect black or perfect grey, and the response of the detector is proportional to the integral

$$\int_0^\infty P_\lambda \, a_\lambda \, a\lambda,$$

where P_λ is the spectral randiant flux and $a\lambda$ is the spectral absorptance. Hence the spectral randiant flux of the calibration source is an important paramete~ if the same detector is used for sources of different spectral energy distribution.

The standard of spectral radiance such as the quartz-iodine lamp, has been calibrated with reference to a blackbody, which in turn has been referred to the international temperature scale. The area under the $P_\lambda - \lambda$ curve, of spectral radiant flux on the y-axis versus wavelength on the x-axis, gives the total radiant flux

$$P = \int_0^\infty P_\lambda \, a\lambda \, .$$

The calibration table of the N.B.S. states that the values of P_λ have a maximum uncertainty ranging from 8 percent at the shortest wavelengths to 3 percent in the visible and infrared (6). It is likely that the inaccuracies are due to random errors in individual measurements rather than systematic errors such as in the standard blackbody. Hence the integration process would tend to cancel out the errors and the value of the total radiant flux may claim a higher accuracy than the values of spectral radiant flux from which it is derived.

There is, however, a major source of uncertainty due to the energy in the wavelength range not covered by the N.B.S. calibration table. As will be seen later, about 20 percent of the radiant flux is in the infrared region of wavelength longer than 2.55 μ. This originates partly from the tungsten coil and partly from the quartz envelope of the lamp. In this wavelength range the absorptance of the surface of different types of detectors shows significant variations so that both the total radiant flux and its spectral distribution should be determined. Direct determination of the energy in this range is not feasible because of the relatively small amount of energy available and the difficulties of making precise measurements in this range. We have followed an indirect method, which besides yielding a value for the radiant flux reveals certain interesting features about these lamps.

A parameter which should first be determined is the temperature of the tungsten filament. This is done by comparing the known spectral radiance function of the lamp for wavelength less than 2.5 μ with computed spectral radiance functions of tungsten coil at different temperatures as viewed through a quartz window.

The computed spectral radiance function is given by the expression $N_{\lambda c} = N_\lambda \epsilon_w \tau_q$ where $N_{\lambda c}$ is the computed spectral radiance, N_λ is the spectral radiance of a perfect blackbody, ϵ_w is the emissivity of tungsten and τ_q is the transmittance of quartz. N_λ and ϵ_w and hence also $N_{\lambda c}$ are dependent on temperature. N_λ is given by Planck's radiation function, and computed values of N_λ at close intervals of wavelength and temperature are available in many standard references, among which Tables of Blackbody Radiation Functions by Pivovonsky and Nagel (12) is the most complete. Data on emissivity ϵ_w covering a sufficiently wide range of wavelength and temperature have been published by deVos (13).

More recent and perhaps more accurate measurements have been reported by Larrabee (14). These values are on an average 2.5 percent lower than those of deVos, but they cover the limited range 0.3 to 0.6 μ and 1600° to 2400°K. The transmittance of quartz is also open to a small degree of uncertainty because of the differences between different types of quartz. The values for the present computations were based on information supplied by the Lamp Division of the General Electric Company and our own independent measurements using a Beckman D K 2 A for the range below 2.5 μ and a Perkin-Elmer I R 7 for the longer wavelength range.

The results of the comparison between the spectral radiance P_λ of one particular lamp, QL 45, and the computed function $N_\lambda \epsilon_w \tau_q$ for six temperatures is shown in Figure B-30. N_λ and P_λ are expressed in different units: N_λ in watts per sq. cm. of emitting area per micron bandwidth per steradian, and P_λ in watts per micron bandwidth per steradian, the area being that of the whole incandescent coil. Since the area of the coil is a constant, for a particular temperature, namely, the temperature at which the ribbon operates, the ratio $P_\lambda/(N_\lambda \epsilon_w \tau_q)$ should be a constant, and the corresponding graph should be a straight line parallel to the x-axis. Actually none of the curves in Figure B-30 is strictly a straight line, but on the short wavelength side the curves for the high temperatures tend to drop since N_λ is relatively high, and those for the low temperatures tend to rise since N_λ is relatively low. The curve for 2900°K seems to be the best fit, and hence we assume that the coil operates at 2900°K. Similar sets of curves for other 200 watt lamps (series QL) also gave 2900°K as the operating temperature. The curves for the 1000 watt lamps (series QM) were significantly different from those of the 200 watt lamps. Their operating temperature was found to be 2980°K. All the curves in Figure B-30 show an upward slope for wavelengths longer than 2.1 μ and a slight curvature in the range 1.5 to 2.0 μ. This may be due to errors in deVos' emissivity function which Larrabee had indicated or to other unknown causes.

Extrapolation of the curve at the appropriate temperature to wavelengths longer than 2.5 μ gives the factor by which the computed radiance function $N_\lambda \epsilon_w \tau_q$ has to be multipled in order to obtain the radiance of the tungsten coil in the long wavelength range not covered by the N.B.S. calibration table. The radiance in the short wavelength range is negligibly small.

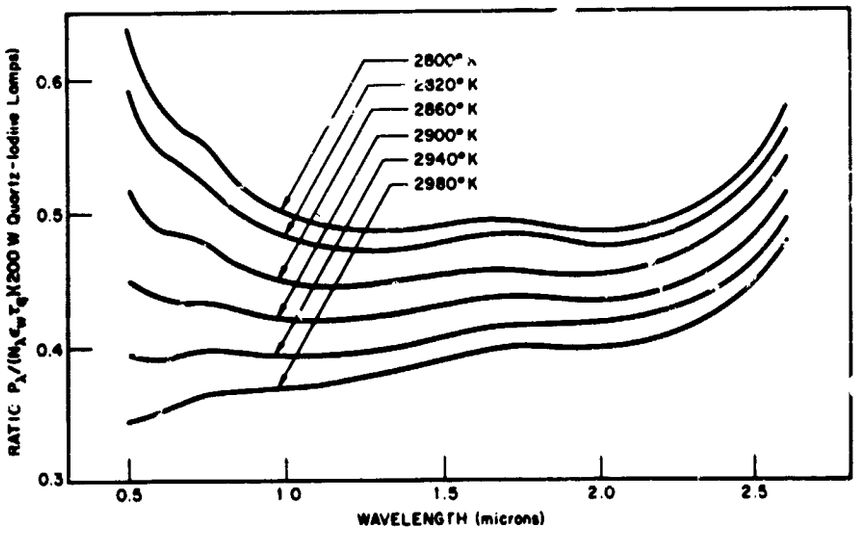


Figure B-30–Ratio of the spectral radiance P_λ of 200 watt standard lamp QL 45 to computed spectral radiance $N_\lambda \epsilon_w \tau_q$ of similar source at different temperatures. P_λ in watts μ^{-1} ster.$^{-1}$; N_λ radiance of a blackbody in watts cm^{-2} μ^{-1} ster^{-1}.

This procedure is applicable also to the earlier standard lamps of tungsten ribbon in pyrex envelope (15). The length of the ribbon is greater than the diameter of the quartz window, and hence for calibrating thermopiles and similar detectors, the lamp should be viewed through the pyrex. A small correction factor is needed for the transmittance of pyrex; the spectral radiance function through quartz should be multiplied by τ_p / τ_q, where τ_p is the transmittance of the pyrex envelope. It should also be multiplied by the area of the ribbon, since the spectral radiance as given in the N.B.S. calibration table is the energy radiated per mm^2 of the ribbon.

RADIANCE OF THE QUARTZ BULB

For radiometric applications it is essential to determine also the radiance of the lamp envelope, which is significant in the case of the quartz iodine lamps, especially when the detector does not have a quartz window. The method of evaluating this function can best be illustrated by the actual calculations which were made in one particular case, lamp QL 45. By integrating the area under the spectral radiance curve, we obtain the energy emitted by the tungsten coil in the direction normal to the length of the lamp, to be 15.683 watts per steradian. This covers the wavelength range from 0.245 μ to 4.65 μ. A set of detailed measurements were made with a thermopile to determine the variation of radiant flux at a given distance r, for different values of the polar coordinates θ and φ. Two independent methods of numerical integration showed that the energy emitted by the tungsten coil over the solid angle 4π is 181.81 watts. The total input energy given by the product of voltage and current is 195.00 watts. The difference, 13.19 watts, which is 6.76 percent of the input energy, is the energy emitted by the quartz envelope. This permits an approximate determination of the average temperature of the outside of the quartz envelope by applying the Stefan-Boltzmann equation, $W = A\sigma\ (\epsilon_q T_1^4 - T_2^4)$, where W is the energy radiated per second, A is the external surface area of the quartz envelope, σ is the Stefan Boltzmann constant, ϵ_q is the total hemispherical emissivity of quartz for its average outside temperature T_1, and T_2 is the temperature of the walls of the room. This indirect method for T_1 is more reliable than direct methods such as by a thermocouple. Approximate values for the emissivity of quartz, both total and spectral, are available in literature (16) (17). Assuming an approximate value 600 to 800°K for T_1 and hence 0.5 for ϵ_q, for $A = 20.34\ cm^2$, T_1 is found to be 701.03°K. A more accurate value, 0.495 of ϵ_q for this temperature yields a second approximation, 705.30°K, which is practically the same. Computations on other QL lamps of 200 w also gave about the same value, near 700°K. The corresponding value for the 1000 w QM lamps is 900°K.

The temperature being known, the energy emitted by the quartz envelope in the forward direction per unit solid angle is obtained from the function $N_\lambda\, \epsilon_{q\lambda}\, A_p$, where N_λ is the spectral radiance of a blackbody at temperature T_1, $\epsilon_{q\lambda}$ is the spectral emissivity of quartz at the same temperature and A_p is the projected emitting area for the forward direction. Figure B-31 shows a comparison between the spectral radiance per unit surface area of a 200 watt QL lamp and that of a blackbody at the same temperature. Below 3 μ where quartz is fully transparent the spectral radiance of the bulb is zero; and it is the same as that of a blackbody above 5 μ where quartz is opaque.

A STANDARD OF TOTAL RADIANCE

Detailed spectral radiance data on one of the 200 watt lamps, QL 182, are presented in Table B-7 Up to 2.5 μ the values are obtained from the N.B.S. tables by direct conversion from units of radiant flux ($\mu w\ cm^{-2}$ nanometer^{-1} at 43 cm) to units of radiance (w micron^{-1} ster.$^{-1}$). In the range beyond 2.6 μ the spectral radiance is partly from the tungsten coil and partly from the quartz envelope; the values have been computed

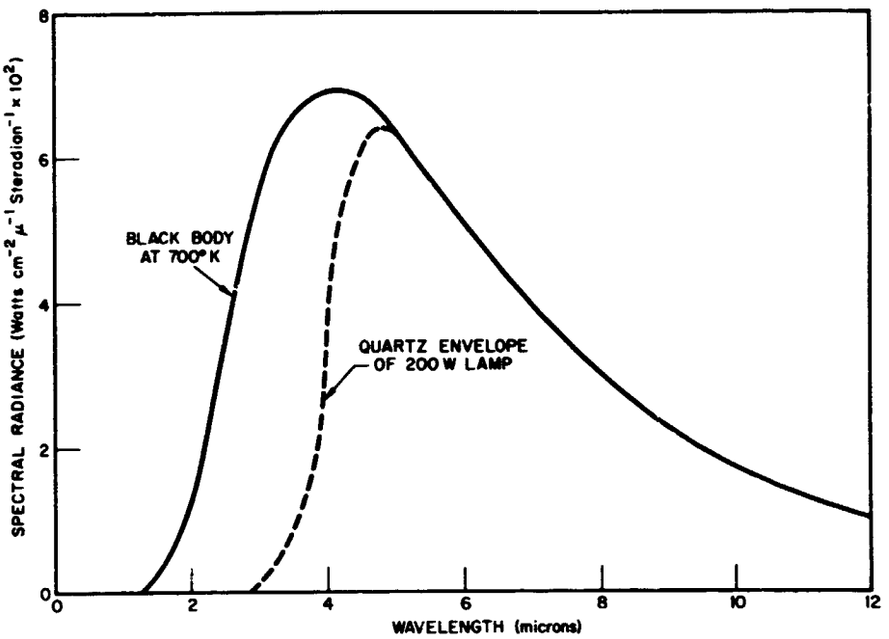


Figure B-31—Spectral radiance of a blackbody at 700° K and that of the quartz envelope of a 200 watt lamp QL 182.

Table B-7

Spectral radiance of quartz-iodine lamp, QL 182, 200 watts. λ, wavelength in microns; N_λ, spectral radiance in watts per micron bandwidth per steradian, from the whole lamp; N_w, spectral radiance from the tungsten coil; N_q, spectral radiance from the quartz envelope.

λ	$N_\lambda = N_w$	λ	$N_\lambda = N_w$	λ	N_w	N_q	N_λ	λ	$N_\lambda = N_q$
.25	.0066	1.1	10.91	2.6	1.886	0	1.886	5.5	.3704
.26	.01239	1.2	9.966	2.7	1.555	0	1.555	6.0	.3302
.27	.02163	1.3	8.949	2.8	1.402	0	1.402	6.5	.2909
.28	.03550	1.4	7.969	2.9	1.257	0	1.257	7.0	.2545
.29	.05492	1.5	7.045	3.0	1.139	.0073	1.146	7.5	.2220
.30	.08025	1.6	6.194	3.2	.9275	.0271	.9546	8.0	.1932
.32	.1583	1.7	5.418	3.4	.7243	.0456	.7699	8.5	.1683
.35	.3735	1.8	4.715	3.6	.5593	.0656	.6249	9.0	.1466
.37	.5880	1.9	4.086	3.8	.4058	.1037	.5095	9.5	.1281
.40	1.022	2.0	3.532	4.0	.2862	.1956	.4818	10.0	.1121
.45	2.089	2.1	3.088	4.2	.1246	.3081	.4327	11	.0865
.50	3.624	2.2	2.718	4.4	.06088	.3646	.4244	12	.0675
.55	5.270	2.3	2.422	4.6	.05016	.3981	.4483	13	.0532
.60	6.915	2.4	2.182	4.8	0	.4090	.4090	14	.0425
.65	8.450	2.5	2.015	5.0	0	.4074	.4074	15	.0342
.70	9.781							16	.0278
.75	10.80							17	.0229
.80	11.44							18	.0189
.90	11.89							19	.0158
1.00	11.61							20	.0133

from the respective blackbody radiation functions and the emissivity functions as described earlier. The data of Table B-7 are also presented graphically in Figure B-32; the inset shows the range of wavelength greater than 3 μ on an expanded radiance scale and contracted wavelength scale.

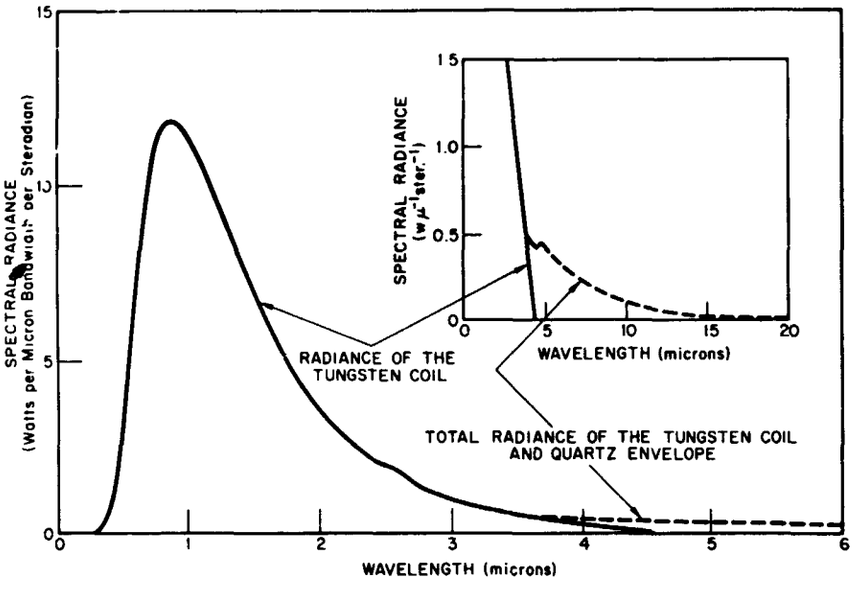


Figure B-32–Spectral radiance of the 200 watt quartz iodine lamp QL 182 in the entire wavelength range from the tungsten coil and from the quartz envelope.

The total radiance of the lamp is given by

$$N = \int_0^\infty N_\lambda \alpha_\lambda ,$$

which for all practical purposes can be approximated by the summation $N = \Sigma N_j \Delta\lambda_j$, where N_j is each of the N_λ's of Table B-7, and $\Delta\lambda_j = 1/2(\lambda_k - \lambda_i)$, λ_i, λ_j and λ_k being three consecutive values of λ. Values of N for six QL lamps and four QM lamps are given in Table B-8. These values are typical of the 200 w and 1000 w lamps. For other calibrated lamps sufficient accuracy can often be obtained by making a numerical integration in the range below 2.55 μ and by assuming that the relative energy in the longer wavelength range is the same as for the lamps listed in Table B-8.

These lamps can be used for calibrating thermopiles in a wide range of radiant flux by simply varying the distance between the detector and the source. The inverse square law gives the radiant flux P at distance r in terms of the radiance N; $P = N/r^2$, provided r is not too large nor too small. At large distances atmospheric absorption introduces a very small source of error which has been studied by Stair and Johnston (3). They have shown that in the case of the carbon filament lamp, a 2 meter pathlength changes the transmitted energy by 0.017 percent per gram of water per meter (3). At very short distances the finite size of the source introduces a source of error which can readily be calculated. For a source of finite size, the radiant flux is less than what it is for a point source of the same total energy output. An approximate expression for the difference is

$$P_o - P = P_o\left(\frac{1}{6}\frac{a^2}{r^2} - \frac{3}{80}\frac{a^4}{r^4}\right),$$

Table B-8

Final results on total radiance of QL (200 w) lamps and QM (1000 w) lamps. N radiance in watts per steradian from the whole lamp and in the entire wavelength range; N_w, radiance from the tungsten coil; N_q radiance from the quartz envelope.

Lamp	N_w 0.245 μ -2.55 μ	N_w 2.55 μ -4.65 μ	N_q 2.95 μ -20 μ	N 0.245 μ -20 μ
QL-45	14.212	1.471	2.046	17.729
QL-46	13.235	1.360	2.046	16.641
QL-74	14.000	1.409	2.046	17.455
QL-147	14.363	1.486	2.046	17.895
QL-181	14.412	1.486	2.046	17.944
QL-182	13.746	1.409	2.046	17.201
QM-17	71.756	7.275	8.768	87.798
QM-18	74.108	7.404	8.768	90.280
QM-10	70.610	7.778	8.768	87.156
QM-20	71.279	7.451	8.768	87.498

where P_o and P are respectively the radiant flux at distance r from a point source and from a source of length a, the width of the source and the dimensions of the detector being considered negligible. Thus, the error is about 4 percent for $r/a = 2$ and 0.04 percent for $r/a = 20$.

Thus, with due precautions indispensable in all precision radiometry, the quartz-iodine lamps and other standards of spectral radiance can be used as secondary standards of total radiance. They have been used over a very wide range of radiant flux varying from a few hundred microwatts per cm^2 to 0.3 watts per cm^2, and for many different types of thermopiles, thermistors, solar cells and other radiation detectors. The results have shown a high degree of internal consistency.

ACKNOWLEDGMENTS

The authors wish to thank Elias Klein, Director of the Summer Workshop Program, Goddard Space Flight Center. Most of the work discussed here was done under the auspices of this program. They are also indebted to A. R. Winker and T. A. Riley for many of the routine measurements. The discussions with Ralph Stair of the National Bureau of Standards were particularly helpful.

REFERENCES

1. W. W. Coblentz, Bull. Natl. Bur. St. 11, 87, 8227 (1914).

2. W. W. Coblentz and W. B. Emerson, Bull. Natl. Bur. Std. 12, 503, 8261 (1916).

3. R. Stair and R. G. Johnston, J. Res. Natl. Bur. Std. (U.S.) 53, 211 (1954).

4. W. J. Condell and F. T. Byrne, J. Opt. Soc. Am., 47, 1135 (1957).

5. R. Stair, R. G. Johnston and E. W. Halback, J. Res. Natl. Bur. Std. (U.S.) 64 A, 291 (1960).

6. R. Stair, W. E. Schneider, and J. K. Jackson, Appl. Optics 2, 1151 (1963).

7. E. G. Zubler and F. A. Mosby, Illum. Engr. 54, 734 (1959).

8. J. A. Moore and C. M. Jolly, General Electric Co. J. 29, 99 (1962).

9. M. P. Thekaekara, Solar Energy, 9, 4 (1965).

10. IGY Instruction Manual, Part VI, Radiation Instruments and Measurements, (Pergamon Press, London 1957) p. 378; A. J. Drummond, W. M. O. Bull. 5, 19 (1956); A. J. Drummond, Quarterly J. R. Met. Soc. 82, 481 (1956).

11. T. P. Jones, J. Sci. Instr. 40, 101 (1963).

12. M. Pivovonsky and M. R. Nagel, Tables of Blackbody Radiation Functions (McMillan Company, New York, 1961).

13. J. C. deVos, Physica, 20, 690 (1954).

14. R. D. Larrabee, J. Opt. Soc. Am. 49, 619 (1959).

15. M. P. Thekaekara, M. S. Ojalvo and A. J. Galli, Calibration of Thermopiles, in Report of the Goddard Summer Workshop Program (Goddard Space Flight Center, Greenbelt, Maryland 1963) X-320-63-264, pp. A-53 to A-63.

16. G. G. Gubareff, J. E. Janssen and R. H. Torborg, Thermal Radiation Properties Survey, A Review of Literature (Honeywell Research Center, Minneapolis, Minnesota, 1960) 2nd ed., pp. 236-240.

17. M. Pirani, J. Sci. Instr. 16, 372-378 (1939).

BLANK PAGE

STANDARDIZATION OF 1000 w QUARTZ-IODINE LAMPS

N. J. Miller and E. I. Mohr

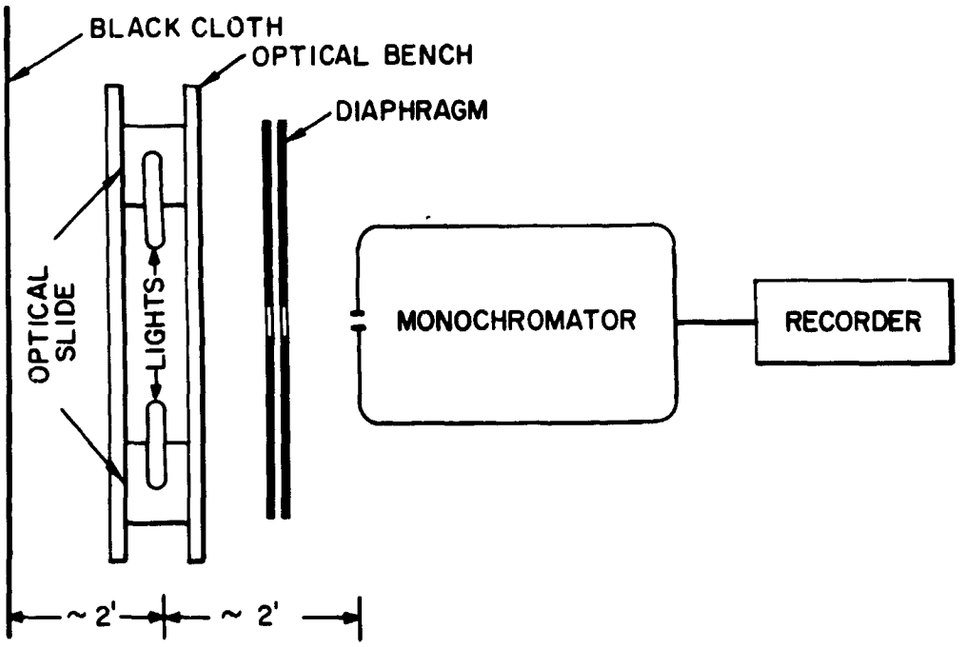
This project was undertaken to formulate a calibration procedure to be used in making working standards of spectral irradiance out of regular stock 1000 w quartz-iodine lamps. Such standards could then be easily provided whenever they might be needed in the workings of the planetary radiations branch.

The calibrations were performed through the use of a Perkin-Elmer model 99 double pass monochromator. The output from a detector contained within the monochromator housing was fed into a recorder. When the monochromator was set at a given wavelength, the spectral energy from a light source produced a visual record and the magnitude of the recorder pen deflection was proportional to the spectral intensity level of the light source. This record could then be compared to that of any similar light source for the same experimental conditions.

The laboratory arrangement was as is shown in Figure B-33. Three 1000 w quartz-iodine lamps were chosen to be calibrated against a similar lamp which had been calibrated as a secondary standard by the National Bureau of Standards. Each lamp was mounted on a holder of the same type that the Bureau of Standards supplies with its 1000 w standards, and the holder was attached to a bench carriage. When the output of a particular lamp was to be detected, the bench carriage which held the lamp was moved along an optical bench to a predetermined measurement position. These measurement positions were at locations such that the output of each lamp produced the maximum response in the detector for a given monochromator slit to lamp distance. First, each lamp was individually oriented parallel to and centered on the monochromator slit using the same slit to lamp distance for all lamps. Then each lamp, in turn, was turned on and its bench carriage was shifted slightly along the optical bench until the monochromator detector gave the maximum response. The position of the bench carriage on the optical bench was noted and this position was the position to be used whenever it was desired to detect the output of the lamp mounted on the carriage. If the slit width of the monochromator, or the slit to lamp distance, or the monochromator detector was changed, then the lamp alignment procedure was repeated and the new measurement position was used for the changed experimental conditions.

In order to cover the entire wavelength range for which the secondary standard was calibrated it was necessary to use more than one detector and more than one slit width with the monochromator. In the wavelength interval from 450 nm to 2500 nm a thermocouple detector was used and for the region from 270 nm to 600 nm a 1P21 photomultiplier was used. The monochromator slit to lamp distance was near 50 cm.

The comparisons of spectral irradiance with the standard lamp were made on a wavelength by wavelength basis, alternately detecting the uncalibrated lamp output and the standard lamp output. After all of the preliminary adjustments necessary for the comparison of a lamp with the standard lamp were made, the monochromator was set to transmit a particular wavelength of energy. Then the monochromator and its attendant electronics were adjusted such that the recorder pen deflection would be near 95% of full scale for whichever of the two lamps had the highest energy output. The recording chart was allowed to run long enough to ensure a clear indication on the chart as to where the intensity level of the lamp should be read. Then the spectral energy output of the remaining lamp was detected without making any further adjustments on the monochromator. This procedure was repeated at each calibrated wavelength over the spectral energy range of the standard lamp. Several lamp comparisons were made at each calibrated wavelength. The data is presented in Table B-9.



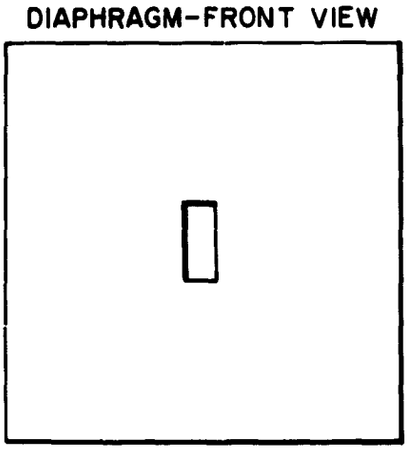


Figure B-33–Laboratory arrangement for lamp calibrations.

Table B-9
Calibration Results

Wavelength nm	Spectral Irradiances, $\mu w/cm^2$-nm			
	Standard	Lamp A	Lamp B	Lamp C
270	0.5838	0.0626	0.0626	0.0656
280	0.0859	0.0996	0.0994	0.104
290	0.130	0.151	0.150	0.158
300	0.187	0.216	0.215	0.227
320	0.357	0.407	0.406	0.426
350	0.820	0.932	0.931	0.960
370	1.27	1.44	1.43	1.49
400	2.21	2.50	2.42	2.45
450	4.41	4.81	4.65	4.84
500	7.37	8.00	7.86	8.06
550	10.7	11.5	11.4	11.6
600	14.2	15.4	15.1	15.4
650	17.5	18.4	18.2	18.5
700	20.2	21.1	21.0	21.3
750	22.2	23.1	23.1	23.3
800	23.4	24.3	24.3	24.5
900	24.3	25.1	24.9	25.2
1000	23.8	24.5	24.2	24.6
1100	22.4	22.9	22.7	23.0
1200	20.5	21.0	20.9	21.1
1300	18.5	18.9	18.9	19.0
1400	16.5	16.8	16.8	17.0
1500	14.6	14.8	14.8	14.9
1600	12.8	13.0	13.1	13.1
1700	11.2	11.3	11.4	11.4
1800	9.73	9.91	9.82	9.92
1900	8.42	8.54	8.50	8.56
2000	7.28	7.39	7.40	7.41
2100	6.36	6.46	6.54	6.47
2200	5.62	5.71	5.69	5.79
2300	5.01	5.09	5.14	5.12
2400	4.52	4.59	4.58	4.62
2500	4.14	4.21	4.18	4.25

The total radiance for each of the lamps involved in the calibration procedure was calculated through a procedure as described by Thekaekara and Mohr.* These results are shown in Table B-10. The spectral data implied that the total directional outputs for all of the lamps tested should be within 1% of each other. This result was verified by detecting the lamp outputs with a thermopile which was located at 40 cm from the lamp filaments. The results are contained in Table B-11.

The chief sources of error are in the lamp alignment, as determined by the measurement position, and the electronic noise which is most significant near the ends of the measurement interval for a detector. An investigation into discrepancies observed between spectral irradiances as measured by a photomultiplier and by a thermocouple led

*Preceding paper, this document

Table B-10
Lamp Energy Output
Total Output (watts/steradian)

Lamp	Filament output		Envelope output	Total output
	0.245-2.55 μ	2.55-4.65 μ	2.85-20 μ	0.245-20 μ
QM-95	70.770	7.262	8.768	86.800
Lamp A	72.837	7.490	8.768	89.095
Lamp B	72.539	7.228	8.768	88.535
Lamp C	73.272	7.316	8.768	89.356

Table B-11
Thermopile Comparisons*

Lamp	Response Volts	Relative Response
QM-95	6.707	1.000
Lamp A	6.740	1.005
Lamp B	6.632	0.989
Lamp C	6.812	1.016

*Lamp comparisons made with lamp filaments at 40cm distance

to the discovery that the determination of the location of the measurement position referred to previously was critical. Failure to determine a new measurement position for a lamp when the monochromator slit width was changed caused errors in the calculated spectral irradiance of as much as 3%. Tests were made which indicated that the errors due to improper alignment of the test lamps with the monochromator slit increased as the slit to lamp distance was increased. The electronic noise near the end of a detector range was on the order of 5%. Smaller sources of error came from current fluctuations within the lamps and from uncertainties in reading the recorder charts. An estimate of the overall error in the comparison is ±1.5% with the primary contributions from the alignment error.

The result of the current investigation has been to obtain a calibration procedure for working standards of spectral irradiance which can be easily carried out with a minimum of error and complications.

BLANK PAGE

A PRISMATIC CUTOFF FILTER FOR TRANSMITTING SHORT WAVELENGTHS

M. P. Thekaekara and S. E. Read

Cutoff filters which transmit radiation of long wavelengths but absorb all short wavelengths are commercially available for a wide range of cutoff wavelengths. Such filters can be readily manufactured by mixing different coloring agents with fused silica or glass. The coloring agents have the property of selectively absorbing light of wavelength below a certain maximum. Cutoff filters which transmit selectively short wavelengths, but not long wavelengths, do not exist since such a result cannot be obtained from chemical properties of color pigments. These filters are highly desirable. They have many applications in optics, as for example, for filtering out unwanted orders in grating spectra, for isolating ultraviolet radiation for specific purposes, for removing the highly intense long wavelengths from the scattered radiation within monochromators, and for filter techniques in radiometry. In the spectrum of many continuous sources such as that of the sun and of solar simulator sources, the energy in the UV is very small compared to that in the visible and IR, and hence measurements in the UV are subject to large errors due to the scattered light of long wavelengths.

The property of total internal reflection exhibited by refracting materials can be used effectively for selective transmission of UV. Fused silica or quartz which has a high transmittance and large variation of refractive index in the UV offers special advantages for this application. When a ray of light travels from quartz to air, if the incident angle is greater than the critical angle, the ray of light undergoes total internal reflection. The critical angle C is related to the refractive index n by the relation $\sin C = 1/n$. The variation of refractive index and critical angle with wavelength for fused silica, Corning Code 7940, is shown in Figure B-34. For our present investigation we are interested in a prismatic cutoff filter which transmits effectively all wavelengths below 4450A. The critical angle corresponding to 4450A is 43° - 0'.

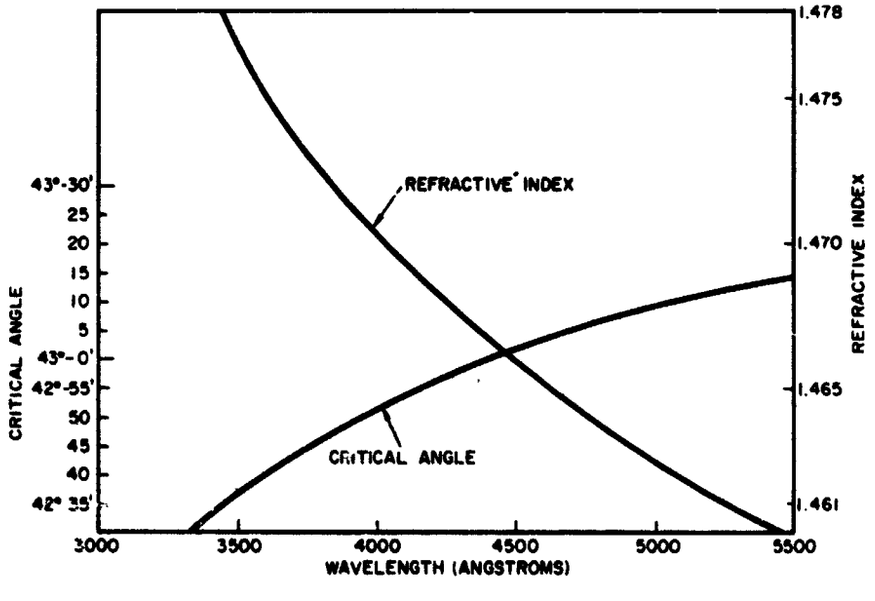


Figure B-34–Refractive index and critical angle of fused silica-Corning code 7940.

Two quartz rhombs of the same shape as the well-known Fresnel's rhomb (1), are machined out from rectangular blocks as shown in Figure B-35. The two rhombs are fused together in optical contact as shown in Figure B-36. It is important that the sides AE' and FB' (Figure B-36) be exactly parallel to each other, as also the sides E'F and AB'. The opposite angles also should be equal, and the two rhombs should be exactly identical. They are enclosed in a metal box with clear windows at opposite ends for incident and emerging light. The inside of the box is painted black for absorbing the refracted long wavelength radiation. The path of a ray of light of wavelength less than 4400A is as shown in Figure B-36. The dimensions of the sides of the rhombs are not of any significance. The rhombs may be made of any convenient size, depending on the possible applications and the availability of quartz of uniform density and refractive index. The degree of tolerance for the parallelism of the pairs of opposite sides and for the magnitude of the angle of the prism is about 1 minute of arc as will be seen from subsequent discussion.

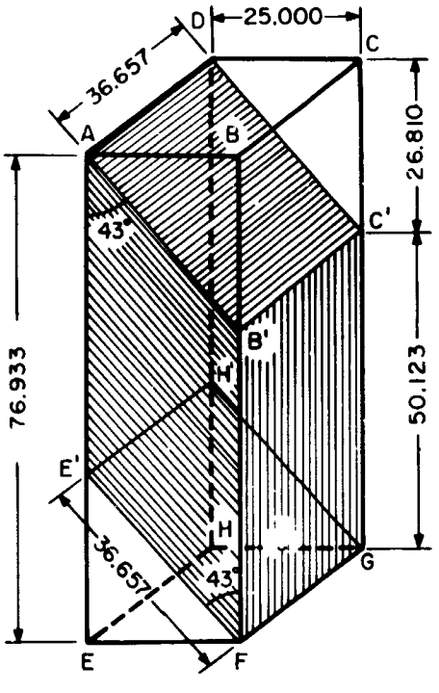


Figure B-35–Quartz rhomb machined from a rectangular block.

The simplest case we may consider is that of strictly collimated light incident normally on the front surface E'F. The light falls on the first refracting surface B'F at an angle of incidence 43°. All wavelengths equal to or less than that for which the critical angle is 43° are reflected internally without any loss of energy. Longer wavelengths suffer partial refraction and partial reflection. The refracted beam is absorbed by the walls, but the reflected beam remains within the quartz and hence affects the sharpness of the cutoff. The energy in the reflected beam can be computed from Fresnel's laws of reflection (2). The laws may be written

$$\frac{R_s}{E_s} = -\frac{\sin(\varphi - \varphi')}{\sin(\varphi + \varphi')}; \frac{R_p}{E_p} = \frac{\tan(\varphi - \varphi')}{\tan(\varphi + \varphi')}.$$

The symbols E and R mean the amplitudes of the electric vector in the incident and reflected light, respectively, the subscripts s and p denote the perpendicular and parallel planes of polarization, φ' and φ are the angle of incidence in quartz and angle of refraction in air. If the incident light is unpolarized, the ratio of reflected energy to incident energy is

$$C = \left[\left(\frac{R_s}{E_s}\right)^2 + \left(\frac{R_p}{E_p}\right)^2\right] \times \frac{1}{2}.$$

The ray of light undergoes two reflections in a rhomb, and hence the ratio of reflected to incident energy is given by C^2.

Values of C^2 have been computed for $\varphi' = 43°$, refractive index that of fused silica, Corning Code 7940 for wavelengths above 4450A at intervals of 200A. The results are presented in Figure B-37, (the curve to the extreme right). They are also given in Table B-12. Since there are two rhombs, the incident light undergoes reflection at four surfaces, and hence the ratio of reflected to incident light is C^4. Thus at 5000A the percentage of reflected energy is 19% and at 6000A it is as low as 8%. The transmittance percentage is slightly less than this since there is an additional loss for all wavelengths of about 8% due to the two quartz-air interfaces at entrance and exit.

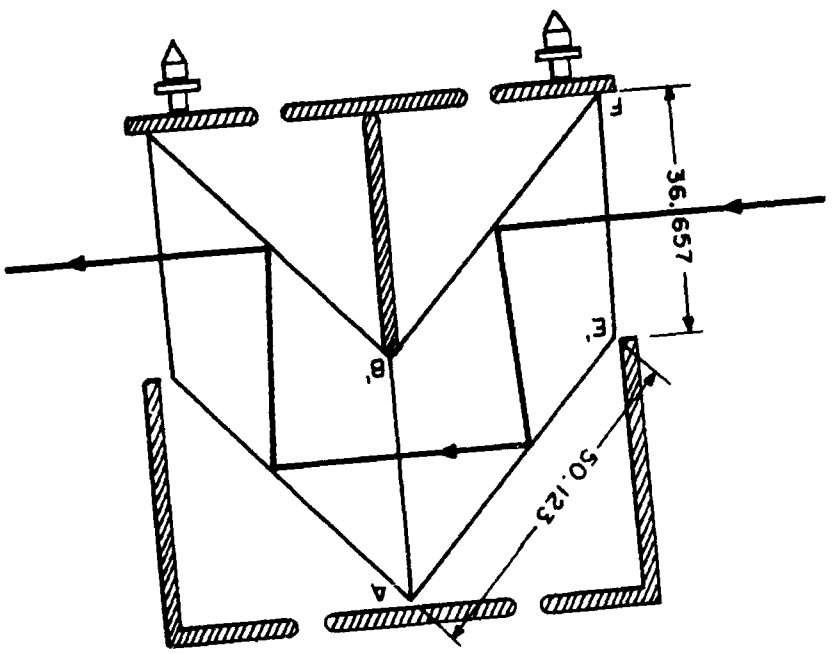


Figure B-36–Quartz rhombs in optical contact.

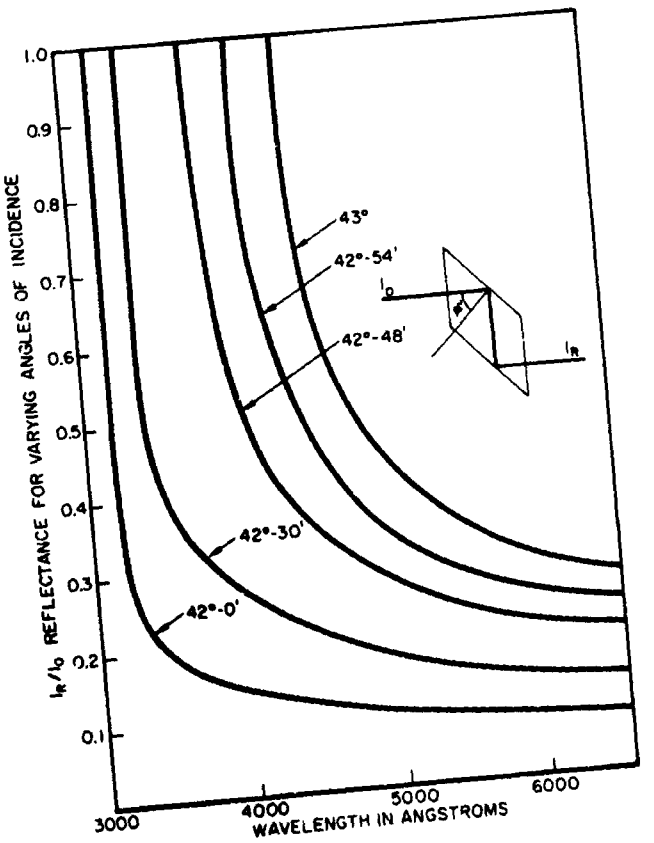


Figure B-37–Internal reflectance for two reflections in a 43° double prism of fused silica for angles of incidence less than critical angle 43° (λ 4450A).

Table B-12
Internal Reflectance for Varying Angles of Incidence
for Two Reflections on the 43° Rhomb

λ \ φ′	43	42-54	42-48	42-30	42
3000	1	1	1	1	1
200	1	1	1	1	1
400	1	1	.2123	.7999	.2123
600	1	1	.1616	.3576	.1616
800	1	1	.1501	.3157	.1501
4000	1	1	.1373	.2712	.1373
200	1	.6877	.1275	.2421	.1275
400	1	.5259	.1203	.2209	.1203
600	.6271	.4323	.1132	.2023	.1132
800	.5052	.3772	.1080	.1894	.1080
5000	.4405	.3392	.1039	.1773	.1039
200	.4115	.3111	.0986	.1688	.0986
400	.3553	.2884	.0960	.1631	.0960
600	.329	.267	.223	.148	.087
800	.307	.254	.213	.141	.083
6000	.291	.245	.20[illegible]	.137	.081
200	.278	.238	.199	.133	.079
400	.269	.233	.197	.131	.078
500	.267	.231	.196	.130	.078

The ideal case of a perfectly collimated incident beam is hardly ever realized in practice except in the study of stellar spectra. The second rhomb inverted with respect to the first ensures a high degree of sharpness of cutoff even for a decollimated beam of light, as may be seen from Figure B-38. Assume that a beam of light of λ = 6000A is incident on the first refracting surface at 45°. It undergoes total internal reflection at this surface and the next. But at the third surface the angle of incidence is 41° which is considerably less than the critical angle and hence only a small percentage is reflected. A further attenuation occurs at the fourth refracting surface. For light incident at angles greater or less than the angle of the prism, the ratio of reflected to incident energy is given by C^2, since partial reflection takes place in one rhomb and total reflection in the other. The computed values of C^2 for four values less than 43° are given in Figure B-37 and Table B-12. The curve for a value 43° - θ is obviously the same for a value 43° + θ. The loss in energy is considerably greater than for 43° incidence. Thus the effect of sharp cutoff is ensured also for a decollimated beam. The curves of Figure B-37 also show that the prismatic cutoff filter makes a decollimated beam more sharply collimated by selectively refracting the oblique rays.

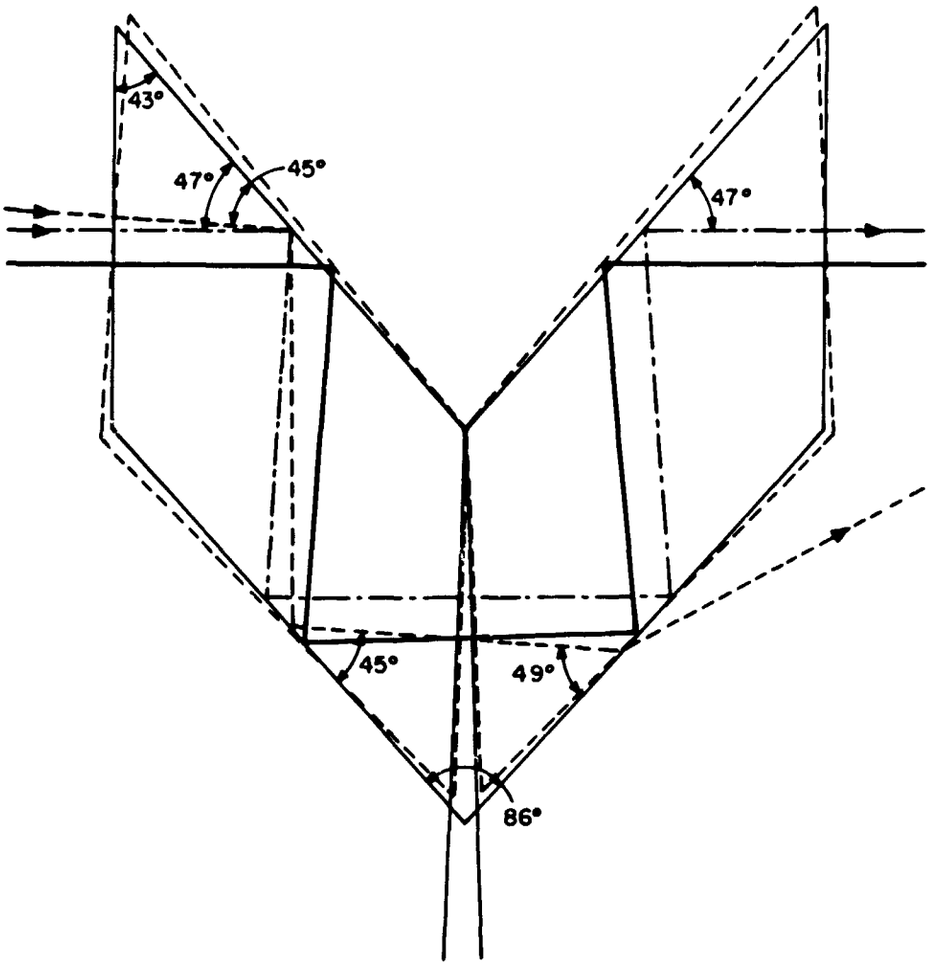


Figure B-38–The second rhomb inverted with respect to the first ensures a high degree of sharpness of cutoff.

A slight modification which enhances a great deal the usefulness of the prismatic cutoff filter is to mount the two rhombs separately so that they can rotate by small equal angles in opposite directions. The position after a slight rotation is shown by the dashed lines in Figure B-38. Let the incident light be in the same direction as previously. The angle of incidence at each of the refracting surfaces is reduced by the same amount, and hence the cutoff occurs at a shorter wavelength. For a rotation of 15 minutes the cutoff changes from 4450A to 3760A as may be seen from Figure B-34. Thus a filter is obtained with a variable cutoff. Separation of the rhombs, however, causes an additional loss of about 8% for all wavelengths due to the two extra quartz-air interfaces.

Loss of energy due to the path length in quartz is relatively insignificant. Hence the filtering off of the long wavelengths and the sharpness of cutoff can be increased to any desired degree by fusing end to end more than one pair of rhombs.

In the quartz prismatic cutoff filter we have a means of isolating the short wavelength end of a spectrum. The instrument can be used for decollimated light. The wavelength of cutoff can be changed by rotating the rhombs. The sharpness of cutoff can be increased by fusing several pairs of rhombs together.

REFERENCES

1. Born, Max, and Wolf, Emil, Principles of Optics, The MacMillan Company, New York, 1964, 2nd Edition, page 51.

2. Jenkins F. A., and White, H. E., Fundamentals of Optics, McGraw Hill Book Company, Inc., New York, 1957, 3rd Edition, page 510.

BLANK PAGE

SPECTRAL RADIANT FLUX IN THE SOLAR ENVIRONMENT SIMULATOR

M. P. Thekaekara, J. D. Rancourt and S. E. Read

INTRODUCTION

This paper discusses five problems investigated by the Summer Workshop in connection with the measurement of the spectral radiant flux in the Solar Environment Simulator. Two different monochromators were used, both Perkin-Elmer Model 112 U, one with a LiF prism and the other with a quartz prism. Three types of detectors were available, thermocouple, 1 P 28 photomultiplier tube and lead sulfide phototube.

The sensitivity of the instruments was measured and the data are presented in graphical form. A new wavelength calibration was made using a large variety of wavelength standards. Special problems were encountered in extending the spectral range below .3μ to .25μ and above 2.5μ to 4.5μ. The data are presented as tables of wavelengths for drum counts at intervals of 10 counts.

The monochromators are used with different slit widths, very narrow slits of 20μ width or less when the source is strong and an f/10 mirror is used to focus the light on the slit and very wide slits of 1.5 mm width when the monochromator views the scanning plate of the SES. The effect of the slit width is a major consideration when line sources are compared with a continuum like that of the tungsten standard lamp. This effect was experimentally studied for two mercury lines 2536A and 4046A and for the continuum.

A source of error in the automatic data reduction program of the in-chamber monitoring of the spectral radiant flux in the SES was the subject of another study. This error arises because of a time lag of about 3 seconds between a gain change and a corresponding change in the signal output.

The final section of the paper discusses the results of a comparison between two major competitors for solar simulation, the Hg-Xe arc which is currently used and the Xe arc which is an alternate possibility. The results are incomplete since due to the difficulties of adapting the automatic data reduction program to this problem, only the UV and visible regions of the spectrum were scanned. The data were reduced by laborious conventional technique. The results are presented in graphical form. The results of a study on the Hg-Xe spectrum in the wavelength range up to 2.5μ are also presented.

SENSITIVITY OF THE PERKIN-ELMER MONOCHROMATOR

The conventional method for measuring the spectral radiant flux of an unknown source consists of making two spectrum charts using the unknown source and a standard source of which the spectral radiant flux is known with sufficient accuracy. If the two sources are focussed on the slit of the monochromator by the same optical arrangement, and the recording system is the same, the ratio of the signals from the two sources at a given wavelength setting is equal to the ratio of the spectral radiant flux.

The sensitivity of the monochromator may be defined as the signal per unit incident energy. The signal may be in inches of the chart paper, millivolts of the thermopile detector, microamps of the photomultiplier, or any other convenient units. The relative

spectral radiant flux of an unknown source can be determined by dividing the signal by the sensitivity.

As part of the Summer Workshop program, some studies were made on the spectral sensitivity of two Perkin-Elmer Model 112 U monochromators.

One of these instruments had a lithium fluoride prism and a thermocouple detector. Its mounting and method of operation have been discussed in detail by N. Zylich (1) who developed the automatic data processing technique used with this monochromator. Spectrum measurements are made by first scanning a 1000 watt quartz-iodine standard lamp and then the collimated beam of the solar environment simulator. The monochromator is mounted permanently in a dark room built outside one of the exit windows of the SES. Light is directed from the vacuum chamber to the monochromator by a sampling flat of effective area one square foot, and a telescopic combination of quartz lenses.

In the normal mode of operation, signals corresponding to the gain factor of the amplifier, the output of the amplifier, and the wavelength setting of the Littrow mirror are fed to a computer at very close wavelength intervals. A complete scan from 0.26 μ to 3.0 μ is made in 20 minutes. Data points are generated and recorded at equal intervals of time every fifth of a second, so that 6000 data points are available for every scan. The recorded scans of the standard lamp and the SES spectrum are combined by the GSFC IBM 7020/7094 with stored data on the spectrum of the standard lamp and the wavelength calibration of the monochromator. The spectral radiant flux of the SES is computed for all the data points, the total energy is integrated and normalized to be equal to one solar constant, and the data are presented as a histogram giving the energy in each 100A wavelength range.

This automatic data processing technique does not require any prior knowledge of the spectral sensitivity of the monochromator, since along with each set of scans of the SES a scan of the standard lamp is made, and the signal to intensity ratio is computed for each data point.

The signal output is also available as in ordinary recording spectrophotometers as a spectrum drawn on a continuous roll of paper. These spectrum charts are helpful for confirming the accuracy of the computer results. They may also be analyzed in the conventional manner to give the spectral radiant flux on occasions when computer time is not available, or the problem on hand does not justify the labor and expense of the computer program.

Spectrum charts of several runs of a 1000 watt standard lamp were available for a study of the spectral sensitivity. The deflection of the recorder pen was measured at close wavelength intervals on these charts. The readings from different charts were in close agreement. The spectral radiant flux of the standard lamp was read from a graph on an enlarged scale based on the NBS data. The reciprocal spectral sensitivity is obtained by dividing the spectral radiant flux by the corresponding signal. The NBS data on the standard lamp are in units of microwatts per (cm^2-nanometer); but since the signal is in arbitrary units (inches of chart paper reduced to a common gain factor 4), the reciprocal spectral sensitivity is also expressed in arbitrary units.

The slit width for these runs was 1.5 mm. A large slit width was necessary partly because of the low sensitivity of the thermocouple detector, but mainly because the source was located 20 ft away from the monochromator in the SES. Even with this wide slit, and using the highest gain setting, at wavelength 4000A the pen deflection was less than 1/2 inch, and at lower wavelengths, the signal readings did not permit sufficient accuracy. Hence the source was brought close to the instrument and the image of the filament was focussed on the slit by an f/10 concave mirror. This increased the signal by a factor of 500 and permitted accurate measurements down to 2500A. The signals in this range were

converted to the same scale as those measured earlier by comparing the charts in the 4000-10,000A range for the two positions of the lamp. The constancy of the ratio of the signals of the two sets of charts showed that errors due to the mirrors and prism being more fully illumined in the second case are rather small.

Scattered light in the monochromator proved to be a major problem in the range below 3000A. Of the total energy of the quartz-iodine lamp, only about 0.02 percent is in the UV range below 3000A. Since the prism has a relatively high dispersion in this range, for a given setting of the Littrow mirror the energy due to the corresponding wavelength band falling on the thermocouple becomes even more insignificantly small compared to the total light entering the monochromator. Correction for the scattered light was made by two UV transmitting, visible absorbing Corning glass filters 9863 and 5850.

The results are presented in Figure B-39. The y-axis is on a logarithmic scale since the value ranges from a maximum of 5 to a minimum of 0.03. The general features of the curve are readily explainable in terms of the known characteristics of the monochromator. The thermopile has practically the same response throughout the range. The rise of the reciprocal sensitivity in the visible and UV is mainly due to the increasing dispersion of the prism. As may be seen from the wavelength calibration, Table B-13, discussed in the next section, the increments in wavelength for a rotation of the littrow mirror of one drum count at wavelength settings of .25, .4, .6 to 1.0μ are respectively 1.4, 6, 23 and 48A. These increments are approximately proportional to the spectral bandwidth incident on the detector for a constant slit width. The product of the wavelength increment for a small rotation of the Littrow mirrow and the reciprocal spectral sensitivity is nearly the same for the whole wavelength range, except for certain spectral regions where other effects predominate.

The small humps in the curve at 1.4μ and 1.9μ and the rapid rise beyond 2.5μ are due to absorption by atmospheric water vapor in the long path length between the source and the entrance slit. The hump at .8μ is probably due to the lower reflectance of the aluminized surface of the mirrors in the monochromator. The rise in the curve at 2.25μ has not yet been explained. This is a range where there is a great divergence between different scans of both the tungsten lamp and the mercury-xenon lamp.

Spectral sensitivity measurements were made also on another Perkin-Elmer model 112 U monochromator which had a quartz prism and a 1 P 28 phototube detector. The phototube voltage was 400 volts and the slit width was 20μ. The signals were considerably stronger than with the thermocouple detector and the LiF prism. In the wavelength range below 2700A Corning glass filter 9863 was used to reduce scattered light.

The results are presented in Figure B-40. The y-axis is on a linear scale. The inset shows the range 3400 to 5700A on an expanded scale for the y-axis. The general features of the curve can be readily explained as due to a combination of the dispersion of the prism and the spectral sensitivity characteristic of phototubes having S-5 response. The wavelength calibration of the P.E. quartz monochromator given in Table B-14 shows the variation with wavelength of the spectral bandwidth seen by the phototube. Data on the spectral response of S-5 surface for equal values of radiant flux at all wavelengths are available in literature (2). The increase in reciprocal sensitivity in the short wavelength range is due to the decrease of both the spectral bandwidth falling on the exit slit and the response of the phototube, and the increase in the long wavelength range at 6000A is due to the response of the phototube alone.

WAVELENGTH CALIBRATION OF THE SPECTROPHOTOMETERS

The calibration of the two Perkin-Elmer model 112 spectrophotometers, one having a LiF prism and the other, a fused silica prism, was broken into 3 ranges: UV, visible,

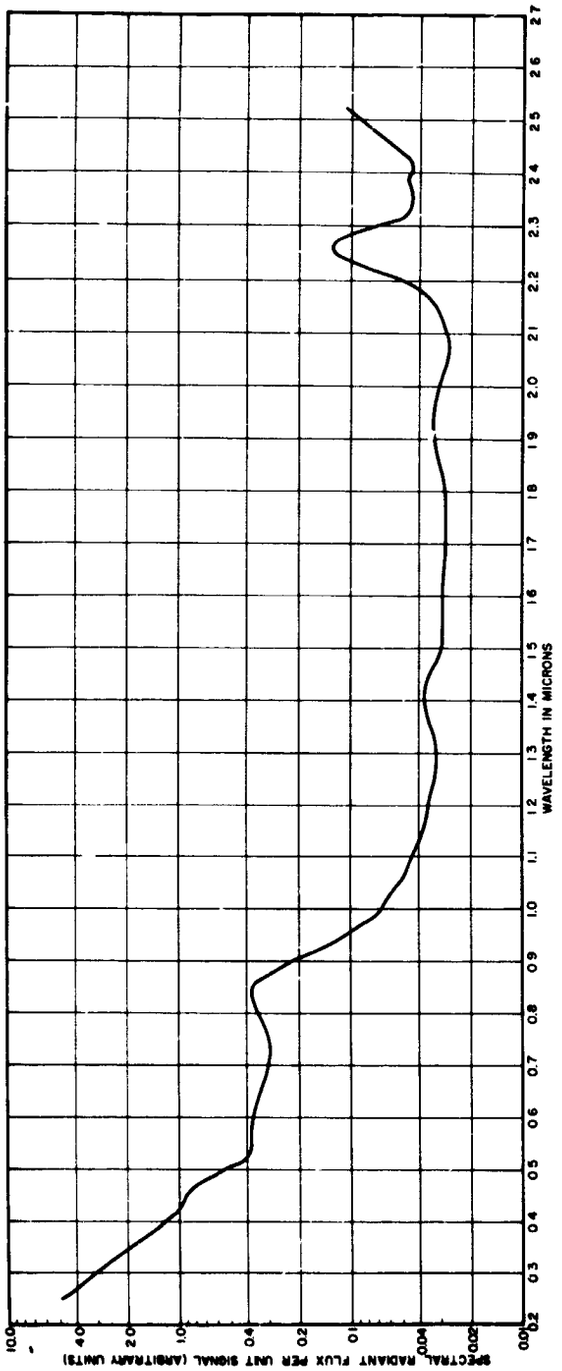


Figure B-39–Reciprocal sensitivity of the Perkin-Elmer model 112 monochromator–LiF Prism using a thermocouple detector.

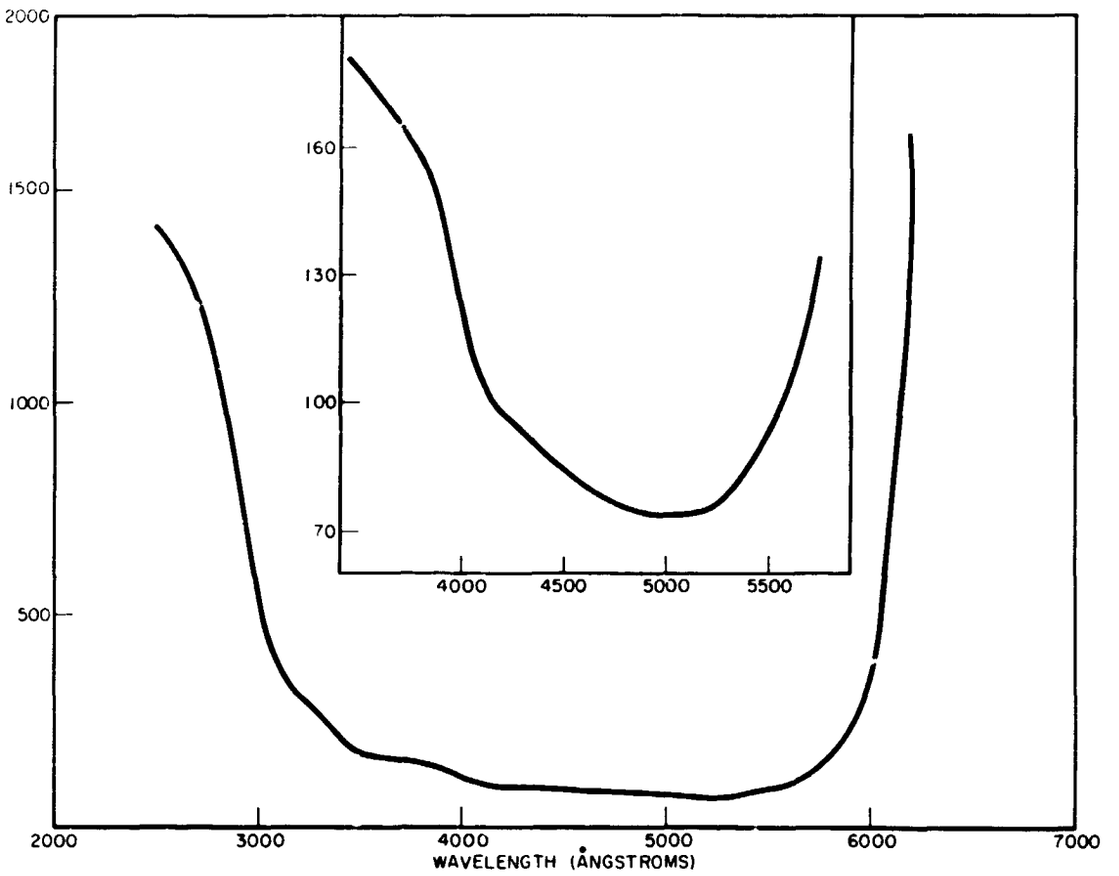


Figure B-40–Reciprocal sensitivity–Perkin-Elmer model 112 1P28–S5 surface–400 V–quartz prism. Spectral radiant flux per unit signal.

and IR. Interference filters, spectral line sources, and various absorption bands were used to correlate the wavelength falling on the detector to the reading on the drum which mechanically moved the Littrow mirror.

To identify the spectral lines of a given known source such as mercury, the following method was used. Two or three lines were tentatively identified on the output of the strip chart recorder. A wavelength scale was then set up along the intensity axis of the recorder and the positions of the identified peaks were plotted on the chart. A line was sketched between the identified peaks; if the positions of intermediate spectral lines, whose wavelengths and intensities were obtained from standard tables, fell near the sketched line, then we were certain that our original identifications were correct. On the other hand if almost all points were very distant from the line and the identification of many lines was impossible, then we realized that the original identification was in error.

In the ultraviolet range, the lines of a low pressure Pen Ray mercury lamp were used to obtain calibration points from 2482A to 3341A. In this range, a 1P21 photomultiplier tube with a quartz envelope was used as a detector.

In the visible range, high pressure mercury and mercury-xenon lamps were used. Since the sensitivity of the 1P28 drops very rapidly for wavelengths greater than about 6000A, a thermocouple detector was used beyond this.

In the infrared range, the lines of a high-pressure mercury discharge were identified out to 19,700A. The lines from a high-power xenon lamp were identified out to 20,262A. Finally, calibrated dielectric narrow passband filters were used to extend the

Table B-13
Perkin-Elmer 112A-SES – Wavelength Calibration – LiF Prism

Count	λ	Δλ	Count	λ	Δλ	Count	λ	Δλ	Count	λ	Δλ	Count	λ	Δλ	Count	λ	Δλ	Count	λ	Δλ	Count	λ	Δλ
2280	2454	15	2050	2831	22	1800	3616	53	1550	6305	270	1290	17842	458	1050	26948	317	800	33952	256	550	39968	220
2270	2469	15	2050	2853	20	1790	3669	53	1540	6575	320	1280	18300	449	1040	27265	313	790	34208	254	540	40188	218
2260	2484	13	2040	2873	27	1780	3722	57	1530	6895	325	1270	18749	440	1030	27578	308	780	34462	254	530	40406	216
2250	2497	14	2030	2900	20	1770	3779	62	1520	7210	320	1260	19189	431	1020	27886	304	770	34716	252	520	40622	213
2240	2511	14	2020	2920	25	1760	3831	55	1510	7530	330	1250	19620	423	1010	28190	301	760	34968	251	510	40835	211
2230	2525	14	2010	2945		1750	3886	57	1500	7860		1240	20043	414	1000	28491		750	35219	250	500	41046	
2220	2539	14	2000		25	1740	3943	67			340	1230	20458	407	990		296	740	35469	249			208
2210	2553	16		2970	24	1730	4010	60	1490	8200	380	1220	20865	401	980	2875	293	730	35718	248	490	41254	206
2200	2569		1990	2994	26	1720	4070	65	1480	8580	390	1210	21266	394	970	2908	290	720	35966	247	480	41460	203
		15	1980	3020	27	1710	4135	95	1470	8970	420	1200	21660		960	29370	286	710	36213	245	470	41663	200
2210	2584	14	1970	3047	29	1700	4230		1460	9390	460			387	950	29656	284	700	36458		460	41863	198
2190	2598	18	1960	3076	28			90	1450	9850	470	1190	22047	382	940	29940	281			244	450	42061	195
2170	2616	16	1950	3104	30	1690	4320	90	1440	10320	480	1180	22429	375	930	30221	279	690	36702	243	440	42256	193
2160	2632	17	1940	3134	30	1680	4410	90	1430	10800	500	1170	22804	371	920	30500	276	680	36945	242	430	42449	189
2150	2649	15	1930	3164	30	1670	4500	95	1420	11300	520	1160	23175	365	910	30776	274	670	37187	240	420	42638	186
2140	2664	16	1920	3194	33	1660	4595	95	1410	11820	540	1150	23540	360	900	31050	272	660	37427	239	410	42824	
2130	2680	17	1910	3227		1650	4690	105	1400	12360		1140	23900	356		31322		650	37666	238			
2120	2697	18	1900		35	1640	4795	110			510	1130	24256	351	890		270	640	37904	236			
2110	2715	18		3269		1630	4905	125	1390	12870	560	1120	24607	348	880	31592	268	630	38140	234			
2100	2733		1890	3292	32	1620	5030	125	1380	13430	550	1110	24955	343	870	31860	267	620	38374	234			
		18	1880	3327	32	1610	5155	125	1370	13930	540	1100	25298		860	32127	265	610	38608	231			
2090	2751	20	1870	3362	35	1600	5280		1360	14520	490			339	850	32392	263	600	38839				
2080	2771	20	1860	3409	38			140	1350	15010	470	1090	25637	326	840	32655	262			229			
2070	2791	19	1850	3438	38	1590	5420	180	1340	15480	470	1080	25963	334	830	32917	261	590	39068	228			
2060	2810	21	1840	3478	40	1580	5600	220	1330	15950	480	1070	26297	328	820	33178	259	580	39296	226			
2050	2831		1830	3522	44	1570	5820	235	1320	16430	464	1060	26625	323	810	33437	258	570	39522	224			
			1810	3567	45	1560	6055	250	1310	16894	479	1050	26948		800	33437	257	560	39746	222			
			1800	3616	49	1550	6305		1300	17373	469					33695		550	39968				
																33952							

Table B-14
Perkin-Elmer 112A – Wavelength Calibration – Quartz Prism

Count	λ	Δλ	Count	λ	Δλ	Count	λ	Δλ	Count	λ	Δλ	Count	λ	Δλ
1220	2488	13	990	2902	24	740	3856	65	490	8570	500	240	23524	527
1210	2501	14	980	2927	25	730	3921	69	480	9070	550	230	24051	521
1200	2505	14	970	2952	26	720	3990	75	470	9620	570	220	24572	515
			960	2978	26	710	4065	80	460	10190	630	210	25087	510
1190	2529	15	950	3004	28	700	4145		450	10820	705	200	25597	
1180	2544	14	940	3032	29			82	440	11525	685			
1170	2558	16	930	3061	30	690	4227	85	430	12210	680			
1160	2572	16	920	3091	31	680	4312	90	420	12890	675			
1150	2588	16	910	3122	32	670	4402	100	410	13565	675			
1140	2604	18	900	3154		660	4502	108	400	14240	665			
1130	2622	17			32	650	4610	115						
1120	2639	18	890	3186	33	640	4725	120	390	14905	655			
1110	2657	17	880	3210	33	630	4845	130	380	15560	625			
1100	2674		870	3252	34	620	4975	135	370	16185	595			
		19	860	3286	36	610	5110	155	360	16780	585			
1090	2693	19	850	3322	39	600	5265		350	17365	580			
1080	2712	18	840	3361	42			180	340	17945	580			
1070	2730	19	830	3403	44	590	5445	190	330	18525	575			
1060	2749	20	820	3447	45	580	5635	212	320	19100	570			
1050	2769	21	810	3492	46	570	5847	223	310	19670	565			
1040	2790	21	800	3538		560	6070	230	300	20235				
1030	2811	22			48	550	6300	280			568			
1020	2833	23	790	3586	49	540	6580	330	290	20803	556			
1010	2856	23	780	3635	52	530	6910	350	280	21359	549			
1000	2879	23	770	3687	53	520	7260	390	270	21908	545			
			760	3740	56	510	7650	450	260	22453	538			
			750	3796	60	500	8100	470	250	22991	533			

range to 2.55 microns on the instrument with a quartz prism, using a quartz-iodine lamp as a source. More of these filters were available to extend the range to about 2.7 microns, but in the quartz prism instrument the water absorption was too great to allow any light of this wavelength region through to the detector. A dessicant was placed in this instrument, but no significant improvement was noted.

In the instrument with the LiF prism, it was possible to extend the range to 4.2 microns. In this case, the water vapor in the 40 cm path between the source and the slit was enough to give us well resolved water vapor absorption lines and these were easily identified. Cellophane and polystyrene films were used to obtain more absorption lines. Finally, the carbon dioxide absorption at around 4.2 microns completed the calibration.

The LiF prism instrument is now calibrated from about 0.25μ to about 4.3μ while the quartz instrument is calibrated from about 0.25μ to about 2.5μ. Tables B-13 and B-14 give the calibrations of the instruments.

Several lead sulphide detectors were available for the very near infrared range, but with the circuits used, their performance was inferior to that of the thermocouple, although in their range they should normally have a higher signal to noise ratio than thermocouples.

The narrowest slit which gave a good signal to noise ratio was always used. This was of the order of 8 microns in the ultraviolet range since the 1P28 is very sensitive in this range. In the infrared range, where we used a thermocouple, the use of a wide slit

was necessary. This did not degrade our results since the prism resolution was much less in this range. A further discussion of slit width effects can be found in the next section.

Table B-15 gives a listing of the wavelengths of the absorption lines of water vapor, polystyrene film, and carbon dioxide which were used for the calibration in the range beyond 2.5μ. The wavelengths were obtained from the charts and tables published by the Commission on Molecular Structure and Spectroscopy, International Union of Pure and Applied Chemistry (3). The wave numbers given in the tables were converted into wavelengths with due correction for the refractive index of air. A slit width of 50μ was found to be the most suitable for giving a large number of easily identifiable water vapor absorption lines. The relative humidity was 40% ± 5% and the temperature 70°F. The path length in the air was 40 cm.

Table B-15
Lines for Wavelength Calibration in the Infrared

Water Vapor (μ)			Polystyrene (μ)	CO_2
2.5324	2.6082	2.7357	3.2150	4.2433
2.5363	2.6274	2.8157	3.2375	4.2587
2.5517	2.6466	2.8285	3.2702	4.2822
2.5652	2.6604	2.8676	3.3044	
2.5759	2.6658	2.8724	3.3182	
2.5879	2.6927	2.9017	3.3202	
2.5961	2.7020		3.5072	

For the reduction of data, a computer program was used with partial success. The first part of the computation employed a curve fitting program already available for the C.D.C. 160A computer. In the wavelength range below .35μ and above 1.5μ the program was quite satisfactory. In the intermediate range where the dispersion of the prism changes rapidly with wavelength, the least square solution gave large differences between observed and computed values, and hence a graphical method was used to determine the corrections to the computer results. The wavelengths for counts at intervals of 10 were obtained by a computer program for evaluation of a polynomial.

THE EFFECT OF SLIT WIDTH ON SPECTRAL LINES AND CONTINUUM

The entrance slit of a scanning spectrophotometer does not only affect the amount of light entering the instrument but also affects the width of the resulting curve drawn by the chart recorder.

Several theories have been advanced on the effect of changing the slit by a given factor on the final curve. To resolve this question, the following study was made.

The experiment was divided into 2 parts: spectral lines and spectral continuum. The various theories predicted different values, depending on the type of light source, for the ratio of the areas under the curves drawn by the chart recorder for various slit widths.

A Perkin-Elmer model 112 spectrophotometer was used. This instrument has adjustable entrance and exit slits which are mechanically coupled. These are always of equal width and are adjusted by a micrometer screw.

A mercury Pen-Ray lamp was used as the line source and a quartz-iodine tungsten lamp was the continuum source. The 4046A and 2536A mercury lines were used because of their very narrow widths.

Because of the extremely high sensitivity of the photomultiplier tube in this range, a narrow horizontal slit had to be introduced between the source and the entrance slit to cut down the light intensity reaching this detector. This slit was necessary only for the mercury lamp and was removed when the tungsten source was used. Another precaution taken was to ensure that no extraneous light entered the instrument. Thus, the measurements were made in a darkroom with only low-level incandescent lighting. The overhead fluorescent lights produced a very large effect, especially when the large slit openings were used, and hence were turned off during the measurement.

A horizontal slit does not affect the resolution of the instrument since only the effective base of the prism enters into the equations(4); a horizontal slit does not affect the effective base. An optical schematic of the instrument is shown in Figure B-41.

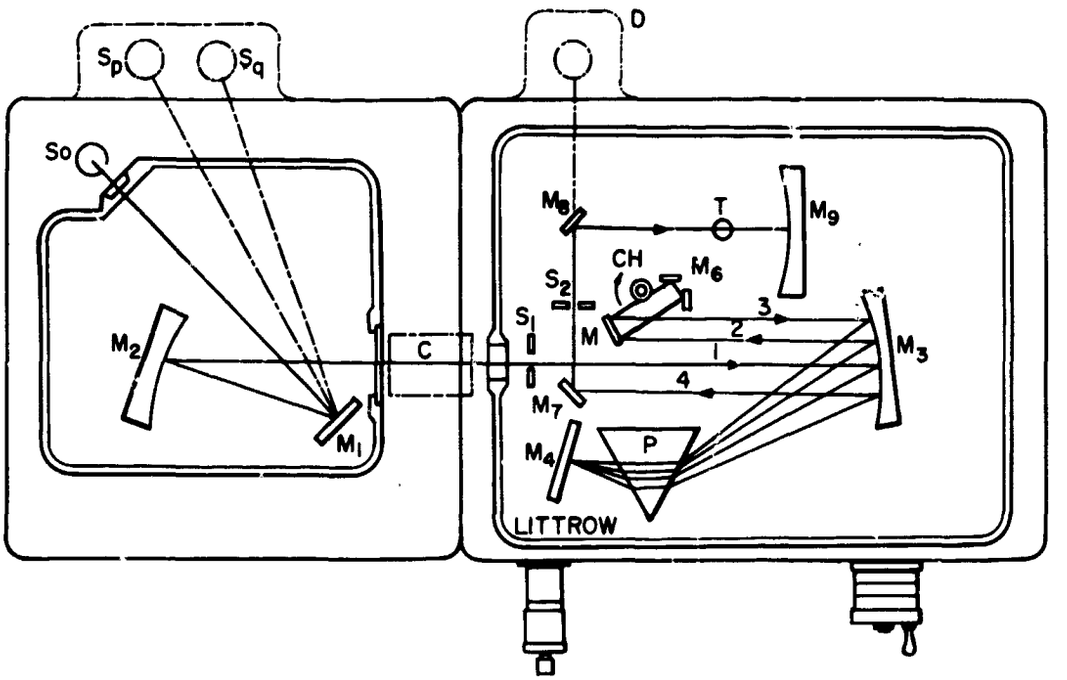


Figure B-41–Optical schematic of Perkin-Elmer monochromator.

A traveling microscope was used to correlate the micrometer readings with the actual slit width. The results of these measurements are listed in Table B-16.

A concave mirror was used to focus the light from the tungsten filament on the entrance slit; the mirror had a speed of f/10 which matched that of the instrument. No mirror was used with the mercury source since it already had enough intensity. Also, if the mercury source had been focussed on the slit at the large settings, its light would not have filled the entrance slit.

A few typical profiles of the 2536A mercury line are shown in Figure B-42. The widths of the lines are 0.47, 1.4, 7.1, and 14.1A for 8, 100, 500, and 1000μ slit widths, respectively.

Table B-16
Correlation Between Micrometer Screw and
Measured Width of Entrance Slit

Micrometer (microns)	Measured (microns)	Micrometer (microns)	Measured (microns)
0.0	0.0	500.0	505.5
10.0	0.0	600.0	612.1
20.0	10.2	700.0	703.6
30.0	20.3	800.0	802.6
40.0	30.4	900.0	896.6
50.0	40.7	1000.0	985.5
90.0	83.8	1100.0	1087.1
100.0	91.0	1200.0	1186.2
200.0	189.2	1300.0	1278.9
300.0	299.7	1400.0	1374
400.0	408.9	1500.0	1473

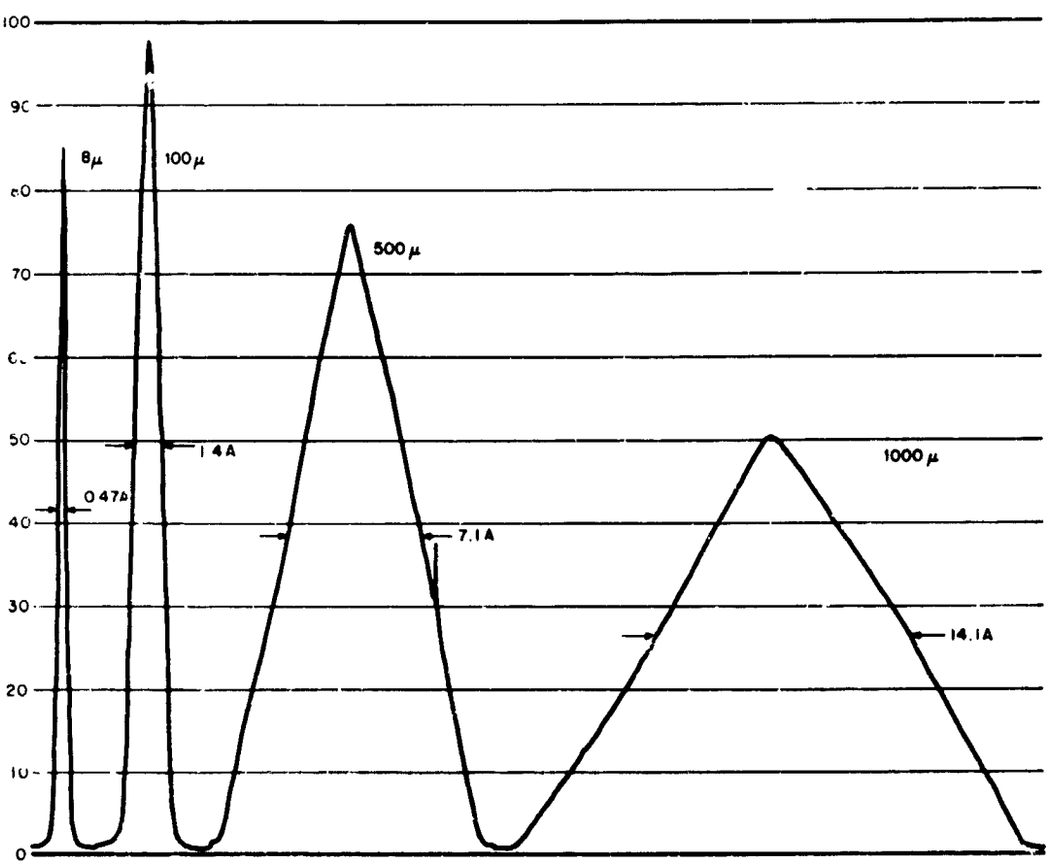


Figure B-42--Recorder output for the 2536A mercury line for slits widths 8μ, 100μ, 500μ and 1000μ, at same scanning speed, but amplification adjusted to keep the line within the scale.

Streiff and Ferriso (5) report that for this type of instrument, we should expect a line width of

$$\Delta\lambda = \Delta\lambda_g + \Delta\lambda_d F(a)$$

where

$\Delta\lambda$ = total spectral slit width

$\Delta\lambda_g$ = "geometrical" spectral slit width

$\Delta\lambda_d$ = diffraction spectral width

$F(a)$ = diffraction contribution function

a = $SD/\lambda f$

s = entrance slit width

D = aperture of instrument

λ = wavelength at which measurement is being made

f = focal length of collimator

The term $\Delta\lambda_g$ is derived by the above authors from geometrical optics.

$$\Delta\lambda_g = \frac{d\lambda}{dx} S = \frac{d\lambda}{dN} \cdot \frac{dN}{d\theta} \cdot \frac{d\theta}{dx} S = \frac{\sqrt{1 - N^2 \sin^2 \phi}}{\frac{dN}{d\lambda}(8 \sin \phi)} \frac{S}{f}$$

where

N = index of refraction

2ϕ = apex angle of the prism.

This report gives, for a slit width of 100μ at 2536A, a value of 1.1A for $\Delta\lambda_g$.

Under the same conditions as above, $a = SD/\lambda f = 4$, and, according to a plot given by these same authors, the diffraction contribution function $F(a)$ is approximately 0.6. The factor $\Delta\lambda_d$ is the limit of resolution imposed by the finite size of the prism and by the index of refraction:

$$\Delta\lambda_d = \frac{d\lambda}{dN} \cdot \frac{1}{4B},$$

where B = base of the prism.

The value of this factor is given as 0.02A. Hence, the second term in Equation (1) is 0.012, which is negligible in this case. Thus, the expected width should be approximately 1.1A. The measured width is 1.4A.

When the second term is negligible and since it varies very slowly with wavelength, we should expect that the width of the spectral line will vary directly with the slit width. This is true for the results quoted above for the 100, 500, and 1000μ slit widths.

Table B-17
Comparison of the Square of the Entrance Slit Ratio to the Ratio of the Areas Under the Respective Curves

2536 A Mercury Line					4046 A Mercury Line					Continuum (~ 4046 A)				
Slit (μ)	Area (sq.in.)	Slit Ratio	$\left(\text{Slit Ratio}\right)^2$	Area Ratio	Slit (μ)	Area (sq.in.)	Slit Ratio	$\left(\text{Slit Ratio}\right)^2$	Area Ratio	Slit (μ)	Area (sq.in.)	Slit Ratio	$\left(\text{Slit Ratio}\right)^2$	Area Ratio
10.2	1.47	2.99	8.88	5.93	10.2	.819	2.99	8.88	7.47	10.2	4.48	2.99	8.88	5.2±2
30.4	8.72				30.4	6.12				30.4	23.36			
30.4	.278	2.37	5.6	5.13	30.4	.025	3.0	9.0	10.6	244.4	5.552	2.068	4.28	4.47±14
72	1.426	1.26	1.60	1.63	91	2.66				505.5	24.80			
91	2.324													
91	1.17	2.08	4.30	4.60	91	.546	2.08	4.30	4.78	985.5	10.40	1.250	1.562	1.557
189.2	5.384				189.2	2.61	2.16	4.67	4.58	1232.5	16.19	1.195	1.428	1.453
					408.9	11.96				1473	23.53			
505.5	4.133	1.95	3.80	3.91±.13	408.9	4.375	1.24	1.54	1.66	1473	13.76	1.067	1.138	1.113
985.5	16.15				505.5	7.264	1.95	3.80	3.85±.09	1572	15.57	1.062	1.127	1.116
					985.5	27.95				1669	17.38	1.058	1.119	1.113
										1765	19.60			
					985.5	10.44	1.49	2.23	2.53					
					1473	26.41	1.17	1.374	1.373	1473	8.448	1.264	1.598	1.602
					1727	36.28	1.08	1.162	1.167	1862	13.54			
					1862	42.34								

When, on the other hand, the diffraction contributions term is not negligible, we get a departure from the linear relation, as can be seen from Table B-16. In all cases with the smallest slit, the square of the slit ratio gives a value larger than the ratio of the areas.

Table B-17 lists the data obtained, along with the calculated values. An examination of these results shows that the area under a spectral line obeys, within the experimental error and when $\Delta\lambda_d$ is negligible, a relation

$$\frac{\text{area under curve for slit width } w_1}{\text{area under curve for slit width } w_2} = \left(\frac{w_1}{w_2}\right)^2$$

In the case of a continuum, the area under the curve between two given wavelength markers was considered. The region chosen was that around 4046A. This continuum region obeys the above relation quite closely, also.

The conclusion that can be drawn from this is that when a spectrum, composed of a superposition of lines and a continuum, is studied, a change in the slit width of the spectrophotometer will not change the relationship of the area under the lines to the area under the continuum, provided the contribution to the line width due to the slit is sufficiently large compared to that due to the other causes such as diffraction and natural line width.

ERROR ANALYSIS OF SES MONOCHROMATOR SIGNAL OUTPUT

The automatic data processing system attached to the Perkin-Elmer monochromator of the SES has a provision whereby the gain factor is changed suitably whenever the signal output tends to go beyond the limits of a preset range. The advantage of this automatic change is that the signal always remains below the maximum for the recorder and associated electronics and above the lower limit required for accuracy of measurement. However, there is a time lag of about three seconds between the change in gain factor and the corresponding change in the signal. The signal output during this interval when multiplied by the new gain factor yields erroneous values of the signal. An attempt has been made to determine the percentage error caused by this time lag.

Figure B-43 shows the signals received by the computer for the standard lamp QM 19 in a short wavelength range. The wavelengths have been obtained by conversion of Epsco counts to Angstroms by an approximate formula. The signal strengths have been plotted for every fourth printout; the values are the product of the signal as given by the printout and the appropriate gain factor. The three peaks correspond to changes in the gain factor. The insets show the peaks on an enlarged wavelength scale; they are a plot of all the data points at intervals of 1/5 second. These show that the signals adjust themselves to the changed gain only after an interval of two to three seconds. Figures B-44 and B-45 show two short wavelength ranges in the spectrum chart of the Hg-Xe arc, similarly reduced to a uniform gain factor. The dashed curve below the peaks and above the dips indicates what the signal would be if the response of the recording electronics were immediate. The automatic data processing program is one that integrates the area under the curve over intervals of 100A, determines the ratio of the areas of the Hg-Xe and QM 19 curves and multiplies the ratio by the spectral irradiance of QM 19 in order to give as the final result the spectral irradiance of the Hg-Xe arc. The computer treats the spurious peaks and dips as genuine, causing a major source of error.

Table B-18 gives a list of the percentage error for each of the peaks and dips and Table B-19 for wide wavelength bands which include the continua. A_1 is the area under the curve of the true signal for a range of 20 printouts (four seconds), A_2 the area under

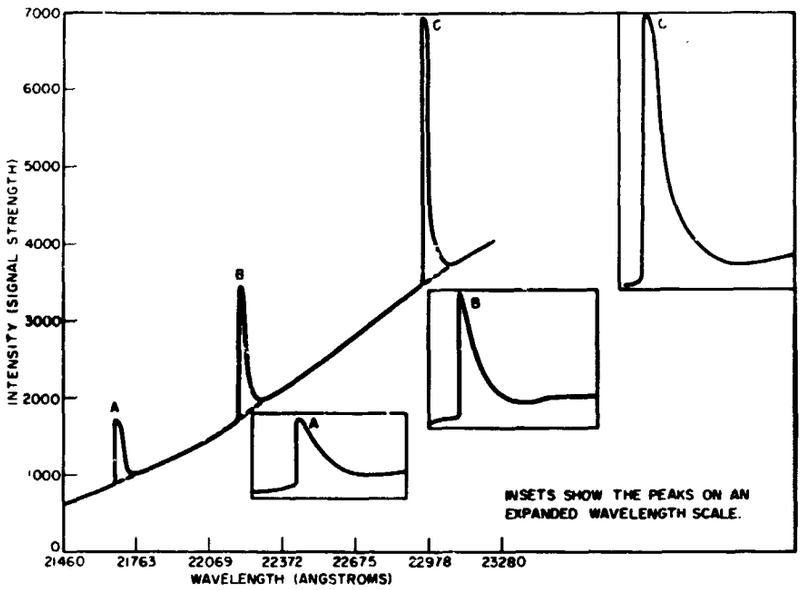


Figure B-43–Signal strength of lamp QM 19 from computer printout data.

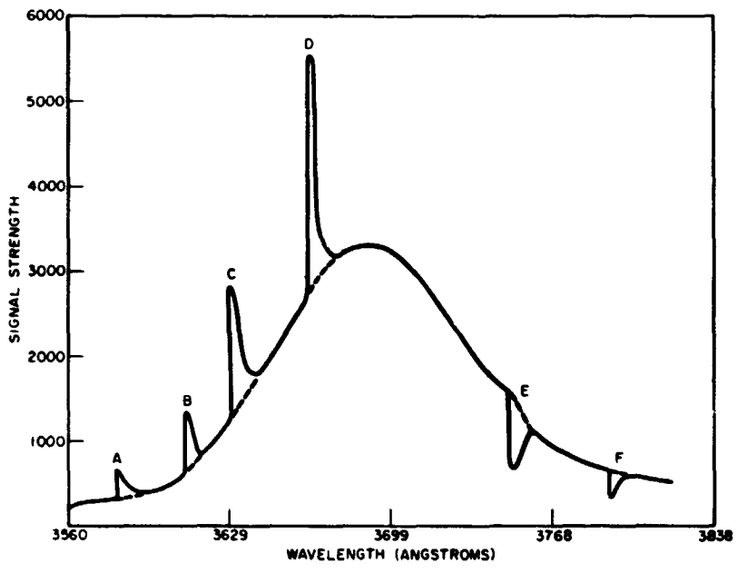


Figure B-44–Signal strength of lamp (Hg–Xe arc) from computer printout data–3560-3838A.

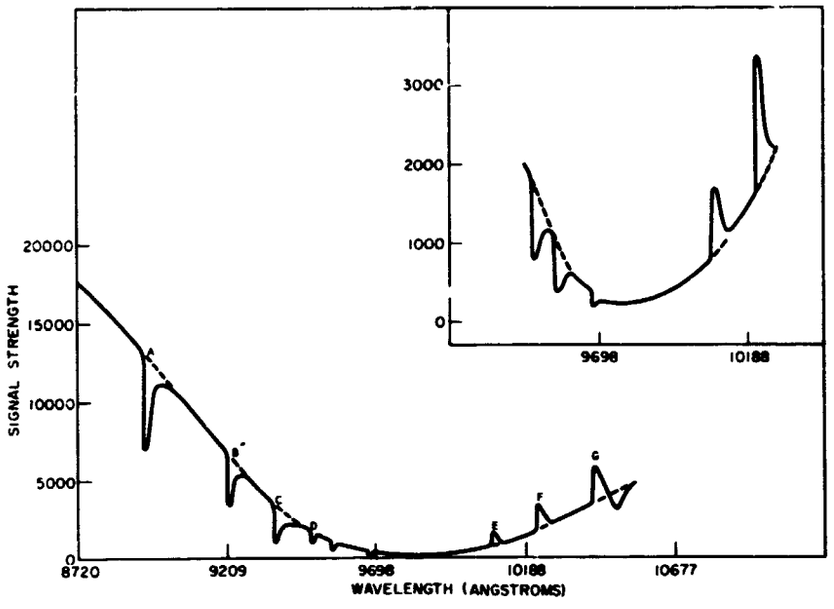


Figure B-45–Signal strength of Hg-Xe arc from computer printout data – 8720-10,677A.

Table B-18
Percentage Error Due to Each Gain Change

Figure	Peak or Dip	Gain Change From	To	A_1	A_2	A_2-A_1	$\frac{A_2-A_1}{A_1}$ Percent
B-43	A	15	14	45	56	11	24.4%
	B	14	13	94	115	21	22.3
	C	13	12	181	236	55	30.4
B-44	A	13	12	18	23	5	27.8
	B	12	11	48	54	6	12.5
	C	11	10	92	100	8	8.7
	D	10	9	180	192	12	6.7
	E	9	10	76	62	-14	-18.5
	F	10	11	37	32	- 5	-13.5
B-45	A	9	10	144	127	-17	-11.8
	B	10	11	65	52	-13	-20.0
	C	11	12	35	29	- 6	-17.1
	D	12	13	14	12	2	14.3
	E	15	14	10	11	1	10.0
	F	14	13	24	26	2	8.3
	G	13	12	48	51	3	6.3

Table B-19
Percentage Error Over Wide Wavelength Ranges

Wavelength Range From	To	A_3	A_4	$\frac{A_4-A_3}{A_3}$ Percent
21460	23280A	1270	1357	6.85
3560	3690	557	588	5.27
3690	3820	587	568	-3.23
8720	9500	600	562	-6.33
9950	10530	99	105	5.05

the curve of the data as processed by the computer, for the same range, both being for individual peaks and dips, A_3 the area under the curve of the true signals for a wide wavelength band, A_4 the area under the curve of the computer data for the same band.

We observe that the percentage error for single gain change may be as high as 30 percent. In an unfavorable case of the same 100A range having a dip in the Hg-Xe curve and a peak in the QM 19 curve the error in spectral radiance may be as high as 60 percent.

Three methods are suggested for correcting this error:

1. Stop the wavelength drive before each gain change and restart it after an interval of three seconds.

2. Modify the electronics to reduce the time lag to less than 1/5 second.

3. Change the computer program to ignore the 15 data points received during the time lag and substitute for them the arithmetical mean of the data points before and after this interval.

COMPARISON OF Hg-Xe ARC AND Xe ARC

The relative merits of the mercury-xenon arc and the xenon arc as sources for solar simulation have been studied by different laboratories. Studies on the mercury-xenon arc have been discussed in earlier Reports of the Goddard Summer Workshop Program (6). Detailed measurements have recently been made by the Thermal Systems Branch of Goddard Space Flight Center on the spectral radiant flux of a high intensity xenon arc. The results have been reported by Gallagher and McIntosh (7). The source used in these measurements was an Osram lamp with the special collimation system and IR filter manufactured by Spectrolab Inc., the X-25L solar simulator. The spectrum of the unfiltered xenon arc shows a large number of strong lines in the wavelength range 0.7 to 1.1μ. The IR filter absorbs a large percentage of the energy in this range, and renders the spectrum a satisfactory spectral match to the sun.

We were interested in comparing the spectra of the Hg-Xe arc and the Xe arc, both as modified by the collimating system of the solar environment simulator and both measured by the same monochromator and recording electronics. It is known that the collimating system of the SES module which has several quartz lenses and reflecting surfaces modifies to some extent the spectrum of the bare arc, especially in the ultraviolet. Precise measurements in the wavelength range below 3200A had hitherto been difficult due to the low intensity of the spectral irradiance standards, the high noise to signal ratio in the instrument, and the high intensity of the peaks of the Hg lines compared to the intensity of the tungsten continuum. The 1000 w quartz iodine lamp became recently available as a spectral irradiance standard, and hence the problem became considerably less difficult.

The two lamps which were compared were a Hg-Xe high pressure lamp, manufactured by Westinghouse, SAHX-2500-E and a Xenon high pressure lamp, manufactured by the Osram Co., Germany, XBO 2500W. Both were rated for 2.5 kw. The lamps were mounted successively in one of the test modules in Building 10. The test module is the prototype from which were copied the modules of the SES. It had been designed specifically for the Hg-Xe arc, the bulb of which is spherical, with about 6 cm outer diameter, and which burns with anode down. The Xe arc has an oval shaped bulb, of outer dimensions 7.5 cm length and 5.5 cm width. Further it is meant to operate with the anode up. The anode of the Xe bulb is cylindrical with a relatively large diameter, 1.6 cm, so that a relatively greater part of the flux is shielded from the ellipsoidal mirror which forms the first element in the collimation module. Further the Xe lamp has a greater overall length, 42 cm, compared to the 25 cm length of the Hg-Xe lamp. Hence a special mount had to be designed in order to fix the Xe lamp in the module so as to have the arc at the first focus of the ellipse. Because of these differences in the dimensions and the mounts of the lamps, the total intensity in the test plane was considerably less for the Xe lamp than for the Hg-Xe lamp.

It seems quite probable that these differences have little effect on the spectral distribution of the radiant flux, and that a comparison is meaningful. It is known that the spectral energy distribution changes slightly as one moves from the center of the test plane where intensity is maximum towards the periphery. The comparison of the two sources was made at the center only. A plane mirror inclined at 45° to the vertical was mounted in the beam to direct the beam in a horizontal direction towards the concave mirror of the monochromator, and a focussed image was formed on the entrance slit.

The Perkin-Elmer monochromator, model 112, with a quartz prism, was used for scanning the spectrum. A photomultiplier tube 1P28 was the detector. The voltage of the PM tube was set at 400 volts to ensure that saturation would not occur at the peaks of the Hg-Xe lines. Tests for linearity of response were made at all the major peaks using neutral density filters. For the xenon spectrum it was sufficient to make continuous runs at three different settings of the amplifier. For the Hg-Xe spectrum several runs at different gain settings varying between 5 and 19 were needed in order to bring all the strong lines within scale and at the same time obtain sufficiently strong signals at the continua. Individual lines were recorded at slow scanning speeds and at suitable gain settings. The area under the curve was read by a planimeter for each spectral range covered by 10 counts on the drum which rotates the Littrow mirror. As the dispersion of the prism decreases rapidly in the range of the PM tube, the spectral bandwidth covered by 10 drum counts increases from 15A in the UV 2500A to about 200A, the orange region of the visible, 6000A.

The area under the curve given by the planimeter and averaged for several charts was reduced to a common gain setting, and then multiplied by the reciprocal sensitivity (spectral radiant flux per unit signal) to give the energy in each of the ranges.

The results are presented as two histograms in Figure B-46. The histogram with a large number of sharp peaks is for Hg-Xe and the other for xenon. Xenon has a few lines superposed on the continuum in the range 4400 to 5000A. The half widths of these lines vary from 8 to 20A. Unlike the Hg-Xe spectrum, these lines contribute less than 5 percent of the energy in the 10 count ranges, and hence the contribution of the lines does not appear in the histogram. The two histograms have been normalized to the same area as the Johnson curve in the range 2400 to 6000A. The Johnson curve is also shown in Figure B-46 for the purposes of comparison.

These curves should be regarded as a preliminary attempt at comparison of the two spectra. A limited range of the spectrum which accounts for only 37 percent of the solar energy has been studied and hence the normalization procedure is not fully significant.

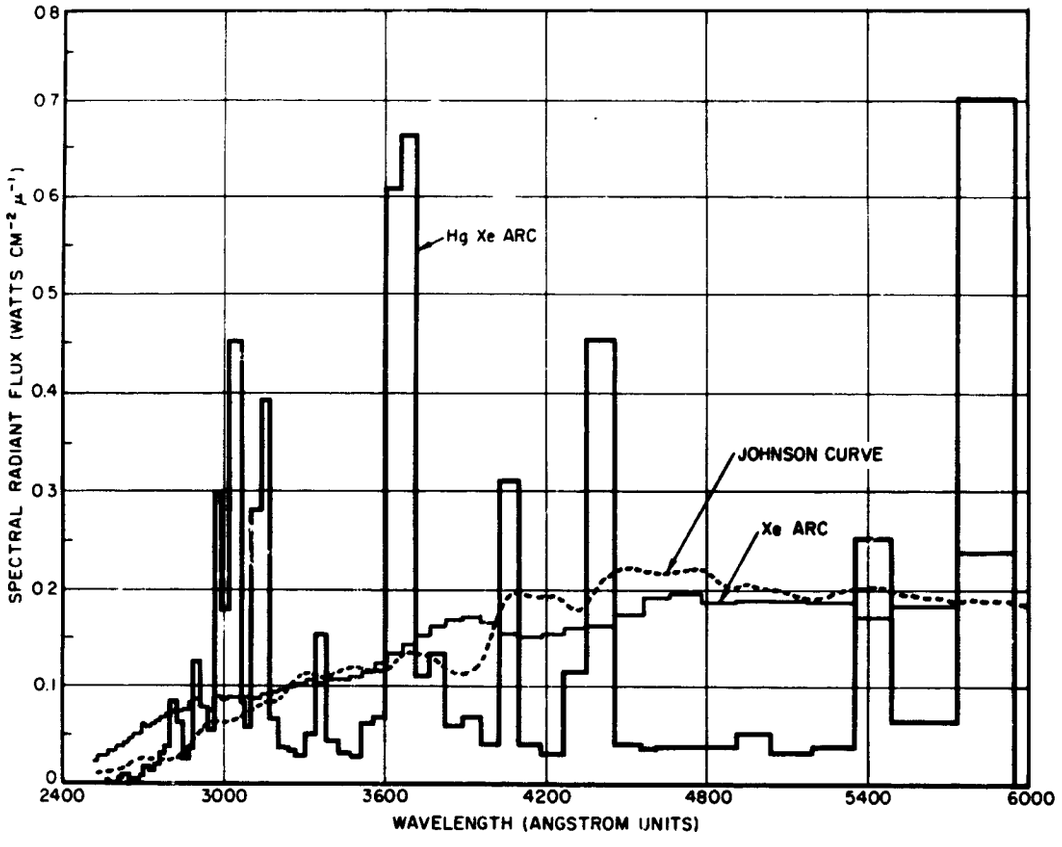


Figure B-46–Comparison of Hg–Xe in module both histograms normalized to same area as Johnson curve in range 2400 to 6000A.

The very strong lines of the xenon spectrum are in the longer wavelength range. The relative strength of the mercury orange doublet 5770, 5790A is perhaps less than that shown in the graph; this is at the limit of the sensitivity of the PM tube. There seems little doubt that in the range below 3000A the xenon spectrum is considerably stronger than the solar and Hg-Xe spectrum. Over the UV and visible range as a whole the xenon spectrum gives a closer match to that of the sun.

The spectral energy distribution of the mercury-xenon arc in the wavelength range 2500A to 25000A is shown in Figure B-47. Unlike in Figure B-46, the radiant flux has been integrated for equal intervals of 100A. The Johnson curve is also shown for purposes of comparison. The area under both curves is the same. This curve is based on a spectrum scan made by the Perkin-Elmer monochromator with the LiF prism on the light emerging from the SES. The spectrum is that of the Hg-Xe 3500 watt arc, as modified by the collimating module of the SES and further modified by the scanning boom of the monochromator. The collimated light of the SES falls on a sampling flat of one square foot effective area, and is reflected towards a relaying system which consists of two lenses, two reflecting surfaces and an exit window. The slit width was 1.5 mm. The scan was made rapidly in 20 minutes. During the scan the gain factor was changed automatically about 40 times whenever the deflection of the recorder pen tended to go above or below a preset range from 2.1 to 8.7 inches. The pen deflection was read at all flexion points of the curve and at sufficiently close wavelength intervals, so as to give an accurate value for the area under the curve. The observed deflections were reduced to a common scale of amplification, and multiplied by the reciprocal sensitivity for the corresponding wavelength. The total area under the curve was computed in order to determine the normalization factor required to make the area equal to 134.5 milliwatts per sq. cm.

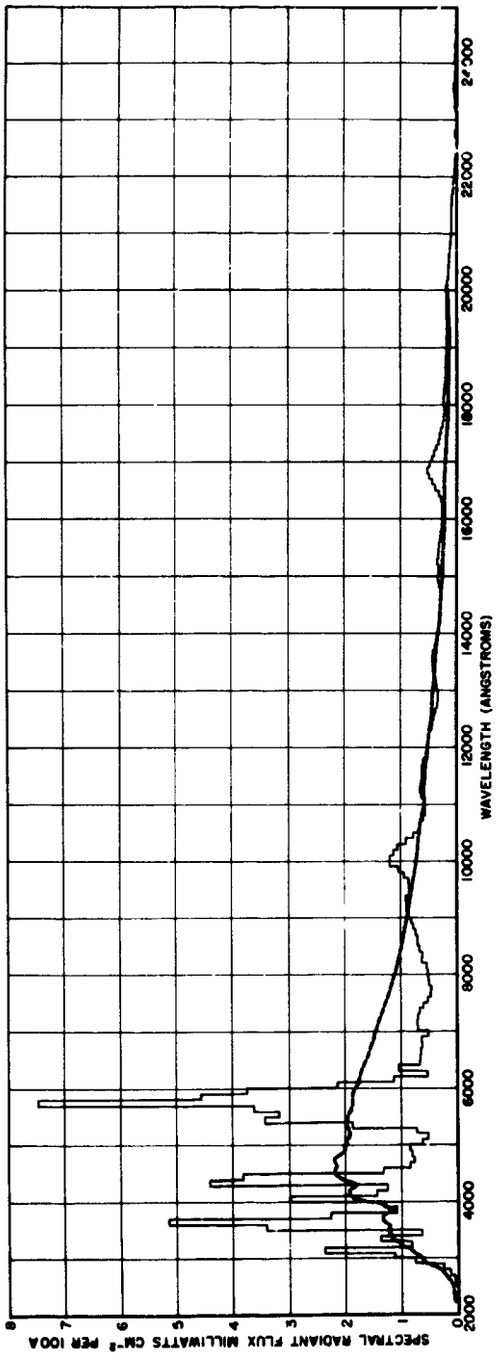


Figure B-47–Histogram of spectral radiant flux of Hg–Xe arc in module compared with solar spectrum.

which is the Johnson value for the solar energy in the range .25 to 2.5 μ. All values of spectral radiant flux were multiplied by this factor, and the energy in each 100A range was computed.

The final curve shown in Figure B-47 is essentially similar to those published in literature. The wavelength resolution is considerably less and hence many discrete lines in the neighborhood of 10000A and 17000A are shown as one broad peak. A detailed comparison between this curve obtained by a manual data reduction method and a similar one based on the same Perkin-Elmer chart and reduced by the computer (1) shows that the peaks occur at the same wavelengths. There are a few significant differences in the relative heights of the peaks, which are probably due to the source of error discussed on page B-69.

REFERENCES

1. N. Zylich, "An In-Chamber Spectral Measurement Technique," paper presented at NASA Headquarters Conference on Solar Simulation, July 14, 1965, due for publication in "Proceedings of the Conference."

2. D. E. Gray, Editor, American Institute of Physics Handbook, McGraw-Hill Book Co., Inc., New York, 1957, pages 6-117.

3. International Union of Pure and Applied Chemistry, Tables of Wavenumbers for the Calibration of Infrared Spectrometers, (Butterworths, Washington, 1961) pages 577, 597, 684.

4. Jenkins, F. A., and White, H. E., Fundamentals of Optics, McGraw-Hill Book Co., Inc., New York (1957), page 301.

5. Streiff, M. L., and Ferriso, C. C., "Spectral Slit Width of a Small Prism Monochrometer," DOD report AD-448724.

6. Elias Klein (Editor) Final Report of the Goddard Summer Workshop Program, X-320-62-193, pages 2-5 to 2-34; X-320-63-264, pages A-1 to A-52.

7. D. Gallagher and R. McIntosh, in Solar Simulation Research, X-633-65-332 Section 3.

BLANK PAGE

SPECTRAL TRANSMITTANCE OF JARRELL-ASH MODEL 82-000 MONOCHROMATOR

H. H. Hoehn and S. M. Shapiro

INTRODUCTION

If one has a spectral standard source and desires a spectral calibration of a light detector, the radiation from the source has to be split into its wavelength components and the response of the detector vs wavelength recorded. A monochromator is used to separate the wavelengths. This will introduce an element into the system which will decrease the amount of light striking the detector. For a valid calibration of the detector it is necessary to know the percentage of light loss for each wavelength when passing through the monochromator.

Little work has been done on measuring the transmittance of the monochromator, because so far as normal spectroscopic use is concerned, there is little need for knowing light loss in the spectrometer, provided it is low enough for light to be detected at the exit slit. One article which dealt with our problem (1) was discovered. This served as a guideline for our work.

EXPERIMENTAL PROCEDURE

The output current from a photomultiplier (P.M.) tube can be written as:

$$I = C(\lambda)\, P(\lambda)\, T(\lambda)\, P(\lambda). \qquad (1)$$

where $P(\lambda)$ represents the power output of the source; $T(\lambda)$, the transmittance of the monochromator; $D(\lambda)$, the spectral sensitivity of the detector, and $C(\lambda)$, the other constants of the system, such as solid angle effects and lens losses. If two monochromators are used in series for a given source and detector, the transmittance of either instrument can be measured. A schematic of the apparatus is shown in Figure B-48.

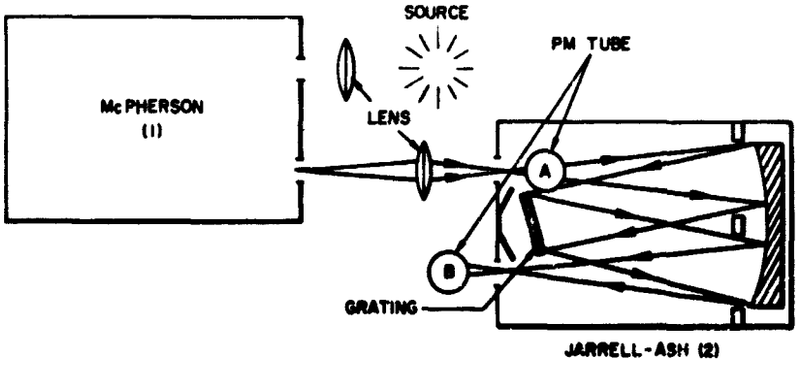


Figure B-48–Experimental arrangement for determining the transmittance of a monochromator.

The detector is first placed in position A and current I_1 which is proportional to the light entering spectrometer no. 2, as recorded. This current is shown as:

$$I_1 = C(\lambda) P(\lambda) T_1(\lambda) D(\lambda), \tag{2}$$

where $T_1(\lambda)$ is the light lost in passing through spectrometer no. 1. Next, the detector is set to position B, and current I_{12} is recorded as:

$$I_{12} = C(\lambda) P(\lambda) T_1(\lambda) T_2(\lambda) D(\lambda), \tag{3}$$

$T_2(\lambda)$ being the transmittance of spectrometer no. 2. This is found by taking the ratio Equations (3) to (2):

$$\frac{I_{12}}{I_1} = T_2(\lambda). \tag{4}$$

In this experiment the spectral transmittance of a Jarrell-Ash Model 82-000, 0.5m, Ebert mounting, electric drive scanning Monochromator was measured from 2000A to 6000A. The primary monochromator was the McPherson Model 225 Monochromator with 1.0m scanning. The gratings in both monochromators are 600 lines/mm, but the McPherson has a bipartite concave grating, and the Jarrell-Ash a plane grating. The reciprocal linear dispersion of the former is 16.7/mm, while that of the latter is 33A/mm.

Since the slits on the McPherson were set at 200μ, we had a bandwidth of 6.7A. Because of the narrow bandpass, no corrections were needed to be made to compensate for this.

The source used for the visible range, 3000-6000A, was a G.E.-microscope-illumination lamp consisting of a glass enclosed tungsten filament. From 2000-4000A, a deuterium source manufactured by Bausch and Lomb was available. The incandescent source was mounted in an air-cooled light box and focused on the slits of the McPherson by a 4-element quartz-CaF lens. The monochromatic beam was then focused on the entrance slit of the Jarrell-Ash by a similar Quartz-CaF lens. The slit width on the Jarrell-Ash was set at 1mm so the entire beam entered the spectrometer. A rectangular stop, 7/16 in. × 1/2 in., was placed over the lens between the two spectrometers so the long dimension was vertical. This was necessary to reduce the size of the incoming beam so the Jarrell-Ash grating was just filled. This insured that no light was lost on any of the baffles. This same precaution will have to be taken when the Jarrell-Ash is used as part of a calibration procedure in order for the following results to be valid:

The photomultiplier tube used was an EMR 541-D-05M (serial no. 6594), with a sapphire window, and operated at 1915 volts. To illuminate the same portion of the cathode, a cover with a 3/8 in. square opening (slightly larger than the light beam), was placed over the circular cathode. The cathode was positioned 2-15/32 in. from the slits when in positions A and B. At both these positions all of the light entered the opening in the P.M. tube aperture.

Figure B-49 shows the apparatus used in a typical test setup to measure I_{12}.

The ammeter was a Kiethly 610BR Electrometer which was fed to a Leeds and Northrup strip chart recorder. These currents were read off the chart after each run. Photo-currents were recorded every 50 Angstrom units. The power supply for the PM tube was a Kiethly 242 Regulated High Voltage Supply. The Instruments and the light source were connected to a Sorenson A.C. Voltage Regulator.

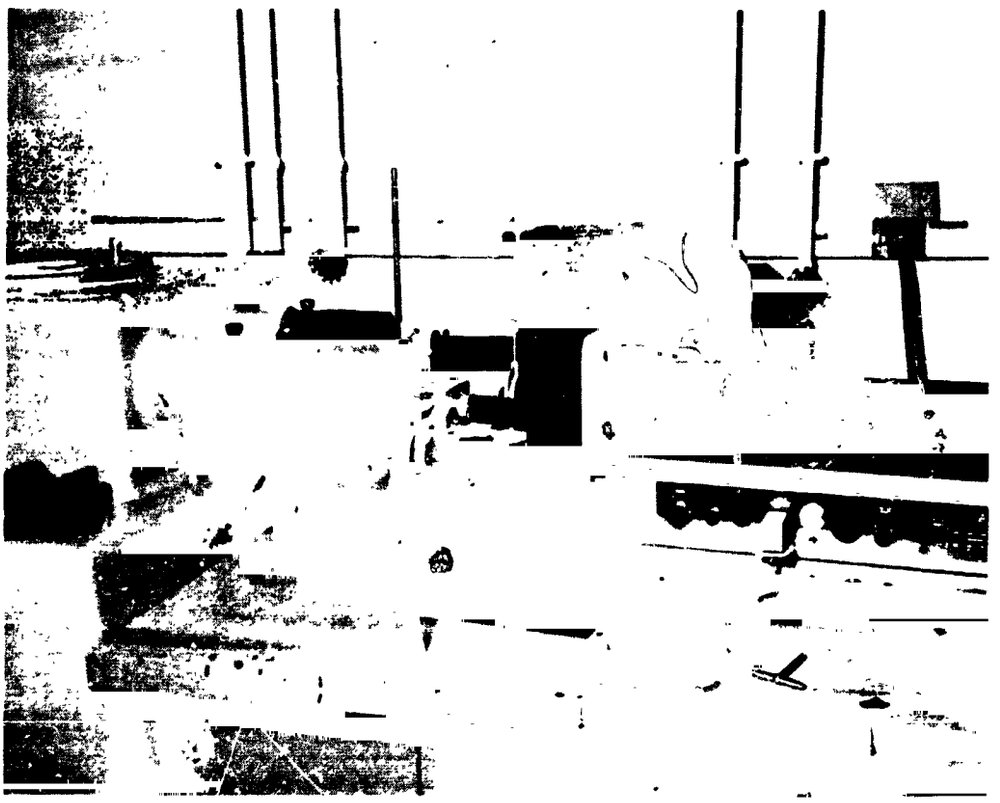


Figure B-49–Arrangement used in measuring transmittance of Jarrell-Ash (right) with PM tube in position B. Incandescent source is in black box in center.

A pair of runs, measuring the photo-currents I_1 and I_{12}, determined the transmittance of the monochromator. After a 1.5 hour warm-up time for both the source and the P.M. tube, the currents were recorded consecutively.

RESULTS

A criterion on which the validity of our results are based, is reproducibility. This was established for the P.M. tube by repeating runs with the tube remaining in a given position. The four current readings at a given λ setting were within 2% of each other. Next, we determined the reproducibility of the experimental arrangement. Between runs, with the visible source, we dismantled the apparatus and reassembled and re-aligned the optics. This was repeated for each of the five trials in the wavelength range 3000-6000A. The average was then taken and the average deviation calculated. The next check was in the wavelength region of 2000-4000A, and since reproducibility was previously established, all but the P.M. tube remained in a given position while the tube was moved from position A to position B. Three trials were run in this wavelength range and the average along with the average deviation was calculated.

The result of the trials, the average values, and the average deviations are given in Tables B-20 and B-21. Figure B-50 shows a plot of the average transmission values from 2000 to 5800A.

Table B-20
Transmittance (%) of Jarrell-Ash 0.5 Meter Monochromator
Slits: McPherson 200μ Jarrell-Ash 1 mm Source: Incandescent

λ	Trial I	Trial II	Trial III	Trial IV	Trial V	% T Ave.	Ave. Dev.	λ	Trial I	Trial II	Trial III	Trial IV	Trial V	% T Ave.	Ave. Dev.
6000	13.52	14.11	13.29	12.18	13.68	13.36	.50	4450	20.40	22.80	20.00	16.34	21.17	20.14	1.58
5950	13.50	13.77	12.84	11.41	13.23	12.95	.66	4400	20.50	22.83	20.21	16.39	20.89	20.16	1.51
5900	13.38	13.81	12.67	11.04	12.88	12.76	.72	4350	20.70	22.73	19.61	16.14	20.56	19.95	1.66
5850	12.69	13.22	12.46	10.99	12.78	12.43	.57	4300	20.70	23.13	20.03	16.77	21.00	20.33	1.54
5800	12.04	13.53	12.27	10.73	12.72	12.26	.70	4250	21.00	23.41	20.35	16.61	21.04	20.48	1.60
5750	12.50	13.54	12.38	10.63	12.68	12.35	.68	4200	21.10	23.23	20.49	17.18	21.13	20.63	1.43
5700	12.50	13.71	12.33	10.71	12.69	12.39	.69	4150	21.10	23.18	20.42	17.11	21.00	20.56	1.44
5650	12.55	14.08	12.33	10.67	12.69	12.64	.77	4100	20.60	23.01	20.20	17.04	20.53	20.28	1.32
5600	12.90	13.94	12.52	10.77	13.15	12.66	.81	4050	20.50	22.72	20.00	16.86	20.31	20.08	1.32
5550	13.49	15.33	13.27	11.37	13.66	13.42	.88	4000	20.21	22.34	19.81	16.66	20.23	19.85	1.29
5500	15.50	17.61	15.26	13.11	15.90	15.48	1.03	3950	20.12	22.40	19.72	16.67	20.13	19.81	1.29
5450	18.70	21.22	18.55	15.79	19.05	18.66	1.19	3900	20.98	23.30	20.50	17.53	21.00	20.66	1.32
5400	20.98	23.80	20.79	17.77	21.62	20.99	1.37	3850	22.33	24.78	21.99	18.74	22.34	22.04	1.34
5350	21.15	23.73	21.05	17.85	21.71	21.10	1.32	3800	23.29	25.99	22.76	19.54	23.31	22.98	1.46
5300	20.35	22.42	20.07	17.17	20.90	20.19	1.25	3750	23.88	26.57	23.35	19.92	23.77	23.50	1.49
5250	19.27	21.42	18.66	16.19	19.88	19.08	1.33	3700	24.05	26.74	23.50	20.00	24.25	23.71	1.57
5200	18.30	20.51	17.98	15.22	18.67	18.14	1.23	3650	24.13	27.02	23.54	20.00	23.87	23.71	1.55
5150	17.83	20.33	17.79	15.07	18.57	17.92	1.23	3600	23.65	26.67	23.00	19.45	23.72	23.30	1.66
5100	18.23	20.34	18.00	15.11	18.52	18.04	1.19	3550	23.38	26.52	23.11	19.53	23.50	23.20	1.51
5050	17.77	20.30	17.87	15.13	18.62	17.94	1.22	3500	23.67	26.32	22.96	19.13	23.49	23.11	1.67
5000	18.39	20.41	18.28	15.12	18.69	18.18	1.22	3450	24.36	26.90	23.57	19.38	23.89	23.62	1.72
4950	18.31	20.71	18.20	15.13	18.83	18.24	1.26	3400	25.00	27.70	24.37	20.27	24.84	24.44	1.69
4900	18.55	20.92	18.18	14.98	19.00	18.33	1.40	3350	25.95	28.89	25.07	20.56	25.68	25.23	1.93
4850	18.75	20.85	18.71	15.03	18.84	18.44	1.36	3300	26.39	29.23	25.37	20.88	26.11	25.60	1.98
4800	18.73	21.35	18.78	15.25	19.46	18.71	1.39	3250	28.10	29.76	25.86	20.91	26.14	26.15	2.22
4750	19.01	21.39	18.71	15.50	19.61	18.84	1.39	3200	26.63	29.55	25.55	20.85	26.03	25.68	2.07
4700	19.32	21.59	19.06	15.59	19.65	19.04	1.38	3150	26.56	29.33	25.10	20.55	25.89	25.49	2.13
4650	19.23	21.98	19.19	15.75	19.65	19.16	1.36	3100	25.71	28.13	24.25	20.07	24.94	24.62	1.97
4600	19.70	22.22	19.73	15.86	20.26	19.55	1.48	3050	24.72	27.98	23.60	19.42	24.13	23.97	1.97
4550	20.07	22.64	19.72	16.38	20.46	19.85	1.32	3000	23.94	27.00	23.00	18.68	24.10	23.34	2.00
4500	20.40	22.70	20.02	16.40	21.06	20.12	1.52								

Table B-21
Transmittance (%) of Jarrell-Ash 0.5 Meter Monochromator
McPherson Slits 200μ, Jarrell-Ash 1mm
Source – Deuterium Continuum (Sylvania)

λ	Trial I	Trial II	Trial III	% T Ave.	Ave. Dev.	λ	Trial I	Trial II	Trial III	% T Ave.	Ave. Dev.
4000	19.28	19.05	19.60	19.31	.193	2950	24.42	24.54	25.27	24.74	.350
3950	19.39	18.82	19.95	19.39	.377	2900	24.21	24.47	25.13	24.60	.350
3900	20.50	19.94	20.73	20.39	.300	2850	24.16	24.13	24.02	24.10	.057
3850	21.95	21.94	22.60	22116	.290	2800	23.93	23.80	24.45	24.06	.260
3800	23.00	22.67	23.44	23.04	.270	2750	23.49	23.49	24.19	23.72	.310
3750	23.61	23.50	24.20	23.77	.287	2700	22.90	23.08	23.75	23.24	.337
3700	23.80	23.71	24.40	23.97	.287	2650	22.34	22.69	23.22	22.75	.313
3650	24.00	24.01	24.61	24.20	.267	2600	21.87	22.01	22.71	22.20	.343
3600	23.30	23.34	23.86	23.50	.240	2550	20.89	21.37	22.00	21.42	.387
3550	23.15	22.97	23.65	23.26	.263	2500	20.42	20.52	21.10	20.68	.280
3500	23.35	22.65	23.62	23.21	.370	2450	19.40	19.57	20.07	19.68	.260
3450	23.60	23.30	24.09	23.66	.283	2400	18.27	18.58	18.93	18.59	.223
3400	24.45	24.03	25.00	24.49	.337	2350	17.01	17.11	17.57	17.21	.220
3350	25.16	25.11	25.86	25.41	.333	2300	15.48	15.84	16.17	15.83	.233
3300	25.55	25.34	26.13	25.67	.303	2250	13.94	14.20	14.71	14.28	.283
3250	25.85	25.42	26.25	25.84	.280	2200	12.53	12.57	13.13	12.74	.257
3200	25.95	25.88	26.48	26.07	.240	2150	10.92	11.21	11.70	11.28	.283
3150	25.80	25.72	26.63	26.05	.387	2100	9.35	9.67	10.00	9.73	.217
3100	25.34	25.47	25.73	25.51	.143	2050	7.62	7.86	8.16	7.88	.237
3050	24.70	24.89	25.69	25.09	.397	2000	6.10	6.48	6.77	6.45	.233
3000	24.60	24.80	25.19	24.86	.217	1950	4.00	4.19	4.66	4.28	.250

DISCUSSION

The fact that the results are independent of the source can be seen from the data for the 3000-4000A range, where the transmittance values determined by two different light sources are in good agreement. The results of the transmittance near 3000A, using the incandescent source, are low because in this region the dark current is a significant percentage of the total signal. This is true because the wavelength cutoff of the glass enclosure is near 3000A. Failure to make accurate corrections for the dark current results in erroneous values for the transmittance. The low values of the transmittance at the 2000A end of the spectrum are due to absorption of the silicon monoxide coatings of the mirrors and gratings, the absorption of air, and absorption of the lamp envelope. It is evident from Table B-20 Trial IV, that the results are independent of the detector. This trial was taken with a different phototube EMR 541-D-05M, (Serial no. 6592).

Since all the light entered the slit of the Jarrell-Ash and none hit the baffles, the light losses are due to reflections and scattering at the mirror and gratings (2). Light is also lost because of distribution into different orders.

The maximum transmittance is in the neighborhood of 3200A. This is the portion of the spectrum for which the grating is blazed. The narrow peaks at 4800A and 5350A are due to anomalies in the rulings of the grating.

The major source of error is the repositioning of the system elements after each trial or run. This is evident because the error in the 3 runs with the invisible source,

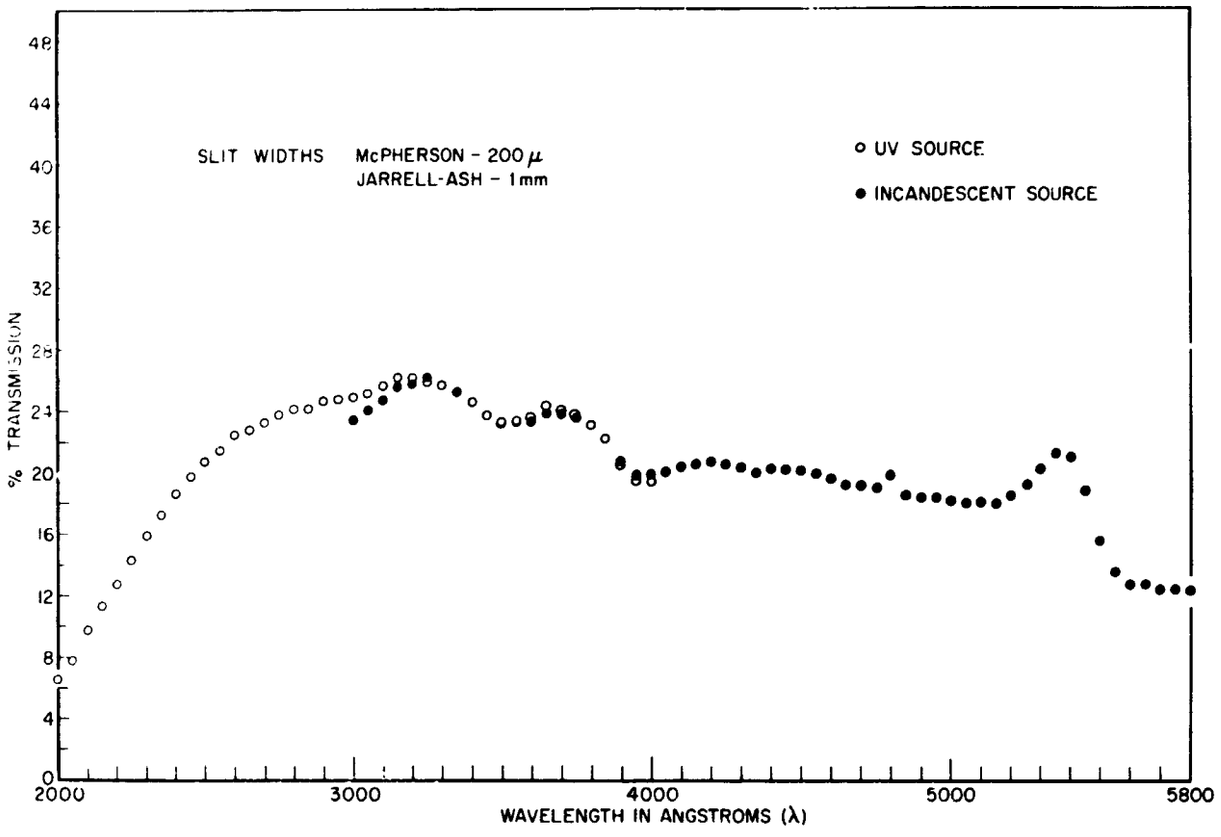


Figure B-50–Transmittance vs. wavelength of Jarrell-Ash Model 82-000 monochromator, slit widths McPherson 200 μ, Jarrell-Ash 1mm.

in which only the PM tube was previously moved, is less than the error in the runs with the visible source, in which the experimental equipment had been previously completely disassembled. After a sufficient warm-up time, fluctuations in the lamp output were insignificant to cause a large error in transmittance determination.

CONCLUSIONS

Figure B-50 illustrates the spectral transmittance of the Jarrell-Ash Model 82-000 monochromater from 2000A to 5800A. However, in order to utilize these results for calibration of light detectors, care will have to be taken to insure that the grating is just filled with light (no light hitting any of the spectrometer baffles). For a precise calibration, using Equation (1), $P(\lambda)$ will be known since it is a standard source. Knowing $T(\lambda)$, the geometry of the system, and transmittance of the lens, $D(\lambda)$ can be calculated. Conversely, the procedure can be reversed to calibrate a source by measuring the output of a calibrated detector, measuring the geometry of the system, transmittance of the lens, and knowing $T(\lambda)$.

REFERENCES

1. Ames, I. and Christensen, R.L., Journal of the Optical Society of America, Volume 51 pages 224-236 (1961)

2. Sawyer, R. A., Experimental Spectroscopy, Dover Publications Inc., New York, (1963) pages 114-123

BLANK PAGE

INFRARED REFLECTANCE SPECTRA OF IGNEOUS ROCKS, TUFFS AND RED SANDSTONE FROM 0.5 TO 22 MICRONS

W. A. Hovis, Jr. and W. R. Callahan

Total reflectance measurements on a number of igneous rocks and other minerals indicate that reststrahlen features may be of use as a remote probe into the nature of the surface of the moon and planets. Varying particle size introduces effects such as reduced spectral contrast and in some cases a new feature that could lead to confusion in identification unless particle size is known. Short wavelength reflectance measurements are seen to offer a method of determining particle size by remote measurements.

INTRODUCTION

Of basic interest to a program of planetary exploration is the surface composition and condition of the planetary body. Infrared analysis offers a promising tool to carry out such an investigation. Although thermal emission analysis seems to be the only feasible method of study from an orbiting vehicle, the present study has involved reflection measurements on igneous rocks and other minerals thought likely to provide basic information about planetary surfaces. The experimental arrangement used herein measured reflectance over 2π steradians with little heating of the specimen. This indirect, but accurate method of determining emissivity, permitted higher spectral resolution than is possible in thermal emission unless the specimen is very hot.

EXPERIMENTAL PROCEDURE

An attempt has been made during sample preparation and measurement, to approximate the natural state of the materials under investigation. Solid surfaces were not polished but were used in a freshly fractured state. The samples with smaller particle sizes were prepared by crushing and then sifting through a series of graded sieves. In this way samples were prepared with particle sizes ranging from 2-4 mm to less than 0.038 mm. Most of the plotted curves contain results for four particle sizes; a solid sample, 1-2 mm, 0.105-0.25 mm, and less than 0.038 mm. For many samples, measurements were also made at intermediate sizes, but these four sizes were found to be representative and results pertaining to other sizes have been omitted to avoid overcrowding of the composite figures.

In the 0.5-2.5 micron wavelength region, the total reflectance of the specimens was measured with a Beckman DK2A spectrophotometer equipped with a total reflectance attachment. Measurements from 2.5-22 microns were carried out on a Cary Model 90 double-beam spectrophotometer equipped with the prototype total reflectance attachment described by White (1).

RESULTS

It might be recalled that igneous rocks are often described by terminology based on the SiO_2 content. SiO_2 is considered the 'acidic' oxide in these rocks, which may be divided into the following categories:

ultrabasic:	less than 45% SiO_2,	e.g., Dunite
basic:	45-53% SiO_2,	e.g., Basalt
intermediate:	53-65% SiO_2,	e.g., Syenite
acidic:	above 65% SiO_2,	e.g., Granite

Acidic rocks are rich in light minerals such as quartz and potash feldspar, while basic rocks are richer in ferromagnetic materials.

Figures B-51 to B-54 contain reflectance measurements for four different igneous rocks which range from the acidic rock, granite, to the ultrabasic, dunite. These spectra indicate the changes in reflectance as the material is reduced from a solid, freshly fractured sample to a sample in which the particles are below 40 microns in diameter.

It has been suggested that tektites originate from the lunar surface. Figure B-55 displays the reflectance spectrum of a S.E. Asia tektite while Figures B-56, B-57 and B-58 provide the spectra of two tuffs and a red sandstone. The tuffs are quite similar at shorter wavelengths but Tuff 6 shows reststrahlen features at 9 microns and 20 microns that are absent in the other tuff.

Figures B-59 to B-62 are composite curves constructed from the respective measurements on the first four materials listed in the figures. These composites are collated for various particle sizes. These curves provide an idea of the complexity introduced when several materials lie within the field of view of the spectrometer.

In order to verify the actual spectrum for composite samples, Figure B-63 displays the reflectance spectra obtained when equal parts by volume of the four igneous rocks, granite, basalt, serpentine and dunite were mixed. These curves follow closely the spectra predicted from composites based on the measurements of the individual materials. A solid sample composed of these four rocks has not yet been constructed.

Figure B-64 is a composite figure that shows the 7-15 micron region for the various samples. This display is primarily intended to show an effect noticed in this work. For many materials, especially quartz, granite, dunite, the tuffs and the sandstone, when the particle diameter is reduced below 100 microns, a new reflectance feature begins to appear at longer wavelengths. For a material such as quartz, when the particle diameter is reduced below 40 microns, this peak is stronger than the original reststrahlen feature. For these four spectra a curve is included that describes the spectrum when the particle size was reduced considerably below 40 microns by means of extended grinding. For quartz, the new feature continued to be enhanced while the other reststrahlen features continued to diminish. For sandstone and dunite, the new features remained approximately constant.

In order to test if this effect was due to selective filtering, Figures B-65 and B-66 contain the reflectance spectra of high purity fused natural quartz and that of synthetic quartz. The persistence of the long wavelength feature in the finely powdered sample for both of these materials indicates that the effect is not one of selective filtering. It may be noted that the synthetic quartz, due to its manufacturing process, shows much stronger water of hydration features at 6.35, 2.7, 2.2 and 1.4 microns than does the fused G.E. quartz, while the reflectance for the fused quartz is much higher in the shorter wavelengths at small particle sizes.

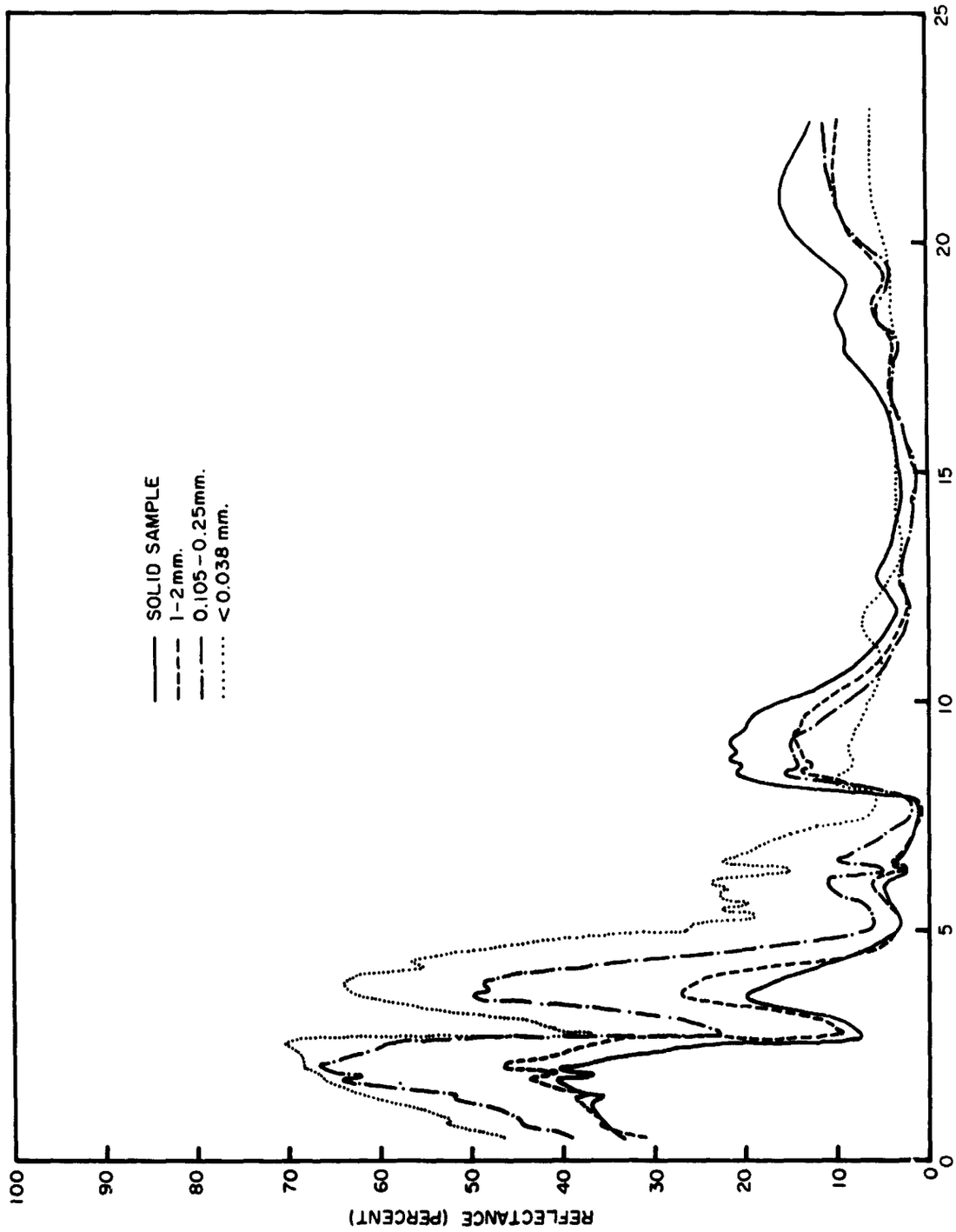


Figure B-51–North American Aviation Standard Granite.

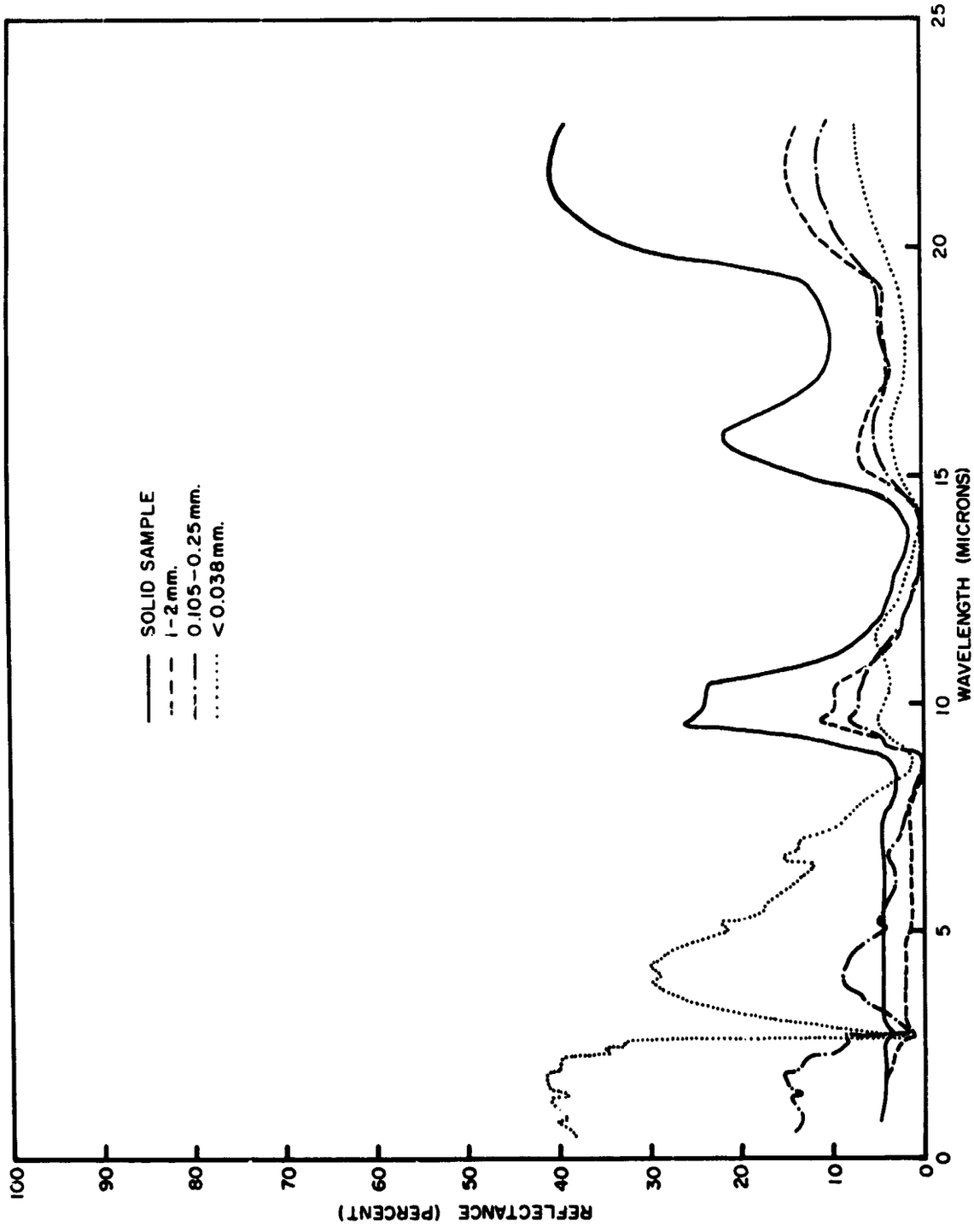


Figure B-52—North American Aviation Standard Oregon Serpentine.

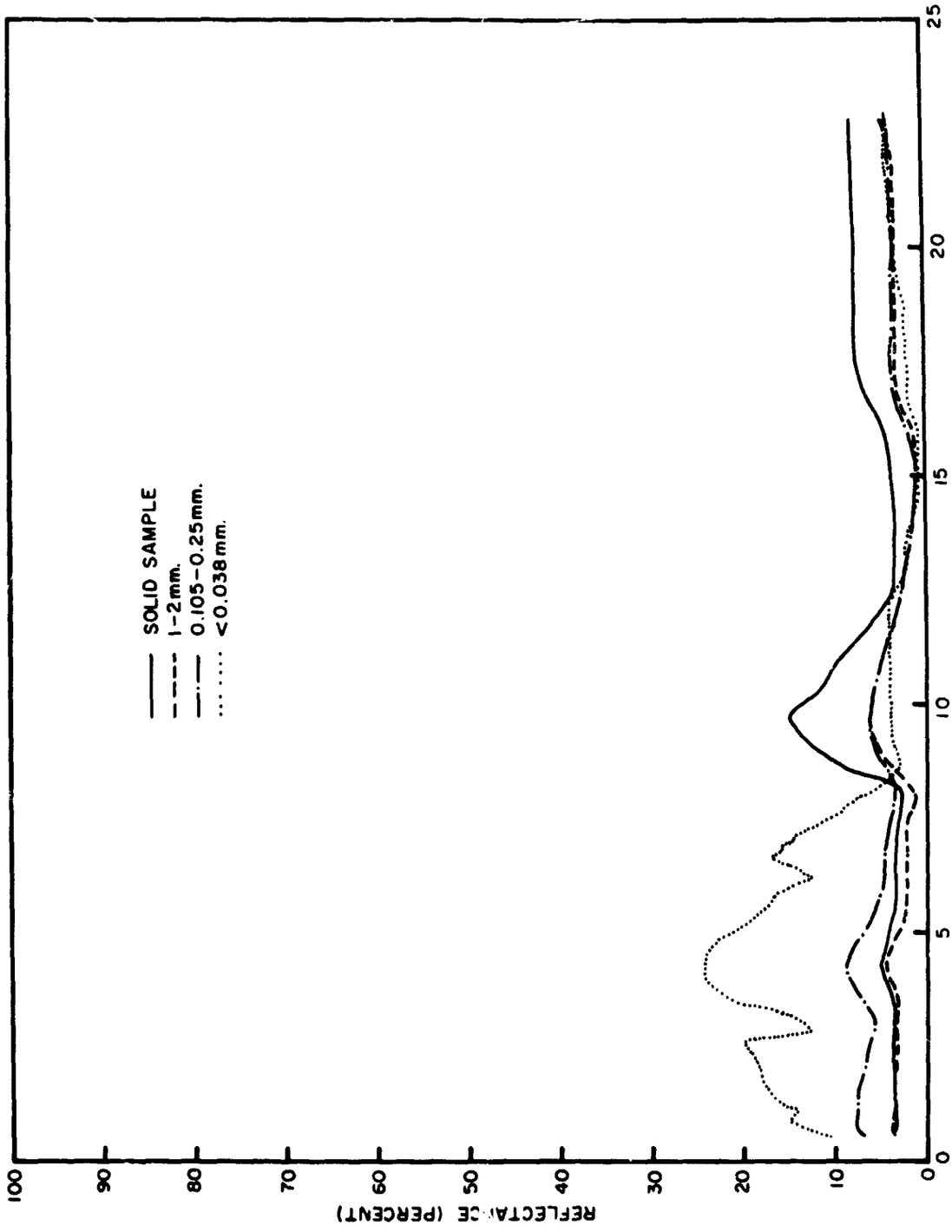


Figure B-53–North American Aviation Standard Oregon Basalt.

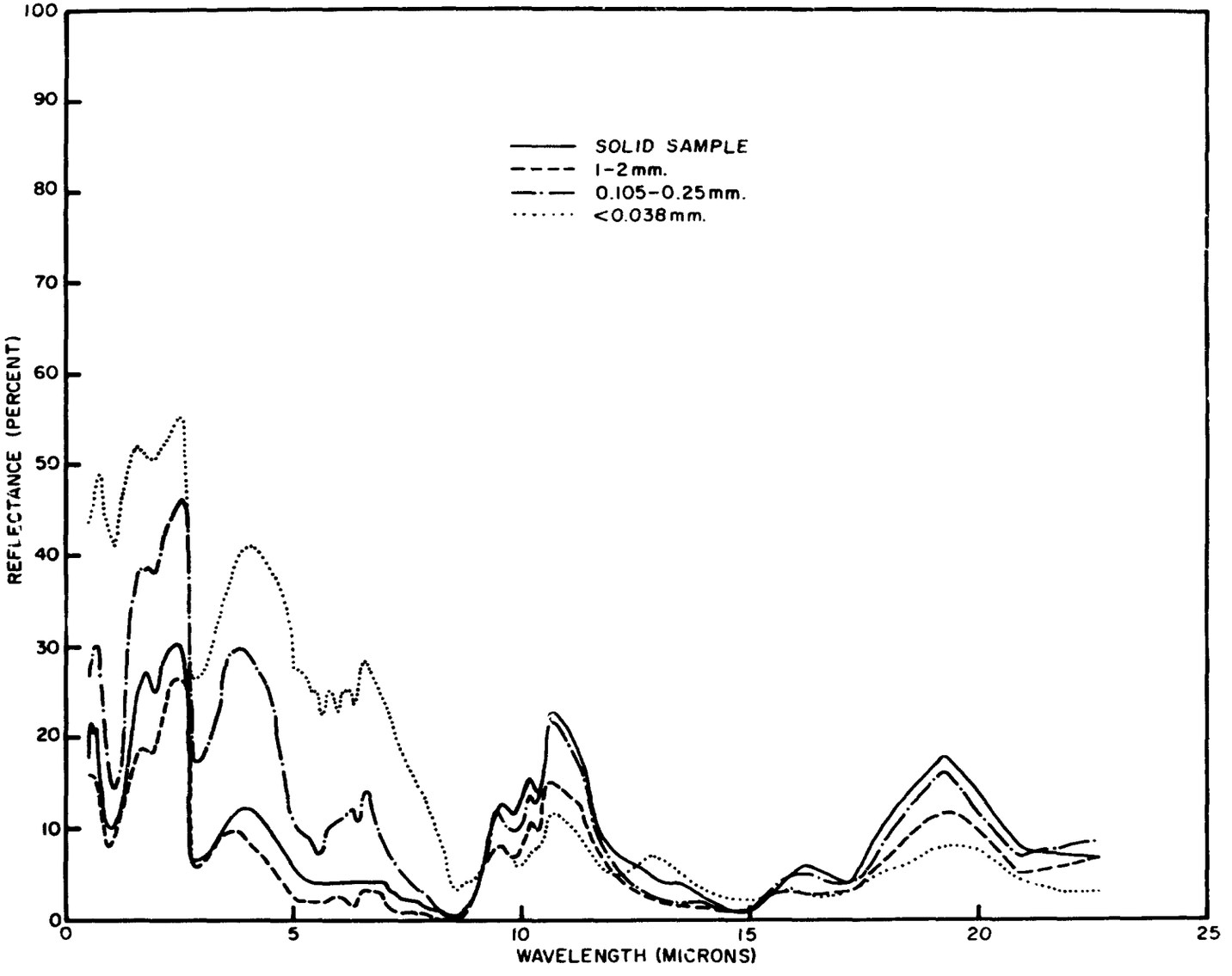


Figure B-54–Dunite, Merenfelder Moar (Germany).

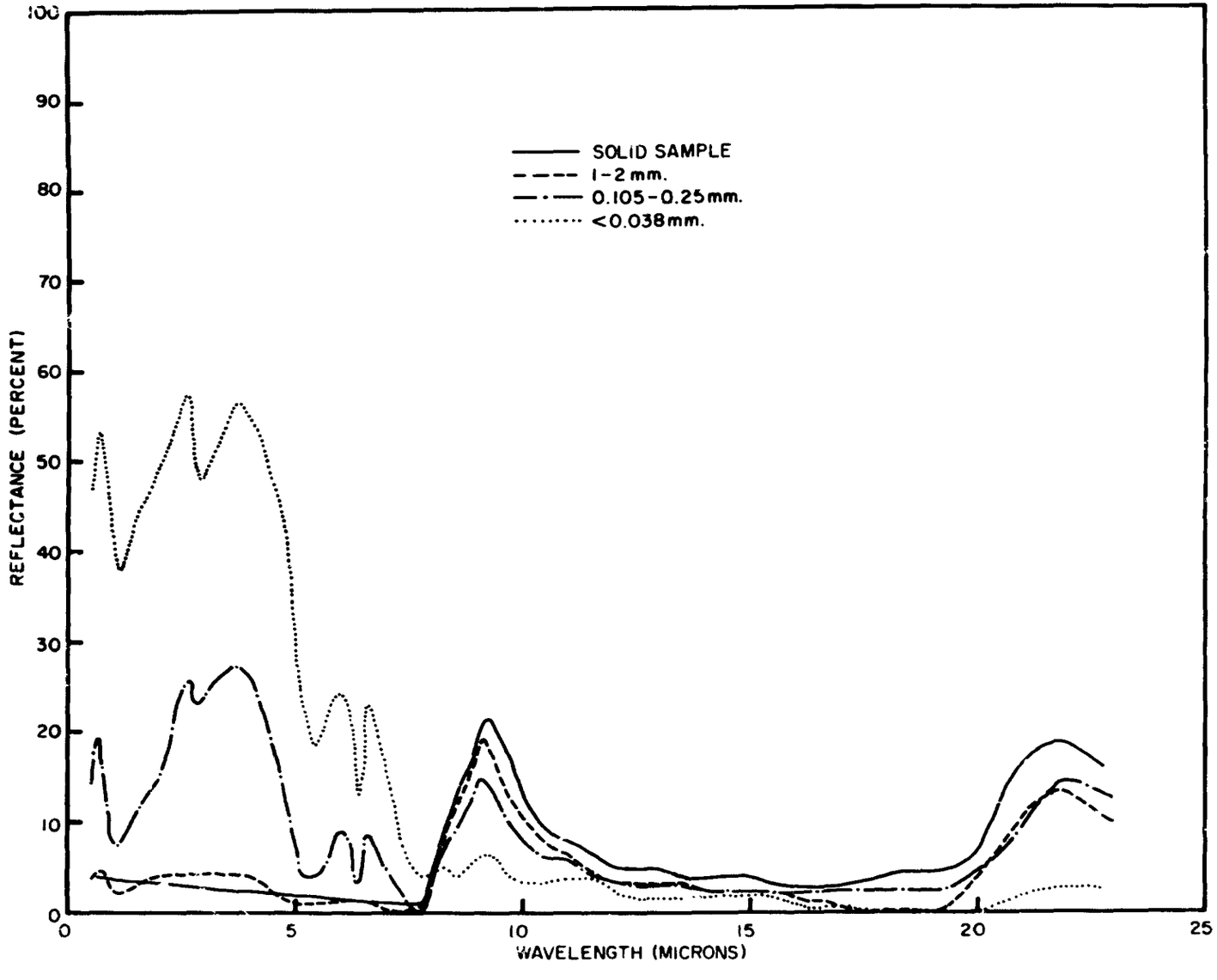


Figure B-55–Tektite (S. E. Asia).

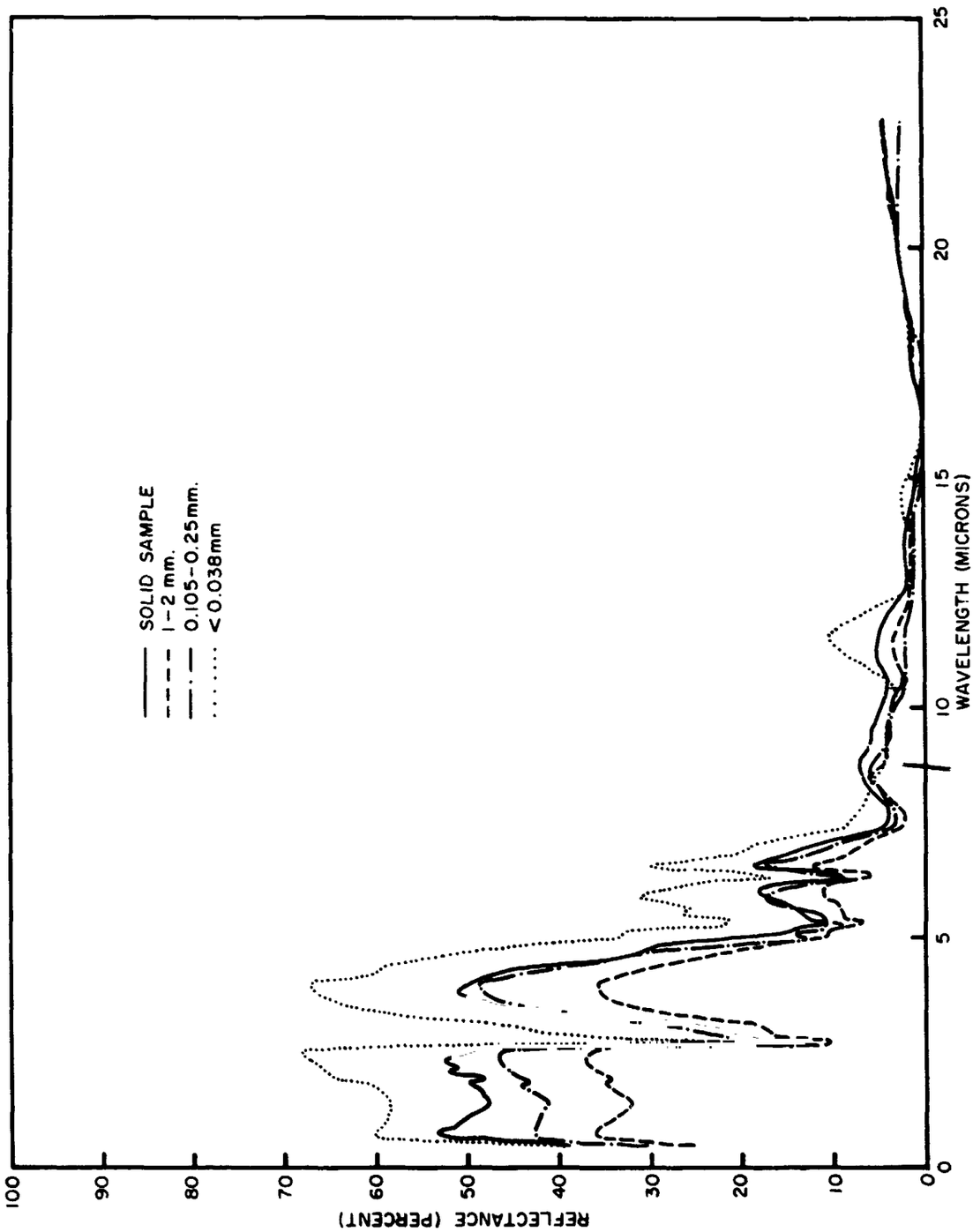


Figure B-56-Yellowstone Tuff #62-13.

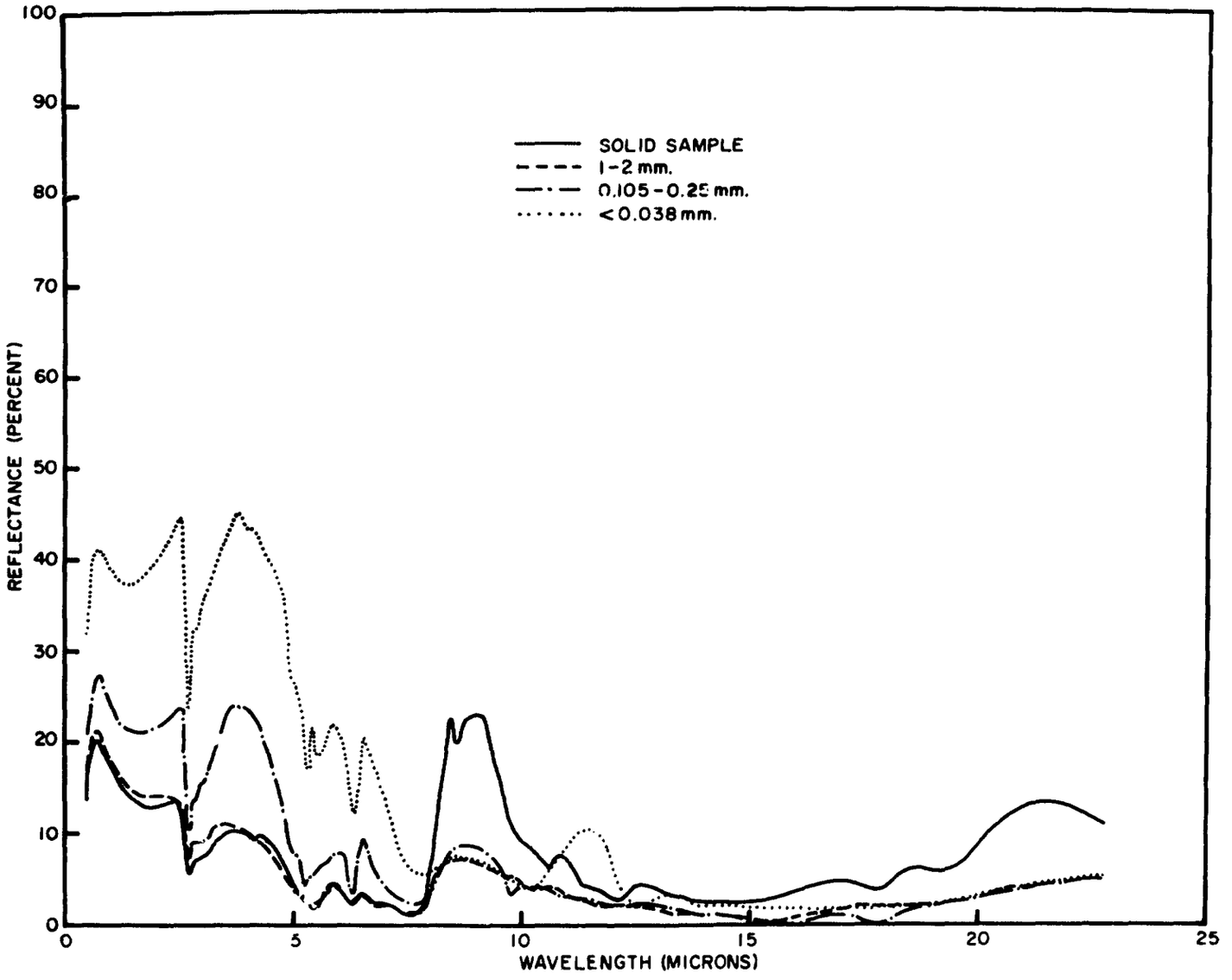


Figure B-57–Yellowstone Tuff #6.

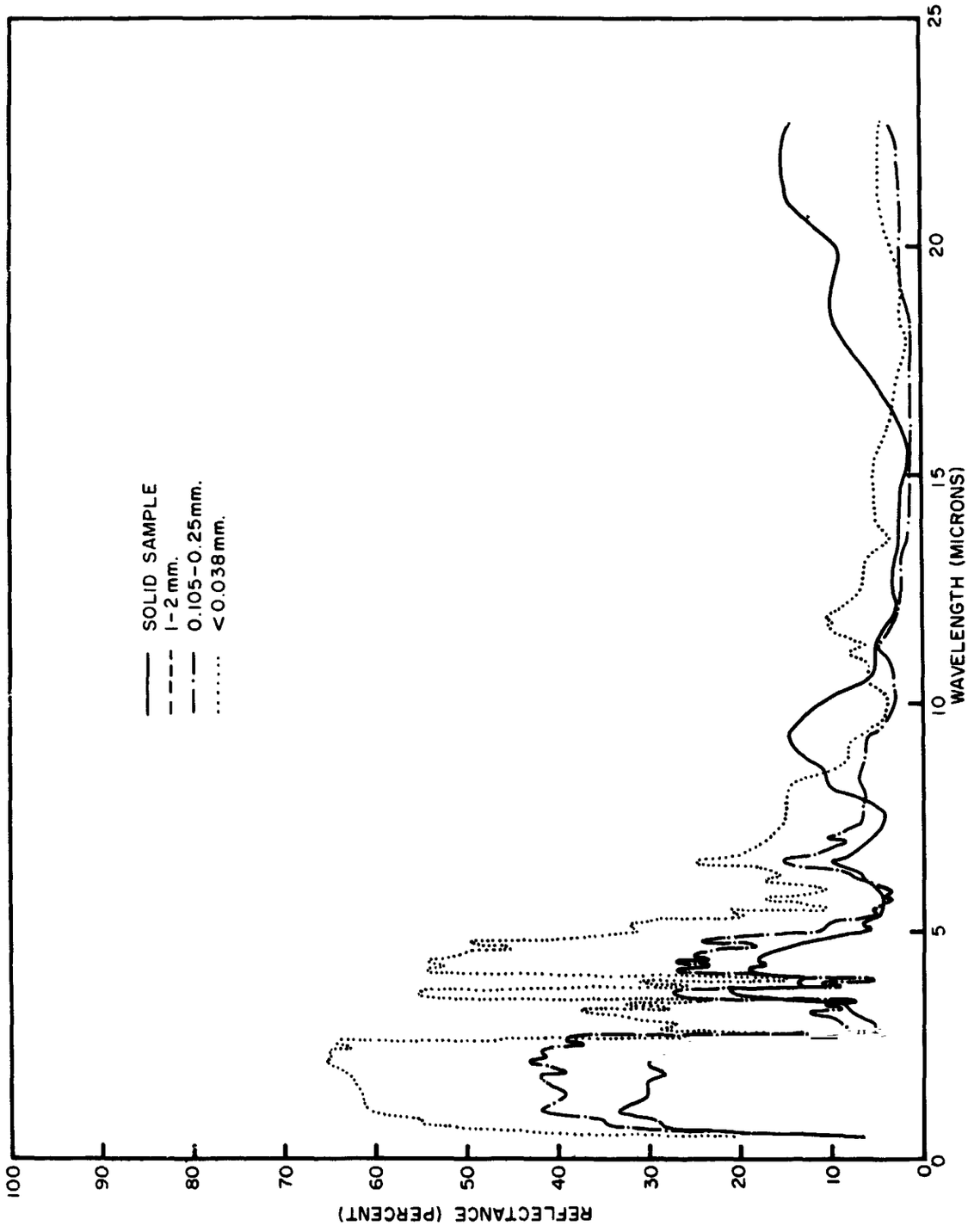


Figure B-58–Red Sandstone.

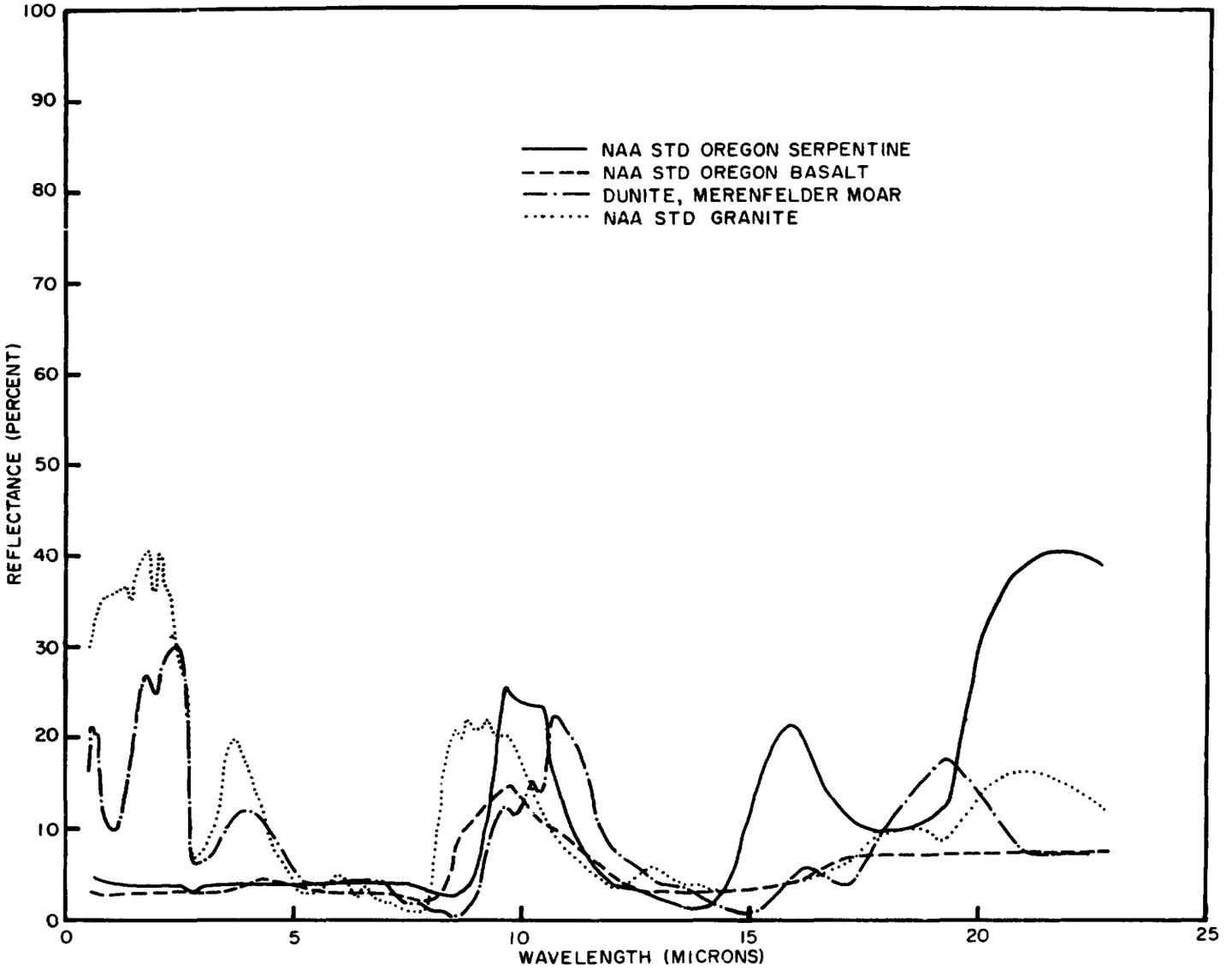


Figure B-59–Igneous rocks, solid samples.

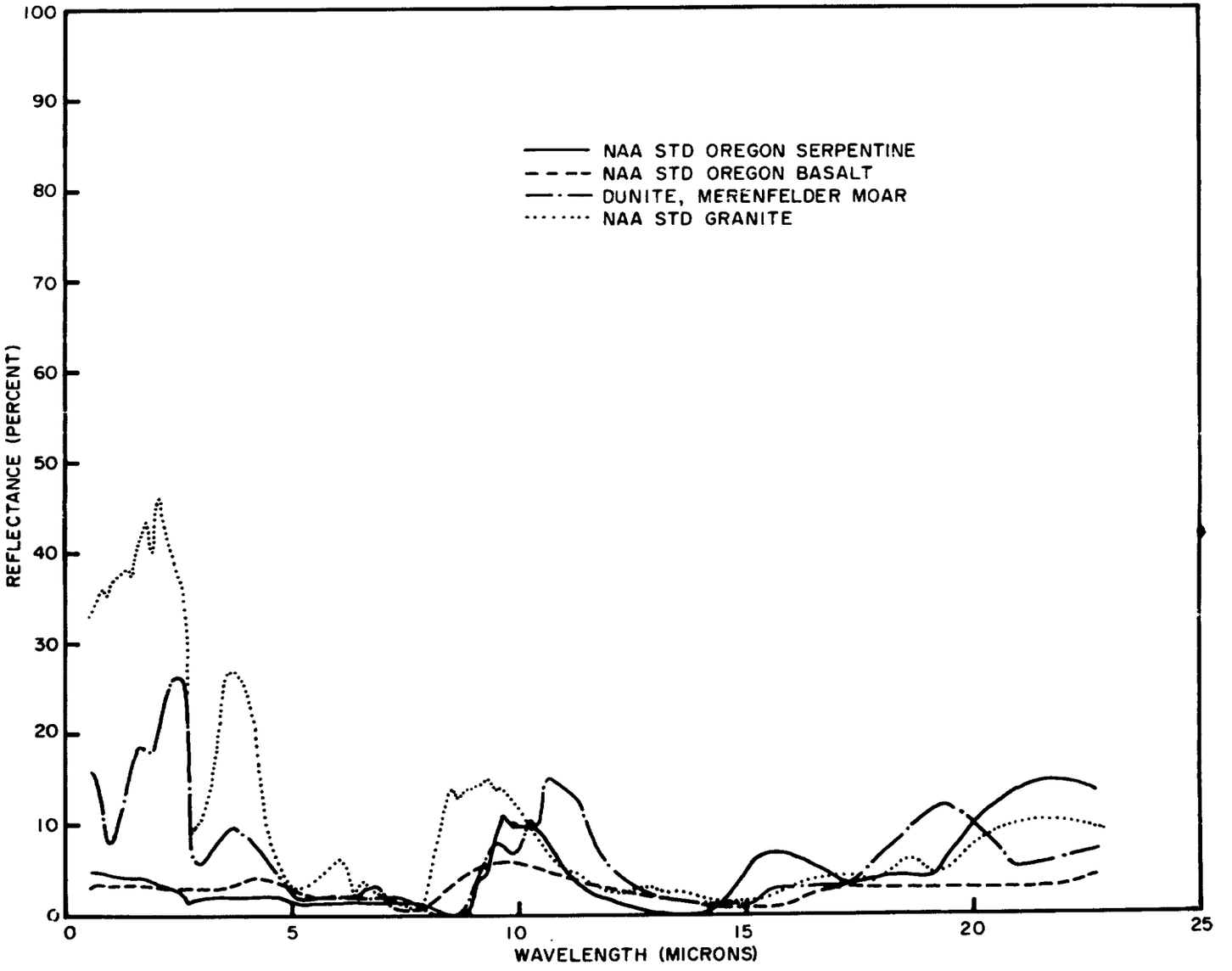


Figure B-60–Igneous rocks, 1-2 mm. samples.

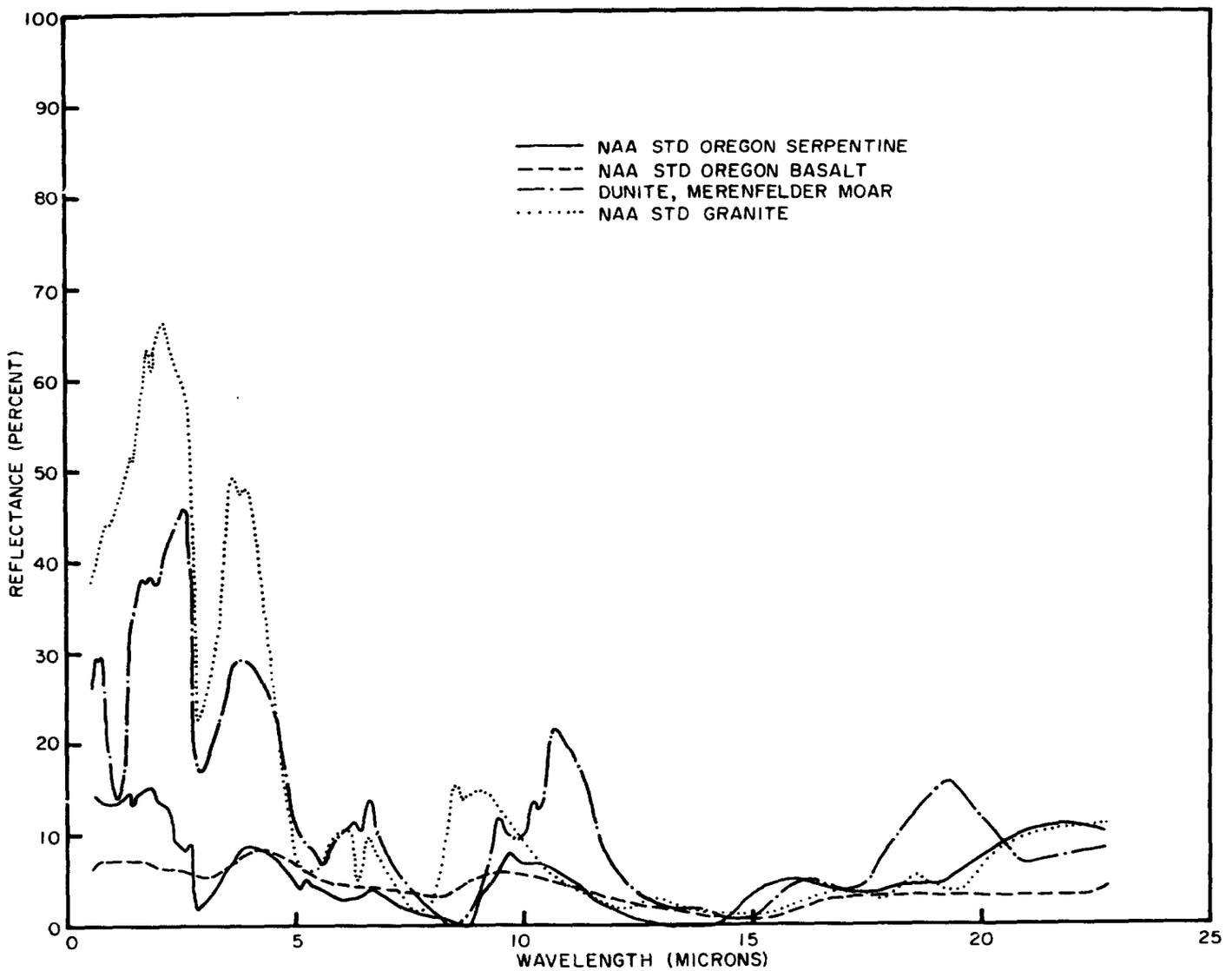


Figure B-61—Igneous rocks, 0.105-0.250 mm. samples.

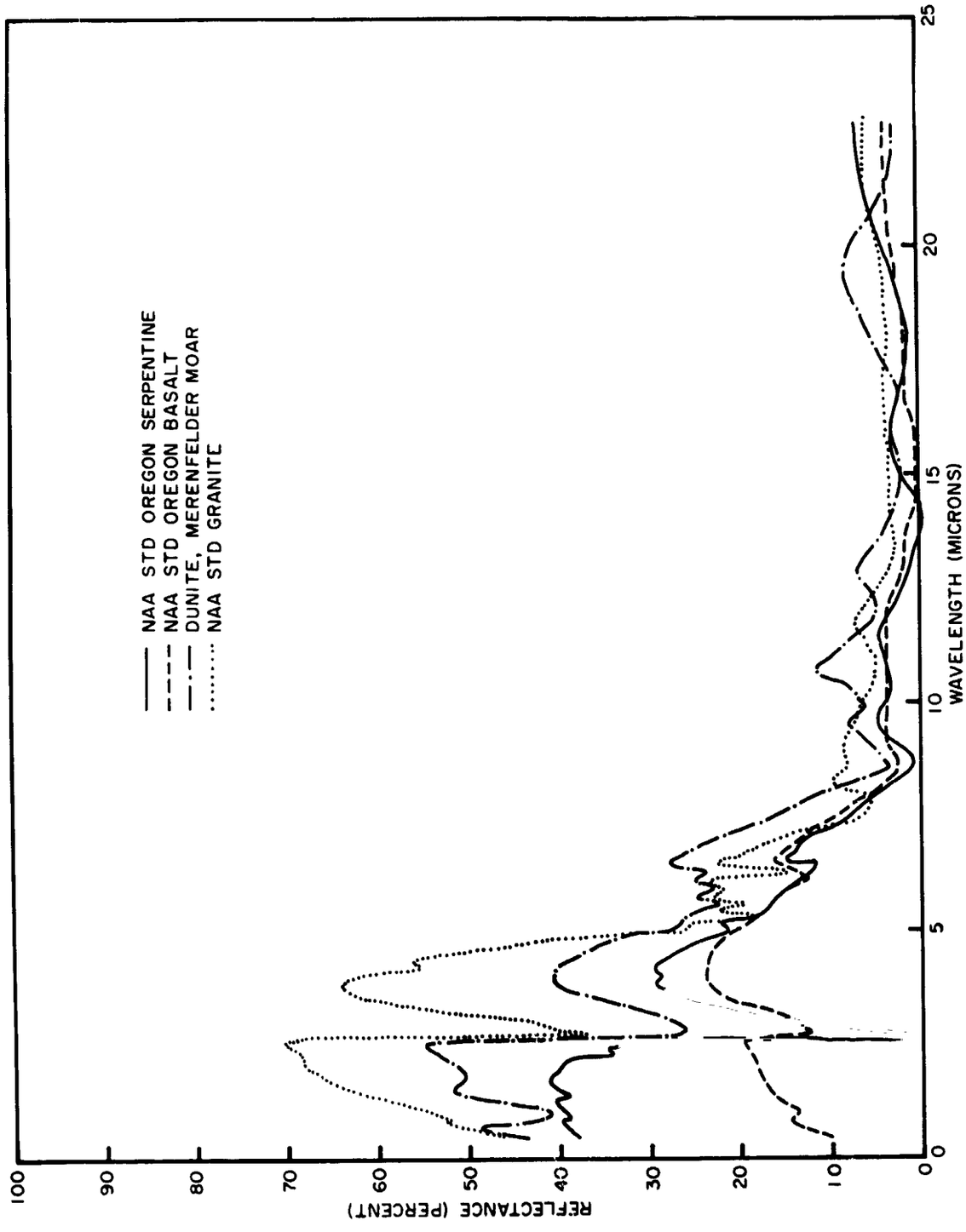


Figure B-62–Igneous rocks, less than 0.038 mm. samples.

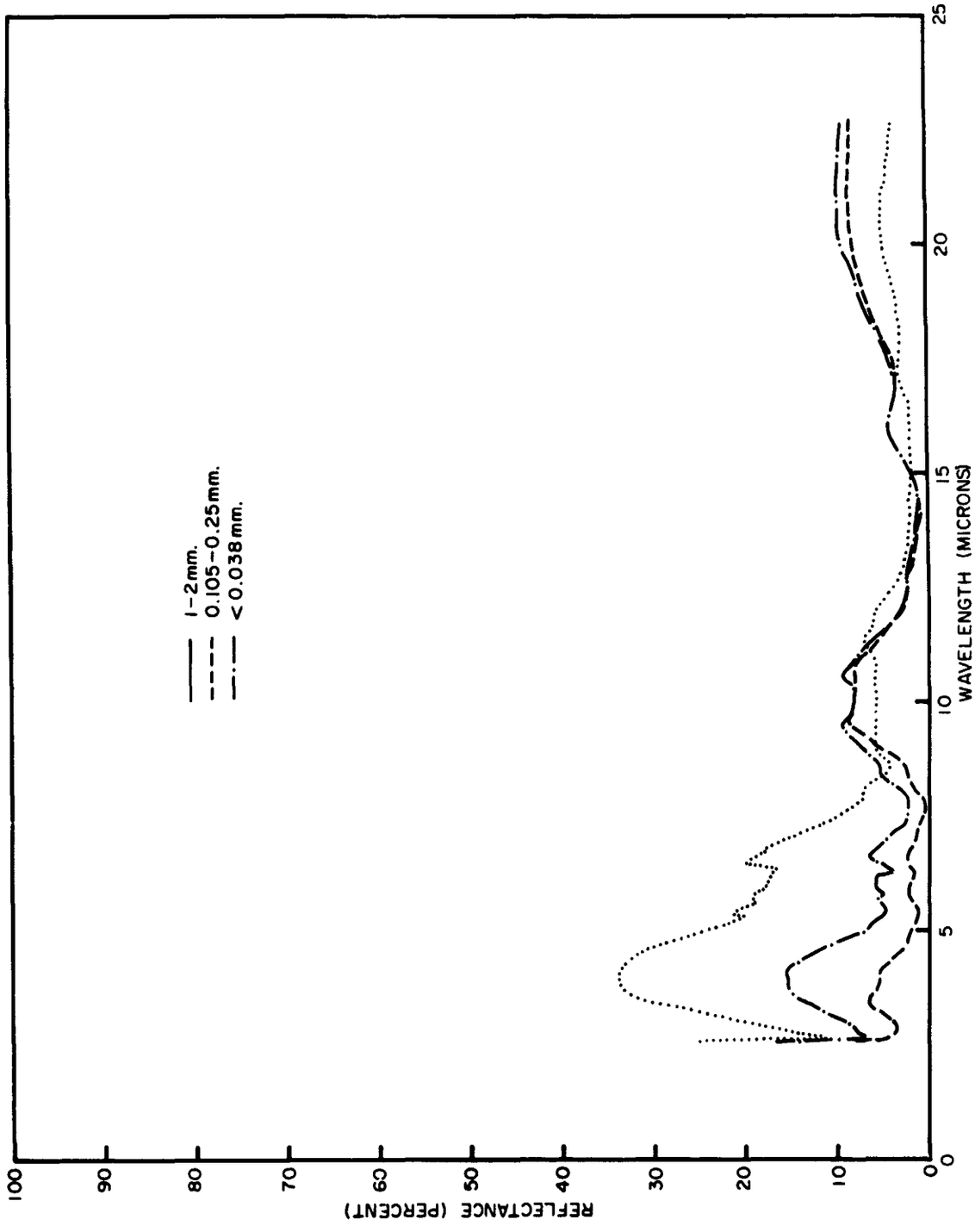


Figure B-63-Igneous rocks, mixed samples.

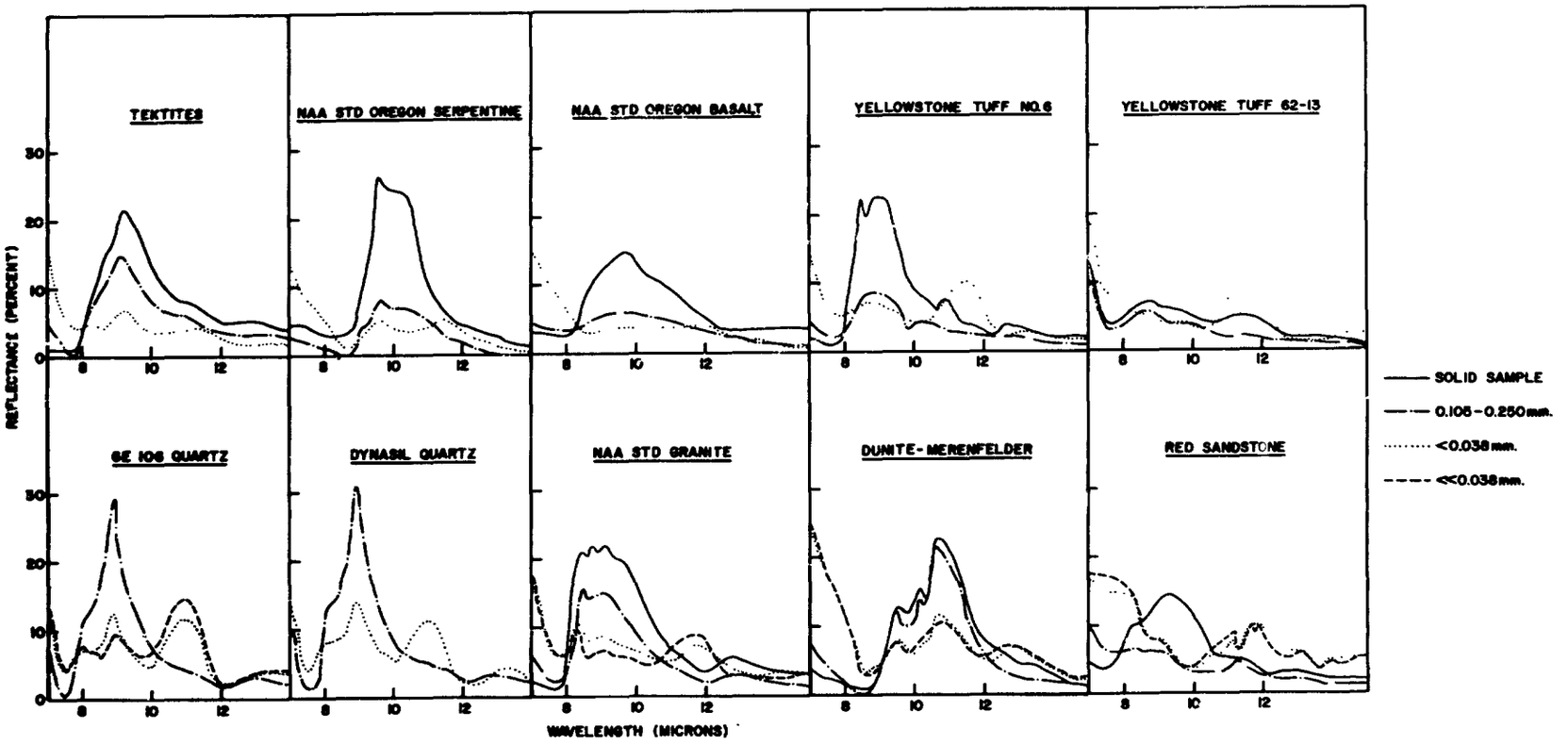


Figure B-64-7-15 micron region for various materials.

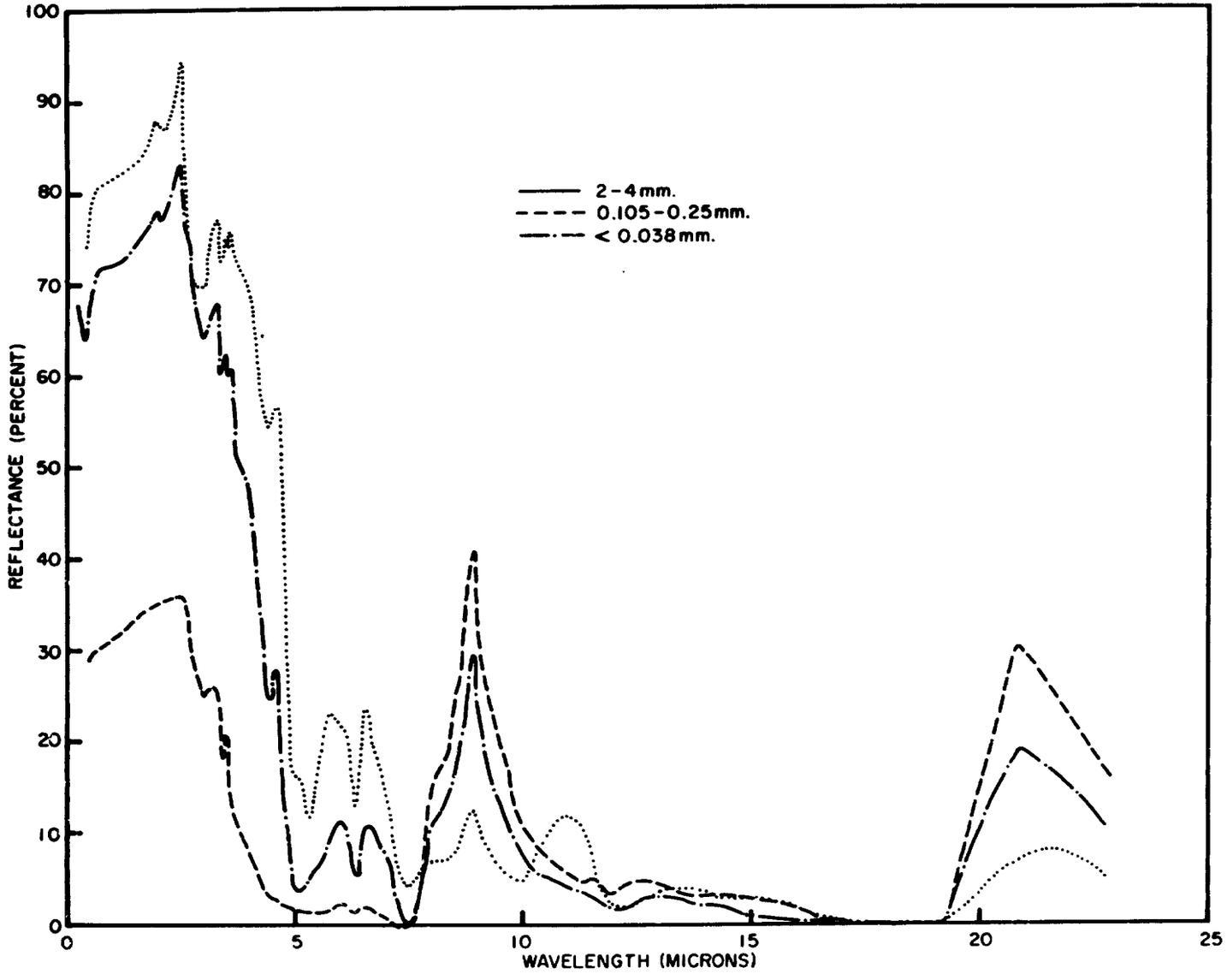


Figure B-65–General Electric 106 Fused Quartz.

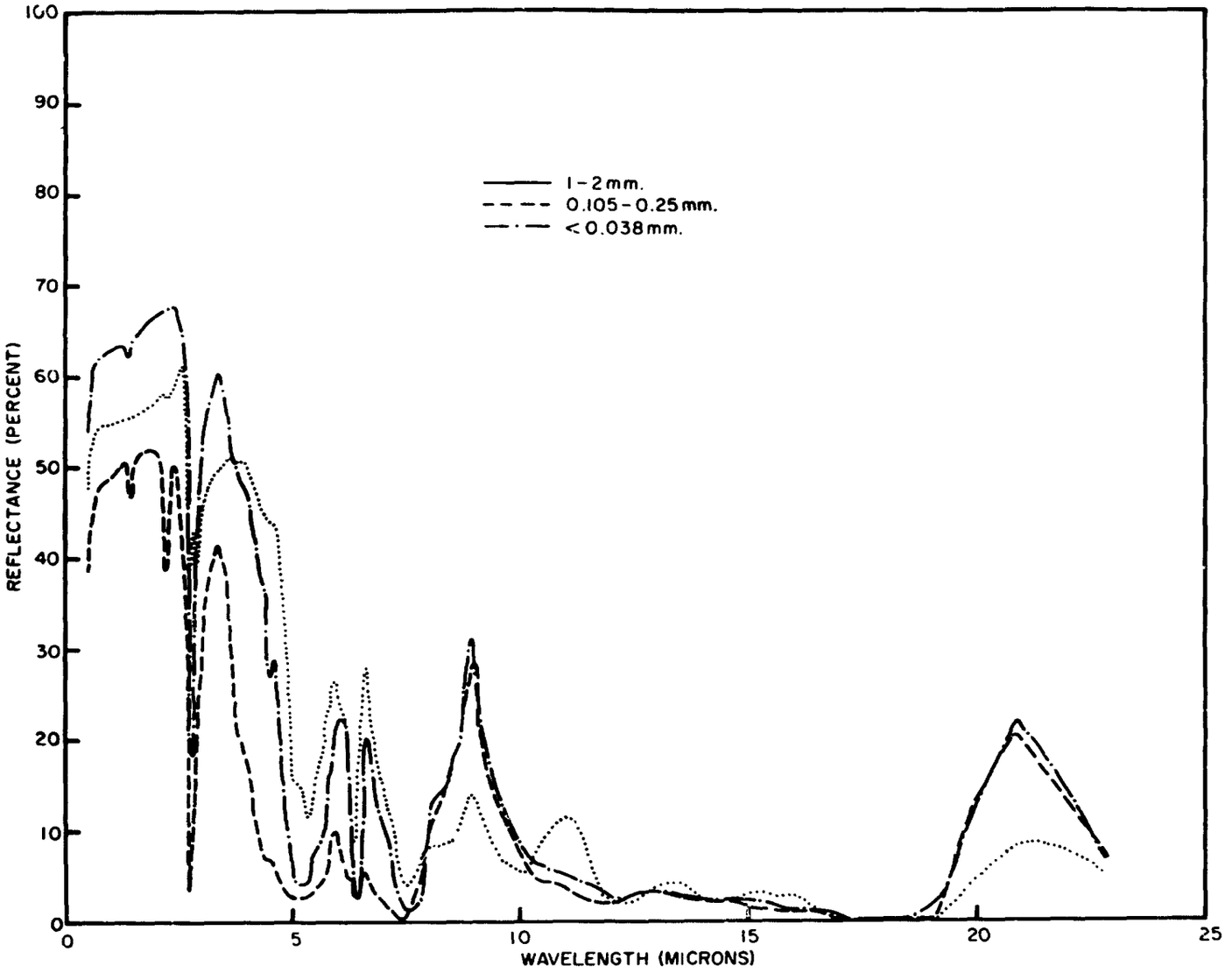


Figure R-66–Dynasil Quartz.

CONCLUSIONS

This work confirms the general observation that the contrast of the reststrahlen features decreases as the particle size of the sample is reduced. Figures B-51, B-58, B-65 and B-66 show this general effect. The sharpest decrease in reflectance at the reststrahlen wavelengths usually occurs between the solid sample and that in which the particle size is between 1-2 mm. The decrease is not too marked between a 1-2 mm sample and that with a particle size of 0.105-0.250 mm, but this is often followed by another sharp decrease in reflectance when the particle size is reduced below 38 microns. Since there is a considerable likelihood that the lunar and martian surfaces are quite dusty, spectral contrast will probably be rather low.

One exception noted is dunite where there is a rise in the reflectance at the reststrahlen wavelength when the particle size is reduced from 1-2 mm down to 0.105-0.250 mm. This may be a selective filtering effect.

This work also confirms the observation (2) that the fundamental Si-O vibration near 10 microns shifts with the type of igneous rock. The composite of Figure B-61 shows this shift. For granite, an acidic rock, and for quartz, this feature occurs near 9 microns, while for the ultrabasic material dunite, this peak occurs near 10.7 microns This effect has been suggested as a test for the general type of material occurring on the surface of the moon. Our observations confirm the possibility but suggest that the situation is highly complex, especially when in actual observation from earth by a telescope or from an orbiting spacecraft, more than one sample will lie in the field of view of the spectrometer, and different particle sizes will be present for each rock. Figure B-65 shows the effect for composite samples of various igneous rocks when the particles are of one size, alone. Such composites suggest that the Si-O shift may be washed out by the presence of several igneous rocks with various particle sizes in the spectrometer's field of view. Obviously this problem would be greatly reduced in analysis performed after soft landing.

The effect which produces a new reflectance feature when the particle size of many rocks is reduced below 100 microns will also complicate this acidic vs. basic test. As Figure B-64 showed, a finely powdered granite dust will appear quite similar to a basic rock, and if a basic rock were also present, the composite curve would appear to peak at the longer wavelengths indicative of an ultra-basic surface. The average shift toward longer wavelengths for the materials measured in this work was 2.3 microns, almost identical with the total shift of the Si-O vibration between acidic and ultra-basic materials. A finely powdered acidic surface might be classified as dunite. This mistaken identity might be avoided if a scan is also made in the 1-4 micron region of the spectrum where the much higher reflectance of finely powdered samples might help distinguish the dust from solid samples.

On almost all of the curves will be noted the water of hydration features at 6.35, 2.7, 2.2 and 1.4 microns. In some cases, such as in tektites, this amount presently lies beyond the limits of geological analysis. Work has begun to calibrate these features for use as a sensitive probe of waters of hydration on planetary surfaces. The water of hydration feature in tektites is the smallest observed in any naturally occurring material.

ACKNOWLEDGMENTS

We would like to acknowledge the assistance of P. Lowman and L. Walters of the Goddard Space Flight Center for providing most of the samples and for their many helpful discussions. We also wish to acknowledge the constant and careful effort of H. Keegan

and V. Weidner of the National Bureau of Standards in performing all of the measurements from 2.5-22 microns. Our special thanks, also, to L. Horrell for his careful reduction of the data.

REFERENCES

1. White, J. U., "New Method of Measuring Diffuse Reflectance in the Infrared," J. Opt. Soc. Amer., Volume 54, page 1332 (1964)

2. Lyon, R.J.P., Econ. Geol. Volume 60, page 715 (1965)

BLANK PAGE

THE GRAVITY REFERENCE SYSTEM

S. M. Shapiro

INTRODUCTION

The problem concerning this phase of the Summer Workshop was twofold:

1. To determine the reliability and usability of the mercury-oil pool; and
2. Evaluation of Building 7 concerning vibrations and their affect on the pool.

More time was spent on the latter problem since in order to do any reliable investigations into the mercury-oil pool, it is necessary to obtain the most stable conditions possible.

EXPERIMENTAL PROCEDURE: EVALUATION OF LOCATIONS

A given mercury-oil pool was used with the monochromatic Hilger-Watts TA-5 autocollimator mounted vertically on the specially designed stand. The mercury-oil pool (1) consisted of a lucite block with the center hollowed out to a depth of 1.25 inches and a diameter of 8.0 inches (Figure B-67). Enough mercury was poured into the container so it just covered the bottom. The meniscus of the mercury approached the curved walls of the container. Molvac A Silicon oil was poured over the mercury. The depths of the fluids were: Mercury – .125 in.; and Oil – .25 in.

Figure B-67–The experimental arrangement for testing the mercury pool. Above the mercury pool is mounted the Hilger-Watts TA-5 autocollimator, in a vertical position, supported rigidly by a massive Unertl stand. To the left is the electronic null detector of the autocollimator.

The TA-5 autocollimator, according to factory specifications, has a 100 arcsec range with a repeatability of 0.01 second and an accuracy of 0.1 second. The output of the electronic readout was connected to a strip chart recorder; either the Hewlett-Packard Sanborn model or the Brush recorder.

Calibration was performed by changing the autocollimator by 1 second and noting the displacement on the recorder. Several 1 second movements were performed and an average taken. Since stability was not obtained at a given angular setting, a straight line was drawn through the markings on the recorder that the experimenter felt represented the dc level.

The mercury-oil pool was placed in the area to be tested and allowed a half hour to stabilize. Turbulence was decreased by covering the mercury-oil with a plate glass window and placing a tube over the air column between the mercury and the autocollimator. Readings were taken for various times while cautions were taken to reduce vibrations. Angular excursions were read directly from the chart paper by finding some mean displacement from the dc average.

All readings were taken in the basement of Building 7. Readings were taken in the Spacecraft Alignment Facility (SAF) located in test cell 029. The SAF is resting on a 300,000 lb. seismic block supported by 20 springs. The loading blocks were positioned at:

N-S block: 4.4 units
E-W block (1): 5.9 units
E-W block (2): 6.4 units

Tests were made on the southeast corner of the block. In addition to the short time readings, an overnight run was made and the relative vibrations of the building compared. Tests for stability were conducted in Room 005, both on the granite block and on the metal table.

Another series of tests was made in Room 006, the VOB room, on the alignment block. In this location attempts were made using different materials as a base for the mercury-oil pool to dampen out vibrations:

1. The pool was placed on some sand. In a rectangular container, the bottom was covered with sand, a layer of paper placed over the sand, then another layer of sand, paper, more sand, and finally a layer of paper. The mercury-oil pool was placed on top of the paper.

2. A 1.0 inch thick sponge pad.

3. .25 in. sponge pad: the species used as packaging material.

RESULTS

The results of the above testing procedures are given below and in Figure B-68. The two numbers in column 2, Table B-22 refer to the minimum and the average peak to peak excursion in arc seconds as indicated by the chart recorder. The average value is near the maximum, but there are several pulses observed whose peak to peak excursion is greater than the average value. The ranges are given along with the frequency values because the frequency of vibration varied.

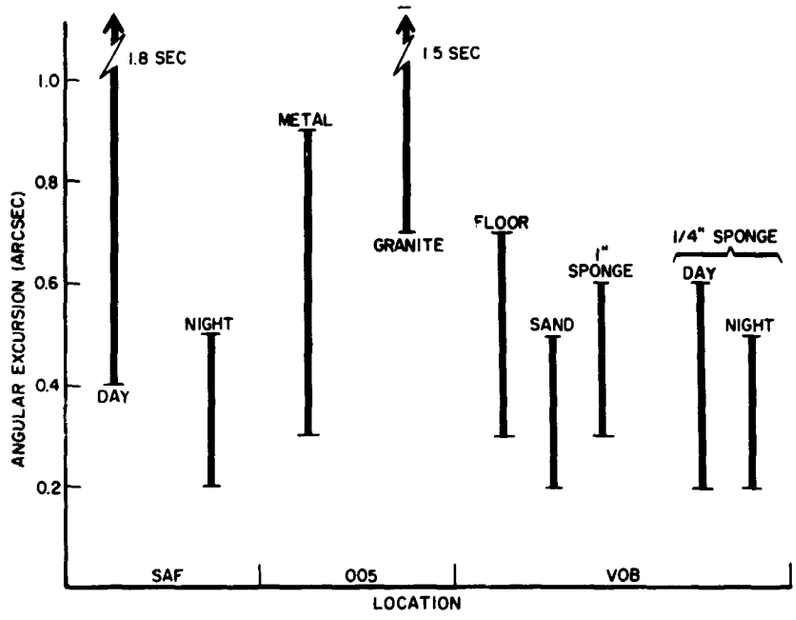


Figure B-68–Range of minimum to average excursion, peak to peak, of normal to mercury surface due to vibration of the building.

Table B-22
Results of Stability Tests at Various Locations in Building 7, GSFC

Location	(p-p) arcsec.	Frequency
SAF		
day	.4 - 1.8	1 ~ 2 cps
night	.2 - .5	1 ~ 2 cps
005		
Metal Table	.3 - .9	
Granite	.7 - 1.5	3 ~ 4 cps
VOB		
Hg on floor	.3 - .7	2 ~ 4 cps
Hg on Sand	.2 - .5	2 ~ 4 cps
Hg on 1" sponge	.3 - .6	2 ~ 4 cps
Hg on 1/4" sponge		
Day	.2 - .6	2 ~ 4 cps
Night	.2 - .5	2 ~ 4 cps

These results show that the VOB room is the quietest of the areas tested. We also revealed that the granite block has a vibration frequency of 3-4 cps. A further investigation of this vibration by use of an accelerometer revealed a frequency of 4 cps. Integration of the output gave us an amplitude of the vibration = 40 microinches.

The tests in the SAF showed a dependence with time of day on the vibration level. As expected it is quieter at night when there is less activity in the building. In the VOB, there is little difference between the various absorbing materials.

An overnight test (17 hours) was made with the .25 in. sponge absorber. This revealed the fluctuation from the dc level observed in all tests in addition to an oscillatory drift of the dc level with amplitude of 1.5 second and period of 16 hours. Another run for 40 hours was made to repeat the above observations. This however revealed no dc drift. The only major change in conditions was that in the 17 hour run, the LTOF was operating, and in the 40 hours test, the LTOF was off. Since the dc shift was not repeated, it was felt that an unusual situation may have been caused by vibrations from the compressors.

DISCUSSION

A question which arises, since we are seeing relative motion, is whether the motion observed is that of the mercury or that of the autocollimator. This is answered by autocollimating off a rigid mirror. This yields a straight line on the chart recorder which suggests either a stable condition or rigid motion of both the autocollimator and the mirror. A further qualitative test to determine rigidity of stand and autocollimator was the "gentle tapping" of the autocollimator and its stand, noting the angular change. This test yielded very little angular change and thus it was concluded the autocollimator was relatively rigid and most of the motion was that of the mercury in attempting to keep its gravity level.

The basic conclusions were that the stablest place for testing the mercury-oil pool is on the block in the VOB room (006). The amplitude of the 3 or 4 cps vibrations varies from .3 to .7 arcsec, peak to peak, and with an absorbent material under the mercury pool; this can be reduced to a range of .2 to .6 arcsec.

RELIABILITY AND USABILITY OF THE MERCURY-OIL POOL

Research into the use of the mercury-oil pool as a gravity reference was begun last year by the summer workshop. A description of the pool is given by Thekaekara and Appler (1). Essentially, the same pool was used; instead of the ashtray as a container, lucite blocks were available with a hollowed out portion whose contours were similar to that of the ash trays used previously.

In order for reliable work to be done, a stable area should be available to test the mercury pool. However, this ideal condition cannot be attained with available facilities. Thus we had to be content with present conditions and develop methods to dampen the vibrations of the mercury.

How flat is the mercury surface and how does the flatness depend on the diameter of the pool? Because mercury does not wet most surfaces, a negative meniscus is formed where the mercury comes in contact with the vertical walls of the solid vessel. This will affect the surface of the mercury in the neighborhood of the edges of the pool. One quick test of the surface is a visual test. If the surface is not flat, the quality of the image will be poor.

Rationalization of the flatness of the surface is obtained by consideration of the forces acting on the surface. Far away from the edges of the pool, the only horizontal force is the surface tension, which is equal in all directions and thus the surface tension forces cancel one another out. Since we can assume that the pressure is constant over the surface of the mercury, the surface should be flat because there are equal

vertical forces acting on the surface (2). A rule of thumb used by astronomers is that a circle, about the center of the vessel, whose diameter is equal to the radius of the vessel can be considered flat (3).

Theoretically, the surface would be flat if it were stable. However, we don't have stability, and motion of the surface is measured with the autocollimator.

The following are a list of ideas to reduce the surface vibration of the mercury:

1. A definite need is to check the various sized containers for the mercury-oil pool. It would seem that with a larger container, waves would not travel to and fro as quickly and would dampen out before being reflected.

2. Effect of the depth of the mercury and oil on surface vibrations should be studied.

3. Increasing the pressure on the mercury surface, either by compressing the air above the oil, or by using a large oil layer, should cause a damping of the surface waves since this will impede any up and down motion of the mercury. There is a question raised as to how this will affect the gravity reference.

4. Floating a smaller mercury-oil pool on a large pool will dampen motion within the smaller pool. If the smaller pool is resting on the surface of the mercury in the larger pool, any motion imparted to the larger pool will disturb the smaller pool. The oil in the larger pool which wets the outer surface of the smaller vessel, will tend to keep the latter in its position and oppose any motion of the vessel. The density of the vessel for the smaller pool has to be low so the weight of the fluid displaced in the larger pool will be greater than the weight of the vessel, mercury, and oil of the smaller pool in order for the latter to float.

5. A magnetic field can be used to dampen vibrations (4). Since mercury is conductive, by placing the pool in a magnetic field, eddy currents will be induced in the moving fluid. The force resulting from this current will be such as to oppose the motion of the mercury. A main problem is the engineering in order to get a field of proper size, uniformity, and direction.

6. Instead of changing the mercury, it is possible to rectify the output so it will exhibit an average, or the dc level. Any shift in the surface will result in a displacement of the dc level.

7. It is also worthwhile to test the results when the mercury is poured into a silver or platinum container; one whose surfaces will be wet by the mercury.

The above experimental suggestions should be performed in quiet surroundings in order to get reliable quantitative data.

A theoretical analysis of the motion of the mercury and the propagation of surface waves as a function of the depth of the mercury and diameter of the pool would be useful.

REFERENCES

1. Thekaekara, M. P., Appler, R. L., Rev. of Sci. Inst. 36, 644 (1965)

2. Champion, F. C. Davey, N., Properties of Matter, Philosophical Library, New York, N. Y. (1959)

3. Appler, R. L., Private communication

4. Feynman, R. P., Leighton, R. B., Sands, M., The Feynman Lectures on Physics. Addison-Wesley Publishing Co., Inc., Reading, Mass. (1964) Vol. II

5. Newman, F. H., Searle, V.H.L., The General Properties of Matter, The Macmillan Co., New York (1933)

6. Stoker, J. J., Water Waves, Interscience Publishers, Inc., New York (1957)

SUMMER WORKSHOP 1965

Program Outline and Team Participants

PROJECT C: Response of Structures to Dynamic Loads and Thermal Stresses

Study Topics

C-1 Analysis of the dynamic response of the Rosman I antenna

C-2 Measurement of structural characteristics of a ground based antenna

C-3 Measurements of the dynamic response of the Rosman I 85-ft dish antenna

C-4 Methods for controlling thermal distortion in structures-a feasibility study

C-5 Devices for controlling thermal gradients in spacecraft structures

C-6 Perturbation torques on a 750-ft extensible boom due to electrodynamic interactions

C-7 Dynamic analysis of the RAE Satellite

C-8 Survey report on possible antenna configurations for a radio astronomy satellite

TEAM A

Academic Personnel

Dr. Yu Chen
Principal Investigator
Dr. J. M. Dalrymple
Mr. R. O. Woods
Mr. M. A. Oien
Mr. J. A. DellaValle
Mr. T. W. Jarrell

Goddard Personnel

George C. Winston
Staff Advisor
Thomas G. Butler
David L. Blanchard
Frank J. Cepollina
Paul G. Fahlstrom
Dr. J. V. Fedor
Thomas W. Flatley
N. A. Haumann
Donald M. Witters
Elliott Woo
Joseph P. Young

NSF Summer Science Student

Robert A. Krauss

BLANK PAGE

PROJECT C: RESPONSE OF STRUCTURES TO DYNAMIC LOADS AND THERMAL STRESSES

CONTENTS

ILLUSTRATIONS

TABLES

BLANK PAGE

ANALYSIS OF THE DYNAMIC RESPONSE OF THE ROSMAN I ANTENNA

M. A. Oien

It was the purpose of this project to determine analytically the dynamic response of the Rosman I antenna to mechanical disturbances incurred while tracking a satellite. This task proved to be too large to be completed by the project group in the time allotted. A large amount of useful work has been done, however, and it is expected that the Goddard staff working on the project will be able to successfully complete the analysis. The major part of the summer was spent in tabulating data necessary for use in the analysis. Models of several substructures of the antenna were successfully analyzed yielding information needed for the synthesis of a complete model. A complete model was assembled. The analysis of this model is incomplete, however, and is to be continued by the Goddard staff. It is expected that this model will yield useful results.

Rosman I is an X-Y mounted, ground based antenna having an 85-foot diameter parabolic reflector. A quadruped structure supports the radio signal receiving elements at the focus of the parabola. During tracking the dynamic response of the structure causes the reflecting surface to deviate from a true parabola and the receiving elements to be displaced from the focus of the parabola. These effects cause distortion of the received signals and limit the tracking accuracy of the antenna. For an overall systems analysis it is desired to find a transfer function relating these distortions to the input to the servo unit controlling the movement of the antenna.

To perform the analysis "A Digital Program For Static And Dynamic Analysis Of Complex Structures" developed by the Martin Space Systems Division was used. This program, written for a high speed digital computer, utilizes a matrix-force method using an element matrix technique. A truss structure such as the Rosman I antenna is simulated by a number of simple structural elements connected at grid points. Each of the elements carries one characteristic load (tension, torsion, bending or shear) and one corresponding deformation. Compliances are specified relating the deformation to the loads in the elements. A set of primary elements comprising a determinate structure is chosen and additional elements are classified as redundants. Vector unit loads are applied at the grid points to define the degrees of freedom. These include both translations (forces) and rotations (moments). Vectors defining any external loads must also be applied at the grid points. Equations are written establishing equilibrium between the internal loads in structural elements, unit loads defining degrees of freedom, and external loads at the grid points. Since there are redundants in the structure, compatability conditions must also be met. These are obtained through application of the principle of virtual work. Having established the structural model and the required equations, the program is able to effect a static analysis, computing the loads and deflections throughout the model due to the externally applied static loads.

The dynamic analysis is based on a modal method. Due to the presence of damping it is an approximate method. The mass properties of the antenna structure are lumped at the grid points of the model. Normal modes are calculated for the undamped system and the equations of motion are transformed into normal coordinates. Viscous damping is then specified for each mode, and the external time dependent loads, simulating the driving forces applied by the servo unit during tracking as applied at grid points of the model are developed into generalized forces in each mode. The responses to these loads can be computed and related to the inputs to yield the desired transfer functions. Transfer

functions must be found for various positions of the antenna since the structural configuration changes during tracking. However, the structural model can easily be altered to accommodate these changes. Using this program it is also possible to find the dynamic response for a given wind loading.

A structural model closely resembling the actual Rosman I antenna would have required about 3000 structural elements – far above the computational capacity of the computer program. The program has a capacity of 1225 primary elements, but practical considerations (debugging and computational time) further limit it to about half that number. It was therefore necessary to simulate the antenna with a model having about 600 primary elements. To straight away produce a model of such limited size would have required so many assumptions as to place the reliability of the results in question. Consequently, it was decided early in the program to perform a detailed analysis of several substructures of the antenna. Using these as a basis for comparison, various simplified models of each substructure could be synthesized and analyzed until a model was found to match the behavior of the detailed substructure in the low frequency range. The simplified models could then be assembled to form a reliable composite model of the entire antenna structure.

The structure was partitioned into six substructures: the foundation, the tower, the X-wheel, the Y-wheel, the reflector, and the quadruped structure. By mid-summer most of the data necessary for a detailed analysis of these substructures had been extracted from the manufacturers drawings and tabulated, the detailed model for the tower substructure assembled, and the assembly of the detailed model for the X-wheel was underway. But at this time it became apparent that the task was placing too much emphasis on the analysis of the detailed substructures. It was felt that while this method of analysis would eventually lead to a reliable model, it would also lead to a prolonged program and require the expenditure of too much time to be practical.

It was then decided that it would be more productive to try to synthesize a model of the complete antenna without completing the analysis of the detailed substructures. It was assumed that the excitation of the very low natural frequencies comprised the major portion of the dynamic response in tracking. It was further assumed that the frequencies important in the dynamic response involved rather simple mode shapes. These assumptions would seem to preclude the necessity for maintaining much detail in a final model for the controls problem. Eventually there will be a need for instantaneous determination of the focus of a distorted reflector. At this time considerable detail will be necessary.

In mid August work was started on a simplified model of the complete structure. The tower was assembled using results from the then completed detailed model of the tower substructure. Simple models of the X-wheel and Y-wheel were obtained rather easily since these substructures are constructed with relatively few main load carrying members braced by a number of small members. Most of the weight in these substructures is concentrated at several points (counterweights, drive units, and bearings) so that a reasonable mass distribution can be obtained with relatively few grid points. The foundation was also easily represented, however, since little is known about the compliance of the soil reacting on the foundation, the nature of these reactions was largely a matter of conjecture. The reflector backup structure and feed box support were both constructed with a number of small members forming a light framed structure, and could not so easily be reduced to simple models. However it was in this part of the antenna structure that detailed information concerning the dynamic response was desired, so that a good deal of care was taken in synthesizing the model in this area. The reflector backup structure was essentially a set of 24 equally spaced radial ribs joined by circumferential bracing. The rigs were identical in sets of 8 and were arranged in an A-B-C-C-B-A-A-B pattern. A separate analysis was performed of detailed two-dimensional models

of each type of rib and circumferential bracing. The results were used as a basis of comparison of corresponding parts of the simplified model. A separate analysis was also performed on the legs of the quadruped structure supporting the feed box, and was used as a basis of comparison for the simplified model.

All of the data necessary for the synthesis of a simple model of the complete antenna had thus been tabulated. This model had 183 grid points, 640 primary elements, and 261 redundant elements, and 552 dynamic degrees of freedom. The data had also been translated into input for the computer program, however, a number of clerical errors in the input prevented the program from successfully being run on the computer.

Although the transfer function could not be obtained until the analysis of this model was completed, several things were learned which are worth noting here.

It had been suggested that the natural frequencies of a model of a highly stable truss structure would not be greatly affected by omitting redundant bending across structural joints. This assumption was checked in the detailed model of the tower, and it was found that while it is true that the inclusion of redundant bending elements did not greatly affect the lower natural frequencies, the magnitude of the relative displacements within the mode shapes were greatly affected. Since the transfer function is a ratio of amplitudes at different points in the structure, the inclusion of redundant bending elements is an important factor in obtaining accurate results.

It had also been assumed that for low natural frequencies, the associated mode shapes would be rather simple. The analysis of the detailed model of the tower indicated that while the lower mode shapes were likely to be simple, they could not easily be predicted. In general, the large masses concentrated at the X-wheel drive unit and bearings tended to oscillate with large amplitudes while the lighter framework of the structure followed its motion. It can be expected that the complete model will behave in much the same way at the low natural frequencies; the heavy counter weights, drive units, and feed box oscillating with large amplitudes and the rest of the structure following. If this is true, it would seem that while the feed box may vibrate with large amplitudes at the low frequencies, the relatively light and rigid reflector backup structure would not be expected to distort greatly, but oscillate as a unit.

It should be mentioned that the progress of the task was greatly hindered by lack of attention given by the group to maintaining a high degree of accuracy in performing the clerical work associated with this problem, especially in the translation of tabulated data into computer input. Although this aspect of the work is quite tiresome and is hardly stimulating, inattentiveness can only lead to frustration and a great waste of time. The need for careful preparation of input data and maintenance of a good checking system cannot be stressed too greatly.

In summary, the work came near to completion in spite of the size of the task. Assuming that the simplified model of the complete structure is acceptable, all data have been assembled, except for clerical checking, for the dynamic analysis of the tower in the vertical position.

ACKNOWLEDGMENT

The author is indebted to this group of workers who helped in the prosecution of this task: L. R. Bruck, J. F. Kuzanek, E. S. Chaney and the NSF High Ability Secondary School Student R. A. Krauss.

MEASUREMENT OF STRUCTURAL CHARACTERISTICS OF A GROUND BASED ANTENNA

N. A. Raumann

PURPOSE OF TEST PROGRAM

At present the 85-foot antennas operated by GSFC are functioning properly and no major problem areas exist that would require a drastic modification of these systems. Consequently, tests which determine the structural characteristics of these antennas are of secondary importance, as far as these existing systems are concerned. However if future tracking requirements are studied it can be seen that certain system changes will be needed. Knowledge of the structural behavior will be essential in analyzing these new requirements, some of which are mentioned below.

At the present time a project is underway to study the possibility of installing a digital computer at each data acquisition facility. This computer is to be used mainly to reduce satellite data at the tracking site, to keep a log on status information of various space vehicles and to generate predictions in local coordinates for positioning of the antennas. This computer could also be used to control the antennas directly and thereby perform most of the functions of the analog servo system. A digital servo could possibly improve reliability and could provide an operation that is more stable.

The operation of antennas is very often influenced by environmental conditions and by the type of targets to be tracked. Wind, rain, and temperature all affect the characteristics of the servo tracking loop and cause a deterioration in performance from some optimum setting. Likewise, satellites, depending on orbit dynamics and signal strength, require varying adjustments of the tracking loop. These adjustments are not provided in the present system configuration. Consequently some adaptive techniques could be used, which could improve this situation by automatically performing these adjustments.

The RF frequency spectrum in which these antennas operate becomes progressively more congested and a switch to higher frequencies may become mandatory. This, however, will also require an increase in tracking accuracy to a point which cannot be handled by present systems. Therefore a more precise control system will have to be designed.

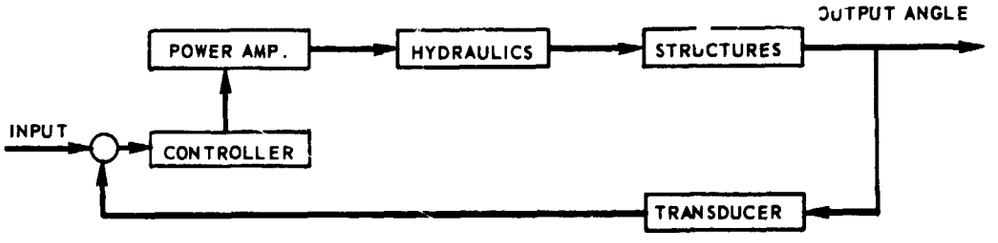
Another area to be investigated is that of reliability. Components of the servo loop are subject to deterioration in performance with time and are subject to eventual failure. It may be possible to construct a system which would automatically compensate for component deterioration and even indicate impending failure.

Most of the antennas operated by GSFC are equipped with hydraulic drives. This may not be the optimum choice as far as reliability and controlability are concerned. Studies of other drives such as eddy current clutches, electric, and variable frequency drives may prove that these are better suited for this application.

All of above mentioned study areas require a good knowledge of the structural dynamics in order to investigate the feasibility and the implementation of these various approaches.

IMPORTANCE OF DYNAMIC CHARACTERISTICS

The studies that will utilize the dynamic characteristics of the antenna structure are generally oriented around control system principles. The antenna control system consists of various blocks, which are interconnected to form a servo loop. When each block in the loop is described by its dynamic input-output relationship or transfer function, then the loop can be analyzed to determine the characteristics of the control system.

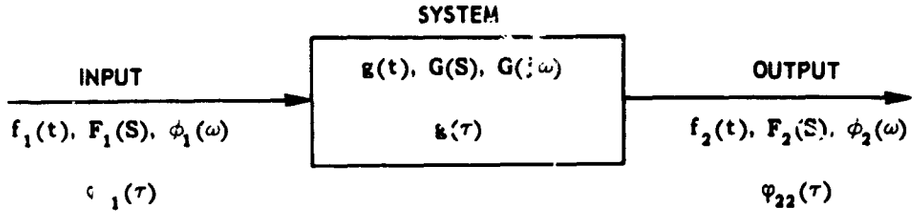


Target dynamics and wind spectra acting on the antenna determine the lower limit of the servo system bandwidth, whereas noise, which is proportional to bandwidth, sets an upper limit. The gain of the servo loop is directly proportional to the accuracy with which the antenna can follow a certain target, but gain is closely related to bandwidth. The shape of the open loop system transfer characteristic together with its associated gain determine the stability of the system and also permit the evaluation of some nonlinear effects. Thus it is very important to have a dynamic description of the structure and furthermore to have this description in terms of a transfer function to perform any of the above mentioned analyses.

TEST AND ANALYSIS METHODS

To test the antenna structure, accelerometers have been placed in strategic positions on the structure, and the antenna drive system has been excited with various test signals. Test signals chosen are square waves, sine waves, impulse functions, and white noise. It is hoped that by comparing accelerometer responses during the application of these test signals to the structure will give a detailed description of the structural dynamics. Accelerometer outputs have been recorded on instrumentation tape recorders and tapes have been brought to GSFC for analysis. Several methods for determining transfer functions will be considered.

Basically, if it is desirable to find the transfer function of an arbitrary system $g(t)$ experimentally, it is required that some signal be applied to the input of the system and the output measured. Input and output wave shapes, when recorded, are in the time domain and are designated $f_1(t)$ and $f_2(t)$, respectively. The functions $f_1(t)$ and $f_2(t)$ can usually be Laplace transformed giving $F_1(s)$ and $F_2(s)$. Input and output waves can also be spectrum analyzed, yielding power density spectra ϕ_1 and ϕ_2, or autocorrelation functions can be found, ϕ_{11} and ϕ_{22}. Thus the arbitrary system looks as follows:



In the time domain, input and outputs are related by the convolution integral.

$$f_2(t) = g(t) * f_1(t) = \int_0^t g(x) f_1(t - x)\, dx$$

This integral, however is often difficult to solve and one may have to resort to the Laplace transform. In this domain the input-output relationship is somewhat simplified:

$$F_2(s) = G(s) F_1(s),$$

and for the special case in which the input is an impulse function, this reduces simply to

$$F_2(s) = G(s)$$

Since testing is usually performed in the presence of noise, an averaging process will improve the accuracy of results obtained from the previous method. This process involves the addition of several output responses obtained from the same input signal. If the measuring noise can be approximated by a stationary Gaussian random process with zero mean and if the hypothesis of ergodicity holds, then this addition process will enhance the actual system response and minimize measurement noise.

Spectrum analysis relates the input and output of a system by

$$\phi_2(\omega) = |G(j\omega)|^2 \phi_1(\omega)$$

This expression can be simplified if white noise, which has a flat spectrum, is used as a test signal:

$$\phi_2(\omega) = |G(j\omega)|^2$$

Spectrum analysis is also very useful in determining non-linearities. If certain fixed frequencies were applied to a linear system, only these frequencies will show up in the response. Frequency components in the output which are not present in the input, are indications of non-linearities.

If the autocorrelation function, of input and output signals is determined, then the following relationship holds:

$$\varphi_{22}(\tau) = \int_{-\infty}^{\infty} g(t_1)\, dt_1 \int_{-\infty}^{\infty} g(t_2)\, \varphi_{11}(\tau + t_1 - t_2)\, dt_2$$

However if the cross-correlation function between input and output is found, a somewhat simpler relationship is obtained.

$$\varphi_{12} = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f_1(t) f_2(t + \tau)\, dt$$

which reduces to simply:

$$\varphi_{12} = g(\tau)$$

when white noise is used as a test signal with $g(\tau)$ being the impulse response of the system.

It must be determined which of these analysis methods will yield the best results when the structure of a large antenna is evaluated. Thus this program may come up with a unique test method which can be used efficiently in the future.

BLANK PAGE

MEASUREMENTS OF THE DYNAMIC RESPONSE OF THE ROSMAN I 85-FOOT DISH ANTENNA

J. Dalrymple

SUMMARY

The Rosman I 85-foot dish antenna was subjected to various types of input signals and the response was measured at several different locations in the antenna system. This was done in order to obtain a more detailed understanding of antenna performance, to aid in any future redesign analysis, and to provide experimental data suitable for checking a theoretical (mathematical) model of the antenna structure which is being programmed for a digital computer.

On the basis of strip charts transcribed from the original tape recordings and of two spectrum analyses, the natural frequencies of the antenna system include at least 2.5, 5.6, 8.0, and 13.5 cps for X angle response to X angle input. Under these conditions, damping factors of approximately 0.1 and 0.05 were obtained for the 2.5 and 8.0 cps modes, respectively. A frequency response diagram, obtained by integration on an analog computer, was in good agreement with these results. A twisting mode applied to the structure gave a fundamental response of 2.0 cps.

Very much more work remains to be done to get the most out of the information contained in the experimental data.

PURPOSE

These measurements were made to serve several purposes:

1. To obtain actual natural frequencies of vibration of the antenna structure
2. To obtain modal damping factors for the complete structure
3. To obtain experimental transfer functions, relating the response of an arbitrary point on the structure to an input at some other point
4. To provide an experimental check on results obtained by computer analysis, and
5. To indicate possible sources of tracking errors.

TEST PROCEDURE

A test signal consisting of a square wave, impulse, white noise, or sine wave was introduced into the input of the antenna servo system by means of an appropriate signal generator. Each type of test signal was applied at various amplitudes and frequencies with the antenna at the zenith position and at the horizon position in the open position loop mode of antenna operation. Two separate series of tests, one with X angle (lower drive gear) excitation and one with Y angle (upper drive gear) excitation, were performed.

In addition to the above types of tests, collimation tower lock-on test sequences were performed, in which the antenna was displaced manually from the direction of the collimation tower and then put in tracking mode. The antenna would return and lock on the the

collimation tower direction. Motion pictures of the collimation tower during lock-on provided a direct means of observing antenna overshoot and tracking error. Lock-on tests were performed at several angular displacements from collimation tower direction and with three receiver frequencies in the tracking mode.

Data were recorded on twenty-one channels of FM magnetic tape at 1-7/8 ips. Data recorded came from the

1. test signal generator
2. servo drive amplifier output
3. hydraulic pump pressure
4. hydraulic pump yoke position
5. antenna position
6. gear box tachometer
7. seismometer
8. accelerometers.

Figure C-1 and C-2 show the antenna structure and a typical accelerometer installation.

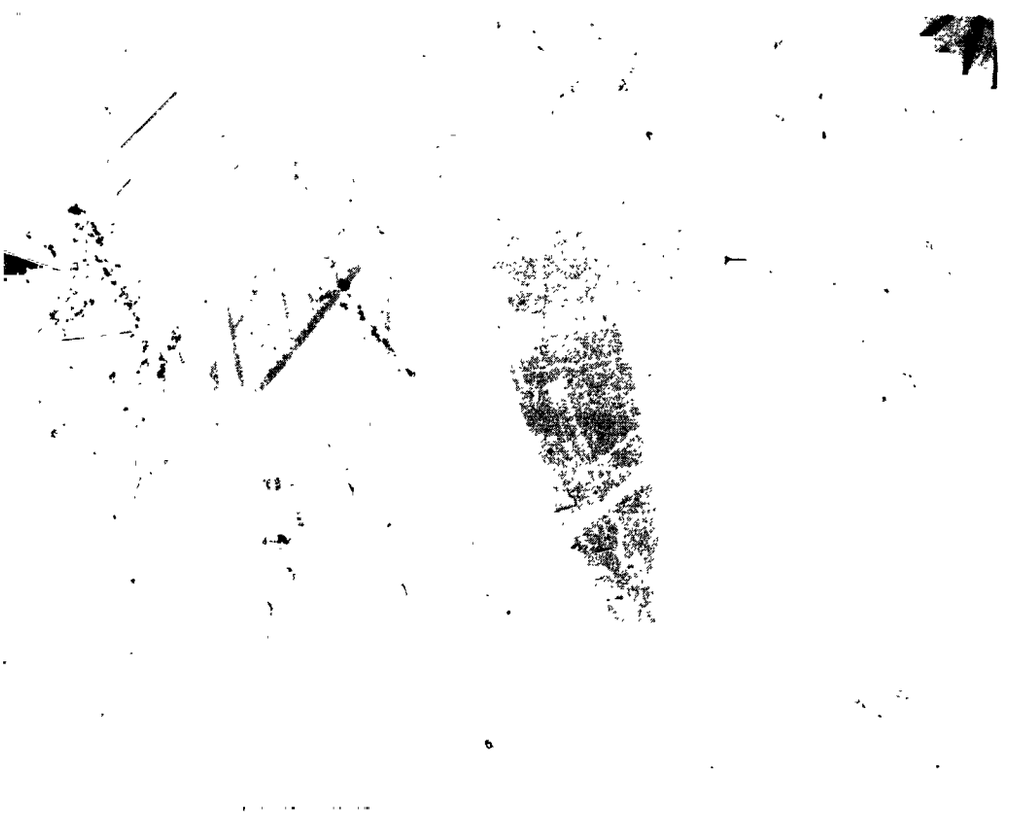


Figure C-1–Rosman I 85 foot dish antenna.

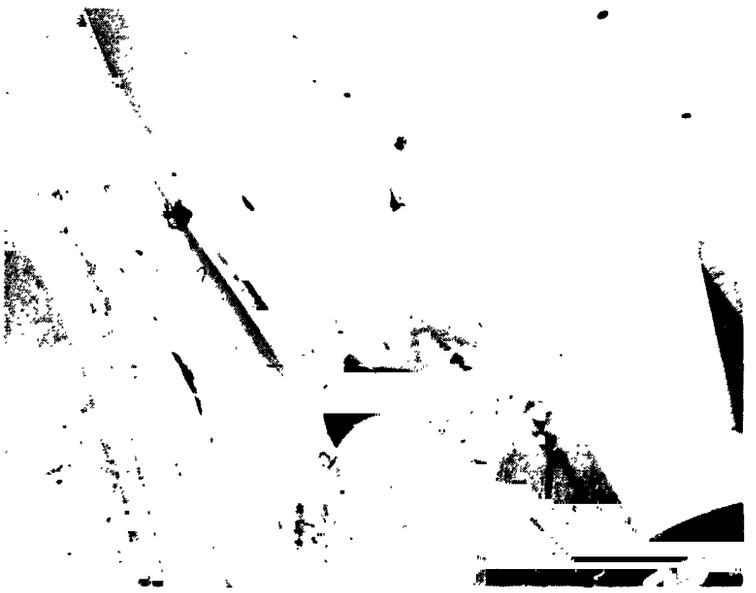


Figure C-2–Typical accelerometer installation.

DATA REDUCTION

Proposed methods of data reduction included:

1. Transfer of data from magnetic tape to strip charts for preliminary visual examination
2. Transfer of data from original magnetic tape to a magnetic tape loop for spectrum analysis, and for determination of transfer function description using the spectral density function associated with white noise excitation
3. Frequency separation of original data by means of a bandpass filter in order to determine the actual wave shape and damping factor associated with each natural frequency present
4. Determination of transfer function description using both analog and digital computer programs operating on the response to an impulse excitation.

TEST RESULTS

Natural Frequencies

Table C-1 lists the natural frequencies which could be approximated from the strip charts for various accelerometers. These were obtained from the impulse and square wave responses for different antenna positions and for X angle and Y angle excitation.

For X angle excitation, a 2.5 cps fundamental frequency is indicated throughout the structure for X angle response, with antenna in zenith position. In addition, a 12.5 cps response is indicated by some of the accelerometers, while the X drive gear indicates a strong 5.6 cps response. Only a 12.5 cps frequency is evident in the Y angle response of the feed box, and there was a 170 millisecond delay before this response was initiated – presumably due to modal coupling. Figures C-3 and C-4 are strip chart recordings of the X angle response at the feed box and the X drive gear, respectively.

Table C-1
Natural Frequencies Estimated from Strip Charts

	Zenith (X=0, Y=0)		Horizon (X=0, Y=-70°)		Horizon (X=+75°, Y=0)	
	X angle	Y angle	X angle	Y angle	X angle	Y angle
X angle Impulse Excitation						
Feed Box	2.5, 12.5	12.5	2.8, 12.5	12.5	no test performed	
*Dish	2.5, 12.5	–	**	–		
Y axis Bearing	2.5	–	2.8, 12.5	–		
X drive Gear	2.5, 5.6	–	5.9	–		
Y angle Impulse Excitation						
Feed Box	**	2.5, 14	†	2.5	†	2.0, 14
Dish	14	2.5, 14	†	2.5	†	2.0, 14
Hub	**	2.5, 5	†	2.5	†	2.0
Y axis Bearing	14	–	2.5	–	2.14	–
Y drive Gear	–	2.5, 5	–	low output	–	4, 5

*Response due to square wave input
**Response frequency masked by 60 cps noise
†No chart available at present

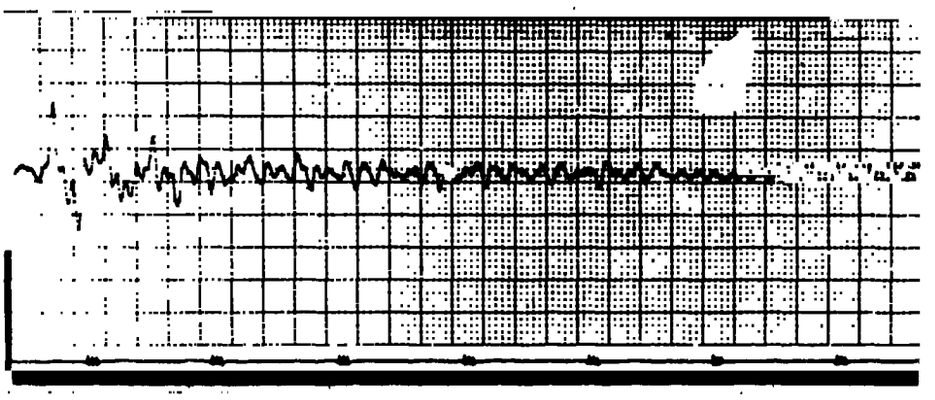


Figure C-3–Feedbox accelerometer response to impulse excitation. Excitation in X, response in X.

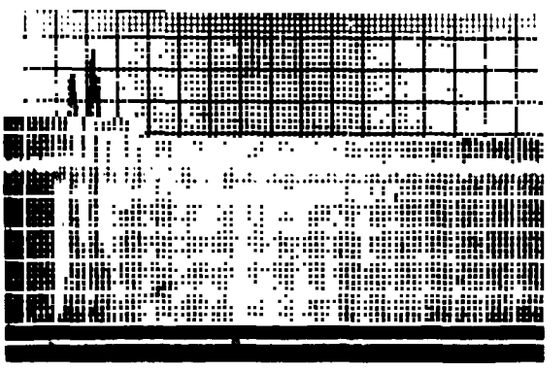


Figure C-4–X drive gear accelerometer response to impulse excitation. Excitation in X, response in X. Tape playback at 1-7/8 ips, chart speed 20 cm/sec.

Spectrum analyses have been run on the X angle response from only two accelerometers at present. These are shown in Figures C-5, C-6, C-7 and C-8. They indicate possible natural frequencies of 2.5, 7.9, 9.2, 11.7, 12.7, 13.3, 13.8, 14.2, and 14.5 cps from the feed box response and 2.5, 5.6, 6.1, 7.6, and 9.2 cps from the X drive gear response. Certainly not all of these represent natural frequencies of the system; some will be due to extraneous electrical noise in the instrumentation. One way of distinguishing between the natural frequencies of the system and the noise signals would be to pass the response through a very selective bandpass filter and examine the waveform.

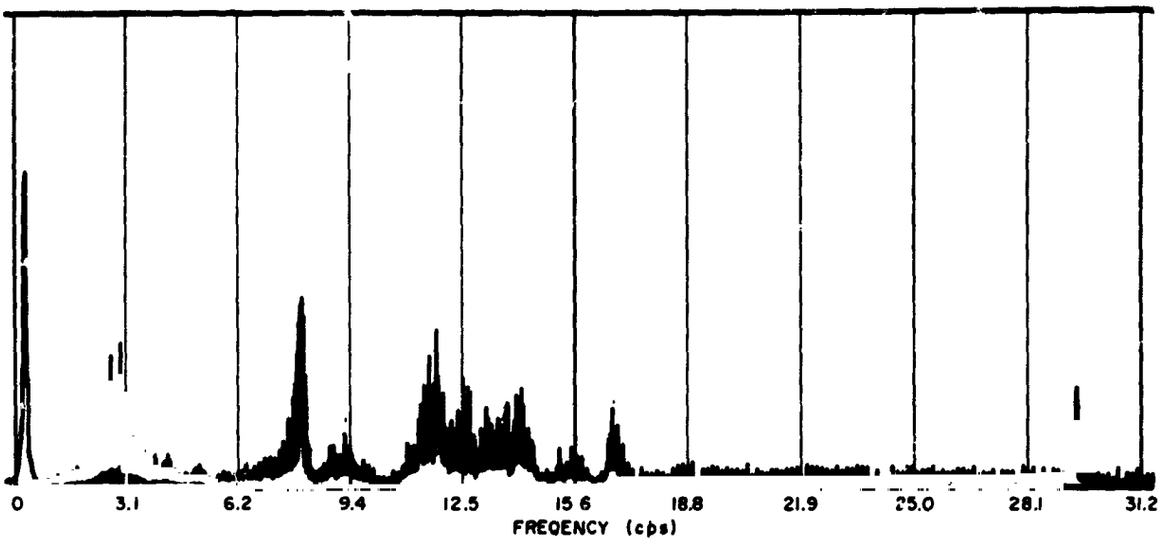


Figure C-5–Spectrum analysis of feed box accelerometer response to impulse excitation. Excitation in X, response in X.

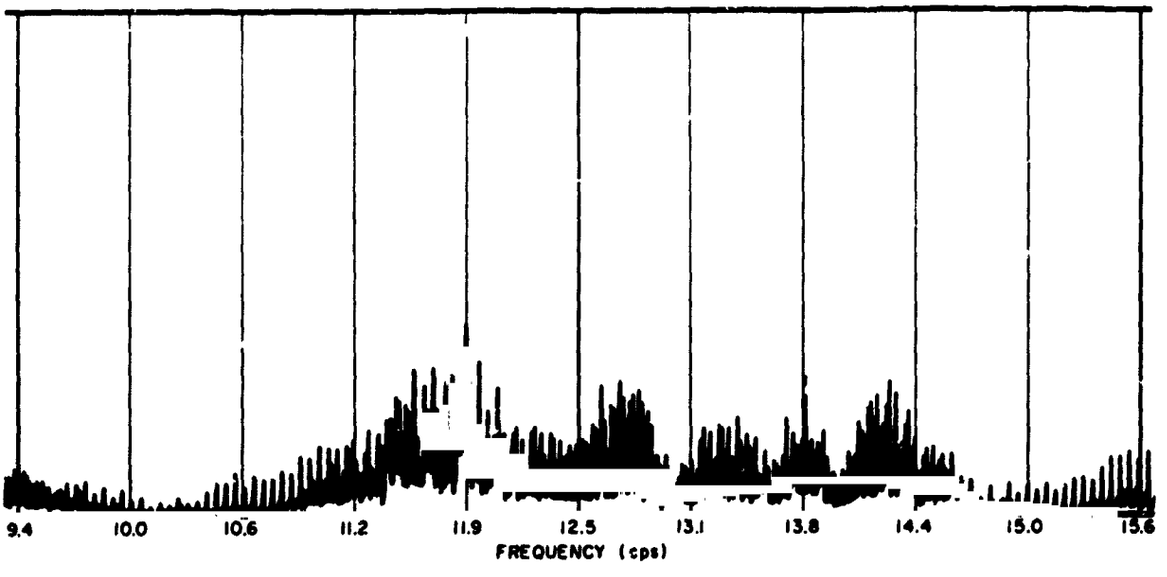


Figure C-6–Same response as Figure C-5. Looking at a narrow portion of the frequency spectrum.

associated with each frequency. The natural frequency should show up as a damped sinusoid, while the noise signal should be an undamped sinusoid. This has been done for some of the above indicated frequencies, as shown in Figures C-9, C-10, C-11 and C-12. Damped frequencies of 2.5 and 7.8 cps for the feed box and 2.5 and 5.7 cps for the X drive gear can be obtained from the filtered waveforms. It is obvious from the wave shapes obtained, however, that a more selective bandpass filter would be desirable – necessary, in fact, for the higher frequencies indicated from the feed box.

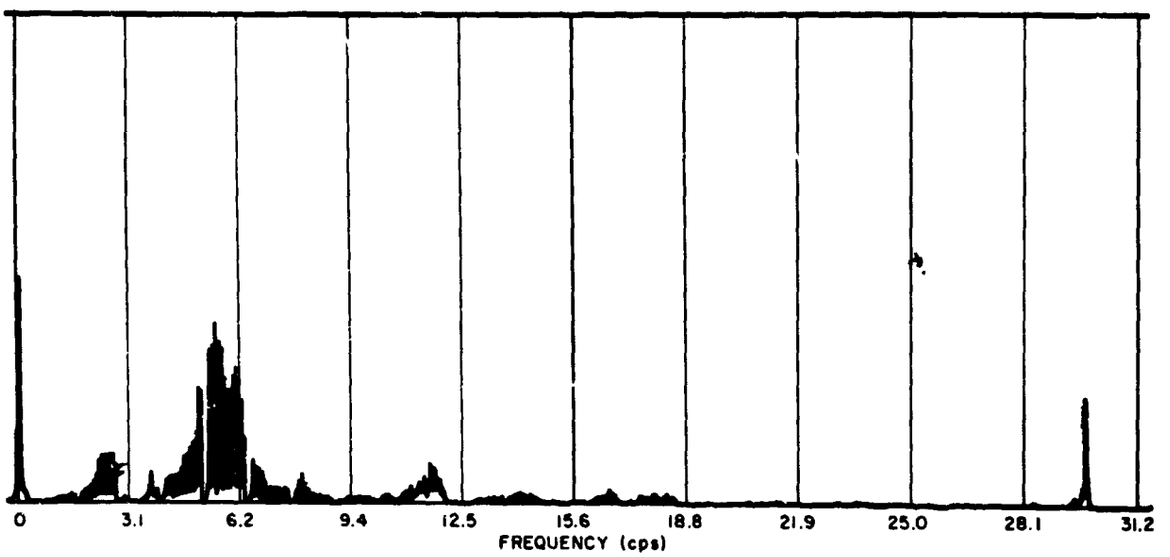


Figure C-7–Spectrum analysis of X drive gear accelerometer response to impulse excitation. Excitation in X, response in X.

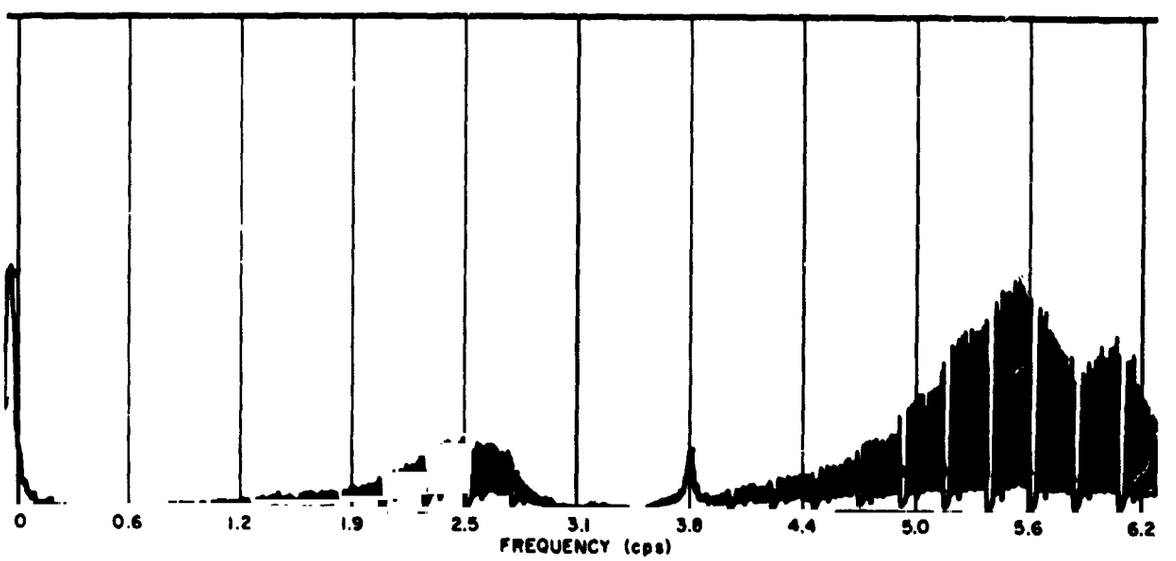


Figure C-8–Same response as Figure C-7. Looking at a narrow portion of the frequency spectrum.

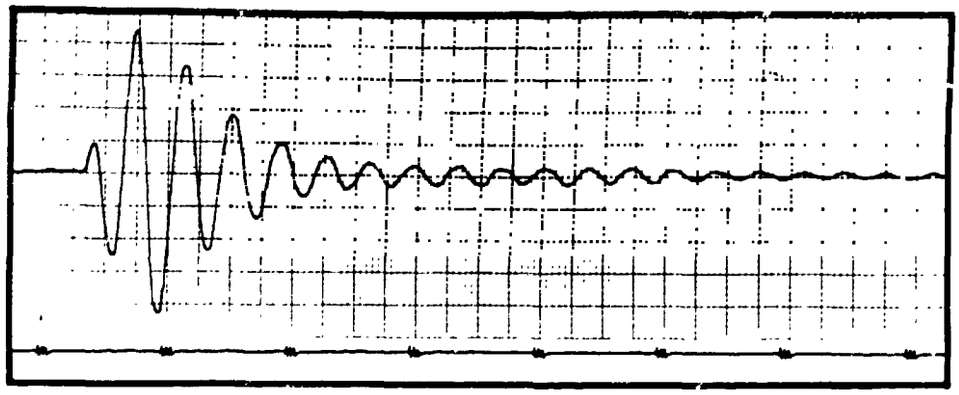
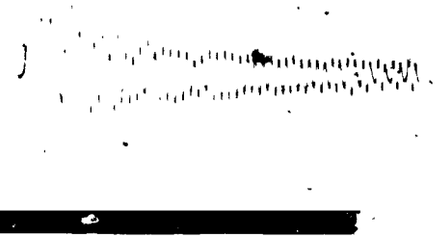


Figure C-9- Same response as Figure C-3, but through a bandpass filter, 1.5-3.5 cps.

Figure C-10–Same response as Figure C-3, playback at 60 ips through bandpass filter 160-260 cps.



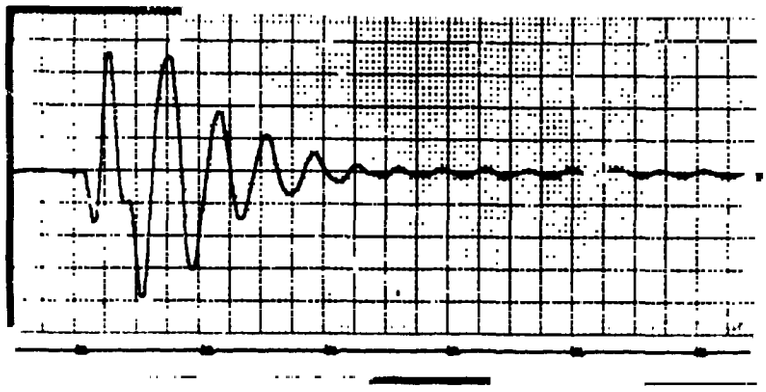
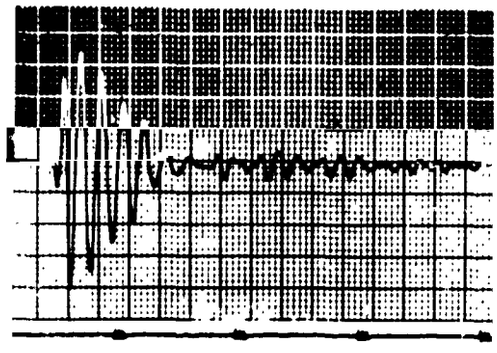


Figure C-11–Same response as Figure C-4, but through a bandpass filter, 1.5-3.5 cps.

Figure C-12–Same response as Figure C-4, but through a bandpass filter, 7-9 cps.



With the antenna in the horizon position (X = 0, Y = -70°), the fundamental natural frequency appears to be slightly higher (2.8 cps); however, the difference may be due to lack of precision in measuring the unfiltered waveforms from the strip charts.

For Y angle excitation in zenith position, a fundamental Y angle response of 2.5 cps throughout the structure is indicated on the strip charts, with a strong 14 cps component above the Y axis and a strong 5 cps component at the Y drive gear and at the hub.

In one horizon position (X = 0, Y = -20°), the fundamental Y angle response of 2.5 cps is evident at Y drive gear, hub, and feed box with a strong 5 cps component at the Y drive. Other frequency components are not distinguishable on the strip charts presently available.

In the other horizon position (X = +75°, Y = 0), Y excitation produces a completely different vibrational mode than any of the previous excitations; the antenna supporting tower is subjected to a twisting action. A fundamental Y angle response of 2 cps occurs at the feed box, dish rim, and hub, with a strong 14 cps component at feed box and dish rim. A 5 cps component, which is quickly damped out, appears at the hub. The accelerometer on the Y drive gear (which is not now at the driving point) indicates a 4 cps fundamental; there appears to be also a strong 5 cps component initially which is quickly damped out.

Damping

Estimates of the damping characteristics of the antenna structure were determined from the filtered waveforms obtained from the feed box and X drive gear accelerometers, with X angle excitation and X angle response (see Figures C-9, C-10, C-11 and C-12). Table C-2 gives amplitudes, measured from positive and negative envelopes, of the 2.5 cps damped sinusoid for the feed box and the X drive gear. The envelope has an exponential appearance at first, becoming a straight line at low amplitudes. Hence, the damping appears to be a combination of viscous type and Coulomb type. The deviation from a true exponential envelope may be due to the Coulomb component present. At any rate, 0.1 would seem to be a reasonable figure for the effective damping factor for this mode.

Figure C-10, showing the second order natural frequency in the feed box response, is not sufficiently well defined to obtain a reliable figure for the damping ratio. Again, however, both types of damping appear to be present. A rather rough approximation to the damping factor was made by sketching in the waveform envelope on a picture taken with a magnified time scale. Measurements to this envelope gave amplitude ratios of 1.5, 1.3, 1.17, with corresponding values of .06, .04, .02 for the damping factors. In Figure C-4 the second order frequency of the X drive gear, there appears to be only Coulomb type damping. In both cases, a more selective bandpass filter might give more definite results.

Transfer Function Description

Due to time limitations, preliminary results were obtained from only one method, analog computer integration. Consider the relation $C(s) = G(s)\,R(s)$, where $C(s)$ is the Laplace transform of the system output, $G(s)$ is the transfer function, and $R(s)$ is the Laplace transform of the signal input. If the input is a unit impulse, then $R(s) = 1$ and $G(s) = C(s) = \int_0^\infty h(t)\, e^{-\sigma t}\, dt$, where $h(t)$ is the transient response to a unit impulse input.*

*See A. R. Teasdale, Jr., "Get Frequency Response from Transient Data by Adding Vectors," Control Engineering, Volume 2, 1955, Pages 56–59.

Table C-2
Damping Factor X angle Response to X angle Impulse Excitation
Fundamental Mode, 2.5 cps

FEED BOX						
Impulse No. 1						
Amplitude	Amplitude Ratio	Damping Factor	Amplitude	Amplitude Ratio	Damping Factor	
21.2	2.14	.12	15.1	2.19	.12	Viscous and Coulomb Damping
9.9	2.02	.11	6.9	2.3	.13	
4.9	2.58	.15	3.0	2.31	.13	
1.9			1.3			
1.1			0.7			Coulomb Damping
0.9			0.5			
0.7						
Impulse No. 2						
20.4	2.26	.13	14.2	2.29	.13	
9.0	1.92	.10	6.2	2.22	.13	
4.7	2.61	.15	2.8	2.8	.16	
1.8			1.0			
X DRIVE GEAR						
17.5	1.95	.11	15.3	2.04	.11	
9.0	1.64	.08	7.5	2.08	.12	
5.5	1.97	.11	3.6			
2.8						
Damping factor = $\ln X_1/X_2/2\pi$						

For steady-state information, it is sufficient to obtain the modified transfer function $G(j\omega) = \int_0^\infty h(t)\, e^{-j\omega t}\, dt$. For any given frequency ω_0, $G(j\omega_0)$ can be considered a vector with X component $A_0 = \int_0^\infty h(t) \cos\omega_0 t\, dt$ and Y component $B_0 = -\int_0^\infty h(t) \sin\omega_0 t\, dt$, then $G(j\omega_0) = G_0\angle-\phi_0$, where $G_0 = \sqrt{A_0^2 + B_0^2}$ and $\phi_0 = \arctan B_0/A_0$. A plot of G vs. ω and ϕ vs. ω will describe the modified transfer function $G(j\omega)$. An analog computer program was devised to obtain A and B and convert them to polar coordinates G and ϕ_1 using as $h(t)$ an accelerometer response to an impulse input. A log-log plot of G vs. ω for the X angle response at the feed box to an X angle impulse input is shown in Figure C-13. With the detail so far obtained, this preliminary plot is consistent with previously determined information.

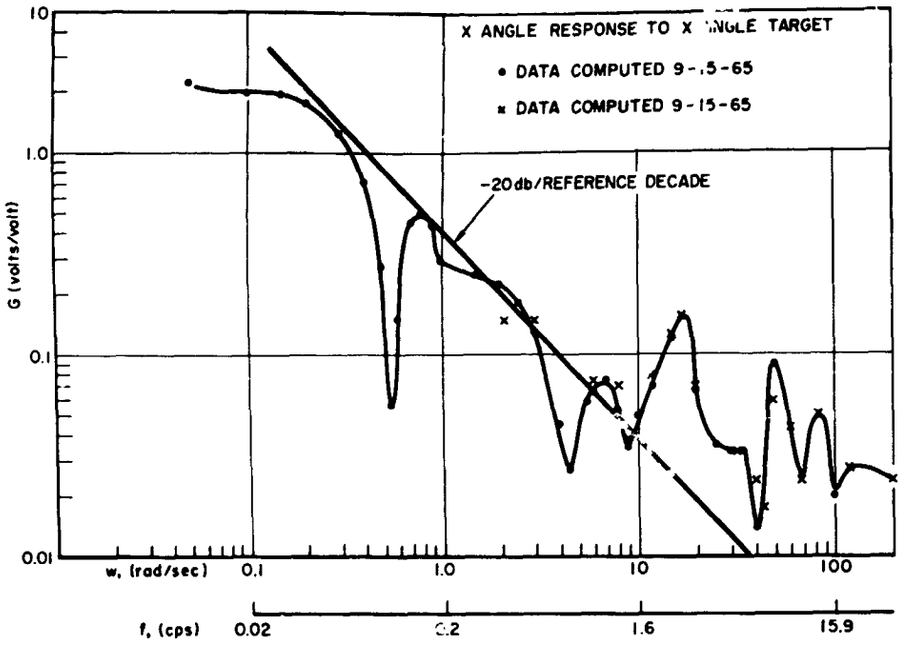


Figure C-13–Frequency response of feed box, gain vs. frequency.
X angle response to X angle input.

Natural frequencies of 2.7 and 8.0 cps damping factors of 0.1 at 2.7 cps and 0.05 at 8 cps. In addition, there is a resonance indicated at 13.5 cps and antiresonances (low amplitude feed box reaction at .09 and .72 cps).

Additional Information Desired

1. Spectrum analysis of data from all antenna positions at each excitation mode to correlate strip chart estimates of fundamental frequencies and to evaluate higher mode frequencies.

2. Complete comparison of fundamental frequencies at various points of the structure for all antenna positions at each excitation mode.

3. Determine and compare damping factors for the various modes of vibration.

4. Determine feed box/input transfer function for each mode.

5. Using correlation analysis equipment, determine block transfer functions from white noise response.

6. From accelerometer outputs, determine antenna dish distortion and vibration modes. Set up a digital computer program using this information to obtain the antenna boresight and determine boresight/input transfer function.

7. Prepare suitable input for summer workshop digital computer representation of antenna structure and compare computer response to experimentally obtain response.

8. Examine data for evidence of system nonlinearities.

9. Determine the significance of the phase reversal of the tachometer output over the frequency range of sine wave excitation used.

10. Investigate the feasibility of using two accelerometers, one at each end of a diameter of the drive gear, or one accelerometer and a synchro output, to provide automatic correction of wind forces.

11. Examine data for possible sources of tracking errors.

12. Evaluate lock-on test data and film.

Discussion

The natural frequencies as determined from the strip charts are necessarily approximate, because of the difficulty in measuring accurately with several mode frequencies and various noise signals superimposed. And because of this superposition, not all the natural frequencies of the system are evident from the strip chart response - as indicated by the few spectrum analyses and filtered responses already obtained.

On the filtered responses so far obtained, there appears to be some distortion of the damped sinusoid, possibly due to incomplete frequency separation. On some of the strip chart responses and, to a lesser degree, on some of the filtered responses, there appears to be some variation in frequency. In some instances, it appears that one frequency is present initially and that this frequency decreases slightly with time. This could be an indication of nonlinear stiffness of the structure, or it could be due to a superposition of frequencies. Because of these, and questions previously raised, it would seem desirable to obtain a more selective bandpass filter.

If a system is linear, the transient response at any point of the system will consist of a superposition of all the natural frequencies of the system, with a specific amplitude associated with each frequency. Depending upon the stiffness - mass distribution of the system, one part of the structure may interact with another in such a way that in the response of a particular point, some of the amplitudes will be zero. A comparison of the frequency spectra for the feed box response and the X drive gear response, given on page C-11, shows only three common frequencies. The combined spectra give a more complete representation of the actual frequency spectrum of the structure than either one above. One might consider that different parts of the structure filter out, or do not respond to, portions of the system spectrum. It almost seems that points above the motion axis respond to higher frequency components; those below the motion axis respond to lower frequency components of the system spectrum.

The analog computer program described on page C-16 provides phase angle information as well as gain. However, there was some question about interpretation of results and time did not permit getting this straightened out and the phase information plotted. Some of the frequencies appearing on the spectrum analysis of Figure C-5 did not show on the G vs. ω plot obtained from the computer (Figure C-12). This may be because ω values were not taken close enough together to indicate these resonances, or it may indicate that these are noise frequencies. One could expect the noise frequencies to show up as sharp spikes on the G vs. ω plot.

Two additional methods of obtaining the G vs. ω plot would be worth investigating. One method would be to perform the operations described on page C-16 with a digital computer rather than with the analog computer. Once the program is made up, results can be obtained with a very close spacing of ω values in a short time with no difficulty; and the gain and phase diagrams can be plotted automatically. The program can be filed away and used later on any other experimental work.

Another method would make use of a spectral density analyzer and the response to a white noise input. This may provide a quick and easy way of obtaining a G vs. ω plot over the entire frequency spectrum, and deserves investigation.

An accelerometer responds to motion in one direction. On the antenna, due to its being mounted on a member which revolves about an axis, the accelerometer output $C(t)$ consists of the sum of an output $M(t)$ due to rigid body rotation and a vibration response $S(t)$, i.e., $C(t) = M(t) + S(t)$. For an impulse input, the rigid body rotation is negligible and $C(t) = S(t)$. For low frequency square wave and sine wave inputs, however, $M(t)$ is significant. For the square wave input, $S(t)$ was obvious as a transient siding on a very low frequency sine wave $M(t)$. For the sine wave input, $M(t)$ and $S(t)$ combined to give results that were at first quite confusing.

It was suggested that an accelerometer might be used to indicate wind forces and provide a means of automatic compensation. To do this, however, the $M(t)$ component would need to be eliminated. This might be done by mounting two accelerometers diametrically opposite and adding their outputs such that $M_1(t)$ and $M_2(t)$ cancel, leaving a combined response $C(t) = S_1(t) + S_2(t)$. As already noted, S_1 and S_2 will be different responses; this may be advantageous or it may lead to difficulties. Another possible way of eliminating $M(t)$ might be by means of a synchro output combined with the accelerometer output.

ACKNOWLEDGMENTS

The writer would like to acknowledge the indispensable help of Mr. G. C. Winston, Mr. N. A. Raumann, Mr. W. H. Long, and Mr. A. Abbott of the Antenna Systems Branch. Detailed information on test procedure and instrumentation, as well as original data information, is available through either Mr. Winston or Mr. Raumann.

BLANK PAGE

ACTIVE METHODS FOR CONTROLLING THERMAL DISTORTION IN STRUCTURES – A FEASIBILITY STUDY

R. O. Woods

In previous orbiting observatories, structural distortion from uneven heating has been controlled only by passive means. Such means usually took the form of a design which was reasonably insensitive to temperature gradients, and extensive application of thermal insulation. This approach is successful when applied to small structures carrying experiments which are tolerant of structural deformation.

In the case of the Advanced Orbiting Solar Observatory (AOSO), the limit of capability of passive techniques has nearly been reached. This is true for two reasons, the first being the large size of the structure. Secondly, the design of the spaceframe as a "flying optical bench" and the extreme pointing accuracy of the sun tracking system impose stringent tolerances upon deflection. As an indication of the sort of quantities involved, we note that the accuracy of the pointing system is within two seconds of arc, but a temperature gradient of one degree across the structure for its entire length would cause an angular difference between opposite ends of approximately ten seconds.

It is the object of the work outlined in the following pages to investigate the feasibility of reducing structural deflections by utilizing active (i.e., power consuming) means. The general approach has been that of attempting to maintain the temperature distribution within tolerable limits, rather than to sense deformation of the structure and compensate it directly.

The problem of temperature control is rendered tractable by the fact that the instrument packages allow for axial growth of the structure. This being the case, the structure need not be compensated for axial variations in temperature or for absolute temperature level; compensation need only be applied for differences in temperature occuring across the structure. This considerably reduces the power requirements by making it necessary to compensate only for the side-to-side unbalance in heat transfer at a given cross section, rather than to balance the entire heat flux.

The portion of the AOSO structure to which temperature control is to be applied is shown schematically in Figure C-14. It is a hexagonal structure one hundred inches in length and approximately fourteen inches across the flats. Construction is of chemically milled aluminum sheet, .050 inches in thickness near the center and tapering toward each end. To the hexagonal structure are mounted a number of experiment packages and the fine sun sensors (FSS). The structure is surrounded by insulation, which is penetrated by struts supporting the FSS and experiments. During flight, the entire assembly is positioned within a cylindrical housing by the three pylons. Since the FSS controls the orientation of the entire spacecraft, it is clearly essential that misalignment between it and the instrument packages be held to a minimum.

Heat flux into the structure has been investigated by Republic Aviation using an extensive computer program. T. W. Jarrell* has analyzed the results in order to establish the magnitude of the imbalance in radial heat flow at various cross sections. It is found that heat is transferred in and out of the structure in three general modes:

*Devices For Controlling Thermal Gradients in Spacecraft Structure

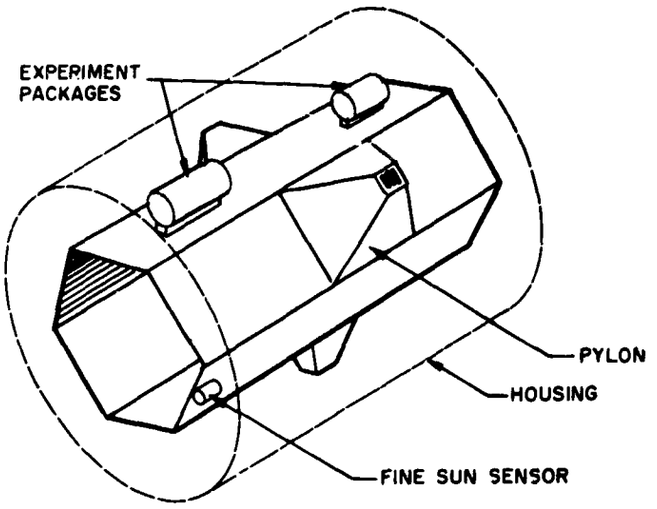


Figure C-14–The portion of the AOSO structure requiring temperature control.

a. Radiation from the housing.

This is small, amounting to approximately a 0.12 watt maximum after steady state has been established.

b. Conduction along the pylons.

This is the largest source, amounting to a sinusoidal flux of 0.8 watt maximum amplitude after steady state has been reached.

c. Conduction from instrument packages.

This has been limited to 0.025 watt per package as a design condition.

It is seen that the net heat flow for which compensation must be provided is on the order of one watt.

As a design objective, it has been decided to attempt to limit temperature variations to .1°F maximum at any given cross section.

The first approach which suggests itself is the use of a number of pad type heating elements applied directly to the structure and used to raise the temperature of the cooler zones whenever an excessive temperature gradient exists. This is the most direct approach, and the greatest part of the present effort has been devoted to a study of the means of implementing it. It is by no means the only solution however, and a number of other solutions have also been investigated.

At the outset before the magnitudes of the heating requirement had been determined, it was thought that an investigation of fluid heat exchangers might be in order. Such devices, while practically unheard of in space application, would have the advantage of moving large quantities of heat for a relatively small price in pump power.

When it was recognized that the power required in the form of heat was small when compared to the load capacity of the solar cells (one watt vs. four hundred), the idea was discarded because of its complexity and probable low reliability. As will be mentioned

later, however, fluid circuits have advantages not immediately obvious and warrant consideration for other reasons.

CHOICE OF ACTIVE ELEMENTS

Area Heating

Assuming that it has been determined to use pad-type heating elements applied directly to the structure, the question arises as to a basis upon which to choose the optimum number of separate pads. This may be approached by noting the effect upon the structure of an axial variation in the side-to-side temperature difference. If we assume that the structure is divided into μ axial bands, each provided with temperature sensors located in the center of the band on each of six faces, and heaters programmed to reduce the temperature difference between opposed sensors to zero, the situation is as shown in Figure C-15, where μ has been taken equal to 3. The important point to note is that, even given a feed-back system capable of reducing the error signal (difference in temperature readings) to zero, the compensation can only be perfect in the vicinity of the sensor. In an extreme case of high heat input rate and small axial conductivity, temperature differences inversely proportional to the first power of the number of bands might exist.

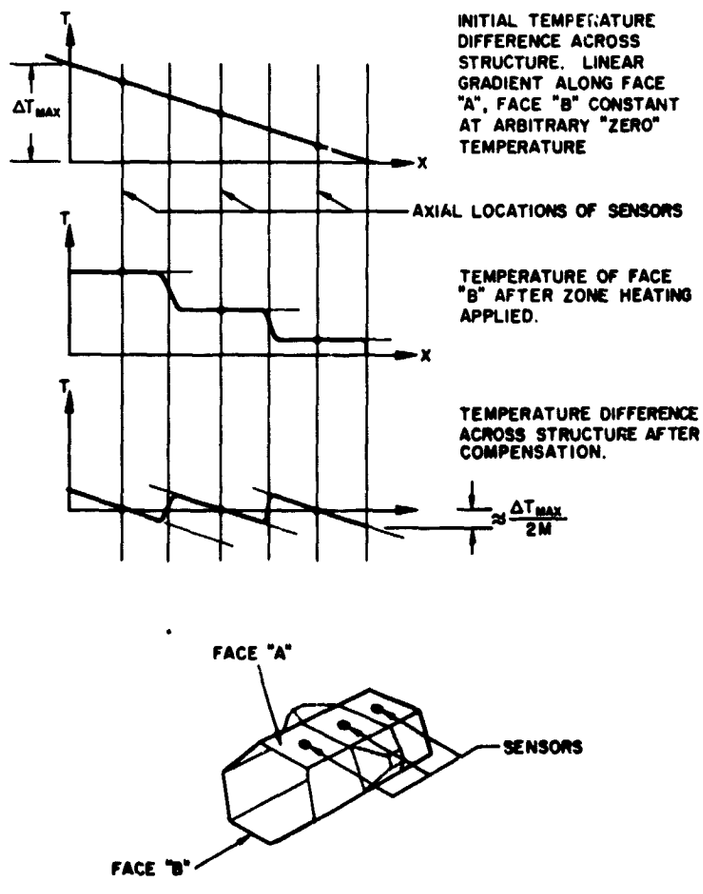


Figure C-15–Temperature variations at sensors before and after application of zone heating and compensation

In Appendix A, an expression for the curvature at any point on the neutral plane due to a temperature variation in the corresponding cross section has been derived. This is a general expression and may be integrated to determine the shape of the elastic curve corresponding to any given temperature distribution. This has been done for the sawtooth waveform shown in Figure C-15 in order to evaluate maximum slopes for an arbitrary number of heated zones.

Thermoelectric Cooling and Heating

Early in the course of this investigation it was recognized that some means of cooling warmer areas would be a desirable complement to the apparatus used for heating the cooler areas. In principle this might be accomplished through use of thermoelectric cooling equipment, and an investigation of the state of the art with these devices has been undertaken by Jarrell.*

It appears unlikely that thermoelectric cooling devices applied directly to the structure would provide a satisfactory means of temperature control for two reasons. The first, that their small size would make it impossible to heat or cool a large area of the structure uniformly. Secondly, the hot and cold surfaces of these devices are normally only a few centimeters apart. This means that instead of removing heat from the warmer portions of the structure and rejecting it at the cooler, it would be necessary to dump any heat extracted from the structure by radiating it to the housing. Doing so would entail penetrating the structural insulation to allow one side of the thermoelectric device to look out. Such an arrangement would constitute a low-resistance leakage path for heat transfer and would probably necessitate constant operation of the device to null the heat flow which would otherwise take place.

It would appear that the characteristics of thermoelectric devices render them most suitable for applications where the temperature of relatively small areas is to be controlled. As it happens, there are portions of AOSO where just such temperature control would be of value. Since the majority of the heat flow in or out of the structure is by conduction along the pylons, temperature control could be accomplished by "short circuiting" this flux at the outboard end of each pylon. By interposing a thermoelectric device between the housing and the pylon end, the pylon could be artificially maintained at the same temperatures as the structure and the heat flux reduced to zero. This also allows larger tolerance on control accuracy in that whatever error occurred would appear as a difference in temperature between the ends of the pylon. As this is a fairly long structure, the temperature gradient and heat flow would be correspondingly small.

Thermoelectric devices could also be used to prevent conduction from the instrument packages, through their supports, to the structure.

TEMPERATURE SENSING

Since the problem at hand is one of controlling the geometry of the structure, the most direct method of generating an error signal would be to sense distortion by use of strain gauges. We have chosen not to do this, but rather to limit the investigation to means of sensing temperature. This was a somewhat arbitrary decision made for the following reasons:

a. A large number of strain gauges would be required to characterize the shape of the elastic curve and, even granted that this information could be obtained, programming the heaters to compensate would not be a simple matter.

*See pages C-39–C-54

b. Skepticism as to the ability of strain gauges to provide reliable data on deflections of such small magnitude for periods up to a year in the absence of independent recalibration.

c. The fact that the gauges are themselves temperature sensitive.

It is conceivable that an optical system could be devised to sense deflections directly. This would have the advantages of measuring the controlled variable (geometry), rather than its antecedent (temperature), and of circumventing the difficulties associated with the more conventional surface strain gauges. The problem of heater programming would still exist, however.

Three methods of temperature sensing have been considered. There appear to be no others possessing the required precision and long-term stability. The instruments under discussion are thermocouples, thermistors, and quartz thermometers. The limitations of each will be described at length below.

Thermocouples

At first glance, the thermocouple seems particularly well suited to the problem. It possesses the great asset of being intrinsically stable over long periods, the only requirement being that the thermocouple wire be of sufficiently large gauge that any chemical changes taking place due to reactions with the insulation (or other environment) be confined to a superficial portion of the wire. A great practical advantage is the fact that thermocouples sense only the difference in temperature between their junctions; this means that instead of measuring two absolute temperatures and subtracting, we can read directly the temperature difference between opposite faces of the structure.

The great, probably conclusive disadvantage of thermocouples is their very low change in signal voltage with temperature. Assuming that a chromel-constantan pair is best suited to our purpose,* we find that the change in emf for our range of temperatures is on the order of 0.03 millivolts per degree Fahrenheit (1). Our aim is to maintain temperatures in the structure within 0.1°F, therefore the temperature sensor should resolve accurately to about 0.025°F. This corresponds to a signal level on the order of a microvolt, which is the noise level in any but the most carefully designed thermocouple circuitry. Disregarding the question of signal-to-noise ratio, we find that accurately sensing such small voltages with flight instrumentation will be an extremely difficult problem. In the opinion of men working in the flight electronics field,† it will be close to – if not beyond – the limit of present day technology to attain the required accuracy. Stability is regarded as less of a problem.

Thermistors

Thermistors are semiconductor devices which have the property of varying greatly in resistance with changes in temperature. A typical unit might show a three or four percent change per degree Fahrenheit near room temperature. Because of the large temperature coefficient, signal strength will be high and designing flight electronics capable of the required accuracy should not be a particularly difficult problem. As an indication of the present state of the art, note that commercial equipment is available off-the-shelf with the capability of 0.1°F accuracy on absolute temperature readings. The ultimate

*Recommended by Mr. D. Thomas, Temperature Section, National Bureau of Standards.
†D. Hepler, Spacecraft Electronics Branch, Spacecraft Technology Div., GSFC, R. Muller, Electronics Instrumentation Section, Advanced Development Div., GSFC.

accuracy on absolute readings is on the order of 0.005°F.* We are interested in differential temperatures, which are substantially easier to measure, and should have no difficulty in attaining the desired sensitivity of 0.025°F.

Unfortunately, thermistors are not without their own peculiar limitations. In this case, the problem is one of long term drift. Comparatively little work has been done on thermistor aging at constant temperature, although there are numerous reports in the literature on the effects of temperature cycling through large ranges.

Reference (2) presents the results of a series of tests conducted by the National Bureau of Standards for one of the leading thermistor manufacturers. Here, the resistances of five nominally similar thermistors were measured at several different temperatures at four, sixteen, and twenty-eight months after manufacture. None of the measurements were made at room temperature, but by interpolating between data taken at 32°F and that taken at 104°F an estimate can be made. It is seen that the greatest difference in temperature indicated by any two units is on the order of 0.01°F at the end of the first twelve month period. If the units were then reset to the same zero, the maximum disparity at the end of the following year would amount to roughly 0.04°F.

Since the thermistors mentioned above were not especially treated or matched, it seems reasonable to suppose that better stability can be obtained using "pedigree" components. In any case, the variation in stock units is not drastically out of our tolerance range.

An unfortunate aspect of the thermistor drift problem is the fact that it is impossible to recalibrate the sensors in flight. The electronics can, of course, be checked against a standard resistance but anomolous performance of a thermistor would be indistinguishable from a temperature change. The problem can be greatly mitigated by installing redundant units at each sensor location. This would have the twofold advantage of giving better accuracy by making it possible to average the readings of a number of units, and any appreciable differential reading would be a positive indication of trouble.

In summation, it would appear that there are no insurmountable problems associated with the application of thermistors, and that their large temperature coefficient makes them appear quite attractive.

Quartz Thermometry

A comparatively obscure branch of thermometry utilizes the change in frequency of an oscillating crystal as an indication of temperature variation. As a technique, this approach appears to have great advantages for space-borne applications. Sensitivity is high,† and the necessary circuitry is conceptually simple. Unfortunately, as indicated in (3, p. 6), some types of crystals show an appreciable monotonic long term drift amounting to as much as 0.01°C in thirty days. Whether this drift is predictable, or could be reduced in the case of differential rather than absolute readings, is not presently known. At least one private concern is currently investigating the problem.‡ Assuming that the drift problem can be solved, these devices are potentially more sensitive and reliable than thermistors.

*R. Harruff, YSI Components Div., Yellow Springs, Ohio

†A recently introduced commercial unit (Hewlett Packard model 2800A) is advertised as having 0.0002° F reproducibility for short terms.

‡Ralph Manies, Hewlett Packard, Palo Alto, Calif., Private Communication.

Control Problem

Use of a proportional feed-back control system will probably not be feasible because, during normal operation, the largest acceptable temperature differences will be on the same order as the limit of the sensors to discriminate. We suggest a relay (on-off") control system (4), which has the added advantage of simplicity.

The design of such a control system should be a reasonably straightforward problem, although experimental evaluation of the controlling parameters will be in order.

The essential requirements upon the system are that temperature fluctuations during steady-state be held to a minimum, and the system respond in an acceptably short time to externally caused temperature variations. We will assume that resonance effects between external temperature fluctuations at orbital period, and system response will be negligible because of the high natural frequency of a system capable of the necessary speed of response.

Design of the control system will be complicated by the fact that temperature differences, not absolute temperatures, are to be controlled. Variations in temperature level affect the rate at which heat is exchanged with the environment, thereby making it difficult to choose an optimum value for the hysterisis (difference between switch-on and switch-off temperatures) built into the system.

Attention will also have to be given to eliminating spurious feed-back loops. For example, a local temperature rise might lead to a heat input elsewhere which, through a series of intermediate stages, resulted in a further temperature increase at the original point. The recognition of such circuits would pose an extremely involved problem in analysis and might well exceed the capacity of any computer. For this reason, much of the design will have to be arrived at by empirical methods using the results of space simulator tests.

FLUID COOLING LOOPS

We have established that the most difficult to implement and least reliable element of an active control system is the temperature sensor. It is interesting to speculate upon the possibility of designing a system which would in effect enhance the thermal conductivity of the structure in the circumferential direction to such an extent that intolerable temperature gradients could not exist at any reasonable value of heat input. This might be accomplished through the use of a series of closed fluid loops surrounding the structure. In Appendix B, such a circuit is investigated and it is found that for the geometry chosen as an example, a single loop can maintain a temperature difference across the structure of less than 0.1°F with a net heat flux of 0.07 watts. The weight of such a loop would be on the order of 0.16 pounds exclusive of circulating pumps. Pump power required at 1 percent efficiency would be on the order of 0.015 watts. It appears that ten loops would be more than adequate to compensate the entire structure at an expense of 1.6 pounds plus the weight of the pumps, and 0.15 watts.

The problem of choosing pumps for this application presents an extremely fertile ground for investigation. That the system must be completely sealed goes without saying. In addition, it must operate reliably for a year and generate no vibration or electrical noise. This suggests a device with no bearings or components subject to fatigue failure – preferably a device with no moving parts.

Somewhat alleviating the stringency of the specifications that the pumps must meet is the fact that power levels are extremely low. This makes it possible to consider some of the more exotic means of moving fluids, including those with relatively low efficiency.

An investigation has been made of two types of pumps which have features well suited to our needs. Of these, one has characteristics which would probably preclude its use. The other would bear further investigation.

Electromagnetic Pumps

In these devices, force is generated internally within a conductive fluid by the interaction between an electric current and a magnetic field (6).

Although they show considerable variation in geometry, electromagnetic pumps have the same basic components in common. These are: a magnet, a duct to convey the fluid through the magnetic field in one or more passes, and electrodes positioned so that a current can be passed through the fluid in a direction normal to the magnetic field. This, according to the classical equation, $\vec{F} = \vec{I} \times \vec{B}$, generates a body force within the fluid which may be used to pump it directly. There are no moving parts in the system other than the fluid itself.

Because current passing through the fluid is essential to the operation of electromagnetic pumps, they are particularly well suited to handle liquid metals. In fact, the first wide-spread application was pumping liquid sodium for nuclear reactor cooling. Unfortunately, our application requires a liquid of high thermal capacity and fairly low thermal conductivity. This puts metals (specifically mercury) at an immediate disadvantage. Calculations were performed to investigate the feasibility of using an electrolyte as a working fluid. It appears (Appendix C) that, because of the high resistivity of any solution compared to a metallic conductor, approximately one watt per circuit would be dissipated at current levels sufficient to produce the necessary pressure rise. This, of course, is far too high to warrant further consideration. Additional calculations were performed assuming mercury as a working fluid and it was found that, because of the low specific heat, weight requirements would be excessive. We conclude that the outlook for electromagnetic pumping in this application is not encouraging.

Ion Drag Pumps

These devices are still in the developmental stage, but appear to have a number of features which are well suited to our problem. As with the electromagnetic pumps, they have no moving parts. In direct contrast however, they will only work with fluids which are good insulators. Efficiencies are not normally high – being on the order of ten percent – which probably accounts for the fact that they have had little practical application to date. Working fluids can be hydrocarbons having a specific heat of 0.3 or higher.

Reference (5) in addition to giving extensive source material, describes an experimental unit which delivers 8 cc of kerosene per second against a head of one meter with an efficiency of ten percent. This is the order of magnitude of the flow rate required for our purpose, and a pressure rise an order of magnitude larger. It has been mentioned elsewhere that power requirements would not be excessive even for efficiencies of one percent or less.

The active elements of an ion-drag pump are completely immersed in the fluid and consist of an electrode (normally pointed) positioned upstream of an orifice. When a sufficiently high potential difference is imposed between these components, ions are formed at the upstream electrode and attracted to the metal surrounding the orifice. In the course of their migration, the ions encounter neutral molecules of the fluid and exchange momentum. The result is a net flux of molecules through the orifice.

The only intrinsic feature of these units which might prove to be a serious handicap for our purpose is the high voltage normally employed. This ranges as high as 20 kilowatts.

No information is available regarding long term performance, but there appears to be no a priori reason to assume that operation for periods up to a year is impossible. The chief sources of deterioration seem to be ablation of the electrodes and chemical changes within the working fluid. Both of these can be reduced, at the expense of efficiency, by operating at lower voltages.

CONCLUSION

In order to summarize the results of this investigation we can make the following general statements:

1. Thermistors appear potentially suitable for fine temperature sensing, but the problem of drift remains to be solved. Experimental work will be required to establish the feasibility of matching pairs of thermistors by, for example, choosing adjacent units in the same production run or measuring drift over a short period and pairing units which have similar properties.

2. An investigation should be made into the long term stability of quartz thermometers. In many respects these devices are even more attractive than thermistors. It is likely that this problem has already been treated in connection with the use of quartz crystals as frequency standards. This implies that any investigation should begin with a literature search treating oscillators rather than thermometers.

3. It appears that thermoelectric devices can be used to advantage where heat is transferred by conduction through a small area. This suggests their use as guards on the pylon ends and between instrument packages and the structure.

4. The use of a series of fluid heat exchanger loops appears practical as regards weight and power requirements. Such an arrangement would dispense entirely with the problem of temperature sensing. We suggest that it be investigated further in order to accurately establish the heat load versus temperature gradient and to better appreciate the problem of pump development.

Appendix A
THERMAL DISTORTION EQUATION

We may, for example, write the equation of the elastic curve

$$\frac{d^2y}{dx^2} = \frac{\epsilon_x}{y}, \tag{1}$$

where ϵ_x is the unit elongation in the x-direction at the distance y from the neutral plane. Now

$$\epsilon_x = \alpha \Delta T + \frac{\sigma_x}{E}, \tag{2}$$

where α is the coefficient of thermal expansion,

ΔT is the temperature differential above some arbitrary zero,

σ_x is the stress in the x-direction,

E is the modulus of elasticity.

Equation (2) may be transposed

$$\sigma_x = E(\epsilon_x - \alpha \Delta T). \tag{3}$$

But from the planar assumption

$$\epsilon_x = G(x)\, y \tag{4}$$

$$\therefore \sigma_x = E[G(x)\, y - \alpha \Delta T]. \tag{5}$$

From statics, we know that the external moment acting on any cross section is zero:

$$\int_A \sigma_x\, y dA = 0; \tag{6}$$

therefore

$$E \int_A (G(x)\, y - \alpha \Delta T)\, y dA = 0,$$

$$G(x) \int_A y^2 dA - \alpha \int_A \Delta T\, y dA = 0,$$

and

$$G(x) = \frac{\alpha \int \Delta T y dA}{\int y^2 dA}. \tag{7}$$

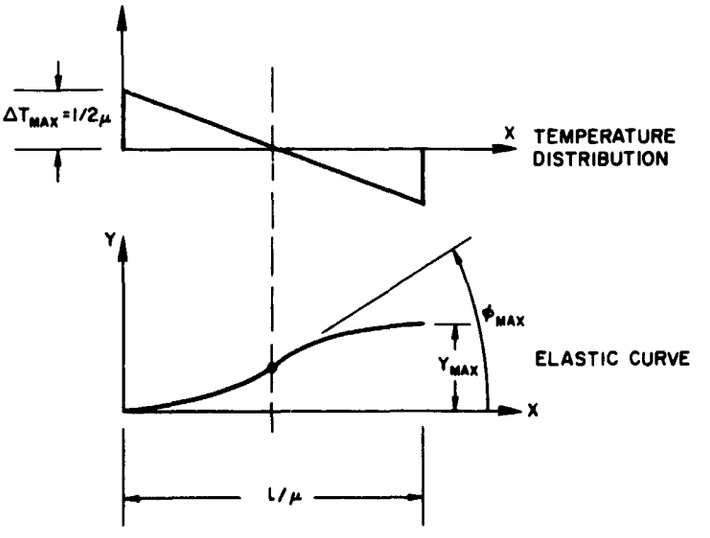
By noting that

$$\frac{d^2y}{dx^2} = \frac{\epsilon_x}{y} \equiv G(x),$$

we recognize that Equation (7) is the equation for the curvature at any point x. This may be successively integrated for the slopes and the deflections.

In order to determine the relationship between the number of heated zones and maximum possible deflections, we will analyze the case of a linear cross sectional temperature difference (ΔT) varying uniformly from zero at one end to unity at the other. This is the case shown in Figure C-15, with ΔT_{max} equal to unity.

From symmetry, the problem can be reduced to that of analyzing μ separate beams, each having zero slope at the ends. This is diagrammed below where ΔT is the temperature above a datum arbitrarily taken as zero on the cooler face and varies linearly in x and y.



The x-dependency of ΔT may be expressed by

$$\left(\frac{1}{2\mu} - \frac{x}{\ell}\right).$$

And, letting r equal the "radius" of the hexagon (i.e., one half the distance across the flats), the linear variation in y may be expressed as $[(y + r)/2r]$. The temperature variation at any cross-section may thus be expressed as a function of x and y:

$$\Delta T\,(x, y) = \left(\frac{y + r}{2r}\right)\left(\frac{1}{2\mu} - \frac{x}{\ell}\right).$$

Letting

$$P(x) = \frac{-\alpha \int_A \Delta T(x,y)\, y\, dA}{\int y^2\, dA},$$

we may write

$$\varphi_{max} = \int_0^{\ell/2\mu} P(x)\, dx$$

For a hexagon of unit thickness, using the expression for ΔT derived above,

$$\int_A \Delta T(x,y)\, y dA = 3.08 \left(\frac{1}{2\mu} - \frac{x}{\ell}\right) r^2$$

and

$$\int_A y^2 dA = 3.85\, r^3.$$

Therefore,

$$P(x) = \frac{-3.08\, \alpha r^2}{3.85\ r^3} \left(\frac{1}{2\mu} - \frac{x}{\ell}\right)$$

or

$$P(x) = .8 \frac{\alpha}{r} \left(\frac{x}{\ell} - \frac{1}{2\mu}\right)$$

and

$$\varphi_{max} = -.1 \frac{\alpha \ell}{r \mu^2}.$$

Letting

r = 7.2 inches

ℓ = 100 inches

α = 13×10^{-6} °F,

we obtain

$$\varphi_{max} = \frac{-18.06 \times 10^{-6}}{\mu^2} \text{ Radians}$$

or

$$\varphi_{max} = \frac{-3.72}{\mu^2} \text{ seconds of arc}$$

where the minus sign results from having taken the upper as the warmer surface.

Appendix B
HEAT EXCHANGER CALCULATIONS

An expression is to be derived for the change in temperature of a fluid passing through a heat exchanger tube applied to the surface of the structure. In order to determine the temperature distribution throughout a closed circuit, a trial value of fluid temperature will be assumed and the expression for temperature change evaluated at each section of the tube (node) in sequence. If the final temperature after completing the circuit does not equal the trial value, a new trial value will be assumed and the process repeated until compatibility is obtained.

Letting

T_N = temperature of N^{th} node,

T_{fNi} = fluid temperature at inlet to N^{th} node,

T_{fNo} = fluid temperature at outlet of N^{th} node,

q_N = heat flow into fluid = external heat flow into node at steady state

ΔT_{fN} = change in fluid temperature during passage through N^{th} node,

We may define a coefficient K such that

$$q_N = K \left[T_N - \frac{T_{fNi} + T_{fNo}}{2} \right],$$

where

$$| T_{fNi} - T_{fNo} | \ll | T_N - T_{fNi} |.$$

Now

$$q_N = \rho A V C_p \, (T_{fNo} - T_{fNi}),$$

where

ρ = fluid density,

A = cross-sectional area of tube,

V = flow velocity,

C_p = specific heat of fluid.

Defining

$$\rho A V C_p = F,$$

we have

$$q_N = F \, (T_{fNo} - T_{fNi}).$$

By equating the q_N's,

$$F(T_{fNo} - T_{fNi}) = K\left[T_N - \frac{(T_{fNi} + T_{fNo})}{2}\right]$$

or

$$\Delta T_{fN} = \frac{K}{F}\left[T_N - \frac{(T_{fNi} + T_{fNo})}{2}\right].$$

But

$$T_{fNo} = T_{fNi} + \Delta T_{fN},$$

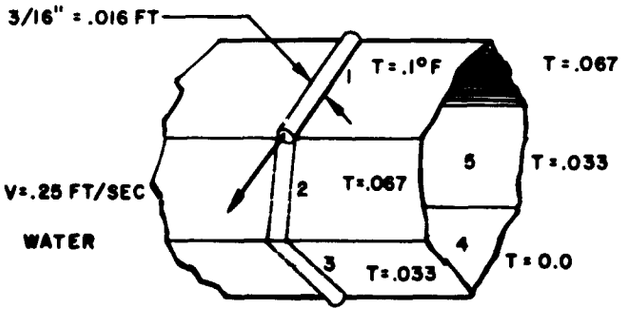
$$\therefore \Delta T_{fN} = \frac{K}{F}\left[T_N - \frac{T_{fNi} + T_{fNi} + \Delta T_{fN}}{2}\right]$$

$$\Delta T_{fN} = \frac{K}{F + K/2}\left[T_N - T_{fNi}\right]$$

which is the desired equation for the change in temperature of the fluid in passing through the node.

In order to obtain an order of magnitude estimate of the amount of heat influx to the structure which could be tolerated without exceeding the .1°F limit, we will insert reasonable values into the expression above and carry out a circuit calculation.

The assumed circuit geometry is shown below:



Noting that

$$K = hA_s,$$

where h = heat transfer coefficient BTU hr^{-1} ft^{-2} $°F^{-1}$

A_s = surface area of tube within node

we may write

$$h = 4.12\,k/d \qquad \text{(6, p. 192)}$$

where k = thermal conductivity of fluid BTU $hr^{-1} ft^{-1} °F^{-1}$

d = tube diameter ft

$\therefore K = 4.12 k / d A_s$.

And, since $A_s = \pi d \ell$ where ℓ is the length of the node,

$K = 4.12 \pi k \ell$ BTU $hr^{-1} °F^{-1}$.

From Reference (6, p. 500),

For water at 68°F: $k = .345$ BTU $hr^{-1} ft^{-1} °F^{-1}$ and $\ell = .693$ ft $\therefore K = 3.09$,

Also for water at 68°F,

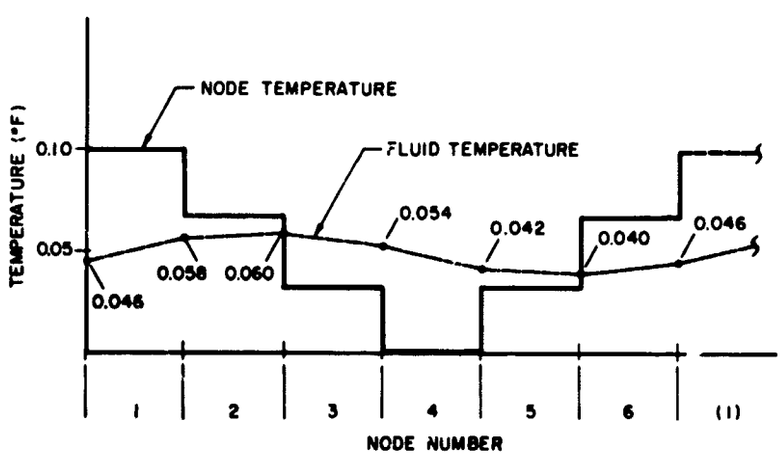
$$F = \rho A V C_p = 62.46 (2.01 \times 10^{-4}) \frac{.25}{3600} (.999)$$

$$F = 11.29 \text{ BTU } hr^{-1} °F^{-1}$$

and

$$\frac{K}{F + K/2} = \frac{3.09}{12.84} = .204.$$

Using this value in Equation (8) and performing the iterative solution, we obtain the following temperature distribution:



Having obtained the temperature distribution, it then becomes possible to evaluate the net heat flux from the warmer side of the structure, through the fluid, to the cooler side. This is found to be on the order of .07 watts (at .1°F difference in temperature).

Weight

The weight of one cooling loop, assuming fluid with a specific gravity of 1.0 enclosed within a 3/16" inside diameter × 1/32" wall aluminum tube, is calculated at .16 pounds.

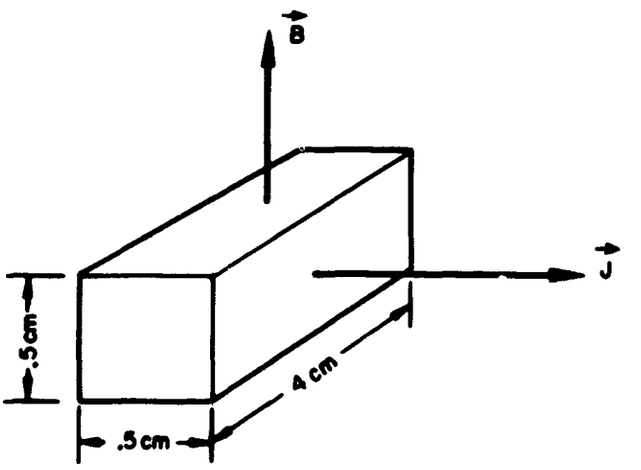
Pump Power

Using the Poisenille expression for pressure drop in laminar flow (7), and the properties of water at room temperature, we compute a pressure difference of approximately two pounds per square foot. Taking this in conjunction with a flow velocity of .25 feet per second, we compute a power input of approximately .15 milliwatt to the fluid stream.

Appendix C
ELECTROMAGNETIC PUMPING

An order of magnitude calculation is to be performed to determine the power required to operate an electromagnetic pump using an electrolytic solution as a working fluid.

Assume pump geometry as sketched below:



The equation for the force on a current carrying conductor in a magnetic field is given in (8, page 246):

$$d\vec{F} = i\, d\vec{s} \times \vec{B}, \text{ newtons}$$

where i = amperes,

S = meters,

B = webers meter^{-2},

We will assume $B = 0.3$ as a reasonable value of field intensity and compute the current required to give the necessary pressure rise of two pounds per square foot. Noting that one pound per square foot equals 478.8 dynes per square centimeter and one newton equals 10^5 dynes, we find that the required force is 0.00237 newton. We can write the scalar equation:

$$i = \frac{F}{S \times B} = \frac{.00237}{.005 \times .3},$$

$$\therefore i = 1.58 \text{ amp.}$$

Consulting the tables of (9, page 2362 et. req.), it appears unlikely that an electrolyte having a resistance of less than 1.50 ohm centimeter^{-3} can be employed. This gives a net resistance of 0.375 ohm in the geometry chosen. The power dissipated in heating the fluid is therefore equal to 0.94 watt. This is to be regarded as an approximation only and might be somewhat reduced by optimizing the configuration.

REFERENCES

1. Shenker, H. et. al., "Reference Tables for Thermocouples," National Bureau of Standards Circular 561, page 84.

2. Yellow Springs Instrument Co., "Report-400 Series Probes," YSI Drawing B-44513 Rev. A. Yellow Springs, Ohio (1964).

3. Hewlett-Packard Dynec Division, "Quartz Thermometers Models 2800A & 2801A." Palo Alto, California, Hewlett-Packard Corporation (1965).

4. Smith, O. J. M., Feedback Control Systems, page 358, McGraw Hill Book Company, New York, (1958).

5. Panholzer, R., "Electrical Engineering," Volume 32, number 2, page 128, (February 1963).

6. Echert, E. R. G., and R. M. Drake, Heat and Mass Transfer, McGraw Hill Book Company, New York (1959).

7. Binder, R. C., Fluid Mechanics, Prentiss Hall, Inc., Englewood Cliffs, New Jersey (1955).

8. Neal, J. P. Introduction to Electrical Engineering Theory, McGraw Hill Book Company, New York (1960).

9. Hodgeman, C. D., et. al., Handbook of Chemistry and Physics, 35th edition, Chem. Rubber Publishing Co., Cleveland, Ohio (1953).

10. Strietzer, O. M., Journal of Applied Physics, Volume 31, number 1, page 136 (1960)

BLANK PAGE

DEVICES FOR CONTROLLING THERMAL GRADIENTS IN SPACECRAFT STRUCTURES

Temple W. Jarrell

INTRODUCTION

General Problem

With the development of the more sophisticated spacecraft comes the requirement for better control of the thermal gradients imposed upon the spacecraft structure and upon associated experiments. The problem being investigated then, is to determine those devices and methods available for control of the thermal gradient on a spacecraft. (The problem of determining which sensing devices are needed to aid the thermal controlling devices for a particular problem has been investigated in this document, by R. O. Woods.)

For the thermal controlling devices to be feasible, they must be light-weight, be able to withstand the rigors of space such as vacuum and weightlessness, and have long term reliability (in particular, at least one year of reliability). In addition, the equipment which controls these devices must satisfy the same requirements since it too will be on the spacecraft. However, it is the requirement that the devices maintain some particular thermal gradient across the spacecraft structure which will form the basis of the report.

The most obvious method for acquiring some particular thermal gradient would be to dump heat into the spacecraft structure at those points or areas which will give the desired gradient. (Note that dumping heat at one point will affect other points in varying degrees. But for simplicity, this investigation will neglect these interactions. To obtain the actual inputs of the devices, a computer program will be needed.) However, at times it may prove more feasible to cool points or areas than to heat them. Yet even more likely, a combination of heating and cooling will be the most economical and feasible method of securing the desired thermal gradient across the structure.

Since it takes more power to pump a given amount of heat from the structure than it takes to dump the same amount of heat into the structure, the number of points that are cooled (while others may be heated) will depend upon the magnitude of the heat to be pumped. To find the magnitude of the heat to be pumped, a numerical analysis or an experiment must first be made to determine the thermal gradient imposed on the structure by passive control methods. Since the gradient is a function of time, it can be expressed instead as a heat flow as can the required gradient. Then the heat to be pumped is determined by subtracting this passively controlled heat flow from the required heat flow. So, the active devices will have the inputs as determined and will thus obtain the thermal gradient as stipulated by the experimenters.

Particular Problem

This investigation will be general wherever possible so that it can serve as a starting point for obtaining any desired thermal gradient across a spacecraft structure. But much of the report will be devoted to the particular problem of a zero gradient (i.e., uniform temperature) over the structure of the AOSO. The Republic Aviation Corporation used a computer program to furnish temperature profiles (due to passive thermal control methods) of the AOSO support structure for three different cases. The first case was

where the cylindrical housing surrounding the support structure was allowed to radiate to the structure. Thus inducing a temperature profile. Next the pylons (see Figure C-16) which connect the support structure to the housing, were allowed to conduct heat either to or from the housing structure. The third case arose when 1/2 watt was dumped into the support structure at one node to simulate the heat dumped into the structure by an experiment mounted on the structure.

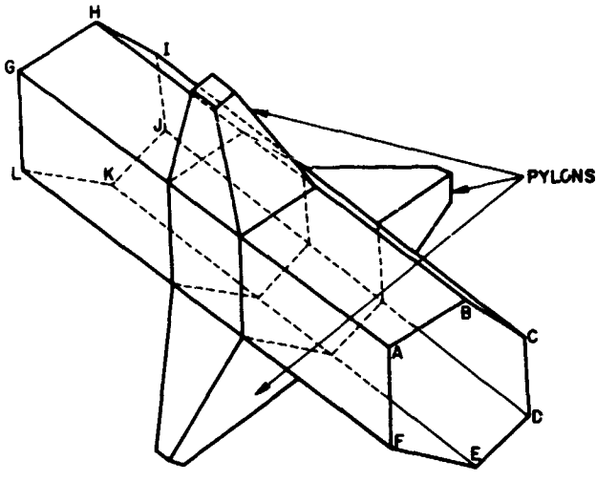


Figure C-16–The Advanced Orbiting Solar Observatory (AOSO) hexagonal support structure.

The three temperature profiles were changed into the form of heat inputs to the structure. The first method for finding the heat input of a device involved using the temperature of the hottest node (see Appendix and Figure C-17) as the basis temperature. Then it was found how much heat was needed to be dumped into each node, except the basis node, to maintain uniform temperature across the entire structure. The results were plotted (see Figures C-18, C-19 and C-20) and, in general, show two peaks per orbit which most heaters should be able to reproduce easily. At most, the largest power input to the structure by a single heater would be 1.41 watts for the second case, 0.02 watts for the third case, and 0.01 watts for the first case. Thus it is the conduction thru the pylons which allows the greatest amount of heat to enter the support structure of the AOSO. However, for this conduction case, the thermal gradient across the five cross-sections are nearly uniform except for the cross-section upon which the pylons rest (cross-section #3). It is the uneven input of heat in the first and third cases which gives rise to the gradient over most of the entire structure as evidenced by the large heating of three nodes for a given cross-section, while the other three are only slightly heated (see Figures C-18 and C-20).

Much of the time it is more economical to cool a few of the hot nodes and heat others than it is to heat all the nodes. To illustrate, a second method where five hot nodes were cooled and the other nodes heated was adopted for the conduction and radiation cases. The power inputs to the structure by the heat pumps were determined and are seen in Figures C-21 and C-22. (The C.O.P. for heating was taken to be 1 and for cooling 0.3. The C.O.P. will be discussed later in the report.) To see the impact of this cooling, the total power drawn from the solar cells to operate the heat pumps, as specified, are drawn for

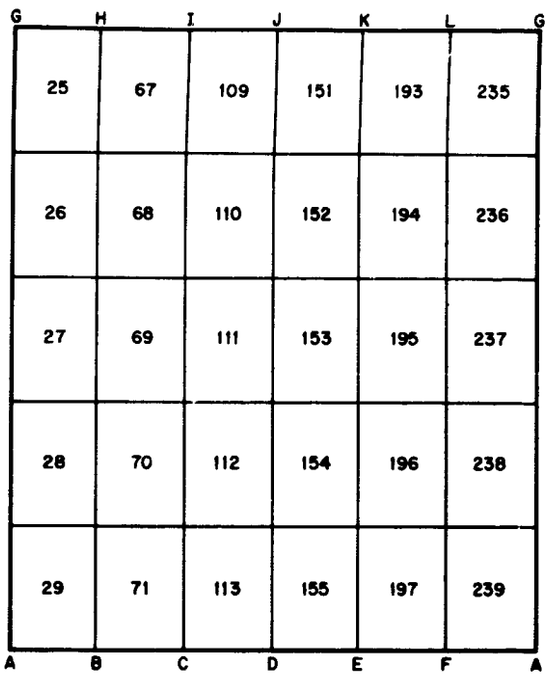


Figure C-17–The nodal program for the support structure for AOSO.

the conduction case (see Figure C-23). The first method, heating all nodes, gives peaks of power of about 12.4 and 20.3 watts, while the second method, both heating and cooling, gives peaks of approximately 2.2 and 3.1 watts. It is readily seen that a great quantity of power is saved by cooling several points.

DEVICES

Resistor

Once the problem has been stated and the requirements established, the problem becomes that of determining what devices are available that actually fit the requirements. So now attention focuses on the resistor and thermoelectric device which are heat controlling devices to be discussed.

When thinking of a heating device, the resistor is the first to come to mind. As electricity passes through the resistor, Joule heat is produced and heats the structure on which it has been placed by conduction and radiation (not by convection in vacuum). The resistor can be made small being used as a point source in the limit and thus putting heat into particular points of the structure or resistors can be continuous (wire) even being formed into rectangular patches one square cm to several hundred square cm in area. Unless there are only a few localized nodes to be heated, the small, single resistors would be of little use. Instead, most thermal control would be by resistors formed into "blanket heaters."

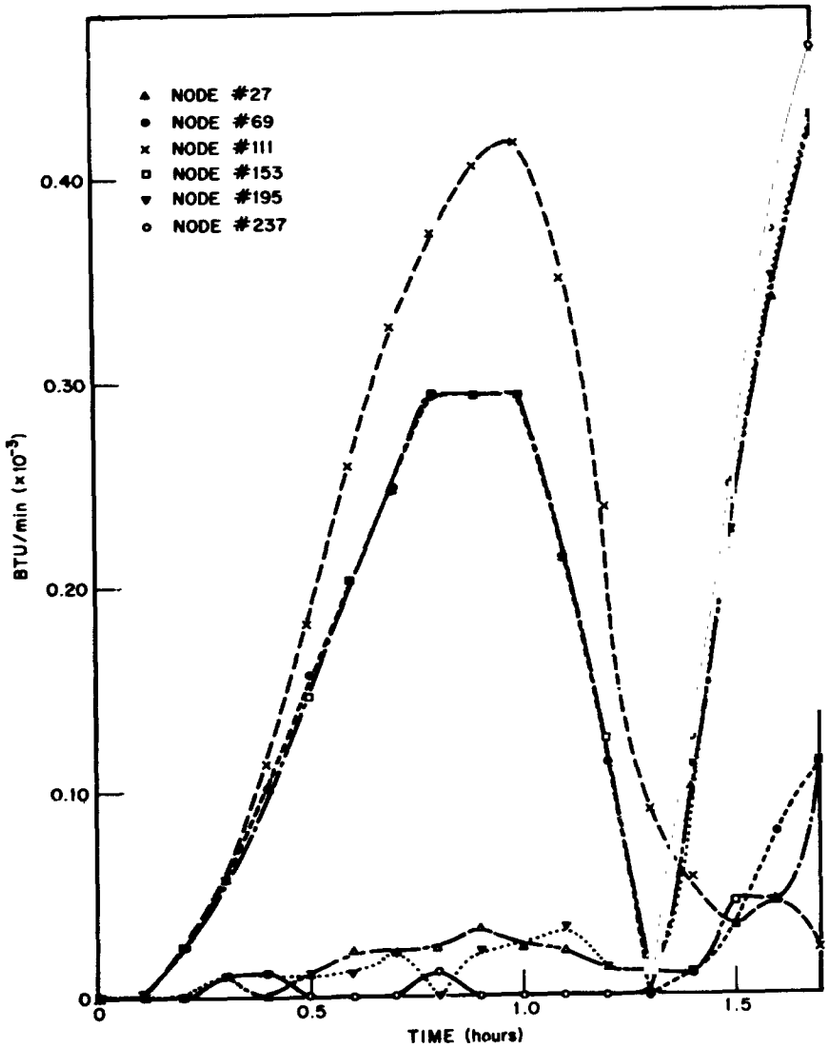


Figure C-18–Radiation from housing. Time varying input of device to structure for cross section # 3.

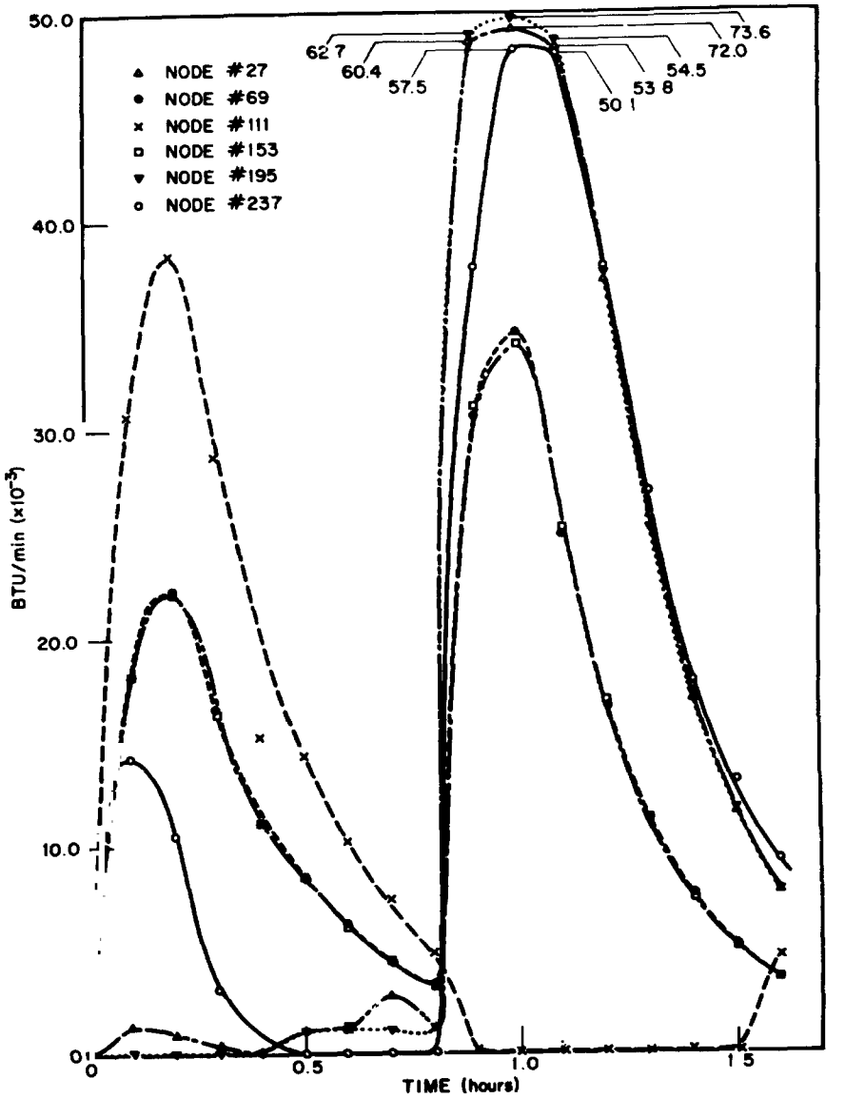


Figure C-19–Conduction through pylons. Time varying power input of device to structure for cross-section #3.

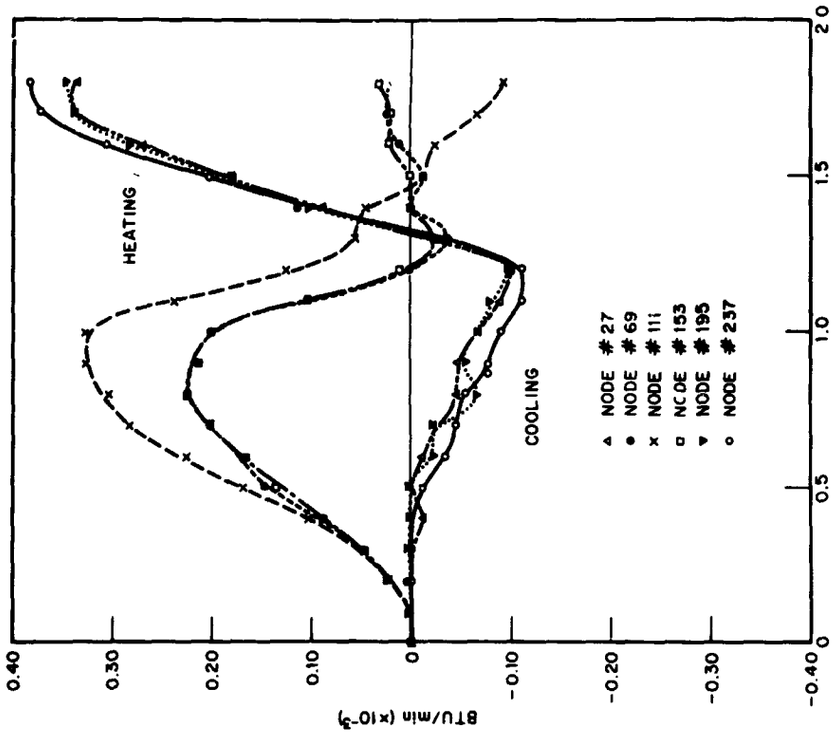


Figure C-21–Radiation from housing. Time varying power input of device which can heat or cool to structure for cross-section #3.

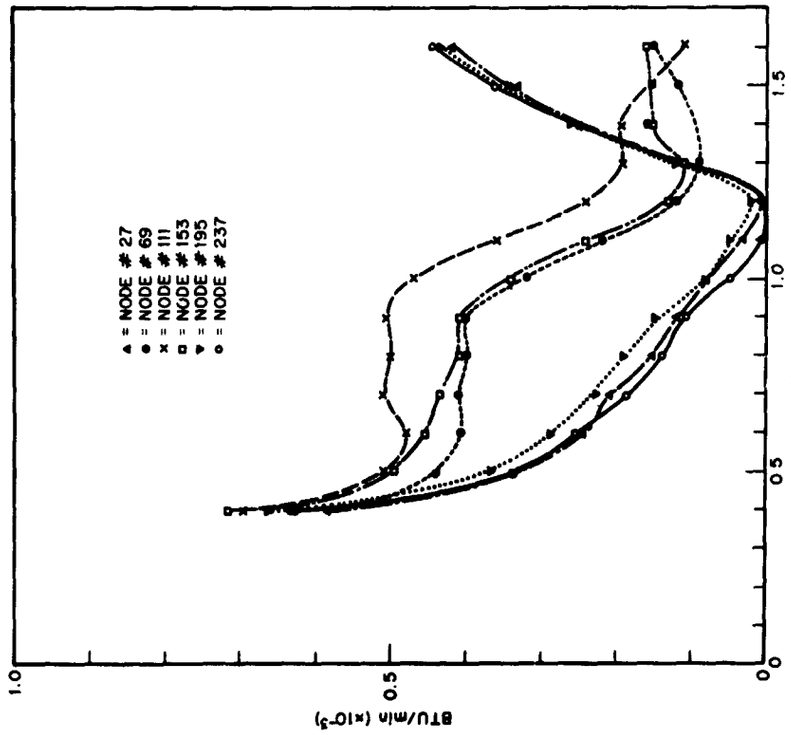


Figure C-20–Heat input needed to keep hexagonal cross-section #3 at uniform temperature.

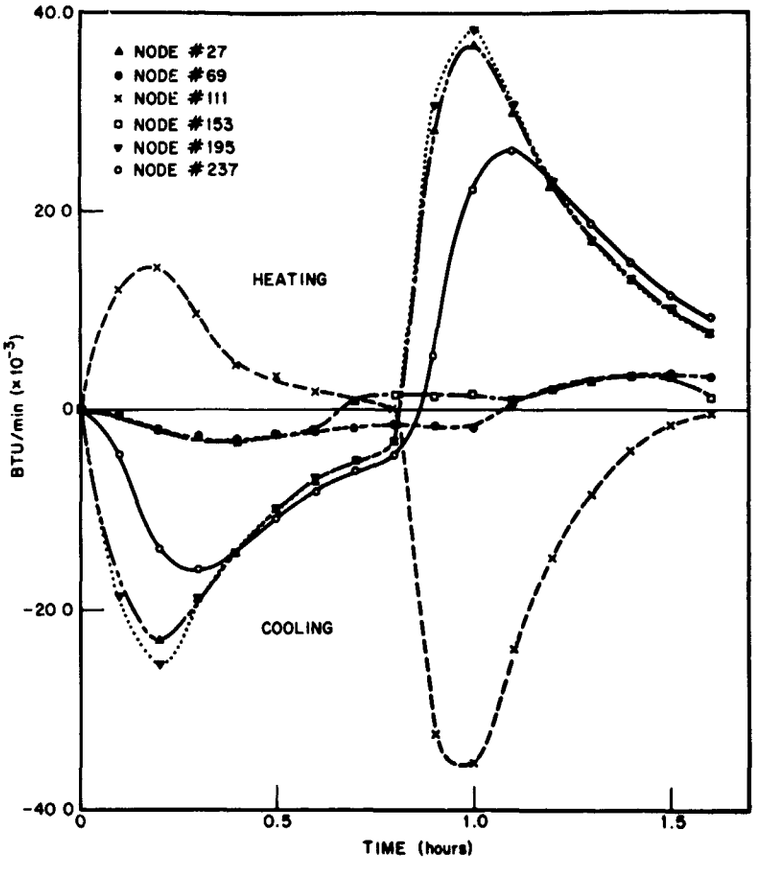


Figure C-22–Conduction through pylons. Time varying power input of heating and cooling device to structure for cross-section # 3.

There are several different kinds of blanket heaters. One called the Electrofilm heater is produced by spraying a resistant coating on a metallic substrate. The coating consists of carbon mixed with an epoxy. The heaters will give uniform heating over evenly sprayed areas, but the resistance of the heaters changes with temperature as seen in Beech Aircraft Corporation Test Report BR 13902-1. In addition, Shapiro at the Goddard Space Flight Center showed that the heaters consistently outgas in vacuum. He found that it was not just surface outgassing but that it comes from the metal interfacing. It was also found that the resistance of the coating changed by 200 ohms indicating a breakup of the conductive paths.

To test OAO experiments in thermovacuum chambers, the Solar Simulation group at Goddard developed a cylindrical housing unit which surrounds the experiments and simulates the temperature profile that the experiments will experience in orbit. They attempted to obtain the profile using strip heaters placed side by side and connected in series to form a node or blanket, but now they are switching to a wire mesh resistant heater.

The strip heaters maintained a gradient of about 15°C over the housing while the gradient over a node was approximately 2°C. Although this gradient will be too large for most spacecraft, better control will result from the use of some passive control in the form of insulation. However, it will take some experimenting to determine what the best gradient is that can be obtained in this manner.

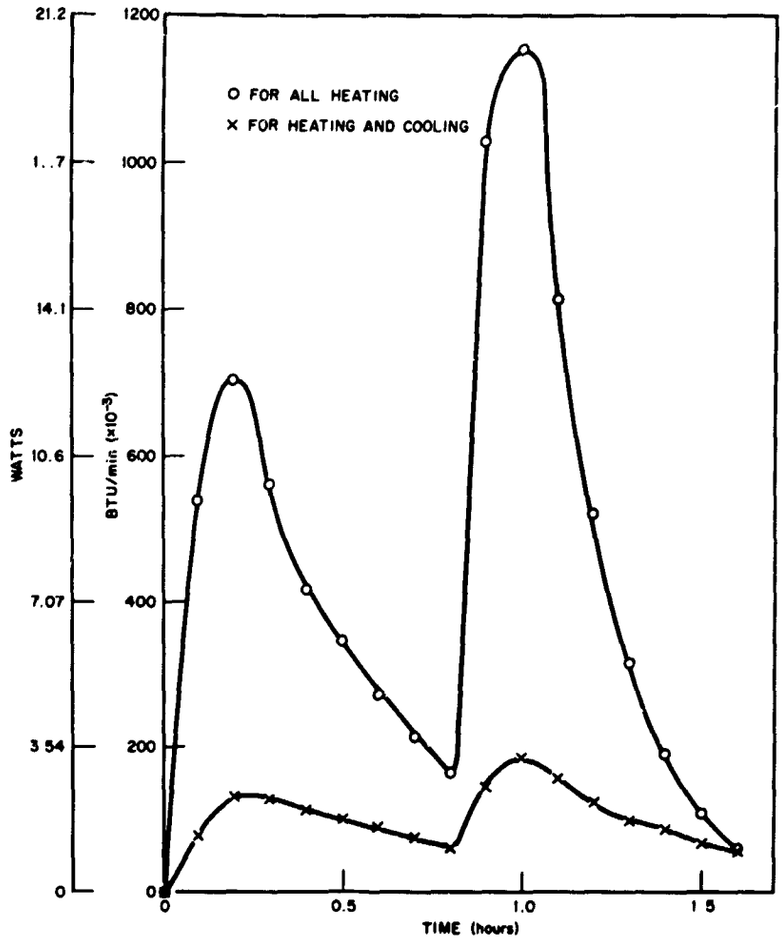


Figure C-23–Power to be put into a thermoelectric device (COP = .3) to obtain uniform temperature.

Tests were run by Republic on three types of blanket heaters to determine the amount of heat produced and the temperature gradient required when each heater was mounted on a surface similar to the AOSO. The first blanket was a nichrome wire grid, while the second consisted of a chemically coated film sandwiched between asbestos sheets, and the third was a neoprene sandwich heater.

Results of the tests showed that an input of 0.6 watts to the first two heaters gave a temperature rise of approximately 4.72°C while the neoprene heater had a temperature rise of 4.97°C for an input of only 0.48 watts. In addition a temperature gradient of 0.26°C was acquired across the nichrome wire grid which was about the size of a node, while the other two heaters had gradients of about 0.43°C. However, even smaller gradients may be obtained through effective use of insulation. The curves giving power input vs. temperature rise, power input vs. $\Delta T_{(max.)}$, and temperature rise vs. $\Delta T_{(max.)}$ are presented in Republic Inter-office Memo by R. Morgenweck.

Republic felt that the installation of the nichrome wire grid heater in the pylon and inner surface area would be costly and time consuming. So the neoprene heaters were

chosen to be used for dry run tests since they had a rugged design and could be removed easier once they had been installed. But for the flight model this last condition is not necessary.

Resistance heaters must be attached to the structure with an epoxy or some other bonding agent. However, the uniformity of the heating depends upon how evenly the bonding agent is spread, and the problem arises that the bonding agent outgasses and can contaminate the experiments.

A problem arises from the fact that geometrical obstructions such as holes in the structure cause heat sinks to occur. Thus the actual design must take these geometrical factors into account.

Resistance heaters will hold up under space conditions, but the resistances will probably change in a predictable manner during the time in space. So the resistor will give the reliability required. They are also rather light and any particular light-weight control system can easily be designed for the resistor.

THERMOELECTRIC DEVICES

The other device to come under consideration is the thermoelectric heat pump which has seen rapid development since the late 1950's. The device is based on the Peltier effect (1834) which states that when a current is passed in one direction through a junction of two dissimilar metals, the junction is cooled but if the current is passed in the opposite direction the junction is heated. A typical single couple thermoelectric heat pump is shown in Figure C-24 below. It consists of a p-type and an n-type semiconductor which connect to conductive metal straps. A current flowing out of an n-type or into a p-type semiconductor cools the junctions while a current in the opposite direction heats the junctions. If a larger area is to be cooled or heated, couples are connected thermally in parallel and electrically in series to give maximum input voltage while requiring minimum input current.

To determine how well the device is doing its job, it is necessary to find its heat pumping capability. For cooling, there is a net useful heat flow Q into the couple at the cold contact.

$$Q = I \alpha T_c - \frac{1}{2} I^2 R - K \Delta T$$

where

I	=	current through the couple	(amp)
α	=	Seebeck coefficient	(V/°C)
T_c	=	cold and temperature of the couple	(°C)
R	=	internal resistance	(ohm)
K	=	thermal conductance	(watt/°C)
ΔT	=	temperature difference across the couple	(°C)

The terms in the above equation are the Peltier, half the Joule heat, and the conductive heat flow from the hot side respectively. The input voltage which causes a current to flow in the couple is

$$V = IR + \alpha \Delta T.$$

Thus the power input to the couple is

$$P = VI = I^2R + \alpha I \Delta T.$$

The capability of the couple to pump heat is then measured by determining its coefficient of performance (C.O.P.) which is defined to be the ratio of the useful heat pumped to the power required to do the pumping. Thus for cooling:

$$\text{C.O.P.} = \frac{Q}{P} = \frac{I \alpha T_c - \frac{1}{2} I^2 R - K\Delta T}{I^2 R + \alpha I \Delta T}.$$

Usually one of two design conditions will be used. One is for maximum C.O.P., and the other is for maximum pumping.

For maximum C.O.P.,

$$\frac{d(\text{C.O.P.})}{dI} = 0$$

simple calculations yield

$$\left(\frac{1}{2} R\alpha \Delta T + \alpha T_c R\right) I^2 - (2KR\Delta T)\, I - \alpha K (\Delta T)^2 = 0.$$

Now define the figure of merit Z to be

$$Z = \frac{\alpha^2}{\rho k}$$

where

ρ = resistivity (ohm-cm)

k = thermal conductivity (watts/cm-°C).

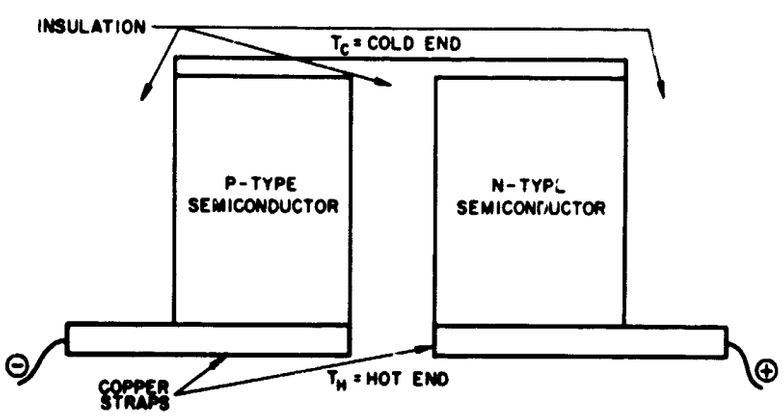


Figure C-24–One stage thermoelectric heat pump.

Since

$$K = \frac{\rho k}{R},$$

then

$$K = \frac{\alpha^2}{RZ}.$$

Substitution yields

$$\frac{\alpha}{RZ}\left[\left(\frac{1}{2}R^2 Z\Delta T + T_c R^2 Z\right)I^2 - (2\alpha R\,\Delta T)\,I - \alpha^2(\Delta T)^2\right] = 0.$$

Note that

$$\frac{1}{2}\Delta T = \frac{1}{2}(T_h - T_c)$$

where T_h = hot end temperature of the couple (°C).

After some elementary calculation we obtain

$$I = \frac{\alpha\,\Delta T}{R\left[\sqrt{1 + \frac{1}{2}Z(T_h + T_c)} - 1\right]}$$

which is the current necessary to be passed through the device for maximum C.O.P. Placing this value of the current in the expression for C.O.P. and letting

$$T_m = \frac{T_h + T_c}{2},$$

we obtain

$$\text{C.O.P.}_{(\text{max.})} = \frac{T_c}{(T_h - T_c)}\,\frac{\left(\sqrt{1 + ZT_m} - \frac{T_h}{T_c}\right)}{(\sqrt{1 + ZT_m} + 1)}.$$

However if maximum heat pumping is desired, then

$$\frac{dQ}{dI} = \alpha T_c - IR = 0.$$

Thus

$$I = \frac{\alpha T_c}{R}$$

is the required current. When this current is used the C.O.P. becomes

$$\text{C.O.P.} = \frac{\frac{1}{2}ZT_c^2 - (T_h - T_c)}{ZT_c T_h}.$$

The most useful thermoelectric cooling pumps have a C.O.P. between 0.3 and 1.0. Since the C.O.P. determines the power needed to cause a certain quantity of heat to flow, then the C.O.P. can be greater than 1.0, but it happens only at low ΔT across the device.

To rid itself of the heat which is built up during the on cycle, fins are used to radiate or conduct heat to a heat sink. The sink must readily receive the heat from the device; otherwise the device will burn up during the cooling process.

It is useful to note that the maximum temperature difference across the device which can be obtained for $Q = 0$ is

$$\Delta T_{(max.)} = \frac{\alpha^2 T_c^2}{2RK} = \frac{1}{2} Z T_c^2 .$$

Since $\Delta T_{(max.)}$, depends upon Z, it can be seen why the low value of the figure of merit retarded the development of thermoelectric devices. Development of semiconductors brought Z's of approximately 3×10^{-3} $^{\circ}C^{-1}$ so that the device could become practical.

If the direction of the current is reversed, the device becomes a heater and is more efficient than a resistor. Since Joule heat (I^2R) aids the heating device instead of hindering it as in the cooling process, the device has a higher C.O.P. equal to

$$\text{C.O.P.} = \frac{I \alpha T_c + \frac{1}{2} I^2 R - K \Delta T}{I^2 R + I \alpha \Delta T}$$

when the useful value for heating lies between 1.0 and 3.0.

The usual commercial thermoelectric devices have 1 to 2 square inches of area for the heating surface (see Figure C-25), but they can be made either larger or smaller. The weight of the devices can range (for typical Sanyo Thermo Modules) from 0.8 to 2.7 ounces. (For other characteristic maximum values, see Table C-3). Typical graphs showing temperature differences vs. hot junction temperature, heat pumping capacity vs. hot junction temperature, and voltage vs. hot junction temperature are presented for the Jepson Thermoelectric Module model 433 in Figures C-26, C-27, and C-28 respectively. It is interesting to note that these devices can hold temperature differences of a fraction of a degree up to at least 90°C (162°F) depending upon the hot junction temperature, the current used, and the amount of heat being pumped.

There are drawbacks in using thermoelectric devices. One is that although the device requires low voltage, it often demands high current which may make it prohibitive for some uses in satellites. However, as the industry develops, it may find ways to lower the current requirements. Another disadvantage arises because the device itself acts as a relatively good heat leak during the off portion of its cycle. The ratio $K \Delta T/Q$ will aid in determining the seriousness of the leak. It should be small compared to unity, otherwise the cooled portion will heat during the off cycle as rapidly as it cools during the on cycle.

The problem of what controlling equipment is needed to activate and regulate the thermoelectric device is also present. The devices run on dc current, so ordinarily ac current must be transformed to dc current. But a satellite generated dc current so no problems arise here. The fundamental systems which are now used for thermal control can also be used for control of the thermoelectric heating and cooling so that the systems can easily be designed and implemented. Weight of the control systems should also cause no problem.

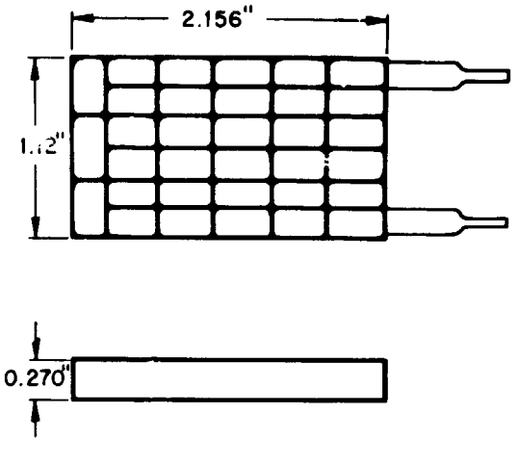
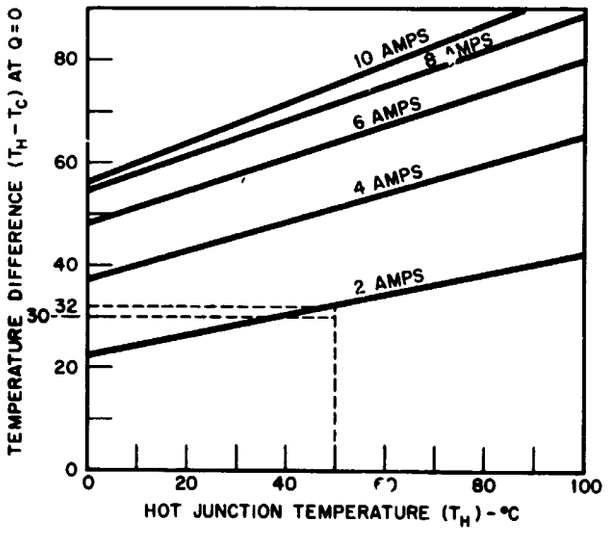


Figure C-25–Frigitronics module F33.

Figure C-26–Temperature difference vs. hot junction temperatures.



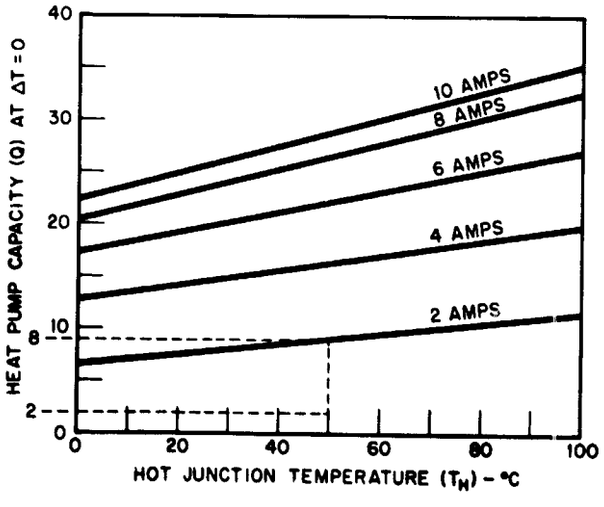


Figure C-27–Heat pumping capacity vs. hot junction temperature.

Table C-3*
Typical Characteristics of Heat Pumps

Symbol	Size	Weight	Number of Couples	Optimum Current	Maximum Temperature Difference	Terminal Voltage	Heat Pumping Capacity
Dimension Type	Inch	Oz.		Amps	°C	Volts	Watts
STM-1025	2.244 × 1.693 × 0.413	2.2	10	25	55	1.1	16
STM-1021	2.244 × 1.693 × 0.413	2.0	10	21	55	1.1	14
STM-1621	2.519 × 2.126 × 0.413	2.7	16	21	55	1.8	20
STM-1012	1.614 × 1.279 × 0.374	1.0	10	12	55	1.1	8.3
STM-1004	1.378 × 1.102 × 0.433	0.8	10	4	50	1.1	3.5
STM-4004	2.244 × 1.968 × 0.472	2.7	40	4	55	4.5	14

*Taken for Sanyo Thermo Modules. The types are given in the table.

Now returning to the particular problem of the AOSO and assuming that stability has been reached, it was found that the support structure had a constant temperature of 68°F (20°C) while the housing cycles between extremes of 32°F (0°C) and 104°F (40°C). If the extreme case of the housing temperature being 104°F is taken and the heat exchanger of the thermoelectric device is taken to be 10°C greater than the housing, then $\Delta T_{(max.)} = 54°F = 30°C$. Thus the hot end of the device will operate at $T_h = 50°C = 122°F$. From the conduction, radiation, and 1/2 watt heater cases, it was found that less than two watts will have to be pumped out of the structure by the devices since this is the amount coming into the structure from the environment. Using Figure C-26 and $T_h = 50°C$, it is seen that for a two amp current, the device will give a temperature difference of approximately 32°C which is greater than the 30°C required. From Figure C-27 it is seen that 9 watts can

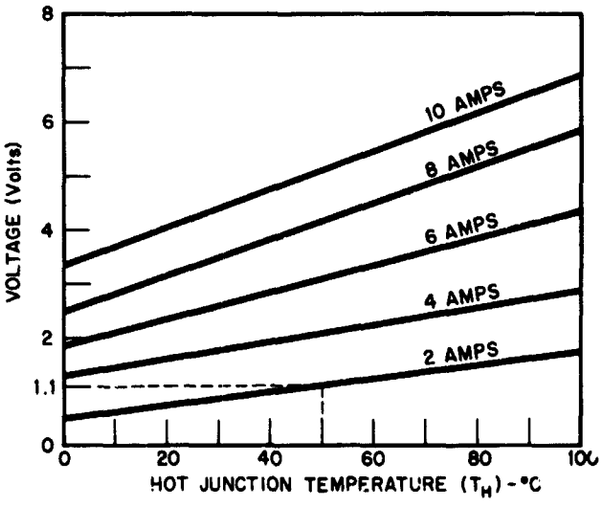


Figure C-28–Voltage vs. hot junction temperature.

be pumped by the device which again is more than required. Finally Figure C-28 shows that 1.1 volts are needed to have the two amps current at $T_h = 50°C$. So this particular thermoelectric device easily fulfills the requirements. Since the extreme case was taken, the current will be even less than the two amps most of the time.

The question of whether a run-away condition exists or not arises. Heat is radiated from the housing to the support structure. But the thermoelectric heat pumps must pump the absorbed heat out of the structure in order to maintain constant temperature. The heat which is pumped out is then radiated back to the housing raising its temperature. So the housing again radiates but now at a higher temperature thus radiating more heat than it did originally. The devices then are pumping larger and larger amounts of heat and this constitutes a run-away condition. But, the housing for the AOSO is large and acts like a reservoir since only two watts are radiated to it. In addition, the housing will be cycling between 32°F and 104°F and will cancel any temperature rise due to the small heat which is radiated. So the run-away condition should not exist for the thermoelectric devices on the AOSO.

As with resistors, the problem of attaching the thermoelectric devices exists. Along with outgassing comes the problem that the epoxy must be non-conductive so that the devices will not be shorted out. The devices may be bolted on, whether this is the best method remains to be determined.

CONCLUSIONS

Passive thermal control coupled with active thermal control is the approach needed for obtaining and holding particular thermal gradients with good accuracy. It was found that a combination of heating and cooling is the best active thermal control method attainable, and that thermoelectric devices appear to be the best devices for maintaining the desired gradient although they are somewhat limited by requirements for high currents and the value of the figure of merit which in turn limits $\Delta T_{(max.)}$. The alternate method is that of heating with resistors which are in the form of blanket heaters. However, they will have higher power requirements and will not give as good control.

For further study, mathematical analysis and actual experiments will be needed. In the particular AOSO problem, their computer program should be used to determine the feasibility of using these active devices. In addition, the problem of attaching the devices to the structure must be studied since outgassing occurs.

Appendix

The entire AOSO spacecraft was broken into nodes, but only the support structure need be presented here. The support structure is a hexagonal cylinder held within a circular cylinder called the housing (Figure C-16). Each of the six sides of the support structure were taken as a basic strip. Then each strip or side was broken into five sections called nodes. Each strip has five nodes and there are six strips giving a total of thirty nodes on the support structure. Each node is numbered in Figure C-17 so it will be easier to interpret the graphs presented in the report.

BIBLIOGRAPHY

Bean, J. E., "Applications of Thermoelectric Cooling," Industrial Electronics, Vol. 1, #3, page 132 (December 1962)

Benedict, R. P., "Thermoelectric Effects," Electrical Manufacturing, Volume 65, number 2, Page 103 (February 1960)

Brickley, R. L., "The Outlook for Thermoelectric Devices," Electronic Industries, Volume 22, Page D-6 (June 1963)

Cadoff, I. B., and E. Miller, Thermoelectric Materials and Devices, Reinhold Publishing Corporation, New York (1960)

Egli, P. H., Thermoelectricity, John Wiley and Sons, New York (1960)

Eickhorn, R. L., "A Review of Thermoelectric Refrigeration," Proceedings of the IEEE, Volume 51, Page 721, May 1963

Eickhorn, R. L., "Thermoelectric Refrigeration," ASRE Conference paper, Minneapolis, Minnesota, Meeting (June 1955)

Finch, D. I., General Principles of Thermoelectric Thermometry, Leed and Northrop Corporation, Philadelphia (1962)

Goldsmid, H. J., Thermoelectric Refrigeration, Plenum Press, New York (1964)

Goldsmid, H. J., Applications of Thermoelectricity, John Wiley and Son, New York (1960)

Goldsmid, H. J., "Thermoelectric and Thermomagnetic Cooling," Industrial Electronics, Volume 1, number 9, Page 467 (June 1963)

Goldsmid, H. J., "Thermoelectric Applications of Semiconductors," Journal of Electronics and Control, Volume 1, Page 218 (1955)

Goldsmid, H. J., and R. W. Douglas, "The Use of Semiconductors in Thermoelectric Refrigeration," British Journal of Applied Physics, Volume 5, Page 386 (1954)

Gramberg, G., and H. G. Plust, "Developments in the Field of Thermoelectricity," The Brown Boveri Review, Volume 49, number 11/12, Page 501 (Nov./Dec., 1962)

Gray, P. E., The Dynamic Behavior of Thermoelectric Devices, Technology Press of the Massachusetts Institute of Technology (1960)

Green, W. B., Editor, Westinghouse Thermoelectric Handbook, Westinghouse Electric Corporation, Youngwood, Penn. (1962)

Guczi, Z., and H. J. V. Tyrrell, "The Maintenance of Controlled Temperature Differences Using a Thermoelectric Heat Pump," Journal Scientific Instruments, Volume 41, number 7, Page 468 (July, 1964)

Heikes, R. R., and R. W. Ure, Jr., Thermoelectricity - Science and Engineering, Interscience Publishers, New York (1961)

Huck, W. V., "Thermoelectric Elements for Heat-Pump Devices," Electrical Manufacturing, Volume 65, number 5, Page 281 (May, 1960)

Ioffe, A. F., Semiconductor Thermoelements and Thermoelectric Cooling, Infosearch London (1957)

Ioffe, A. F., "The Revival of Thermoelectricity," Scientific American, Volume 199, number 5, Page 31 (November 1958)

Jaumat, F. E., Jr., "Thermoelectric Effects," Proceedings of the IRE, Volume 46, Page 538 (1958)

Kelley, J. C. R., Jr., "Applying Thermoelectric Devices - System Design Considerations," Automatic Control, Volume 17, number 5, page 34 (December, 1962)

Lefferts, P., "Thermoelectric Cooling," Machine Design, Volume 37, Page 118 (January 7, 1965)

Lindenblad, N. E., "Thermoelectric Heat Pumping," Electrical Engineering, Volume 77, number 9, Page 802 (September, 1958)

MacDonald, D.K.C., Thermoelectricity, an Introduction to the Principles, John Wiley and Sons, Inc., New York (1962)

Ott, L. H., "Electronic Cooling and Heating," Radio-Electronics, Volume 33, number 1, Page 26 (January, 1962)

Rittner, E. S., "On the Theory of the Peltier Heat Pump," Journal of Applied Physics, Volume 30, number 5, Page 702 (May, 1959)

Staebler, L. A., "Thermoelectric Refrigeration," ASHRAE Journal (August, 1959)

Stil'bens, L. S., E. K. Jordanishvili, and T. S. Stravitskaza, "Thermoelectric Cooling," Izvest. Akad, Nauk. SSSR. Ser. Fiz., Volume 20, Page 31 (1956)

Stubstad, W. R., "Thermal Control with Thermoelectrics," Control Engineering, Volume 8, number 9, Page 178 (September, 1961)

Tave, J., Photo and Thermoelectric Effects in Semiconductors, translated by Margaret Laner, Pergamon Press, Oxford, New York (1962)

Taylor, S., and A. Mulicia, "Thermoelectric Coolers for Electric Components," Electrical Manufacturing, Volume 65, number 5, Page 284 (May, 1960)

Wolfe, R., and J. H. Wermik, "Thermoelectric Devices and Materials," Bell Laboratories Record, Volume 40, Page 191 (June, 1962)

The following are references used other than books or papers written for periodicals.

Frigitronics, Incorporated, Bridgeport, Conn., "Instructions and Precautions for Using Frigitronics Thermoelectric Modules."

Frigitronics, Inc., Bridgeport, Conn. Technical Bulletin #331-334

Jepson Thermolectrics, Inc., "General Instructions for the Use of Jepson Thermoelectric Modules (Mode #433)."

Republic Aviation Corporation, "AOSO Final Report - Phase I," Volumes I and III

Sanyo Electric Co., Osaka, Japan, "Sanyo Thermo Modules."

Rice, P. W., "Ten Inch Sample Apollo Heaters," Beech Aircraft Corporation Test Report BR 13902-1 (January, 1963)

Shapiro, H., "Analysis of Electrofilm 'Spray-on' Heaters," Goddard Space Flight Center Vacuum Research and Technology Section Memorandum (March, 1965)

BLANK PAGE

PERTURBATION TORQUES ON A 750-FOOT EXTENSIBLE BOOM DUE TO ELECTRODYNAMIC INTERACTIONS

J. A. DellaValle

Project RAE (Radio Astronomy Explorer) is an effort to orbit a spacecraft with two V-shaped antennas. One of the V's will scan the heavens to receive celestial radio signals and the other will pick up background noise from the earth. This mission should enable radio astronomers to receive signals below 10 Mc which do not reach the earth because of strong atmospheric attenuation. The antennas are four hollow beryllium-copper booms about 1/2 inch in diameter and 750-1000 feet long, extending from the satellite and making an angle of approximately 60° at the vertex of each V. The booms will be deployed after the satellite achieves the desired orbit. The satellite will be in a circular orbit 6000 km above the earth (Figure C-29).

The booms are light, weak structures and can be bent easily. Thus, owing to their length, a small force acting near the end of the boom can produce a considerable bending torque about an axis through the satellite. These bending torques can have several undesirable effects on the mission. First of all, if the torque is large enough it can snap the boom, thereby permanently distorting it.

Secondly, since the general shape of the antennas effects the orbital dynamics and stabilization techniques of the spacecraft, it is desirable to avoid certain distortions.

Finally, and most important to the experimenter, is the effect of antenna shape on its radiation pattern. In order to interpret the experimental data the shape of the antenna must be known at all times, hence, all the forces tending to change the shape of the antennas must be known as a function of space and time.

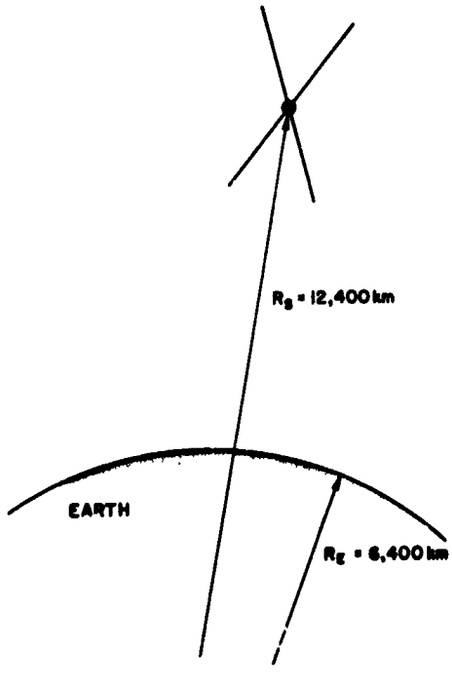


Figure C-29–Height of satellite above the earth (circular orbit).

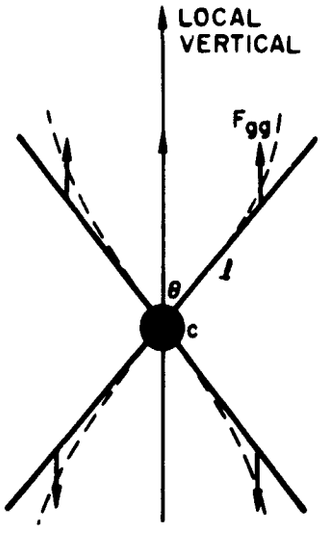
The purpose of this report is calculation of the electric and magnetic forces on the booms. These forces are believed to be small compared to the so-called gravity-gradient bending force discussed below but a reasonable order of magnitude calculation must be made in order to verify this.

The dimensions of most satellites are small enough so that they may be treated as a point mass orbiting the earth. A circular orbit is obtained when the centrifugal and gravitational forces acting on the satellite cancel, if the magnitude of the forces are considered constant over the satellites extension. With the RAE spacecraft the situation is different. The centrifugal force increases linearly with the distance from the earth while the gravitational force decreases as an inverse square. These forces cancel at the center of mass of the satellite but not along the boom. During the length of the boom the imbalance

of forces is not negligible and gives rise to what is sometimes called the "gravity-gradient" force. This force per unit length is given by*

$$F_{gg} = 3\rho\omega^2 x \tag{1}$$

directed along the local vertical as shown below. The bending torque produced by this force about an axis through c normal to the page is given by



$$T_{gg} = \rho\omega^2 \sin\theta \cos\theta \ell^3 \tag{2}$$

$$= .024 \text{ ft. lb.} = .032 \text{ newton-meter.}$$

We shall now consider the electric and magnetic effects on the booms. Each boom constitutes a conductor cutting the earth's magnetic field; hence, there is a force on each electron in the boom given by

$$F = -e\vec{v} \times \vec{B}, \tag{3}$$

where e = charge on the electron,

$\vec{v}$ = velocity of the electron (boom),

$\vec{B}$ = earth's magnetic induction.

This force moves all the free electrons to one end of the rod and a potential difference is thus created (Figure C-30). For a rod moving at right angles to the field and the zero potential taken to be at the rod's end, we have the difference in potential along the rod:

$$\varphi \equiv \int_0^{\ell} \frac{dW}{q} = \int_0^{\ell} \frac{qBv}{q}\, d\ell = Bv,\ell = \mu_o Hv\ell. \tag{4}$$

By substituting the values given in the appendix, we have $\varphi = 11.4$ volts.

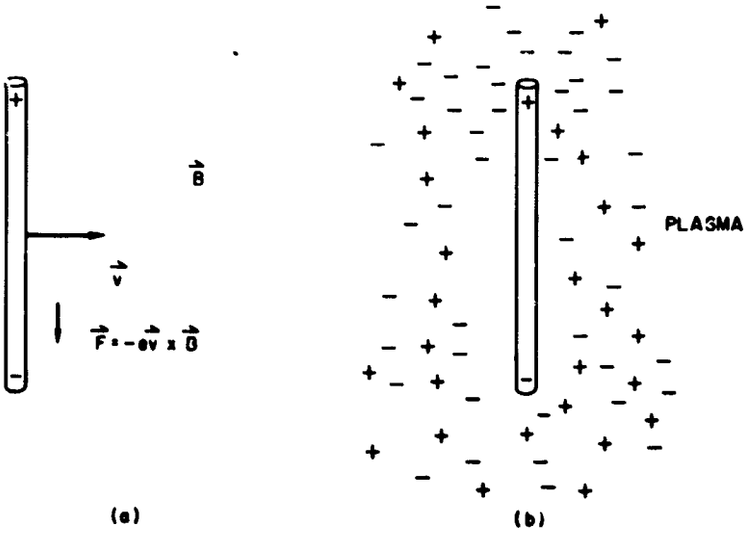


Figure C-30—Polarization of a boom in a magnetic field. (a) Showing how boom becomes polarized moving thru magnetic field; (b) Showing how one end is neutralized while other is not.

*Master's Thesis, T.W. Flatly, "Equilibrium Shape of an Array of Long, Elastic Structural Members in Circular Orbit," Catholic Catholic U. of America, April 1965

The $e\vec{v} \times \vec{B}$ force moves the electrons to one end of the rod thereby polarizing it. Owing to this polarization, an electrostatic field is set up which opposes the effective $\vec{v} \times \vec{B}$ field. If we were in free space, polarization would build up until the fields would cancel and we would have a boom with an induced dipole moment. However, since this occurs in a plasma, we have a somewhat different situation. At this altitude the primary neutral atoms are hydrogen and the ions are protons and electrons. The existing literature indicates that the ion energies are such that the electrons are traveling much faster than the satellite and the protons are traveling much slower. Because the electrons are attracted to the positive end of the boom and are fast moving - relative to the boom, they catch up and neutralize that end. On the other hand, the only protons which strike the boom are those occupying the volume swept out by the boom (Figure C-30b). As the positive ions neutralize the negative end of the boom, the $e\vec{v} \times \vec{B}$ force acts on the electrons and tends to make it negative. This electron flow in the boom constitutes a current, the magnitude of which is limited by the rate at which protons are intercepted by the boom. The number of protons swept out per unit time is just the number contained in the volume swept out by the boom in unit time, i.e., the current in the boom is

$$I = A_b v_s n + e \approx 10^{-6} \text{ amp.} \tag{5}$$

where A_b is the cross-sectional area of negative portion of boom.

The force on a conductor carrying a current I at right angles to a magnetic field is

$$F = BI\ell. \tag{6}$$

Assuming this force to be uniform along the boom and substituting the values given in the appendix, we have

$$T_i = F\frac{\ell}{2} = \frac{BI\ell^2}{2} \approx 10^{-6} \text{ n.m.} \tag{7}$$

for the torque about one end of the boom.

This T_i is called the "induction torque" and is negligible compared to the gravity-gradient torque. The force calculated in Equation (7) is for the worst possible conditions.

The derivation for the induction torque was carried out for a uniform field but actually the earth's field drops off as an inverse cube (magnetic dipole). This would cause the part of the boom nearest the earth to experience a greater force than the part farther away. To calculate this difference in the force we have from Equation (6):

$$F = BI\ell = \frac{K}{r^3}, \text{ since } B \approx \frac{1}{r^3} \tag{8}$$

Therefore the difference in the force from one end of a current carrying rod of length ℓ to the other is

$$\Delta F = K\left[\frac{1}{(r+\ell)^3} - \frac{1}{r^3}\right], \tag{9}$$

where r is the distance from the origin of the nearest end of the rod. Expanding this we have

$$\Delta F = K\left(r^{-3} - 3r^{-4}\ell + \ldots - r^{-3}\right)$$

$$\Delta F \approx \frac{3K\ell}{r^4}$$

$$\Delta F \approx \frac{3F\ell}{r}\ , \quad \text{neglecting higher terms}, \tag{10}$$

where $F = K/r^3$ is the original force calculated.

The torque about one end of the rod produced by this difference in the force along the rod is given approximately by

$$\Delta T \approx \frac{3F\ell^2}{r} = T_{io}\frac{6\ell}{r}\ ,$$

where $T_{io} = F\ell/2$ is the original induction torque calculated. Hence, the net induction torque acting on the boom is

$$T_i = T_{io} \pm \frac{6\ell}{r}T_{io} \tag{11}$$

$$= 10^{-6} \pm 10^{-9}\ \text{n. m.},$$

where + is for the lower booms - is for the upper booms.

The next effect we shall consider is the so-called coulomb drag. Since more electrons strike the boom than protons the boom attains a net negative charge which causes it to attract and collide with more protons than if it were neutral case. The coulomb drag is the drag due to the increase in the number of proton collisions. To calculate the change in momentum due to these collisions we must know the number of collisions. In order to determine this number we must determine the potential of the satellite.

The following are from the derivations of Chang and Smith.* Here, the boom is taken as the reference for both protons and electrons. The constants of the motion for protons are the angular momentum

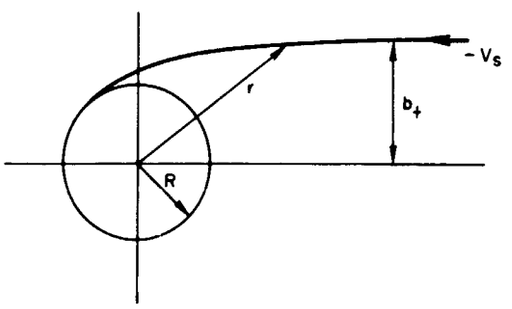
$$L = m r^2 \dot{\theta}\ , \tag{12}$$

and the energy

$$E = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) + e(-\varphi). \tag{13}$$

*Chang, H.H.C. and Smith, M.C. "On the Drag of a Spherical Satellite Moving in a Partially Ionized Atmosphere," J. of British Interplanetary Soc., Vol. 17 (1959-60).

Since the ionic angular momentum about the center of the boom is conserved, it remains equal to its initial value at the ion's last collision. Hence, if V_s is the magnitude of the ion velocity to that of the satellite after the last collision before it encounters the satellite, then



$$L = mV_s b_+ = mr^2\dot{\theta} = mb_+ \sqrt{\frac{2E}{m}}, \tag{14}$$

where the parameters are defined below and E is the kinetic energy of the ion relative to that of the satellite (assuming that the last collision occurs at distance far enough from boom so that the potential energy is negligible). We then have from Equation (14)

$$\dot{\theta}^2 = \frac{b_+^2}{r^2}\left(\frac{2E}{m}\right); \tag{15}$$

substituting this into Equation (13), we obtain

$$E = \frac{1}{2}m\dot{r}^2 + \frac{b_+^2}{r^2}E - e\varphi.$$

Since we are considering only those ions which strike the boom, $\dot{r} = 0$ at $r = R$ and

$$E = \frac{b_+^2 E}{R^2} - e\phi. \tag{16}$$

Solving for b_+ we get

$$b_+ = R\left(1 + \frac{e\varphi}{E}\right)^{1/2} \tag{17}$$

where b_+ is the impact parameter; that is, all ions coming within this distance of the boom are attracted, strangely enough, by the negative potential collide with it. Likewise for the electrons, we can show

$$b_- = R\left(1 - \frac{e\phi}{E}\right)^{1/2}. \tag{18}$$

The positive ion current to the boom is

$$I_+ = en_+\bar{V}_s A_+ = en_+\bar{V}_s 2\ell R\left(1 + \frac{e\varphi}{E}\right)^{1/2} \tag{19}$$

where A_+ is the effective area for protons and $\bar{V}_s$ is the mean velocity of satellite relative to that of the ions approximately equal to satellite velocity since $V_s \gg V_+$. Since the electrons are striking the boom from all sides the calculation of the electron current is not as simple as the proton current. To simplify things we shall consider only a cross-sectional plane of the boom. From kinetic theory we have that number of electrons striking both sides with speed v per unit area per unit time $= 2 \times (1/r) v\, dn_v$ where dn_v

is the number of electrons with speeds between v and $v + dv$. Since the boom has a negative potential φ only electrons with speeds greater than $\sqrt{2e\phi/m}$ reach the boom. Therefore the total number of electrons striking boom per unit area per unit time is

$$\frac{1}{2}\int_{\sqrt{2e\varphi/m}}^{\infty} v\,dn_v$$

and

$$\text{current density} = J = \frac{1}{2}\int_{\sqrt{2e\varphi/m}}^{\infty} en_- v\,dn_v,$$

$$\text{electron current} = A_- J = \frac{1}{2}\int_{\sqrt{2e\varphi/m}}^{\infty} en_- 2\ell R\left(1 - \frac{e\varphi}{E}\right)^{1/2} v\,dn_v.$$

By assuming a Maxwellian distribution for the electrons the electron current is given by

$$I_- = C\int_{\sqrt{2e\varphi/m}}^{\infty} v^3\left(1 - \frac{2e\phi}{m_- v_-^2}\right)^{1/2} e^{\frac{-mv_-^2}{2kT}}\,dv, \tag{20}$$

where

$$C = 4\pi e n_- \ell R\left(\frac{m_-}{2\pi kT}\right)^{3/2}$$

At equilibrium there is no net current to the boom, i.e., $I_- + I_+ = 0$. This equation would then determine ϕ. The integral for I_- is difficult to evaluate, however, we know that for a physical situation we must have

$$\left(1 - \frac{2e\varphi}{m_- v_-^2}\right) \geq 0, \tag{21}$$

$$\therefore \varphi \leq \frac{mv^2}{2e} = .05 \text{ volts}.$$

This is, of course, a very rough estimate, and represents an upper limit for the potential.

Substituting this value for φ into Equation (17) we get

$$b_+ \leq 300R.$$

A rough calculation will show that the transfer of momentum by electron collisions is negligible compared to that by protons, for

$$\frac{\Delta P_-}{\Delta P_+} = \frac{m_- v_-}{m_+ v_+} \approx 5 \times 10^{-2}.$$

Therefore, we need only consider proton collisions:

$$\text{Number of protons integrated per sec} = n_+ \bar{V}_s \, 2\ell b_+ ;$$

$$F_c = \frac{\Delta P}{\Delta P} = 2 m_+ \bar{V}_s^2 n_+ \ell R \left(1 + \frac{2e\varphi}{m_+ v_+^2}\right)^{1/2} \approx 10^{-11} \text{ newton.} \tag{22}$$

The "coulomb-torque" is then approximately

$$T_c = F_c \frac{\ell}{2} \approx 10^{-9} \text{ newton - meter.} \tag{23}$$

Compared to the gravity-gradient torque this is completely negligible. An ion density of approximately $10^{15}/m^3$ would give $T_c \approx 10^{-3}$ newton-meter which would be approaching the gravity-gradient torque, but densities of this kind are not realized at 6000 km.

There is also a transfer of momentum due to protons which do not actually strike the boom but whose trajectories are affected by the negative potential. According to all calculations done by other authors, this effect is negligible compared with those which actually collide with the boom. Since this last effect is negligible we can make the assumption that the first is therefore entirely negligible too.

The possibility of eddy currents has been considered. If the boom were spinning the magnetic forces on an electron would be such that it could complete a path around the boom thus creating an eddy current. This eddy current could then interact with the magnetic field and produce a torque on the boom. However, since the boom is not spinning or rotating no eddy currents are produced.

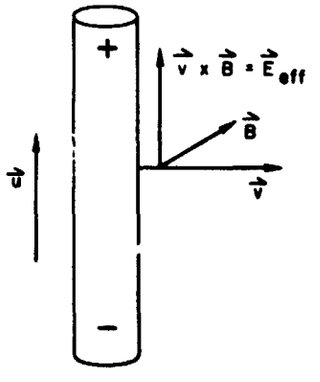
The increase in effective area of the boom due to the negative potential will also have the effect of increasing the induction drag by a factor of about 300. However the estimated value of 300 R for the impact parameter does not seem realistic and should be treated as a very rough estimate. Its actual value is probably much less thereby ensuring that both the induction and coulomb drag are still insignificant.

As shown before, the boom becomes polarized in an effective electric field of $E_{eff} = \vec{v} \times \vec{B}$. The question arises, is there a torque on the dipole due to this field?

To answer this, consider the fact that the boom is in an effective electric field

$$\vec{E}_{eff} = \vec{v} \times \vec{B} .$$

It has an induced dipole moment $\vec{\mu}$ which is in the direction of $\vec{v} \times \vec{B}$, as shown below. The torque on a dipole in a field is given by



$$\vec{T} = \vec{\mu} \times \vec{E}.$$

Since $\vec{\mu}$ is parallel to $\vec{E} = \vec{E}_{eff}$, $T = 0$.

Data obtained just before this paper was printed indicates certain assumptions in the work are incorrect. The electron and proton temperatures at 6000 km has been estimated* to be about 5000°K. This gives an ion velocity of

$$v = \sqrt{\frac{2kT}{m}},$$

*Brase, L. H., Private Communication, G.S.F.C.

or

$$v_+ \approx 9 \times 10^3 \text{ m/sec},$$

$$v_- \approx 4 \times 10^5 \text{ m/sec}.$$

From this we see that the velocity of the electrons is very much greater than that of the protons and satellite but that the velocity of the protons is about the same as that of the satellite. This information was obtained at too late a date to affect a change in this paper but should be taken into account for more accurate calculations. However, according to the calculations done in this paper, the electric and magnetic effects on the booms are negligible compared to the gravity-gradient effect.

Appendix

Symbols and Physical Values

Mass of electron, $m_- = 1.67 \times 10^{-27}$ kg

Mass of proton, $m_+ = 9.1 \times 10^{-31}$ kg

Electron density, $n_- \approx 2 \times 10^9/m^3$

Proton density, $n_+ \approx 2 \times 10^9/m^3$

Neutral particle density, $n_v \approx 10^9/m^3$

Charge on electron and proton, $e = 1.6 \times 10^{-19}$ coul

Velocity of satellite, $v_s \approx 6 \times 10^3$ m/sec.

Orbital angular velocity of satellite, $\omega \approx 2 \times 10^{-7}$ rev/sec.

Velocity of electrons, $v_- \approx 3 \times 10^5$ m/sec.

Velocity of protons, $v_+ \leq 10^3$ m/sec.

Linear mass density of boom, $\rho \approx 5.3 \times 10^{-4}$ slug $= 2.54 \times 10^{-2}$ kg/m

Radius of orbit (from center of earth), $r = 12{,}400$ km

Magnetic field intensity, $H = 4.5 \times 10^{-2}$ oersted

Permeability of free space, $\mu_v = 12.57 \times 10^{-7}$ henry/meter

BIBLIOGRAPHY OF ELECTROMAGNETIC EFFECTS ON SATELLITES

1. Chopra, K. P., "Interactions of Rapidly Moving Bodies in Terrestrial Atmosphere," Rev. Mod. Phys., Volume 33, pages 153-189 (1961)

2. Jastrow, R., and Pearse, C. A., "Atmospheric Drag on the Satellite," J. Geophys. Res. Volume 62, number 3, pages 413-423 (1957)

3. Beard, D. B., and Johnson, F. S., "Charge and Magnetic Field interaction with Satellites," J. Geophys. Res., Volume 65, pages 1-7 (1960)

4. Chang, H. H. C., and Smith, M. C., "On the Drag of a Spherical Satellite Moving in a Partially Ionized Atmosphere," J. Brit. Interplanet. Soc., Volume 17, pages 99-205 (1960)

5. Gringauz, K. N., and Zelikman, M. Kh., "Measurement of the Concentration of Positive Ions Along the Orbit of an Artificial Satellite," Usp. Fiz. Nauk, Volume 63, pages 239-252 (1957); Transl. in The Russial Lrt. of Satellites, Part II (Internal. Physical Index, New York, 1958) pages 133-147

6. Lehnert, B., "Electrodynamic Effects Connected with the Motion of a Satellite of the Earth," Tellus, Volume 3, pages 408-409 (1956)

7. Hohl, F., and Wood, G. P., "The Electrostatic and Electromagnetic Drag Forces on a Spherical Satellite in a Rarefied Partially Ionized Atmosphere," Rarefied Gas Dynamics; edited by J. Saurmann, Academic Press, Inc., New York (1963), pages 45-64; a shorter version appeared in NASA SP-25, pages 33-41 (December 1962)

8. Davis, A. H., and Harris, I., "Interaction of a Charged Satellite with the Ionosphere," Rarefied Gas Dynamics, edited by L. Talbot, Academic Press, Inc., (1961) pages 691-699

9. Chopra, K. P., "Review of Electromagnetic Effects on Space Vehicles," J. of Astronaut. Sci. Volume IX, pages 10-17 (1962)

10. Epstein, P. S., "On the Resistance Experienced by Spheres in Their Motion through Gases," Phys. Rev., Volume 23, pages 710-733 (1924)

11. Chopra, K. P., and Singer, S. F., "Drag of a Sphere Moving in a Conducting Fluid in Presence of a Magnetic Field," Proc. 1958 Heat Transfer and Fluid Mech. Inst., Stanford University Press, Stanford, Calif., (1958) pages 166-175

12. Wyatt, P. J., "Induction Drag on a Large Negatively Charged Satellite Moving in a Mag.-field-free Ionosphere," J. Geophys. Res. Volume 65, pages 1673-1678 (1960)

13. Licht, A. L., "The Drag on a Charged Satellite," J. Geophys. Res., Volume 65, pages 3493 (1960)

14. Jefimenko, O., "Effect of the Earth's Magnetic Field on Motion of an Artificial Satellite," Am. J. Phys., Volume 27, pages 344-348 (1959)

15. Kraus, L., and G. N. Watson, Phys. Fluids, Volume 1, page 480 (1958)

16. General Dynamics/Astronautics, Unclassified Study of Veristat Orientation for Communication Satellite, GD/A Report No. AE 62-0808 (15 Sept. 1962) pages 4.8-4.19

17. Hohl, Frank & Wood, George P., "The Electrostatic & Electromagnetic Drag Forces on a Spherical Satellite in a Rarefied Partially Ionized Atmosphere." For presentation at Third International Symposium on Rarefied Gas Dynamics, Paris, France (June 26-30, 1962)

18. Chang, H. H. C. and Smith, M. C.: "On the Drag of a Spherical Satellite moving in a Partially Ionized Atmosphere," Jour. Brit. Interplanetary Soc., Volume 17, number 7, pages 199-205 (Jan.-Feb. 1960)

DYNAMICS ANALYSIS OF THE RAE SATELLITE

Yu Chen

INTRODUCTION

The Radio Astronomy Explorer (RAE) has the mission of studying galactic electromagnetic radiation in the 1 to 10 Mc range, which is shielded from ground-based radio telescopes by the earth's atmosphere. To accomplish this a pair of V-shaped antennas, each consisting of two extensible rods, are placed back to back as shown in Figure C-31. These extensible rods, 750 to 1,000 ft. long are made of beryllium copper tape, of the "STEM" (self-storing tubular extensible member) type designed and developed by DeHaviland Aircraft of Canada, Ltd. The unit is a preformed tape which tends to coil into a long tube when unwound from the storage drum. Tapes are .002 inch to .005 inch thick and 2 inches to 5 inches wide. These long rods not only are used as antenna but also serve as attitude control devices.

It has been a known fact that the variation in the gravitational field over an earth satellite produces torques which tend to keep the axis of minimum inertia of the satellite coincident with the local vertical. Therefore, when the satellite attains its proper orientation the axis of the V-antenna, which coincides with the axis of the minimum moment of inertia, should be coincident with the local vertical.

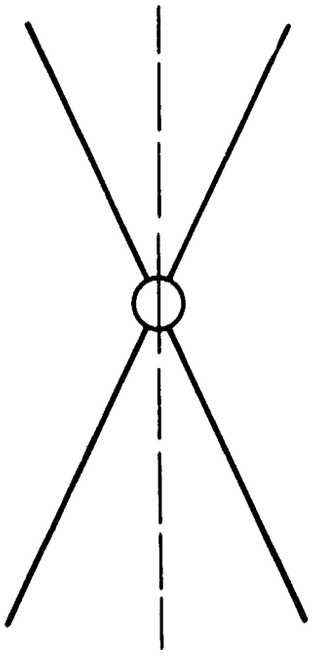


Figure C-31–V-antenna.

The initial acquisition of an earth-pointing orientation is to be achieved by a magnetic attitude control system. After the initial orientation is achieved the antenna booms are deployed and the magnetic system deactivated. The deployed booms interacting with the gravitational field cause libration of the satellite about the local vertical which is to be damped out by a damper.

Damping of vibrational motions due to gravity gradient torque or any other transient perturbation torques can be provided by the "joint-type" damper. This type of damping device is based on the connection between the satellite through a dissipative joint to an "anchor." The relative motion between the satellite and the "anchor" causes the joint to operate and generate dissipation. The "anchor" can be a system of gyros, or a second rigid body. Another type of damping utilizes the geomagnetic field as an "anchor." The motion of a satellite equipped with rods made of materials such as supermalloy causes hysteresis losses which damp out the vibration.

The primary problem in the dynamics study is concerned with the motion of the satellite about its center of gravity. Besides the torques due to gravity gradient and inertia, the booms are subjected to the influence of solar pressure, thermal bending, and other sources of disturbances such as electrodynamic interactions, etc. The flexural motion of the boom created by the total environmental excitations must be determined in order to define the instantaneous configurations of the antenna which are essential to the interpretation of the experimental data collected by RAE.

Two methods are available in establishing the equations of motion of a dynamical system, Lagrange's equation and Newton's Law. With the inclusion of an elastic body in the satellite, the motion of which when referred to proper coordinate system is directly describable by partial differential equations in the displacement variables, the latter method is quite convenient.

The system of booms and the central body are dynamically connected through the reaction forces and torques at the physical boundaries. Since the booms are essentially cantilever beams fixed to the central body of the satellite, reaction forces and torques are readily calculable once the external loadings are defined. The angular motions of the central body are governed by Euler equations and the flexural motions of the booms are describable by the dynamical beam equations.

COORDINATE SYSTEMS

Figure C-32 shows a diagrammatic view of all coordinate systems to be defined in this section.

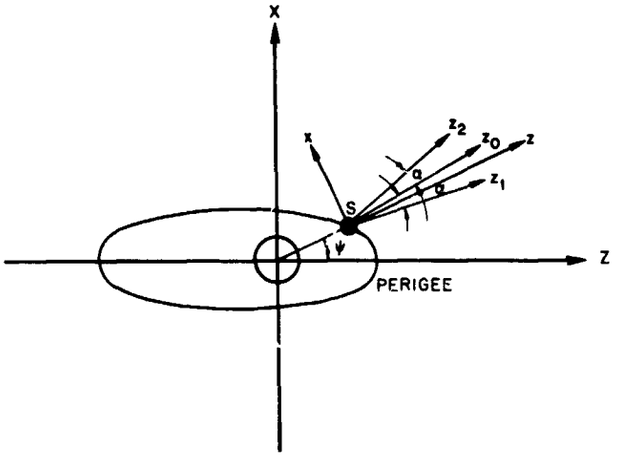


Figure C-32–Coordinate Systems .

The Inertial Systems, IS

The IS is a nonrotating system of geocentric coordinate axes O-XYZ, with its Z-axis passing through the perigee of the orbit, and its Y-axis normal to the orbital plane. The X-axis forms a right-handed triad with the Y- and Z-axes.

The Orbital System, OS

The OS is a rotating system of "sateocentric" (origin at the center of mass of the satellite, S) coordinate axes S-xyz. The z-axis coincides with OS, OS being the local vertical, making an angle ψ with OZ. The y-axis is parallel to OY. Sx is such that S-xyz is a right-handed triad.

The Central-Body System, CBS

The central-body system is a rotating sateocentric coordinate axes S-x y z_0, fixed to the satellite with Sz_0 coincident with the axis of minimum inertia I_3. Euler parameters (ξ, η, ζ, χ) are employed to the rotation of S-x_0yz_0 relative to X-syz.

The Antenna Systems, A_1S, A_2S

Two sets of rotating sateocentric coordinate axes are defined to describe the shapes of the V-antenna in orbit. S-$x_1 y_1 z_1$ is fixed to the satellite such that S_{x_1} and S_{z_1} are in the plane $x_0 S y_0$, with the positive z_1-axis coinciding with one of the undeformed leg of the V-antenna, making an angle α with Sz_0, α being the half angle of the V-antenna. S_{y_1} is the remaining axis of a right-handed triad S-$x_1 y_1 z_1$. Similarly we define the second set S-$x_2 y_2 z_2$ with Sz_2 making an angle α with S_{z_0}, etc.

Let [A] be the transformation matrix between OS and IS. The relationship between the unit vectors is

$$\begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix} = \begin{bmatrix} C\psi & 0 & -S\psi \\ 0 & 1 & 0 \\ S\psi & 0 & C\psi \end{bmatrix} \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix}, \tag{1}$$

where S and C are abbreviations of sine and cosine. Hence, we have

$$[A] \equiv \begin{bmatrix} C\psi & 0 & -S\psi \\ 0 & 1 & 0 \\ S\psi & 0 & C\psi \end{bmatrix}. \tag{2}$$

The transformation matrix [B] between CBS and OS is defined such that

$$\begin{bmatrix} \hat{x}_0 \\ \hat{y}_0 \\ \hat{z}_0 \end{bmatrix} = [B] \begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \end{bmatrix}, \tag{3}$$

where [B] is given by

$$[B] \equiv \begin{bmatrix} \xi^2 - \eta^2 - \zeta^2 + \chi^2 & 2(\xi\eta + \zeta\chi) & 2(\xi\zeta - \eta\chi) \\ 2(\xi\eta - \zeta\chi) & -\xi^2 + \eta^2 - \xi^2 + \chi^2 & 2(\xi\chi + \eta\zeta) \\ 2(\xi\zeta + \eta\chi) & 2(-\xi\chi + \eta\zeta) & -\xi^2 - \eta^2 + \zeta^2 + \chi^2 \end{bmatrix}. \tag{4}$$

The Euler parameters are related by

$$\xi^2 + \eta^2 + \zeta^2 + \chi^2 = 1. \tag{5}$$

The transformation matrix $[C_1]$ between A_1S and CBS is given by

$$[C_1] = \begin{bmatrix} C\alpha & 0 & S\alpha \\ 0 & 1 & 0 \\ -S\alpha & 0 & C\alpha \end{bmatrix}. \tag{6}$$

Similarly we define the transformation matrix $[C_2]$ between A_2S and CBS.

To transform a vector in one system to another it is only necessary to calculate the appropriate product of transformation matrices. For example, a vector in the $A_1 S$ can be transformed to IS by premultiplying it with $[A]\,[B][C_1]$, etc.

The Equations of Motion of the Central Body

We shall write dynamic equations in vector form such that they are valid in all coordinate systems. One fundamental assumption in this type of study of satellite dynamics is the separability of the motion of the center of mass from the motion about the center of mass. The long term coupling effects of the vibration of the satellite on the orbital motion of an earth satellite is negligible (1).

The rate of change of angular momentum $\vec{L}$ about the sateocenter, with respect to the CBS is governed by the equation

$$\dot{\vec{L}} + \vec{\omega} \times \vec{L} = \vec{T}, \tag{7}$$

where $\vec{\omega}$ is the angular velocity of the satellite, and $\vec{T}$ is the resultant torque about the sateocenter,

$$\vec{T} = \vec{T}_g + \vec{T}_i + \vec{T}_s + \vec{T}_t + \vec{T}_d, \tag{8}$$

where $\vec{T}_g$ is the gravity-gradient torque, $\vec{T}_i$ the torque due to inertia forces, $\vec{T}_s$ the solar pressure torque, $\vec{T}_t$ the torque due to transient thermal loading, and $\vec{T}_d$ the damping torque.

For the central body we have

$$\vec{L} = \vec{I} \cdot \vec{\omega}_s, \tag{9}$$

where

$$\vec{I} = I_1 \hat{x}_0 \hat{x}_0 + I_2 \hat{y}_0 \hat{y}_0 + I_3 \hat{z}_0 \hat{z}_0. \tag{10}$$

If $\vec{\omega}_s$ is the satellite angular velocity relative to the OS coordinates, we have

$$\vec{\omega} = \hat{y} \dot{\psi} + \vec{\omega}_s, \tag{11}$$

where ψ is the polar angle of the sateocenter, measured from the perigee in IS, satisfying the equation

$$\dot{\psi} = \Omega(1 + \epsilon \cos \psi)^2/(1 - \epsilon^2)^{3/2}, \tag{12}$$

where ϵ is the eccentricity of the orbit, and Ω is the orbital frequency, $\Omega = 2\pi/T_0$, T_0 being the orbital period.

THE TORQUE LOADINGS ON THE CENTRAL BODY

At the time of reporting only $\vec{T}_g$ and $\vec{T}_i$ have been defined. Other torques are to be worked out later.

Gravity-gradient Torque, $\vec{T}_g$

The gravity force acting on a mass particle dm at point $P(x_k, y_k, z_k, t)$, $k = 1, 2$, is given by

$$d\vec{F}_g = - \left(\frac{\mu dm}{R_p^3}\right) \vec{R}_p \tag{13}$$

where $\mu = g R_E^2$, g being the gravitational acceleration at the surface of the earth, and R_E the earth radius. R_p is the distance between 0 and P, $\vec{R}_p$ is the position vector of P, thus

$$\vec{R}_p = \vec{R}_s + \vec{r}_p \tag{14}$$

where $\vec{R}_s$ is the position vector of the sateocenter, and is computed from the orbital differential equation

$$\ddot{\vec{R}}_s = \mu \frac{\vec{R}_s}{R_s^3}, \tag{15}$$

and

$$\vec{r}_p = x_k \hat{x}_k + y_k \hat{y}_k + z_k \hat{z}_k. \tag{16}$$

Let the linear density of the boom be ρ, we have

$$dm = \rho (1 + x_k'^2 + y_k'^2)^{1/2} \, dz_k \tag{17}$$

where the prime denotes differentiation with respect to z_k'. Hence Equation (13) reduces to

$$d\vec{F}_g = - \left(\frac{\mu \rho}{R_p^3}\right) (1 + x_k'^2 + y_k'^2)^{1/2} \, dz_k \, \vec{R}_p. \tag{18}$$

The elementary torque due to $d\vec{F}_g$ about the sateocenter is given by

$$d\vec{T}_g = \vec{r}_p \times d\vec{F}_g,$$

and the total torque is obtained by integrating $d\vec{T}_g$ between 0 and L. Hence we have

$$\vec{T}_g = \mu \rho \int_0^L \frac{(1 + x_k'^2 + y_k'^2)^{1/2}}{R_p^3} \vec{r}_p \times \vec{R}_p \, dz_k, \quad k = 1, 2. \tag{19}$$

In the study of deployment of the boom, the length L is a predetermined function of time t.

Inertial Torque, $\vec{T}_i$

The inertial load at P applied to dm can be written as

$$d\vec{F}_i = - dm \, \ddot{\vec{R}}_p \tag{20}$$

where

$$\vec{R}_p = \vec{R}_s + \vec{r}_p,$$

which upon differentiating twice yields

$$\ddot{\vec{R}}_p = \ddot{\vec{R}}_s + \ddot{\vec{r}}_p + 2\vec{\omega} \times \dot{\vec{r}}_p + \dot{\vec{\omega}} \times \vec{r}_p + \vec{\omega} \times (\vec{\omega} \times \vec{r}_p). \tag{21}$$

The torque $\vec{T}_i$ is given by

$$\vec{T}_i = -\int_0^L \vec{r}_p \times \ddot{\vec{R}}_p \left[\rho(1 + x_k'^2 + y_k'^2)^{1/2}\right] dz_k, \quad k = 1, 2. \tag{22}$$

THE DEFLECTIONS OF THE BOOMS

Since bending is the major contribution of the deflections, deflections due to two-plane bending will be the basis of the calculation. The deflection x_k is describable in Euler-Bernoulli equation of beam with curvature correction as follows.

$$EI_b \frac{\partial^2 x_k}{\partial z_k^2} = M_{y_k} \left[1 + \left(\frac{\partial x_k}{\partial z_k}\right)^2\right]^{3/2} \tag{23}$$

and

$$\frac{\partial^2 M_{y_k}}{\partial z_k^2} + \rho \frac{\partial^2 x_k}{\partial t^2} = p_{y_k}, \quad k = 1, 2.$$

where the subscript y_k is used to specify the plane of bending, that being normal to $\hat{y}_k$.

The deflection y_k is governed by a pair of equations

$$EI_b \frac{\partial^2 y_k}{\partial z_y^2} = M_{x_k} \left[1 + \left(\frac{\partial y_k}{\partial z_k}\right)^2\right]^{3/2}$$

$$\frac{\partial^2 M_{x_2}}{\partial z_k^2} + \rho \frac{\partial^2 y_k}{\partial t^2} = p_{x_k}, \quad k = 1, 2. \tag{24}$$

Since both Equation (23) and Equation (24) are actually stating the equilibrium of a beam in the $\hat{z}_k$-direction, the load intensity is taken to be in the direction normal to $\hat{z}_k$. Hence we have

$$P_{y_k} = (\vec{P}_g + \vec{P}_i) \cdot \hat{x}_k, \tag{25}$$

and

$$P_{x_k} = (\vec{P}_g + \vec{P}_i) \cdot \hat{y}_k, \tag{26}$$

where $\vec{P}_g$ is given by

$$\vec{P}_g = -\left(\frac{\mu\rho}{R_\beta^3}\right) (1 + x_k'^2 + y_k'^2)^{1/2} \vec{R}_p \tag{27}$$

and $\vec{P}_i$ is given by

$$\vec{P}_i = - \rho(1 + x_k'^2 + y_k'^2)^{1/2} \ddot{\vec{R}}_p . \tag{28}$$

Numerical integration schemes are at present under investigation to solve Equation (23) and Equation (24).

SURVEY REPORT ON POSSIBLE ANTENNA CONFIGURATIONS FOR A RADIO ASTRONOMY SATELLITE

J. A. DellaValle

Because radio signals from celestial sources below 10 Mc are strongly attenuated in the earth's atmosphere, radio astronomers wishing to investigate such sources would like to place an antenna above the atmosphere. Such a project is RAE, Radio Astronomy Explorer, which will put an antenna-bearing satellite into orbit about the earth.

The antenna used for this mission should have the following properties:

1. unidirectional
2. constant directivity and gain over the 1-10 Mc range
3. mechanically simple and easy to deploy
4. electrical properties not influenced by the environment.

The antennas listed below are briefly described in this report as possible configurations for the project:

1. dipole
2. array of dipoles
3. single long wire
4. V-antenna
5. rhombic
6. parasitic array
7. Yagi-Uda array
8. helical
9. log periodic
10. fishbone.

In the following discussions we are interested in the receiving pattern of the antennas. However, since the receiving pattern is the same as the transmitting pattern due to the reciprocity principle and the latter is easier to determine, we will discuss the antennas in terms of their transmitting pattern.

DIPOLE ANTENNA

The center-fed quarter wave or half wave dipole antenna has a figure 8 pattern as shown in Figure C-33. The radiating dipole is a classical problem in electromagnetic theory. The dipole fields are given by

$$\vec{B} = k^2(\vec{n} \times \vec{P}) \frac{e}{r}^{ikr} \left(1 - \frac{i}{ikr}\right).$$

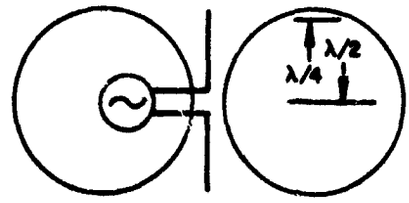
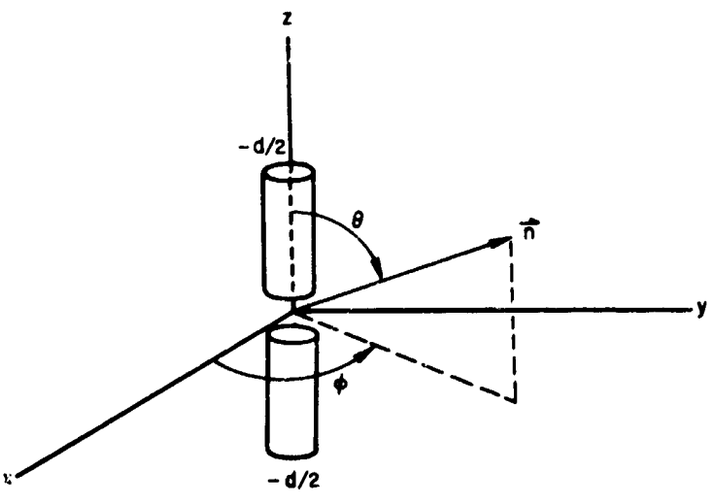


Figure C-33–Simple dipole antenna.

$$\vec{E} = k^2(\vec{n} \times \vec{P}) \times \vec{n} \frac{e^{ikr}}{r} + \left[3\vec{n}(\vec{n} \cdot \vec{P}) - \vec{P}\right] \cdot \left(\frac{1}{r^3} - \frac{ik}{r^2}\right) e^{ikr},$$

where $k = 2\pi/\lambda$ and $\vec{P}$ is the dipole moment. The coordinates are shown below. We note that the magnetic induction is transverse to the radius vector at all points, but the electric field has components parallel and perpendicular to $\vec{n}$ (1). The radiation pattern has rotational symmetry. Since the simple dipole is not very directional it is not very desirable for our purposes.



DIPOLE ARRAY

By properly arranging several dipoles we can construct a very directive, high gain antenna. The simplest array is two identical dipoles separated by a distance, measured perpendicular to their respective axes. Three cases are of interest:

1. When the elements are driven in phase with equal currents.
2. When the two are driven 180° out of phase with equal currents.
3. When the elements are driven with equal currents and with a phase difference equal to 180° less the element spacing expressed in electrical degrees.

The third case is of most interest to us because of the directivity of the pattern (2) as shown in Figure C-34. This pattern results from the familiar phenomenon of interference

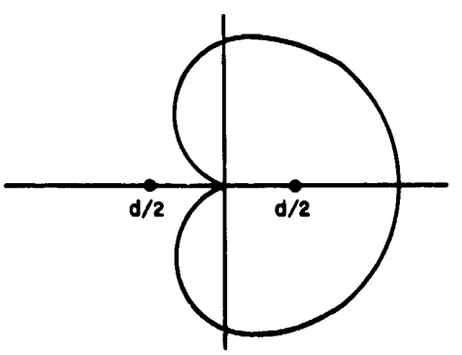
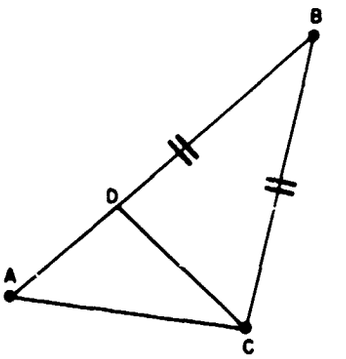


Figure C-34–Elements have equal currents but differ in phase so that $\Delta\phi = 180° - d$.

which is common to all wave motion. In the figure below points A and B are sources of waves. At point C the waves from A and B meet. Since DC = BC, if AD is half a wavelength the waves from the two sources are 180° out of phase at C and cancel each other. If AD is zero or an integral multiple of a wavelength, then the waves are in phase at C and reinforce each other. A phase relationship intermediate between these extremes results in an intermediate intensity.



A multi-element array can also be formed with the radiation pattern determined by the relative phase and spacing of the elements. When the currents in the arrays are all in phase and of the same magnitude the array is known as a broadside array, since the main beam will be perpendicular to the line of the array. When the phase progression between elements is equal to the element spacing in electrical degrees, the array is known as an end fire array, since the main beam is directed along the line of the array. One can also excite the elements in such a way that they have different currents. This is known as a tapered array.

The broadside array produces the greatest gain for a given length; it also produces relatively high side-lobe levels. It is possible, in principle, to reduce the side-lobe levels of an array to as low as is desired by choosing an appropriate current distribution along a linear array. This is obtained at the expense of reduced gain and larger beamwidth (3).

Further details concerning arrays of discrete elements will not be considered since it is not a practical system for a spacecraft. The main reason for this is that for a directional beam the spacing between elements and the phase and magnitude of their currents are all critical parameters determining the properties of the array. If one is transmitting all these parameters can be adjusted to obtain the desired beam. However, the antenna for RAE must be a broadband receiving antenna and all these parameters cannot be adjusted to give optimum operation for every frequency; hence, an array of dipoles is not well suited for RAE. Mechanical deployment presents other problems which would be very complicated to handle for a dipole array.

SINGLE LONG WIRE RADIATOR

We consider now radiation from a straight wire of any length carrying a sinusoidal current. We examine the cases in which one end of the wire is free and the waves are reflected, and when one end is terminated in a resistance equal to the characteristic impedance of the wire so that all the power is absorbed in the resistor and none reflected. In the first case we have standing waves and in the latter travelling waves. We note the following basic facts (2) about the radiation pattern of long wires isolated in free space with idealized natural currents:

1. There is a lobe in the radiation pattern for each half wavelength of wire. Each lobe is a cone of radiation centered on the wire.

2. With respect to the middle of the wire as center of the polar radiation pattern half of the lobes are tilted forward and half of them backward.

3. The direction of the electric field reverses in each successive lobe. This is analogous to the reversals of phase of the currents in successive half-wavelength portions of the wire.

4. Between successive lobes are regions of little or no radiation called nulls or zeros. Practically all these are minima because the field strength at these angles never

goes to zero. These zeros are related analogously to the current zeros in the standing wave current distribution or to the phase reversals in a travelling wave distribution.

5. The angles of the zeros are symmetrically distributed about the plane normal to the middle of the wire.

6. The largest lobe in the radiation pattern is the one forming the smallest angle to the wire. For the travelling wave case, this lobe is in the direction of current flow only. For the standing wave case the radiation pattern is symmetrical with respect to the middle of the wire, and so has a complementary major lobe toward the other end of the wire.

7. With travelling waves the field-strength pattern has lobes of diminishing amplitude, the smallest being in the direction opposite the current.

Figure C-35 shows the radiation pattern for the cases described above. Though the single long wire is mechanically simple and has good gain properties, neither the terminated nor the unterminated wire is unidirectional; hence they lack the most important requirement for an RAE antenna.

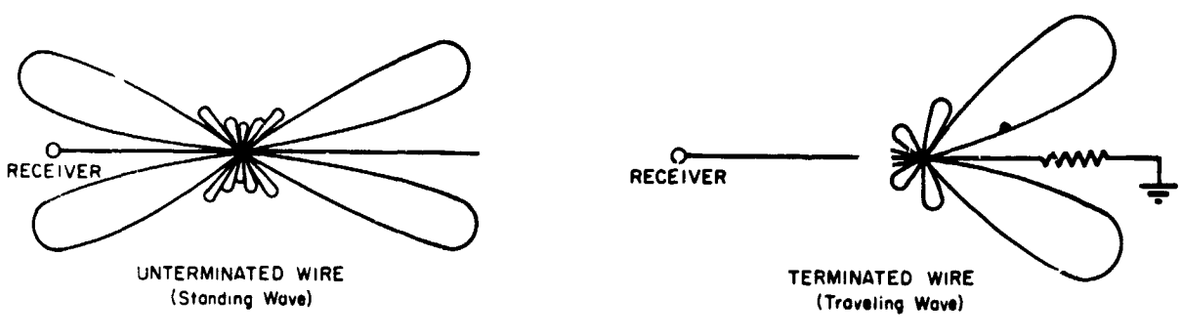


Figure C-35–Radiation Pattern

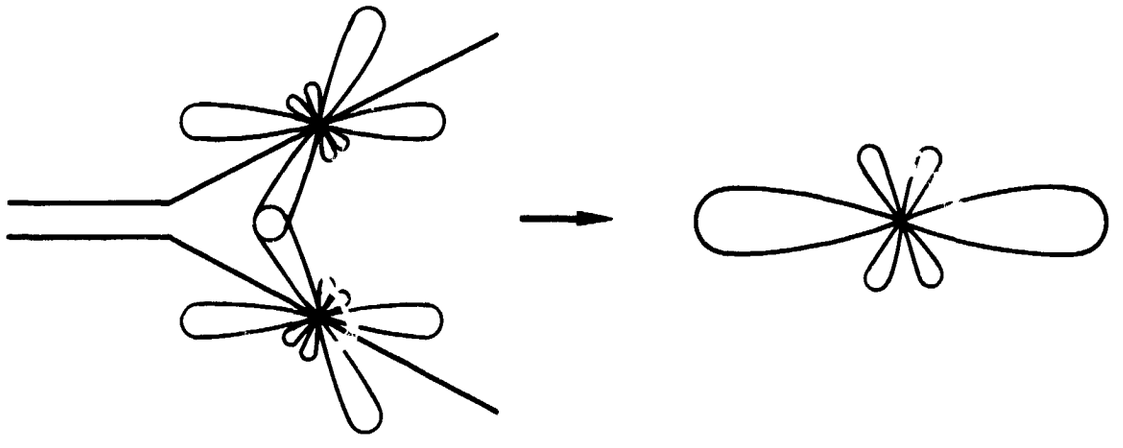
V-ANTENNA

The best use of long wires is realized in those antennas where the main lobes from two or more long wires add in phase and other lobes are reduced or eliminated by constructive wave interference. Two of these configurations are the rhombic and the V-antennas.

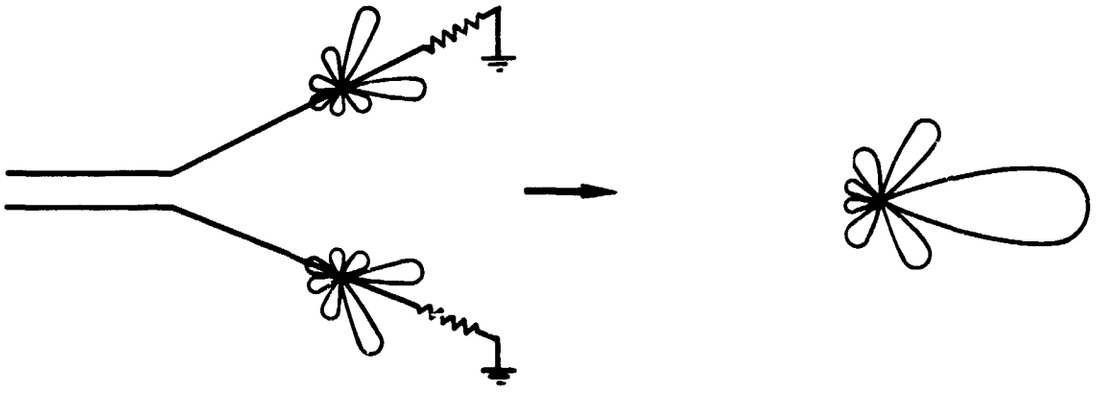
As with the straight wire, we may have travelling waves or standing wave situations (sometimes referred to as non-resonant and resonant cases). For the travelling wave case the ground is usually achieved by an actual wire connection to ground. This is impossible for an orbiting antenna but an equivalent method is to place the resistor an odd number of quarter wavelengths from the end. The pattern for the travelling wave case is unidirectional as shown in Figure C-36. More detailed properties of the V-antenna in the 1-10 Mc range have been given by the Avco Corporation (4).

RHOMBIC ANTENNA

A diamond or rhombic antenna is made with four wires terminated in a matched load; the radiation pattern is shown in Figure C-37.



UNTERMINATED "V"



TERMINATED "V"

Figure C-36–Terminated and unterminated V antenna patterns.

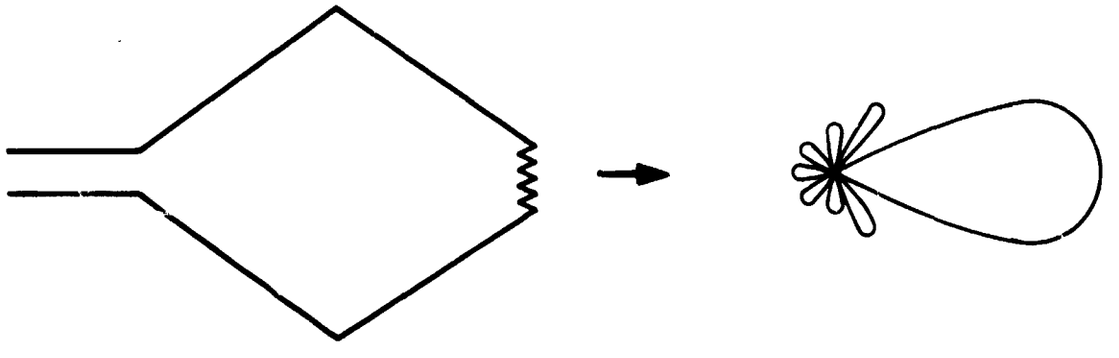


Figure C-37–Non-resonant rhombic antenna.

Rhombic antennas have relatively high gain and broad frequency response. The power gain will be approximately 8 times that of a wire of length equal to one side. The contributions from the four sides will add in phase, but the input impedance will be doubled. Attenuation reduces the contributions from the terminated sides, and in practice a smaller value is found – on the order of 6 times a single wire (3).

It seems that a rhombic antenna would be a satisfactory antenna for RAE if some means could be found to deploy it.

PARASITIC ARRAYS

Because of feedline problems it is sometimes desirable to feed certain antennas of an array parasitically. In these cases, the currents appearing in the several, parasitic antennas are derived by radiation from a nearby driven antenna. A particularly simple case is a driven antenna together with a single parasite. The magnitude of the current in the parasite falls off rapidly with separation from the driven element, hence there is a tendency toward small separations. On the other hand the phase of the parasitic current relative to the driven element increases with separation so that in order to obtain the desired phase change, it may be necessary to operate with considerable spacing – on the order of a quarter or half wavelength. Often there is a need to compromise between these two variables, thereby making their separation somewhat less than might be expected. A few polar diagrams are shown in Figure C-38 to illustrate the principle.

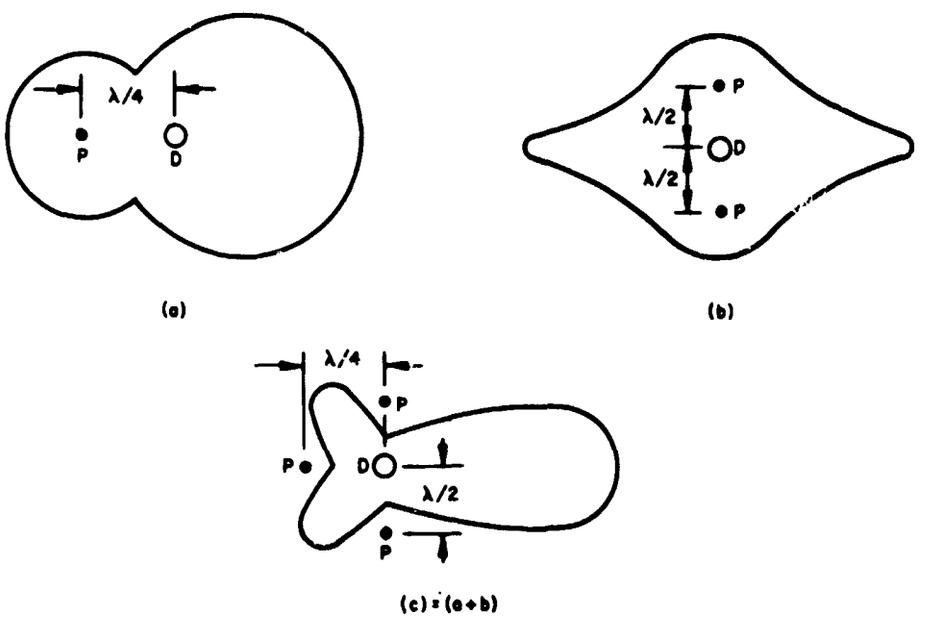
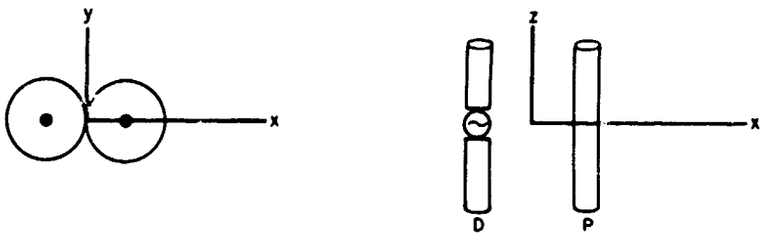
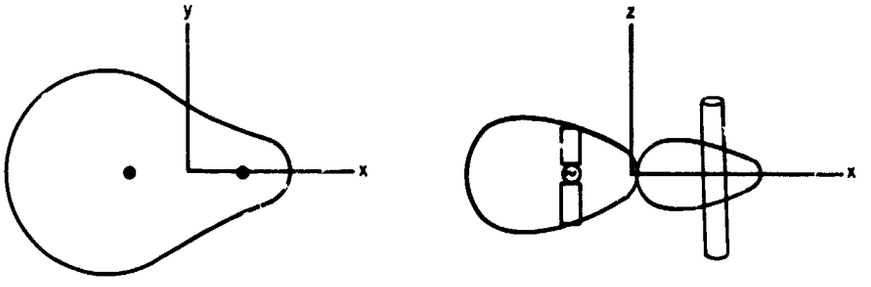


Figure C-38–Some polar diagrams.

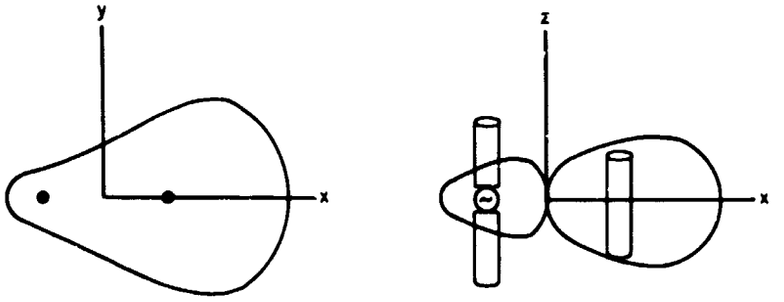
In addition to the nominal quarter and half-wave spacings between a driven antenna and its parasite, an interesting result may be obtained from relatively close spacings, especially when the parasite is detuned relative to its normal free-space resonant length. Figure C-39 represents patterns to be expected from a driven antenna together with a single closely spaced parasite.



(a) BOTH ELEMENTS IN PHASE



(b) PARASITE LENGTHENED BY 5% TO ACT AS REFLECTOR



(c) PARASITE SHORTENED BY 5% TO ACT AS DIRECTOR

Figure C-39–Patterns to be expected from a driven antenna together with a single closely spaced parasite.

In (a) the current in the parasite is substantially that in the driven element, and since they are oppositely phased there is a null along the y-axis as well as the expected null along the z-axis. Though the front-to-back ratio is unity, there is nevertheless a directive gain of 3 db relative to a single element.

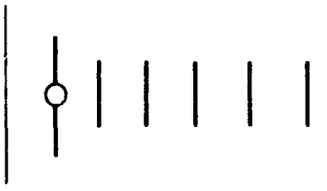
By increasing the length of the parasite 5%, still keeping the spacing the same, the pattern is changed to that in (b). The parasite acts now as a reflector giving a directive gain approaching 6 db. By detuning the parasite an approximately equal amount in the opposite sense, the pattern will be essentially the same but will be reversed in direction. In this case the parasite is called a director. As might be expected both a reflector and a director may be combined with a driven antenna to form a three element array, but because of interaction between elements the gain now falls somewhat short of that to be expected. A gain of 8 db is representative (2).

YAGI ARRAYS

It is found from experiment that little is gained by adding more reflectors to a parasitic array but that adding numerous directors increases the directivity and gain. When several directors are added we have what is called the Yagi antenna. Each director receives its power parasitically from the next preceding director, and each in its turn radiates a portion of its power. The respective currents are progressively less as we proceed along the array, and there is a measure of tapering. Also there is a practical limit to the total number of elements; usually less than a dozen should be used, though arrays of 30 and 40 have been built. A total length of 6 wavelengths is taken as a reasonable limit.

The lengths of the reflector and various directors, as well as their respective spacings, are for the most part compromises with such factors as gain, input impedance front-to-back ratio, magnitude of minor lobes, and bandwidth. Spacings range from 0.15λ to 0.30λ or even 0.40λ are representative (2).

The Yagi antennas have a very directive beam and are much more compact than broadside arrays, but suffer from larger sidelobes. An example of a polar diagram obtained from a multi-element Yagi antenna is shown below (3).



(a)

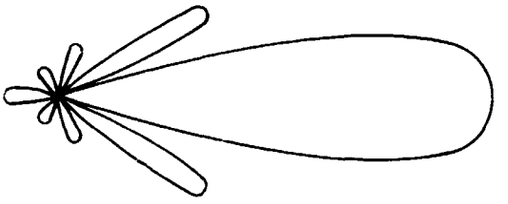
YAGI ANTENNA WITH

DIRECTOR SPACING = 0.34λ

OVERALL LENGTH = 10.5λ

DIRECTOR LENGTH = 0.4λ

REFLECTOR SPACING = 0.30λ



(b)

YAGI RADIATION PATTERN IN

PLANE OF ELEMENTS

Theoretically the maximum power gain of a long Yagi of length ℓ is given approximately by

$$G = 9.2\,\ell/\lambda$$

In practice a gain of only about $5\ell/\lambda$ is found. A director length of about $.4\lambda$ is recommended depending on the number of directors in use. A reflector exactly half a wavelength long is normally used, spaced a quarter wavelength behind the driven element.

The main advantages of the Yagi are its compactness and lightness, but these are countered by the disadvantages of low input impedance and narrow bandwidth. Both these quantities may be comprised against other factors to moderate the difficulty. Bandwidths of 2% are representative (3).

The bandwidth of a Yagi is a complex problem to analyze. Qualitatively, the narrowness of the bandwidth is due to the fact that for a given Yagi, as the frequency increases, both the relative spacings and element lengths increase, while the optimum element length decreases with the increase of spacing. Although a capacitor may be added to effectively shorten the length, there is no passive reactance network with a capacitive reactance increasing with frequency. Therefore, the narrow bandwidth is an inherent property of Yagi antennas.

The foregoing argument is based upon the optimum gain condition. If, however, the directors are deliberately made shorter and the reflector longer, the bandwidth can be considerably widened with sacrifice in peak gain (2). There are, in general, three classes of broadband Yagis:

1. Dual-channel – covers two channels which in most cases are in the same band but are not adjacent.

2. Single band – those which receive two are more adjacent in one band.

3. high and low band – all VHF channels.

Details for constructing these broadband Yagis can be found in Jasiks' Antenna Engineering Handbook (2).

HELICAL ANTENNA

The helical antenna produces the highest directivity possible for an antenna of its size and does this over a considerable bandwidth. Depending on the frequency of operation or the dimensions of the helix, the beam can be omnidirectional or axial (see Figure C-40). The antenna operates as an end-fire beam antenna generating circularly polarized waves.

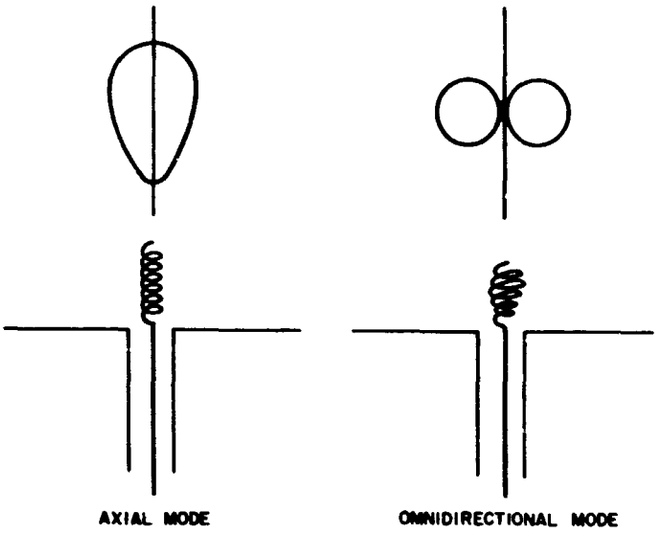


Figure C-40–Helical antenna.

Jasik describes an optimum helix as one having a pitch angle of 14°, an overall axial length of 1.65λ and a conductor diameter of 0.017λ at the center frequency. He finds that a pitch angle of 14° is close to optimum for helices that are considerably shorter or longer or of somewhat different conductor diameter. An optimum helix having 6 turns and a 14° pitch has been constructed. The field pattern measurements show that the axial mode of radiation occurs for frequencies between 290-500 Mc. The bandwidth at the half power points is given by

$$\beta = \frac{52}{\frac{c}{\lambda}\sqrt{\frac{ns}{\lambda}}} \text{ degrees,}$$

where

λ = free space wavelength,

s = spacing = c tan α,

c = circumference,

n = number of turns,

and the gain is given by

$$G \approx 15 \left(\frac{C}{\lambda}\right)^2 \frac{ns}{\lambda}$$

The frequency range for axial mode operation (290-500 Mc) is far from that desired in RAE (1-10 Mc). Also, the optimum length of 1.68λ is, for say 1 Mc, approximately 450 meters making it rather complicated to deploy.

LOG PERIODIC

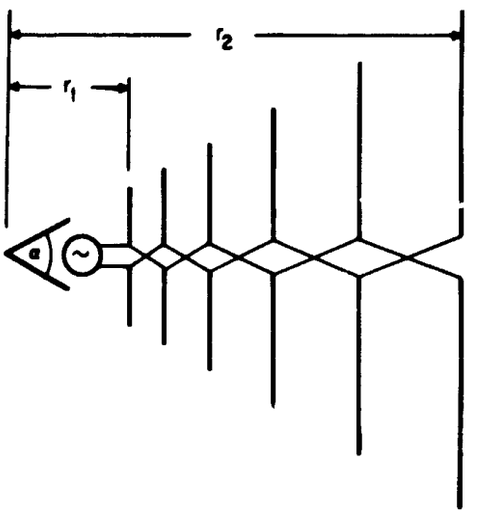
Recent advances in antennas have shown how it is possible to make antennas which are essentially frequency independent. One of the ideas evolves from the observation that the properties of an antenna, impedance and pattern, are determined by its shape and dimensions. If by an arbitrary scaling the antenna is transformed into a structural equal to the original one (measured in wavelengths) its properties will be independent of the frequency of operation. The antenna then satisfies the angle condition, which means that its form can be specified entirely by angles only and not by any particular dimension. There are two classes of antennas satisfying this condition. Conical antennas made up of infinite cones of arbitrary cross-section having a common apex, and equiangular antennas having their surfaces generated by equiangular spirals with a common apex and the same defining parameter.

The second idea is that if a structure becomes equal to itself by a particular scaling of its dimensions by some ratio, τ, it will have the same properties at a frequency f and at the frequency τf. It follows from this that the impedance or any other characteristic, is a periodic function, with period log τ of the logarithm of the frequency. The antennas obtained from this principle are called log periodic. By making τ close enough to unity, good frequency-independent behavior is observed (2).

The antennas using the equiangular principle may be planar or non-planar. The planar structures take the form of a spiral and their radiation pattern is bidirectional. The non-planar antenna results when the spiral is wrapped on the surface of a cone. The resulting pattern becomes unidirectional, with the major lobe in the direction of the apex. For both cases the radiation is practically circularly polarized over the entire frequency band.

The geometry of a log periodic antenna structure is chosen so that the electrical properties must repeat periodically with the logarithm of the frequency. A recent and rather simple type of log-periodic antenna is shown on the following page. It consists of an array of dipoles with lengths and spacings arranged in a log periodic manner. The dipoles are excited by a uniform two-wire line with the line transposed between adjacent dipoles. The antenna may be fed with a balanced line connected to the feed point or a coaxial line entering from the right and passing through and inside one of the balanced lines to the feed point. This is illustrated on the following page.

To be truly frequency independent the antenna must be infinitely long. Hence the lengths r_1 and r_2 determine the upper and lower cutoff frequency and the degree to which the antenna is frequency independent. Physical reasoning leads to the conclusion that log periodic antennas are unidirectional as indicated by Figure C-41 which shows the pattern for the antenna on this page.



An antenna very similar to the log periodic dipole array is the trapezoidal tooth antenna of Figure C-42. Array of log periodic antennas can also be constructed. The pattern of an array of periodic elements may be considered as the simple superposition of the patterns of the individual elements if it is assumed that the presence of other elements does not affect the pattern of an element. A schematic representation of a two element array is given in Figure C-43. The radial lines separated by the angle ψ represent the two elements and d is the distance to the phase center (origin of radiation pattern). The direction to a distant field point is given by ϕ. Typical element patterns are shown in the schematic. Figure C-44 below gives the E-plane patterns for a wire trapezoidal tooth structure (2).

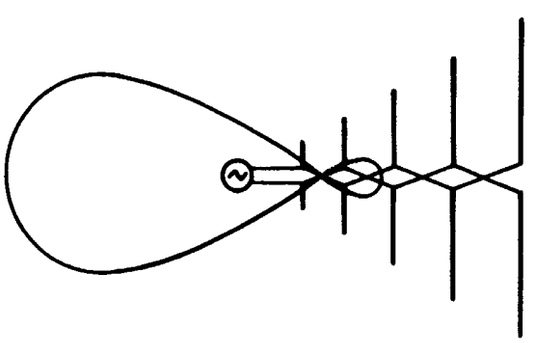


Figure C-41–Antenna field pattern.

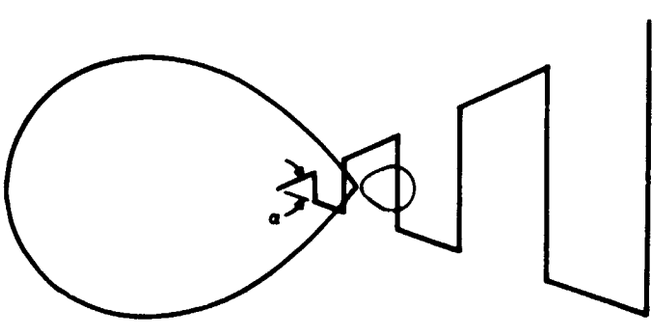


Figure C-42–Log periodic trapezoidal tooth antenna.

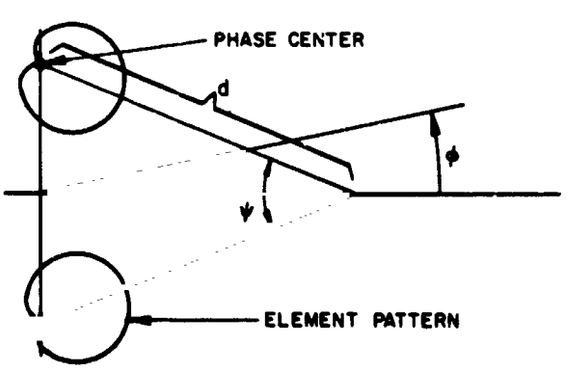


Figure C-43–Schematic representation of a two element array.

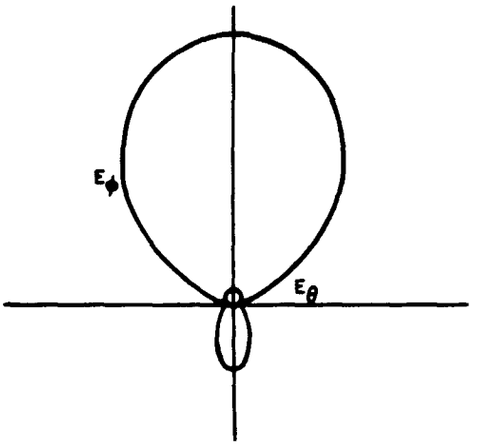


Figure C-44–E-plane patterns for a wire trapezoidal tooth structure with $\alpha = 60°$, $\tau = 0.6$, $\psi = 35°$, f = 60° Mc (from Jasik).

FISHBONE ANTENNAS

A good receiving antenna is the fishbone. This antenna is very effective in discriminating against noise and functions over a broadband of frequencies differing by as much as 2:1. Because of its inefficiency, it is seldom used at transmitting stations. The fishbone antenna consists essentially of a two wire transmission line, perhaps four wavelengths long, to which are attached at closely spaced intervals horizontally disposed lateral wires, each of the order of a quarter wavelength long. Often these wires are coupled to the transmission line by small capacitances as shown in Figure C-45.

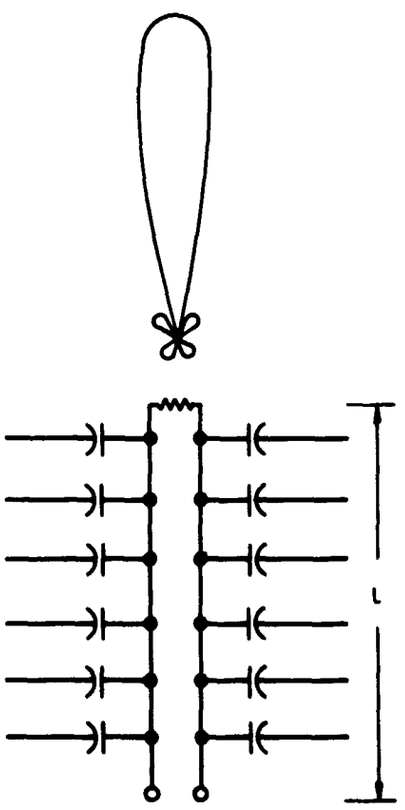


Figure C-45–The fishbone receiving antenna and the horizontal pattern for the two bay fishbone antenna (from Jasik).

The gain relative to an isotropic radiator of a fishbone antenna of length ℓ is given (2) approximately by

$$G = 4 \rho AB,$$

where

$\rho = \ell/\lambda$, and

A, B are experimentally determined parameters.

This concludes the survey of antennas which could possibly be used on project RAE. The log periodic antennas have the most desirable electrical properties, i.e., broad bandwidth and unidirectionality, and for this reason are the most desirable. However, the log periodic antennas are also complicated geometrically and thus hard to deploy. The V-antenna seems to be the best all-round antenna for a satellite since it is unidirectional rather broadbanded, and mechanically simple.

REFERENCES

1. Jackson, J. D., Classical Electrodynamics, New York:John Wiley & Sons, Inc. (1962)

2. Jasik, Henry, Antenna Engineering Handbook, New York: McGraw-Hill (1965)

3. Smith, R. A., Aerials for Metre and Decimetre Wavelengths, Cambridge (Eng.):The University Press (1949)

4. Avco Corporation, Progress Report on RAE presented at Goddard Space Flight Center (July 8, 1965)

BIBLIOGRAPHY

American Radio Relay League (ARRL) Antenna Book, West Hartford, Conn.

Jordan, E. C., Electromagnetic Waves and Radiating Systems, Prentiss Hall, Inc., Englewood Cliffs, New Jersey (1950)

King, R. W. P., The Theory of Linear Antennas, Harvard University Press, Cambridge, Mass. (1956)

Kraus, J. D., Antennas, McGraw-Hill, Inc., New York (1950)

Ramo, S., and J. R. Whinnery, Fields and Waves in Modern Radio, John Wiley & Sons, Inc., New York (1953)

Schelkunoff, S. A., Früs, H. T., Antennas: Theory and Practice, John Wiley & Sons, Inc. New York (1952)

Williams, H. Paul, Antenna Theory and Design, Sir Isaac Pitman & Sons, Ltd., London (1950) 2 volumes

SUMMER WORKSHOP 1965

Program Outline and Team Participants

PROJECT D: Magnetic Fields and Plasma Studies

Study Topics

D-1 Plasma chambers and scaling

D-2 A static theory of a beam generated plasma in a cylinder

D-3 Various means of generating plasmas

D-4 Ionosphere simulation through gas discharges in larger volumes

D-5 Measurements in a plasma

D-6 Magnetic properties of materials

TEAM A

Academic Personnel

Dr. D. L. Waidelich
Principal Investigator
Dr. D. H. S. Cheng
Dr. H. W. Jones
Rev. J. H. Kinnier
Mr. J. P. Boright
Mr. A. V. Dralle

Goddard Personnel

Robert E. Gebhardt
Staff Advisor
Thomas L. Aggson
Donald R. Burrowbridge
Dr. K. W. Ogilvie
Dr. A. C. Aiken
Robert E. Bourdeau
Joseph C. Boyle
Donald E. Frericks
Dr. J. P. Heppner
William M. Jackson
James A. Munford
Gideon P. Serbu
Jesse E. Stern
Dr. R. G. Stone

NSF Summer Science Student

James B. Moore

PROJECT D: MAGNETIC FIELDS AND PLASMA STUDIES

CONTENTS

ILLUSTRATIONS

TABLES

PLASMA CHAMBERS AND SCALING

D. L. Waidelich

INTRODUCTION

Team D of the 1965 Goddard Summer Workshop was given three main problems for investigation. The first was concerned with the simulation in the laboratory of plasmas such as those found in the ionosphere, magnetosphere and interplanetary space. Such simulated plasmas would be useful in the testing of spacecraft and of instruments and components carried by a spacecraft. Information was sought on the characteristic properties of the plasmas to be simulated, on plasma experiments which have been performed in large chambers at low pressures, and on the possibilities of scaling plasma experiments so that smaller dimensions for the chambers could be employed. A theoretical study was started on the generation of a plasma in a cylindrical geometry by a beam of electrons. An investigation was made of the possible ion sources that might be useful in the simulation of plasmas. Experimental studies of the screen cathode and brush-cathode dc discharges and RF discharges in a bell jar are reported. Also, the results of RF discharge and screen-cathode dc discharge work in a large chamber are presented. The various methods of measuring the properties of a plasma were considered especially with regard to probe methods. A number of experimental results obtained with probes in a bell jar are presented.

The second problem was that of determining the magnetic properties of nearly nonmagnetic material at low temperatures and at low magnetic field strengths. This information would be useful in the construction of spacecraft and components which are to be used in magnetic field measurements in space or which should be made in such a way as to have a low magnetic torque while in orbit near the earth. The methods of measurement and some results of the measurements on typical alloys are presented. The third main problem was that of assembling a handbook on the production of large volumes of uniform magnetic fields for the magnetic testing of spacecraft and components. An annotated bibliography of the work done and a presentation of some of the most useful information was assembled. This work was not completed but an interim report will be made and published separately as a supplement to this document.

There is a need for the capability of producing laboratory plasmas to simulate those of the ionosphere and magnetosphere of the earth interplanetary space and the regions surrounding the planets and other bodies of the solar system. Such a facility should prove very useful in the calibration of instruments such as electron and ion probes and electric field meters used in testing the component parts of a satellite and of the whole satellite while immersed in a plasma, and in the performance of experiments, i.e., antenna characteristics and radio wave propagation in a plasma.

To obtain some idea of the electron and ion energies, electron and ion densities to be simulated, a number of papers and books on the characteristics of the ionosphere, magnetosphere and interplanetary space were consulted (1–13). No attempt was made to summarize this large amount of material, but it was felt that one goal should be to try to achieve a low electron temperature in the plasma, of about 1500°K. Another goal was to obtain a range of charged particle densities of say 10^9 to 10^5 particles per cubic centimeter. It was also felt that pressures in the vicinity of one torr or less should be considered. Other objects were to try to achieve as uniform a plasma as possible and also to look for the effects of different gases. The results of the investigation to find

reports of experiments in large plasma chambers with low pressures is presented here. Also the results of a start in the investigation of plasma scaling is given.

PLASMA CHAMBERS

A search was made for plasma experiments carried out in rather large chambers and particularly those concerned with low pressures and low charged particle densities. Two experimental investigations of the interaction of a plasma on a body immersed in a plasma (14-16) were carried out at low pressures of the order of 10^{-6} torr although the chambers themselves were relatively small in size. The production of a plasma by means of ultraviolet sources (17), (18) was also considered and later a large chamber (19) was constructed for experimental work. Plasma experiments in a large chamber (2) were performed to simulate the plasma wind using various gases in a 4 by 8 foot chamber (21) with a back pressure of 2×10^{-6} torr. Work on the neutralization of the beam from an ion thrustor engine (20), (22) had been carried out in a smaller chamber of dimensions 2 by 4 meter, with a pressure of about 10^{-5} torr and in a larger chamber of dimension 8 by 25 meters. An electric field meter was also tested in a plasma beam in a vacuum chamber (20).

PLASMA SCALING

If it were possible to perform plasma experiments on a smaller scale than that of the actual happenings, the experiments would be easier to perform and less expensive. A search was made for relationships between the parameters of a plasma that would aid in scaling the experiments. One such relationship is that of a law of similitude (23) in which

$$\frac{P_1}{P_2} = \frac{d_2}{d_1} = \frac{E_1}{E_2}$$

where P is the pressure, d is a dimension and E is the electric field intensity.

Since there seemed to be quite a number of references on plasma scaling, some bibliographies on this subject were sought and two of these were found. The first one (24) has 96 references which pertain to a wide variety of the applications of plasmas to the modelling of geophysical phenomena. The second (25) has 39 references specifically concerned with the scaling of plasmas.

The effect of a change in plasma density on a model study of plasma-vehicle interaction (16) has been considered. The problem of scaling for a model study of the electric drag on a satellite (14) has been discussed. The use of plasma scaling in model experiments of cosmical phenomena (26), (27) and in re-entry simulation (28) has been presented. A lack of time prevented further work in this direction.

REFERENCES

1. Parker, E. N., Interplanetary Dynamical Processes, Wiley (1963)

2. Brundin, C. L., Effects of Charged Particles on the Motion of an Earth Satellite, AIAA Journal, Vol. 1, pages 2529 to 2538, (November 1963)

3. Aarons, J., Radio Astronomical and Satellite Studies of the Atmosphere, Wiley (1963)

4. Chapman, S., Solar Plasma, Geomagnetism and Aurora, Gordon and Breach (1964)

5. Thrane, E. D., Editor, Electron Density Distribution in Ionosphere and Exosphere, Wiley (1964)

6. Rossi, B., Editor, Space Exploration and the Solar System, Academic Press (1964)

7. Bleil, D. F., Editor, Natural Electromagnetic Phenomena Below 30 kc/s, Plenum Press (1964)

8. Bourdeau, R. E., "Research within the Ionosphere," Science, Volume 148, pages 585 to 594, (30 April 1965)

9. Gringauz, K. I., "Study of Interplanetary Gas and Ionospheres of Planets with the Help of Traps of Charged Particles," NASA STAR N65-16468

10. Johnson, F. S., "The Interplanetary Medium," IAA A65-14378

11. Bourdeau, R. E. et al, "Ionospheric Research by Means of Rockets and Satellites," IAA A65-22384

12. Seikel, G. R., "Plasma," Space/Aeronautics, Volume 43, pages 40 to 45, (April 1965)

13. "Data Book for Environmental Testing and Spacecraft Evaluation," Volume I, Test and Evaluation Division, Goddard Space Flight Center, Greenbelt, Md.

14. Knechtel, E. D. and Pitts, W. C., "Experimental Investigation of Electric Drag on Satellites," AIAA Journal, Volume 2, pages 1148 to 1151, (June 1964)

15. Pitts, W. C. and Knechtel, E. D., "Experimental Investigation of Electric Drag on Spherical Satellite Models," NASA TN D-2619

16. Hall, D. F., Kemp, R. F. and Sellen, J. M., Jr., "Plasma-Vehicle Interaction in a Plasma Stream," AIAA Journal, Volume 2, pages 1032 to 1039, (June 1964)

17. Katzin, M. et al, "Experimental Investigations Related to Ionospheric Probing," Report No. NAS 8-2, Goddard Space Flight Center, (30 September 1964)

18. Electromagnetic Research Corporation, "Quarterly Progress Report No. 1 on Contract NASw-835 Amendment No. 3 for period 1 October to 31 December 1964," Goddard Space Flight Center.

19. Electromagnetic Research Corporation, "Quarterly Progress Report No. 2 on Contract NASw-835, Amendment No. 3 for period 1 January to 31 March, 1965," Goddard Space Flight Center.

20. Sellen, J. M., Jr. and Kemp, R. F., "Investigation of Ion Beam Diagnostics," NASA STAR N65-22617.

21. Hall, D. F., Kemp, R. F. and Sellen, J. M., Jr., "Generation and Characteristics of Plasma Wind Tunnel Streams," AIAA Journal, Volume 3, pages 1490 to 1497 (August 1965)

22. Sellen, J. M., Jr., Bernstein, W. and Kemp, R. F., "Generation and Diagnosis of Synthesized Plasma Streams," Review of Scientific Instruments, Volume 36, pages 316 to 322, (March 1965)

23. Cobine, J. D., Gaseous Conductors, Theory and Engineering Applications, pages 159 to 160

24. Bachynski, M. P., "Simulation of Geophysical Phenomena in the Laboratory," AIAA Journal. Volume 2, pages 1873 to 1882 (November 1964)

25. Bein, G. P., "Bibliography of Plasma Physics," NASA STAR N65-24842.

26. Chapman, S., "Notes on Aurorae and Magnetic Storms, Considerations as to a Model Experiment," Journal of Atmospheric and Terrestrial Physics, Volume 1, pages 189 to 199, (1951)

27. Lehnert, B., "Plasma Physics on Cosmical and Laboratory Scale," Nuovo Cimento, Supplement, Volume 13, pages 59 to 110 (1959)

28. Hollister, D. D., "An Electrodeless Technique for Full Scale Simulation of the Re-entry Environment," IAA A65-23634

BLANK PAGE

A STATIC THEORY OF A BEAM GENERATED PLASMA IN A CYLINDER

David H. S. Cheng

INTRODUCTION

In order to simulate a plasma environment similar to various regions of the ionosphere for experimental study, a beam generated plasma seems to afford a good possibility. It has the advantages of being easy to control and simple to operate. The purpose of this study then is to investigate a plasma generated by a beam of electrons in the cylindrical geometry. A static theory concerning such a plasma based on a modification of Tonks and Langmuir's complete plasma sheath equation (1) is attempted. Such a modification was first attempted by Dunn and Self (2) for rectangular coordinates.

The configuration in question is shown in Figure D-1. It consists of a cylinder of radius a with a column of electron beam coaxial with the cylinder of radius d. A potential function V is assumed to be zero at the center of the beam and decreases monotonically in the radial direction. Electrons generated in the plasma have a Maxwell-Boltzmann distribution corresponding to a temperature T_e; whereas ions created according to certain

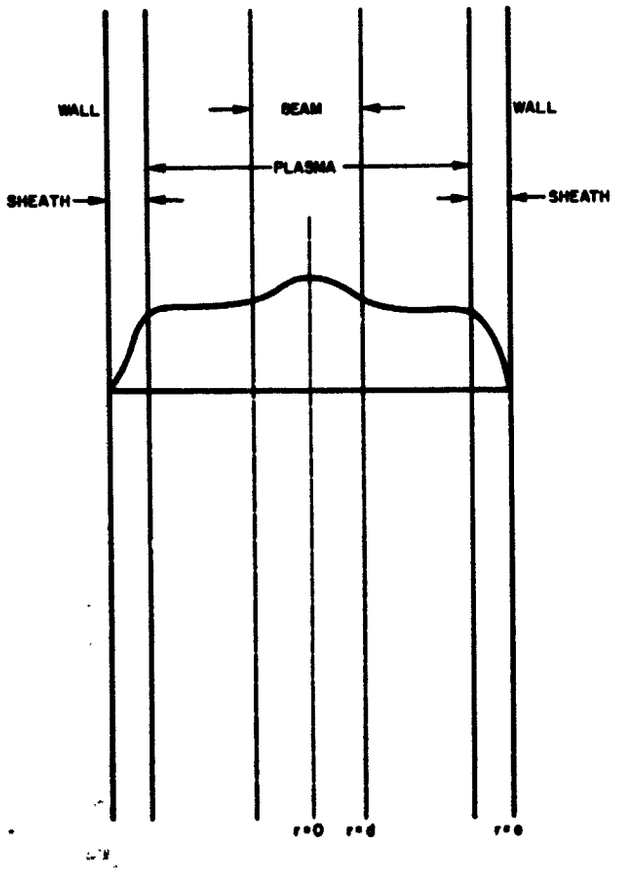


Figure D-1--Simulation of a Plasma Environment

specified generation functions fall without collision to the wall. At the wall, where the potential is V_w, the ion and electron current densities are zero. The pressure ranges from 10^{-4} to 10^{-2} mm Hg where the mean free path is relatively long.

THEORETICAL EQUATIONS

Poisson's equation under consideration for cylindrical geometry is:

$$\frac{d^2V}{dr^2} + \frac{1}{r}\frac{dV}{dr} = -4\pi e(n_i - n_e) \tag{1}$$

It will be assumed that the potential falls monotonically at $r = 0$. The plasma electron density under thermal equilibrium at temperature T_e is given by the Maxwell-Boltzmann distribution and the total electron density is

$$n_e = n_{eo} \exp(eV/kT_e) + n_b \tag{2}$$

where n_{eo} is the density at $r = 0$, n_b is the electron density of the beam, and k is the Boltzmann's constant.

The ions generated at any point ξ acquire a certain velocity V_ξ by the time they pass some other point r. If the number of ions generated per second per unit volume at ξ is N_ξ, their density when they pass r is, in the cylindrical case,

$$N_\xi \left(\frac{\xi}{r}\right)\frac{d\xi}{V_\xi} \tag{3}$$

The total contribution to n_i at r due to the ions at ξ less than r is given by

$$n_i = \frac{1}{r}\int_0^r N_\xi \xi \, d\xi / v_\xi \tag{4}$$

In the case where ions fall freely from the point of generation,

$$v_\xi = [2e(V_\xi - V)/m_i]^{1/2} \tag{5}$$

where V_ξ and V are the potentials at ξ and r respectively. Following Harrison and Thompson (3) one may write

$$N_\xi = \lambda n_\xi = \lambda n_{eo} \exp[\gamma(eV/kT_e)] \tag{6}$$

where $\gamma = 0, 1, 2$ indicates the three cases of ion generation: uniform, proportional to electron density and proportional to the square of electron density. And λ denotes the number of ions generated by an electron in one second.

Substituting various equations into Equation (1) and letting

$$\eta = - eV/kT_e \tag{7}$$

we have

$$= (kT_e/4\pi e^2 n_{eo}) \left(\frac{d^2\eta}{dr^2} - \frac{1}{r}\frac{d\eta}{dr}\right)$$

$$= \lambda \left(\frac{m_i}{2kT_e}\right)^{1/2} r^{-1} \int_0^r \xi \exp(-\eta_\xi) (\eta - \eta_\xi)^{-1/2} d\xi - \exp(-\eta) - \frac{n_b}{n_{eo}} \tag{8}$$

But

$$s = \lambda(m_i/2kT_e)^{1/2} r \tag{9}$$

Equation (8) can be further reduced to the following form

$$\frac{d\eta}{ds}\left[1 + s\frac{d}{d\eta}\left(\frac{d\eta}{ds}\right)\right] = \frac{2\eta^{1/2}}{a_1} \sum_{j=0}^{\infty} \beta_j \eta^j$$

$$+ \frac{2\eta^{1/2}}{a_1} \sum_{j=0}^{\infty} \eta^j \left[\sum_{\ell=0}^{j} \beta_{j-\ell} \sum_{m=0}^{\ell} \frac{2m+1}{2(\ell-m)+1} a_{\ell-m}\beta_m\right] \tag{10}$$

Let

$$h = (kT_e/4\pi e^2 n_{eo})^{1/2}, \quad a = \lambda(m_i/2kT_e)^{1/2}$$

and $h/a = E$, then the complete plasma sheath equation can be written as follows:

$$E^2\left(\frac{d^2\eta}{ds^2} + \frac{1}{s}\frac{d\eta}{ds}\right) = s^{-1} \int_0^s \zeta \exp(-\eta_\zeta) (\eta - \eta_\zeta)^{-1/2} d\zeta - \exp(-\eta) - \frac{n_b}{n_{eo}} \tag{11}$$

Equation (11) applies inside the beam for $0 \leq \eta \leq \eta_c$, $0 \leq s \leq s_c$, where s_c and η_c are the values of s and η at $r = d$ where the generation is cut off. Outside the beam $d < r \leq a$ there is no electron beam, or n_b/n_{eo} is zero; moreover, there is no contribution to the integral because there is no generation. Thus, for $d < r \leq a$ the complete plasma-sheath equation becomes

$$E^2 \left(\frac{d^2\eta}{ds^2} + \frac{1}{s}\frac{d\eta}{ds}\right) = s^{-1} \int_0^{s_c} \frac{\zeta \exp(-\eta_\zeta)}{(\eta_c - \eta_\zeta)^{1/2}} d\zeta - \exp(-\eta) \tag{12}$$

Tonks and Langmuir solve the plasma equation (4) by letting $E = 0$ in Equation (11). The solution was made possible by regarding η rather than s as the independent variable. They solved this in series form as follows.

$$s = a_1 \eta^{1/2} + a_3 \eta^{3/2} + a_5 \eta^{5/2} + \cdots \tag{13}$$

Using the same series form, a solution to Equation (11) is attempted. Now multiplying both sides of Equation (11) by s, we obtain

$$E^2 \left(s\frac{d^2\eta}{ds^2} + \frac{d\eta}{ds}\right) = \int_0^s \frac{\zeta \exp(-\eta_\zeta)\, d\zeta}{(\eta - \eta_\zeta)^{1/2}} - s \exp(-\eta) - s\frac{n_b}{n_{eo}} \tag{14}$$

Following Self's lead (5) we let

$$\frac{d^2\eta}{ds^2} = \frac{1}{2}\frac{d}{d\eta}\left(\frac{d\eta}{ds}\right)^2 = \left(\frac{d\eta}{ds}\right)\frac{d}{d\eta}\left(\frac{d\eta}{ds}\right) \tag{15}$$

Then the left hand side of Equation (11) becomes

$$E^2 \left\{\left(s\frac{d\eta}{ds}\right)\frac{d}{d\eta}\left(\frac{d\eta}{ds}\right) + \frac{d\eta}{ds}\right\} = E^2 \frac{d\eta}{ds}\left\{1 + s\,\frac{d}{d\eta}\left(\frac{d\eta}{ds}\right)\right\} \tag{16}$$

On the other hand, following Tonks and Langmuir, by letting

$$\eta = \rho^2, \quad \eta_\zeta = \rho_\zeta^2, \quad \rho_\zeta = \rho \sin\theta \tag{17}$$

we convert the integral term in Equation (11) to the form

$$\int_0^s \zeta \exp(-\gamma\eta_\zeta)\,(\eta - \eta_\zeta)^{-1/2}\, d\zeta = \int_0^\phi \zeta' \exp(-\gamma\rho_\zeta^2)\frac{d\zeta}{d\rho_\zeta}\, d\theta \tag{18}$$

where $\phi = \pi/2$ for $0 \le r \le d$, and $\phi = \phi_c$, corresponding to the cutoff value of η_c.

By virtue of Equation (13) the recursion formula as developed in the Appendix is given by

$$\left\{a_1^2\sigma_\ell \left(1 + \frac{1}{2\ell+1}\right) - \frac{a_1}{2\ell+1}(1+\nu) + \frac{2E^2}{a_1}\left[(2\ell+3) - \frac{1}{2\ell+1}\right]\right\} a_\ell$$

$$= -\frac{2E^2}{a_1} \sum_{m=1}^{\ell-1} \left\{ (2\ell + 3)\, a_m \beta_{\ell-m} - \beta_{\ell-m} \sum_{j=1}^{m} \frac{2j+1}{2(m-j)+1}\, a_{m-j}\, \beta_j \right\}$$

$$- \frac{2m+1}{2(\ell-m)+1}\, a_{\ell-m}\, \beta_m + d_1 \sum_{j=1}^{\ell} \frac{(-)^j}{j!}\, \frac{a_{\ell-j}}{2(\ell-j)+1}$$

$$- \sigma_\ell\, a_1^2 \left\{ \sum_{i=1}^{\ell-1} \frac{a_i\, a_{\ell-i}}{2i+1} + \sum_{k=0}^{\ell-1} \frac{(-)^{\ell-k}}{(\ell-k)!}\, (\gamma)^{\ell-k} \sum_{i=0}^{k} \frac{a_i\, a_{k-i}}{2i+1} \right\} \tag{19}$$

Equation (19) holds for both regions $0 \leq r \leq d$ and $d < r \leq a$. In the first region σ_ℓ corresponds to $\phi = \pi/2$, and $\nu = n_b/n_{eo}$; whereas in the second region $V = 0$, and the value of σ_ℓ corresponds to $\phi = \phi_c$ is used.

Numerical evaluations using digital computer were attempted, but so far the program has not yielded useful results. Nevertheless, once the potential profiles η vs s are known, all the other relevant discharge parameters could be calculated.

1. Ion current density

$$J_i = nev_i = \left(\frac{2kT_e}{m_i}\right)^{1/2} n_{eo}\, eI \tag{20}$$

where

$$I = \frac{1}{s} \int_0^s \zeta \exp(-\gamma\eta)\, d\zeta \tag{21}$$

2. Electron current density

$$J_e = \left(\frac{kT_e}{2\pi m_e}\right)^{1/2} n_{eo}\, e \exp(-\eta) \tag{22}$$

3. Normalized space charge density

$$(n_i - n_e)/n_{eo} \tag{23}$$

4. Wall potential and the sheath

By equating Equations (20) and (22) at $\eta = \eta_w$, once can solve for η_w

$$\eta_w = \left(\frac{m_i}{4\pi m_e}\right)^{1/2} \frac{I}{I(\eta_w)} \tag{24}$$

from which s_w can be determined. From tne knowledge of potential profiles and the wall potential, the sheath region can be investigated.

It is believed that a plasma so generated is controllable by the beam voltage and current.

If J_b is the beam current density and V_b is the beam voltage, then

$$n_b = J_b/e\,[2(e/m_e)\,(V_b - V)]^{1/2} \tag{25}$$

where V is the space potential in the plasma. If $V_b \gg V$, then

$$n_b \cong J_b/e\,[2(e/m_e)\,V_b]^{1/2} \tag{26}$$

Now, if we assume that all the ionization is due to the beam electrons, and N is the number of electron ion pairs produced per beam electron per unit length per mm Hg., then the ion current outside the beam is given by

$$J_i = J_b N p d \tag{27}$$

From Equation (20),

$$J_i = \left(\frac{2kT_e}{m_i}\right)^{1/2} n_{eo}\, eI \tag{28}$$

where

$$I = \frac{1}{s}\int_0^s \zeta \exp(-\gamma\eta)\, d\zeta \tag{29}$$

From Equations (26), (27), and (29), we can solve for n_{eo}/n_b in terms of the beam voltage or current.

By virtue of Equation (13) it follows that

$$\frac{ds}{d\eta} = \frac{a_1}{2}\,\eta^{-1/2}\left(1 + 3\frac{a_3}{a_1}\eta + 5\frac{a_5}{a_1}\eta^2 + \cdots\right) \tag{30}$$

And

$$\frac{d\eta}{ds} = \frac{2\eta^{1/2}}{a_1}\left(1 + 3\frac{a_3}{a_1}\eta + 5\frac{a_5}{a_1}\eta^2 + \cdots\right)^{-1} \tag{31}$$

But

$$a_0 = 1,\ a_n = (2n+1)\frac{a_{2n+1}}{a_1} \quad \text{for} \quad n = 1, 2, 3, \ldots \tag{32}$$

$$\frac{d\eta}{ds} = \frac{2\eta^{1/2}}{a_1}\,(a_0 + a_1\eta + a_2\eta^2 + \cdots)^{-1} \tag{33}$$

Consider

$$(a_b + a_1 \eta + a_2 \eta^2 + \cdots)(\beta_0 + \beta_1 \eta + \beta_2 \eta^2 + \cdots) = 1 \tag{34}$$

or

$$\sum_{n=0}^{\infty} a_n \eta^n \sum_{m=0}^{\infty} \beta_m \eta^m = 1$$

Put $\ell = m + n$, we have

$$\sum_{\ell=0}^{\infty} n^{\ell} \left[\sum_{m=0}^{\ell}\right] a_{\ell-m} (\beta_m) = 1, \tag{35}$$

It follows that

$$a_0\beta_0 = 1, \quad \text{or} \quad \beta_0 = 1: \quad \sum_{m=0}^{\ell} a_{\ell-m}\beta_m = 0, \quad \text{for} \quad \ell \geq 1. \tag{36}$$

And Equation (31) becomes

$$\frac{d\eta}{ds} = \frac{2\eta^{1/2}}{a_1}(1 + \beta_1 \eta + \beta_2 \eta^2 + \cdots) \tag{37}$$

and

$$s \frac{d}{d\eta}\left(\frac{d\eta}{ds}\right) = \left(1 + \frac{a_3}{a_1}\eta + \frac{a_5}{a_1}\eta^2 + \cdots\right)(1 + 3\beta_1\eta + 5\beta_2\eta^2 + \cdots) \tag{38}$$

By virtue of Equation (32) where

$$a_n = (2n + 1)\left(\frac{a_{2n+1}}{a_1}\right)$$

thus

$$\frac{a_{2n+1}}{a_1} = \frac{a_n}{2n + 1} \tag{39}$$

Thus Equation (38) becomes

$$s \frac{d}{d\eta}\left(\frac{d\eta}{ds}\right) = \sum_{\ell=0}^{\infty} \sum_{m=0}^{\ell} \frac{2m + 1}{2(\ell - m) + 1} a_{\ell-m}\beta_m \eta\ell \tag{40}$$

and

$$\left[1 + s\frac{d}{d\eta}\left(\frac{d\eta}{ds}\right)\right]\frac{d\eta}{ds} = \frac{2\eta^{1/2}}{a_1}\sum_{k=0}^{\infty}\beta_k\eta^k + \frac{2\eta^{1/2}}{a_1}\sum_{k=0}^{\infty}\beta_k\eta^k\sum_{\ell=0}^{\infty}\sum_{m=0}^{\ell}\frac{2m+1}{2(\ell-m)+1}\alpha_{\ell-m}\beta_m\eta^{\ell} \quad (41)$$

Consider the second term in Equation (41)

$$\frac{2\eta^{1/2}}{a_1}\sum_{k=0}^{\infty}\beta_k\eta^k\sum_{\ell=0}^{\infty}\sum_{m=0}^{\ell}\frac{2m+1}{2(\ell-m)+1}\alpha_{\ell-m}\beta_m\eta^{\ell}$$

$$= \frac{2\eta^{1/2}}{a_1}\sum_{j=0}^{\infty}\eta^j\left\{\sum_{\ell=0}^{j}\beta_{j-\ell}\sum_{m=0}^{\ell}\frac{2m+1}{2(\ell-m)+1}\alpha_{\ell-m}\beta_m\right\} \quad (42)$$

Thus Equation (41) becomes

$$\left(\frac{m_i\lambda^2}{8\pi e^2 n_{eo}}\right)\left(\frac{d^2\eta}{ds^2} + \frac{1}{s}\frac{d\eta}{ds}\right)$$

$$= s^{-1}\int_0^s \zeta\exp(-\eta_\zeta)(\eta-\eta_\xi)^{-1/2}\,d\xi - \exp(-\eta) - \frac{n_b}{n_{eo}} \quad (43)$$

or

$$E^2\frac{d\eta}{ds}\left[1 + s\frac{d}{d\eta}\left(\frac{d\eta}{ds}\right) = \frac{2E^2}{a_1}\sum_{j=0}^{\infty}\eta^{j+1/2}\beta_j + \sum_{\ell=0}^{j}\beta_{j-\ell}\sum_{m=0}^{\ell}\frac{2m+1}{2(\ell-m)+1}\alpha_{\ell-m}\beta_m\right] \quad (44)$$

And the coefficient of $\eta^{j+1/2}$ is

$$\frac{2E^2}{a_1}\left[\beta_j + \sum_{\ell=0}^{j}\beta_{j-\ell}\sum_{m=0}^{\ell}\frac{2m+1}{2(\ell-m)+1}\alpha_{\ell-m}\beta_m\right] \quad (45)$$

Following Tonks and Langmuir, by letting

$$\eta = \rho^2, \quad \eta_\zeta = \rho_\zeta^2, \quad \rho_\zeta = \rho\sin\theta \quad (46)$$

we convert the integral term in Equation (11) to the form

$$\int_0^s \zeta\exp(-\gamma\eta_\zeta)(\eta-\eta_\zeta)^{-1/2}\,d\zeta = \int_0^{\phi}\zeta'\exp(-\gamma\rho_\zeta^2)\frac{d\zeta}{d\rho_\zeta}\,d\theta \quad (47)$$

where $\phi = \pi/2$ for $0 \le r \le d$, and $\phi = \phi_c$, corresponding to the cutoff value of η_c.

Since

$$\zeta \frac{d\zeta}{d\rho_\zeta} \exp(-\gamma\rho_\zeta^2) = \sum_{\ell=0}^{\infty} \rho_\zeta^{2\ell+1} \left\{ \sum_{k=0}^{\ell} \frac{(u)^{\ell-k}}{(\ell-k)} (\gamma)^{\ell-k} \sum_{i=0}^{k} [2(k-i)+1]\, a_{2(k-i)+1}\, a_{2i+1} \right\} \tag{48}$$

By virtue of Equation (46)

$$\int_0^{\varphi} \zeta' \exp(-\gamma\rho_\zeta^2) \frac{d\zeta}{d\rho_\zeta} \, d\theta$$

$$= \sum_{\ell=0}^{\infty} \rho^{2\ell+1}\, \sigma_\ell \left\{ \sum_{k=0}^{\ell} \frac{(-)^{\ell-k}}{(\ell-k)!} (\gamma)^{\ell-k} \sum_{i=0}^{k} [2(k-i)+1]\, a_{2(k-i)+1}\, a_{2i+1} \right\} \tag{49}$$

where

$$\sigma_\ell = \int_0^{\phi} \sin^{2\ell+1}\theta \, d\theta$$

$$= \frac{(2)^{2\ell}(\ell!)^2}{(2\ell+1)!} \left\{ 1 - \cos\phi \left[1 + \frac{\sin^2\phi}{2} + \frac{1.3}{2.4}\sin^4\phi + \ldots + \frac{1.3\ldots.(2\ell-1)}{2.4\ldots.(2\ell)} \right] \sin^{2\ell}\phi \right\} \tag{50}$$

And the remaining two terms on the right hand side of Equation (11) give

$$s \exp(-\rho^2) = \sum_{\ell=0}^{\infty} \rho^{2\ell+1} \left[\sum_{j=0}^{\ell} \frac{(-)^j}{j!}\, a_{2(\ell-j)+1} \right] \tag{51}$$

$$s \frac{n_b}{n_{eo}} = \frac{n_b}{n_{eo}} \sum_{\ell=0}^{\infty} a_{2\ell+1}\, \rho^{2\ell+1} \tag{52}$$

Outside the beam where $d < r \leq a$, Equation (52) vanishes. The right hand side of Equation (11) yields

$$\sum_{\ell=0}^{\infty} \rho^{2\ell+1} \left\{ \sigma_\ell \sum_{k=0}^{\ell} \frac{(-)^{\ell-k}}{(\ell-k)!} (\gamma)^{\ell-k} \sum_{i=0}^{k} [2(k-i)+1]\, a_{2(k-i)+1}\, a_{2i+1} \right.$$

$$\left. - \sum_{j=0}^{\ell} \frac{(-)^j}{j!}\, a_{2(\ell-j)+1} - \left(\frac{n_b}{n_{eo}}\right) a_{2\ell+1} \right\} \tag{53}$$

Again, knowing

$$\alpha_n = (2n + 1) \left(\frac{a_{2n-1}}{a_1}\right)$$

$$a_{2n+1} = \frac{\alpha_n}{2n + 1} a_1$$

Then, Equation (53) becomes

$$\sum_{\ell=0}^{\infty} \rho^{2\ell+1} \left[\sigma_\ell a_1^2 \sum_{k=0}^{\ell} \frac{(-)^{\ell-k}}{(\ell - k)!} (\gamma)^{\ell-k} \sum_{i=0}^{k} \frac{\alpha_i \alpha_{k-i}}{2i + 1} \right.$$

$$\left. - a_1 \sum_{j=0}^{\ell} \frac{(-)^j}{j!} \frac{\alpha_{\ell-j}}{2(\ell - j) + 1} - a_1 \left(\frac{n_b}{n_{eo}}\right) \frac{\alpha_\ell}{2\ell + 1} \right] \tag{54}$$

By virtue of Equations (11), (45), (46), and (54), the recursion formula is given by

$$\left\{ a_1^2 \sigma_\ell \left(1 + \frac{1}{2\ell + 1}\right) - \frac{a_1}{2\ell + 1} (1 + \nu) + \frac{2E^2}{a_1} \left[(2\ell + 3) - \frac{1}{2\ell + 1}\right] \right\} \alpha_\ell = -\frac{2E^2}{a_1} \sum_{m=1}^{\ell-1} \left\{ (2\ell + 3) \alpha_m \beta_{\ell-m} \right.$$

$$- \beta_{\ell-m} \sum_{j=1}^{m} \frac{2j + 1}{2(m - j) + 1} \alpha_{m-j} \beta_j - \frac{2m + 1}{2(\ell - m) + 1} \alpha_{\ell-m} \beta_m + a_1 \sum_{j=1}^{\ell} \frac{(-)^j}{j!} \frac{\alpha_{\ell-j}}{2(\ell - j) + 1}$$

$$\left. - \sigma_\ell a_1^2 \left\{ \sum_{i=1}^{\ell-1} \frac{\alpha_i \alpha_{\ell-i}}{2i + 1} + \sum_{k=0}^{\ell-1} \frac{(-)^{\ell-k}}{(\ell - k)!} (\gamma)^{\ell-k} \sum_{i=0}^{k} \frac{\alpha_i \alpha_{k-i}}{2i + 1} \right\} \right. \tag{55}$$

which is Equation (19).

REFERENCES

1. Tonks, Lewis and Irving Langmuir, "A General Theory of the Plasma of an Arc," Phys. Rev., Volume XXXIV, No. 6 (September, 1929)

2. Dunn, D. A. and S. A. Self, "Static Theory of Density and Potential Distribution in a Beam-Generated Plasma," Journal of Applied Physics, Volume 35, No. 1, pages 113-122 (January 1964)

3. Harrison, E. R. and W. B. Thompson, "The Low Pressure Plane Symmetric Discharge," Proc. Phys. Soc. (London) Volume 74, pages 145-152 (1959)

4. Tonks and Langmuir, Loc. cit.

5. Self, S. A., "Exact Solution of the Collisionless Plasma-Sheath Equation," Phys. of Fluids, Volume 6, No. 12, pages 1762-1768 (December 1963)

VARIOUS MEANS OF GENERATING PLASMAS

John Kinnier

INTRODUCTION

To reproduce in the laboratory the plasmas found in the ionosphere and magnetosphere for purposes of instrument calibration and development, requires, as one of its essential constituents, an ion source capable of providing ions of well defined composition, temperature, and energy. In the interests of efficient experimentation it is desirable that the plasma be produced uniformly over a large volume.

The ion energy requirements for ionosphere simulation, 1 ev to 10 kev, are by the standards of the particle accelerator physicist very low, in fact, charged particles with energies below 100 ev are not readily achieved since an electron, neutral atom, or positive ion must have an energy of 30-60 ev to be an effective ionizing agent.

IONOSPHERE, D-REGION

This region occurs in the altitude range above the surface of the earth from 50 to 85 km. In terms of atmospheric models it has a particle density of $10^{21}/cm^3$, a collision frequency of 10^5/sec and an average free path length of 10^{-2} cm. The mean free path of neutral atoms in their own gas is, in terms of the atomic diameter d, and density n,

$$\lambda = \frac{1}{\sqrt{2}\, n \pi d^2} \tag{1}$$

An atom, relatively speaking, loses little of its total mass when ionized and it is to be expected that a positive ion will have approximately the same mean free path as its parent atom. It may be slightly less in those cases where a positive ion is able to polarize a neutral molecule. If one assumes that the electrons freed in the ionization process acquire a mean energy nearly that of the gas, then their velocity will be very much larger and their mean free path longer by a factor

$$\lambda_{electron} \approx 4\sqrt{2}\, \lambda_{ion} \tag{2}$$

In the D-region the mean free path is of the order of 1/100 of a centimeter. Therefore the mechanism chosen for the establishment of a large volume plasma cannot be dependent on the transport of ions over any appreciable distance. To achieve a large volume, plasma ions must be generated simultaneously throughout the volume. Following the directive of nature, photon ionization processes, by irradiation with ultraviolet or soft x-ray radiation, suggest themselves as a means for producing a plasma. Photons with wavelengths below 1000 A will produce on the average one ion pair per photon, while those below 400 A two or more ion pairs.

A glow discharge in hydrogen run at a current of approximately one ampere, produces the many lined ultraviolet spectrum of hydrogen. If radiation from the discharge passes through a cooler layer of surrounding gas most of the radiation below 1000 A will be absorbed. This may be prevented by using a ceramic capillary arc source and radiation down to 150 A has been realized (1).

Romand described a triggered vacuum spark source yielding emission at wavelengths < 80 A. The source is energized by a storage capacitor and has achieved a rate of 20 discharges per second (2).

It appears probable that some portions of the D-region will be successfully simulated by photo-ionization processes.

IONOSPHERE, E, F-REGIONS

In these regions the pressure has decreased and the mean free path increased to values where it is possible to cause ions produced in one area to diffuse into a large volume. In a rather limited portion of the F-region, where pressures range from a few hundred to ten microns, one can employ cold cathode or normal glow discharges to produce an ionized volume. Some experiments have been conducted on the screen cathode, brush cathode and RF discharge with a view to determine ion densities and temperatures. As the experiments progressed it became increasingly evident that the geometrical arrangement of the electrodes with respect to each other and the chamber walls, were of critical importance, particularly at pressures above 100 microns. Below this pressure the discharge as it appears visually spreads more or less uniformly throughout the bell jar used in these experiments. Because of the nature of the normal glow discharge only a limited amount of control may be exercised over ion densities and energies. For example, the discharge cannot be maintained below a potential difference of approximately 400 volts and this accelerating potential determines the ion energies produced. Measurements made upon these discharges indicate an ion density of 10^8 ions/cm^3.

In the higher altitudes of the F-region and in the E-region, the particle densities have become so low that ion production at these pressures becomes inefficient. Accordingly one may consider the production of ions in the controlled region of an ion source and then inject these into the experimental volume.

Ion beam production has been the subject of extensive investigation for many years by those doing research with particle accelerators and atomic collision processes. The general property sought for in these ion sources is a beam ranging in current from a few microamperes to several hundred milliamperes, the beam must be of high ion purity, it must be well focused and of small cross sectional area. Such a device is not ideally suited for ionosphere simulation where one requires a uniform ion density over a large volume. It is true that one may use a divergent electrostatic lens to spread the beam over a larger area.

ION SOURCES

As representative of 25 or 30 different forms of ion sources, several representative types will be examined briefly.

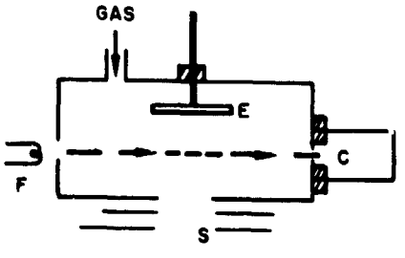


Figure D-2–Neir source.

(1). Nier Source (3)

This source (shown in Figure D-2) will operate over a wide range of pressures. In the range of 10^{-4} to 10^{-6} mm Hg, it performs as a confined beam electron impact device. In the higher pressure range 10^{-3} to 10^{-2} mm Hg it acts as a low voltage arc. Electrons from the filament, F, are confined by an axial magnetic field as they pass across the main chamber into a collector C.

Transverse to the beam of electrons an electric field is established by the electrode E. Gas is allowed to leak into the chamber at an optimized rate. Those gas atoms ionized by the electron beam are extracted from the chamber through the slit S. Very often additional beam defining slits are used. The currents produced by this source are usually in the microampere range though models having larger beam currents have been constructed. It is an important feature of this source that ions be produced by electrons having a narrow energy distribution.

(2) Radio Frequency Ion Source

A radio frequency ion source developed by Thonemann (4) and others is frequently used as a proton or deuteron source in particle accelerators. It yields beams of up to 70% pure H^+ as a proton source at currents of several hundred microamperes (see Figure D-3).

Under typical operating conditions the source pressure is between 10^{-3} and 10^{-1} mm Hg. A direct current extraction potential of 2-3 kv is provided across the source. Radio frequency excitation is supplied by an oscillator in the range 20-100 Mc/sec at a power of 30 to 60 watts and an axial field of 300-1000 gauss surrounds the lower part of the source. The average life of the exit canal, which is subject to positive ion erosion, is approximately 500 hours.

Continued use gradually darkens the glass bottle to a rich brown color and causes certain impurities to collect on the inner surface which decreases the beam current.

(3) Duoplasmatron Source

A low pressure arc ion source of exceptional efficiency was described by van Ardenne in 1942 (3, 4) and is shown in Figure D-4. A very dense plasma is formed in the arc region under the constraint of a magnetic field concentrated by the shaped pole pieces. At nearly 100% efficiency in gas utilization the source is capable of delivering beam currents up to 500 ma. In general it is a superior source to the RF type and may be operated at extraction potentials from 200 to 10 kv.

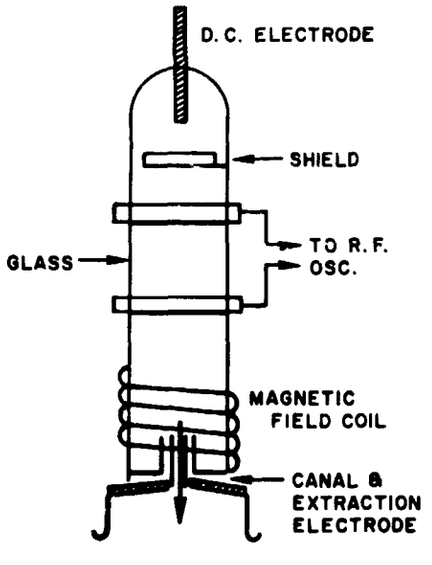


Figure D-3–Radio frequency ion source.

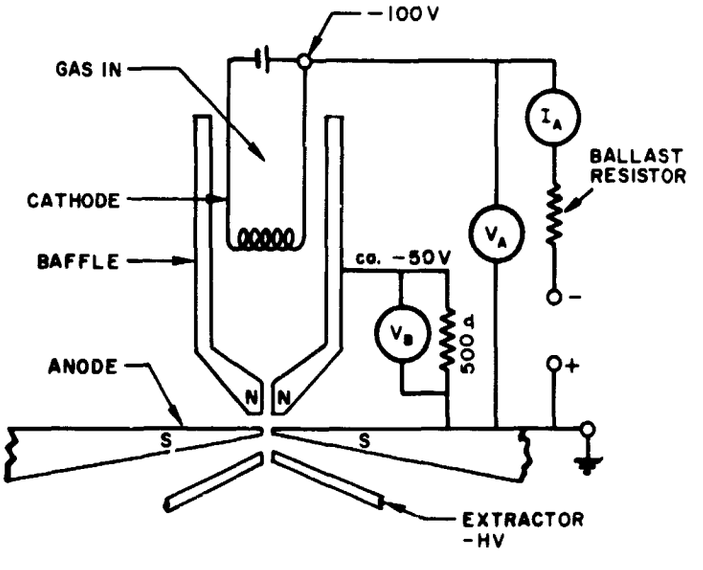


Figure D-4–Duoplasmatron source.

(4) Surface Ionization Sources

The surface ionization source used many years ago by Langmuir has been the object of extensive development in the past few years as a means of providing ion propulsion for spacecraft. In its present form it consists of a heated tungsten disk covered by a monotonic layer of cesium. The work function of tungsten is 4.5 ev and for pure cesium 1.8 ev, however, the work function of a cesium covered tungsten surface is only 0.7 ev. The explanation of this exceedingly low work function lies in the fact that a cesium atom on a tungsten surface experiences an image force between its nucleus and the free electrons within the tungsten metal. As a result it is distorted into a dipole with its positive pole near the surface. This induced polarization results in a strong bond between the cesium and tungsten nuclei and reduces the force binding an electron to the cesium atom. An electron with only 0.7 ev of energy may be released from the surface. Ionization of cesium atoms may be described by a two stage process. A cesium atom adhering to the tungsten surface is polarized. For a sufficiently high tungsten temperature (1200°K) the cesium atom may evaporate. In this process the evaporating atom will retain its outer electron if the ionization energy Q of the electron is greater than the work function ϕ of the metal. When the work function is larger than the ionization energy, $\phi > Q$, the outer electron is retained by the metal and the absorbed atom leaves the surface as an ion. The proportion of positive ions produced by the source may be expressed as follows:

$$\frac{n_+}{n_0} \alpha\, e^{\frac{-(Q-\phi)}{kt}}$$

Element	Ionization Energy (electron volts)
Cs	4.16
Rb	3.87
K	4.3
Na	5.1
Li	5.36
N	14.5
O	15.8

Thus the elements Cs, Rb, K can be ionized by a heated tungsten surface while Na, Li, N, and O will not be. Because of the high ionization potentials of nitrogen and oxygen the use of this method for plasma simulation is not possible.

LARGE AREA ION SOURCES

Motivated by the requirement that the plasma for ionosphere simulation fill a large volume, it seemed worthwhile to investigate the existence or possibility of constructing ion sources having a large emission area compared to accelerator as mass spectrometer type sources. Two sources of this type will be described, the oscillating electron source and an electron bombardment type ion engine.

1. Oscillating Electron Source

An oscillating electron source for plasma production, based on the Penning discharge, was constructed by Meyerand (7, 8, 9).

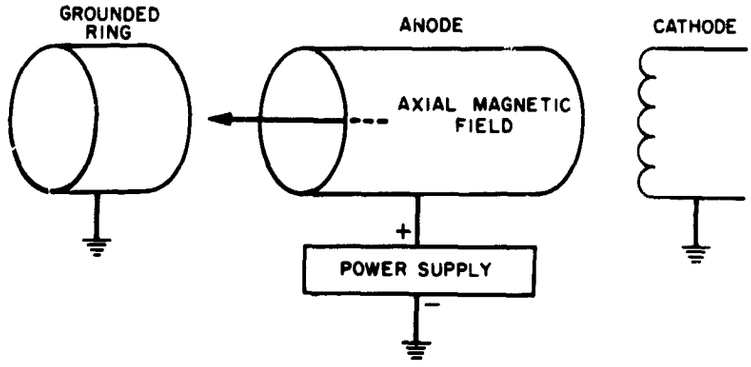


Figure D-5–Minus-plus-minus source.

The operating principle of the source may be understood with reference to Figure D-5. Two cylindrical electrodes are located on a common axis. The axial cathode serves as a hot filament electron source. Electrons emitted by the cathode, at ground potential, are accelerated by the positively biased anode. A magnetic field produced by an external solenoidal winding constrains the electrons to move along paths parallel to the cylinder axis. As they begin to emerge from the anode, they experience a decelerating force and are drawn back into the anode cylinder. Thus the electrons oscillate back and forth along the source axis. The source is so constructed that it may be housed in a confining envelope whose function is to permit a gas pressure of one micron or so to be maintained in the anode region. The oscillating electrons are now able to ionize the gas in the anode region from which they are ejected by the same potential that retains the electrons. The plasma potential is shown in Figure D-6 and the applied positive anode potential may vary from a few hundred to several thousand volts. The required axial magnetic field is 1500 gauss. Various gases have been used in the source including the noble gases, nitrogen, carbon dioxide and mercury vapor. In terms of overall efficiency a figure of 30% has been achieved. For purposes of ionosphere simulation it is important to note that this source produces a charge neutralized beam.

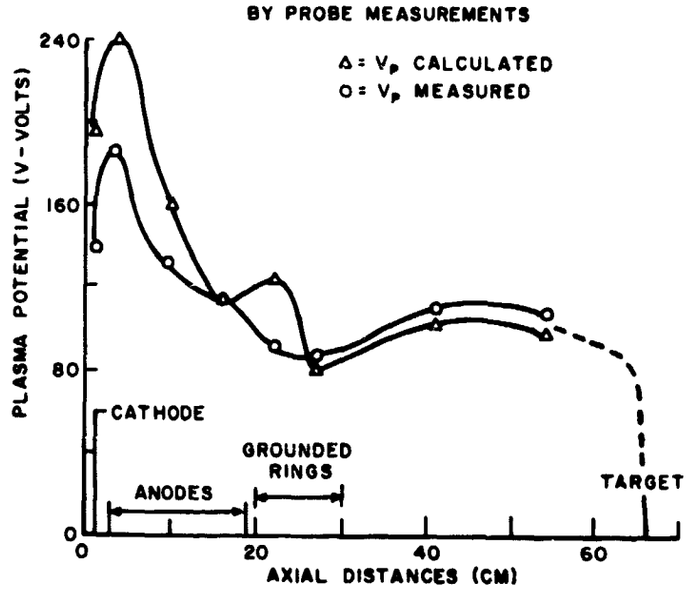


Figure D-6–Plasma potential in oscillating-electron plasma source.

2. Electron Bombardment Source

A source whose operating principles are closely related to the oscillating electron source was developed by Finkelstein (10) and perfected for purposes of ion propulsion by Kaufmann and Reader (11, 12, 13). Sources were constructed to achieve a high efficiency and large cross sectional area (50 cm). A schematic view of the source is shown in Figure D-7.

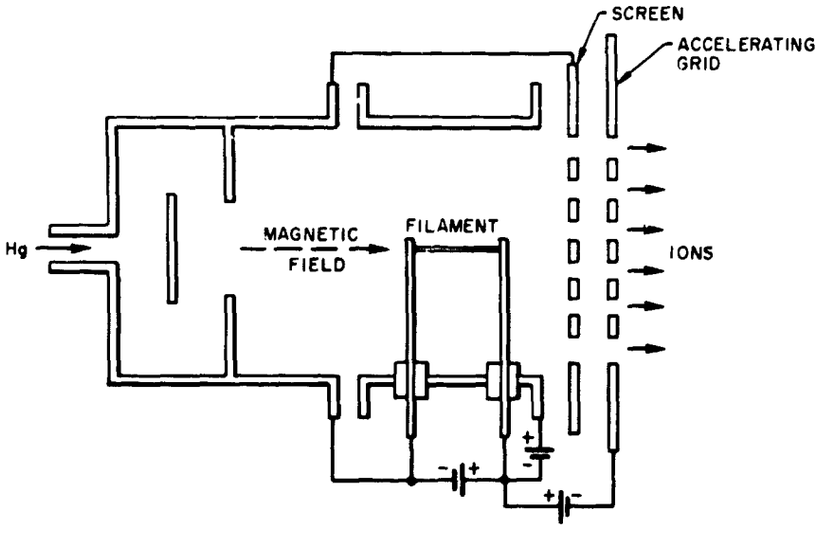


Figure D-7–Electron bombardment source.

Electrons emitted from the filament are confined by an axial magnetic field of 20-30 gauss while being accelerated by an anode potential of 20-100 volts. Within the cylindrical chamber the accelerated electrons make ionizing collisions with gas molecules. Positive ions formed in this region pass through the screen at the end of the chamber, which is at the same potential as the anode, and are accelerated by the negatively biased grid just beyond. A grid potential of 2-3 kv is nominally used. It has been found that even for efficient operation in a vacuum chamber some form of beam neutralization is required. In its simplest form this consists of a single hot wire filament stretched transverse to the ion beam at a distance of 10 inches from the exit grid. Though originally designed as a mercury vapor engine the source has been used to accelerate various gases including nitrogen. It was found, however, that helium could not be used. In a source of this type it is obvious that its operating characteristics depend upon many geometric factors such as, the chamber length, the screen and grid diameter, the filament length, the axial field strength and accelerating potentials; as many as nine sources of this type have been arranged in a mosaic structure indicating that it is possible to operate the sources in close proximity without interference.

AN ION SOURCE FOR IONOSPHERE SIMULATION

To create an artificial ionosphere of practical value an ion source capable of providing beam currents in the milliampere range of various gases is a useful and perhaps essential means. The design of an ion source for this purpose is sketched in Figure D-8. The source is an oscillating electron or minus-plus-minus type and is capable of being scaled to a larger model. It has the advantage of simplicity of construction and may be easily disassembled. The electrodes are three coaxial sections of tubing mounted within

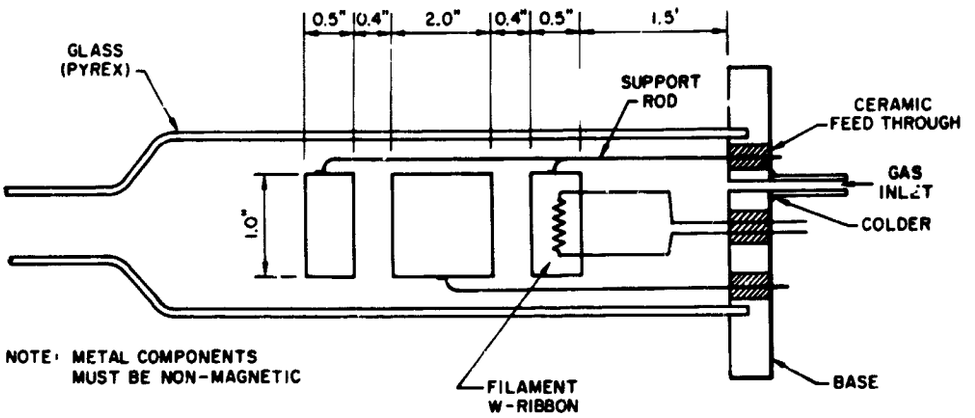


Figure D-8–Oscillating electron ion source, cross section.

a glass envelope which permits adjustment of the gas composition and pressure in the source without adversely affecting other vacuum requirements. The electrodes are supported by a dual rod system from the base plate and terminate in insulated connectors. This provides for bringing the appropriate potentials to the grounded ring and the anode. The tubing section enclosing the filament serves as a grounded heat shield. Under operating conditions the source must be placed within a solenoid capable of producing an axial magnetic field of approximately 1500 gauss. Depending upon the power output demanded of the source and its immediate environment, cooling coils may be required. The axial magnetic field constricts the beam to the source axis. The length of beam exit tube should be adjusted to provide a sufficiently low gas conductance path so that a differentially pumped vacuum system will be feasible.

This source in addition to providing large currents of various heavy ions over a comparatively large exit orifice has the further advantage for ionosphere simulation of providing a neutralized beam.

In the discharge used in the experiment the frequency is high so that the electrons lose little energy to the gas molecules between cycles of the field with the result that the mean energy will be nearly constant. The spread of electrons over a given volume will be substantially a diffusion process. Let u_0 be the mean energy of the electrons acquired from the rms value of the field which is effectively a direct field of $E = E_0/\sqrt{2}$; then the instantaneous energy of the electron is composed of a large steady component upon which is superimposed a smaller sinusoidal variation as shown in Figure D-9.

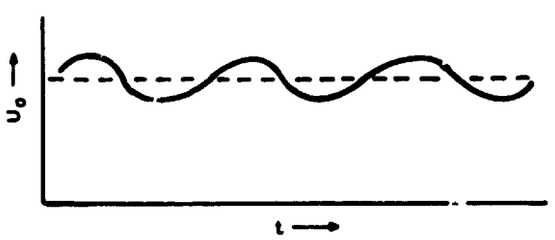


Figure D-9–Sinusoidal variation superimposed on steady component.

The two basic physical effects which control the nature of the discharge are the ionization produced by electrons colliding.

CONCLUSIONS

1. For D-region simulation, the most promising method appears to be the use of photoionization by ultraviolet and soft x-ray radiation sources.

2. In the micron pressure range a glow discharge may be useful in limited applications.

3. In the E and F regions for pressures less than the micron range ion gun or ion engine type sources should be used for plasma production.

4. In view of the large volume plasma requirement oscillating electron sources or electron bombardment ion engines of large exit orifice are suggested.

5. To achieve very low energy plasmas (< 100 ev) decelerating systems and charge neutralizers will have to be added to the ions sources of #4.

6. To ensure that the plasma produced is of the proper energy and quality, accurate diagnostic methods will be required.

REFERENCES

1. Weissler, G. L.. Vol. XXI Handbuch der Physik, Springer Verlag (1956) Berlin
2. Romand, J., J. Quant. Spect. Rad. Transfer 2, 691 (1962)
3. Nier, A. O., Rev. Sci. Instr., Volume 11, page 212 (1940)
4. Thonemann, P. C., Progress in Nuclear Physics, Volume 3, page 219, London, Pergamon Press, (1953)
5. Ardenne, M. von "Tabellen der Electronenphysik, Ionenphysik and Ubermikros kopie," Deutscher Verlag der Wissenschaften, Berlin, (1956)
6. French, P., "Experiments with Arc Ion Sources for Electric Propulsion," Thompson-Ramo-Wooldridge Report ER-4124, Cleveland, (June 1960)
7. Penning F. M. and J. H. A. Maubis, "Eine Neutronenroehre ohne Pumpvarrichtung," Physica, Volume 4, page 1190 (1937)
8. Meyerand, R. G. and S. C. Brown, "High Current Ion Source," Rev. Sci. Instr. Volume 30, Number 2, page 110, (Feb. 1959)
9. Meyerand, R. G., "The Oscillating Electron Plasma Source," ARS Electrostatic Propulsion Conference, Monterey, Calif., page 81 (Nov. 1960), "Electrostatic Propulsion," "Progress in Astronautics and Rocketry," Volume 5, Academic Press, Inc., New York (1961)
10. Finkelstein, A. T., "A High Efficiency Ion Source," Rev. Sci. Instr., Volume 11, page 94, (1940)
11. Kaufmann, H. R., One Dimensional Analysis of Ion Rockets, NASA Tech. Note D-261 (March 1960)
12. Kaufmann, H. R., and P. D. Reader, "Experimental Performance of Ion Rockets Employing Electron Bombardment Ion Sources," ARS Electrostatic Propulsion Conference, Monterey, Calif., (November 1960)
13. Reader, P. D., "50 Centimeter Diameter Electron Bombardment Ion Engine," AIAA Paper 64-689, A64-24903, 20-27

BLANK PAGE

IONOSPHERE SIMULATION THROUGH GAS DISCHARGES IN LARGER VOLUMES

John H. Kinner and David H. S. Cheng

INTRODUCTION

For a number of years the properties of the earth's ionosphere have been studied by scientists using earth bound instrumentation from which there gradually emerged the basic structure of the ionosphere and the variation in its characteristics as a function of time and altitude. Its daily fluctuations were observed and values for the density and abundance of its charged particle species were evaluated from these indirect measurements. Satellite borne instrumentation capable of immersing itself in the ionosphere for extended periods of time, provides a unique opportunity to investigate this hitherto inaccessible region, at the same time, however, it precludes the use of the earth itself as an electrical reference point leaving the satellite to provide its own reference potential.

To more fully understand the implications of this procedure and to determine in detail the influence of a charged particle environment on the satellite itself, its instrumentation for ionosphere and non-ionosphere experiments and its effect on the ionosphere in its immediate vicinity, a laboratory facility for ionospheric simulation would provide at once three means of increasing our understanding of the ionosphere.

It would enable the study of the effects of external forces and fields on the ionosphere itself under measurable and controlled conditions; it would permit a study of the effect of the ionosphere on the behavior of a satellite as a whole and allow the detailed response of non-ionospheric experiments to be evaluated; finally it would offer a ready and known environment for the development and testing of ionospheric experiments. With these goals as an ideal, an experimental program guided by the theory available, was undertaken to produce an ionized region of predictable properties and of sufficient volume to accommodate a real size test object.

METHOD

The method chosen in these experiments to simulate the ionosphere was an investigation of the properties of direct and alternating current gas discharges of large volume and to determine how closely these approach the known charged particle densities and energies of the ionosphere. For this purpose, it is important that the basic parameters of a gaseous discharge be known and the atomic mechanisms be understood.

In many gaseous discharges initiated and sustained by the application of a direct or alternating electric field, there exist ionized regions wherein the density of charged particles is high and the relative number of positive and negative charged particles nearly equal. Even if neutral gas particles are present, as in the usual case, or if the gas is not fully ionized, such a region is often called a plasma. A familiar example of a plasma is the positive column of a glow discharge. A plasma in thermal equilibrium, by reason of the presence of large numbers of charged carriers of both sign, is highly conducting and only relatively low voltage gradients exist. The negative carriers are principally electrons; the positive carriers are ionized molecules or atoms of the gas present. The application of an electric field of sufficient magnitude to exceed the breakdown potential of a gas results in the production of positive ions having a temperature greater than that

of the un-ionized gas and of the electrons of considerably higher temperature. The energy distribution of both ions and electrons is usually assumed Maxmellian except in those regions where large potential gradients may force it to be otherwise. In a plasma region the current flux is ideally random. Often in practice there exists a much smaller drift current which, while it does not seriously disturb the Maxmellian distribution, raises the electron and ion temperature. The free electrons in the discharge–because of their small mass and high mobility–acquire more energy in scattering processes from the applied field than do positive ions, consequently the electron temperature may be much higher than the positive ion temperature and the positive ion temperature higher than that of the neutral gas.

The energy necessary to maintain the discharge comes primarily from the applied field through the acceleration of free electrons. Ionization of a neutral atom may be due to electron impact, collision with a positive ion or the absorption of radiant energy. In an inelastic collision the kinetic energy of the incident particle is converted into the potential energy of excitation of an atomic electron or it may remove an electron completely producing a positive ion. The probability of ionization due to single electron impact with an unexcited atom is zero when the energy of the electron is just equal to the ionization potential and rises rapidly with increasing electron energy. The ionization efficiency of electrons rises linearly with energy reaching a broad maximum at 100 ev and afterwards decreases.

In general, neutral atoms or positive ions in the discharge are much less effective ionizing agents than are electrons of the same energy. A process that often affects the mobility of ions in their own gas is the charge transfer process

$$A^+ + B \rightarrow A + B^+$$

There is some experimental evidence to indicate that the critical energy for positive ions to ionize neutral atoms of their own gas is twice the ionization potential. Compared to ionization by electrons of equal energy the probability of this process occurring is even less than that of charge transfer.

Photoionization, when considering unexcited, may take place when $h\nu \geq eV_i$ which implies that the wavelength of the incident light must be $\lambda \leq 12{,}400/V_i$ where V_i is the ionization potential in volts. Thus ultraviolet radiation of $\lambda < 900$A will ionize nitrogen. The amount of photon produced ionization in a gas ordinarily is much less than that produced by electron collisions.

In a plasma in equilibrium containing electrons and positive ions, the faster moving electrons tend to diffuse outwards more rapidly from the source, the excess positive charge left behind then retards the electron flow and accelerates the diffusion of positive ions with the final result that electrons and ions diffuse at the same rate tending to maintain a uniform density.

A gaseous discharge excited by an alternating electric field in the megacycle frequency range may differ in several aspects from the discharge produced by a steady field. The periodic reversal of the exciting field under certain circumstances will tend to contain within the discharge charged particles which in a direct current discharge, under like conditions, would soon reach the walls of the containing vessel. With this reduction in losses, the field intensity required for a self sustaining discharge is reduced. Furthermore secondary processes taking place at the electrodes and walls have considerably less effect on the discharge than in the direct current case. Likewise if secondary electrons are produced at the walls, these will not affect the discharge unless they are emitted when the field direction is such as to add them to the discharge.

There are three important factors which contribute to the characteristics of an alternating current discharge:

1. The gas pressure, P, which determines the mean free path, λ, and the collision frequency ν of the charged particles.

2. The frequency, f, of the applied electric field.

3. The dimensions of the discharge container.

The behavior of the discharge is dependent on the inter-relation of these three quantities. For the discharge used in these experiments the mean free path, λ, was of the order of a few millimeters, the excitation frequency 20-30 megacycles, and the collision frequency approximately 10^5/sec. Thus the applied frequency is a hundred times the collision frequency and the mean free path negligible compared to the size of the discharge chamber. Under these circumstances an electron collision during a period of one cycle will be a rare occurrence. The equation of motion of an electron starting at a point $x = 0$ with a velocity $V = V_0$ becomes from Newton's second law

$$m\frac{d^2x}{dt^2} = eE_0 \sin(\omega t + \phi) \tag{1}$$

where ϕ is the phase angle of the field E at $t = 0$. Assuming these boundary conditions the velocity at a time t becomes,

$$v = v_0 + \frac{eE_0}{m\omega}[\cos\phi - \cos(\omega t + \phi)] \tag{2}$$

and the displacement,

$$x = v_0 t + \frac{eE_0 t}{m\omega}\cos\phi + \frac{eE_0}{m\omega^2}[\sin\phi - \sin(\omega t + \phi)] \tag{3}$$

The first two terms of Equation (3) indicate that there is a steady uniform motion in the direction of v_0 upon which is superimposed a sinusoidal displacement whose amplitude is $eE_0/m\omega^2$. Equation (3) shows that only those electrons having $v_0 = 0$ will oscillate with respect to their initial position. Secondly, those electrons moving perpendicular to the field continue to do so. The velocity acquired by an electron due to the oscillating electric field for values in the neighborhood of 500 volts/meter will be approximately that of an electron having a few eV of energy throughout the vacuum enclosure. Losses by recombination will be small and the walls of the discharge vessel will not exert any appreciable influence. Electrons gain energy from the field upon making collisions with gas molecules and having their ordered oscillatory motion changed at random. The energy absorbed is proportional to the square of the electric field and independent of its sign. The discharge becomes self sustaining when the gain in electron density by ionization becomes equal to the loss by diffusion.

BELL JAR EXPERIMENTS

Screen Cathode

Two different electrode sets were used to excite a direct current discharge and an air core coil was used to excite a discharge at radio frequencies. Experiments were initiated using an electrode configuration, hereafter designated as the screen cathode, patterned after a plasma appartus at Brooklyn Polytechnic Institute. In its initial form the apparatus was placed in an 18" bell jar as shown in Figure D-10.

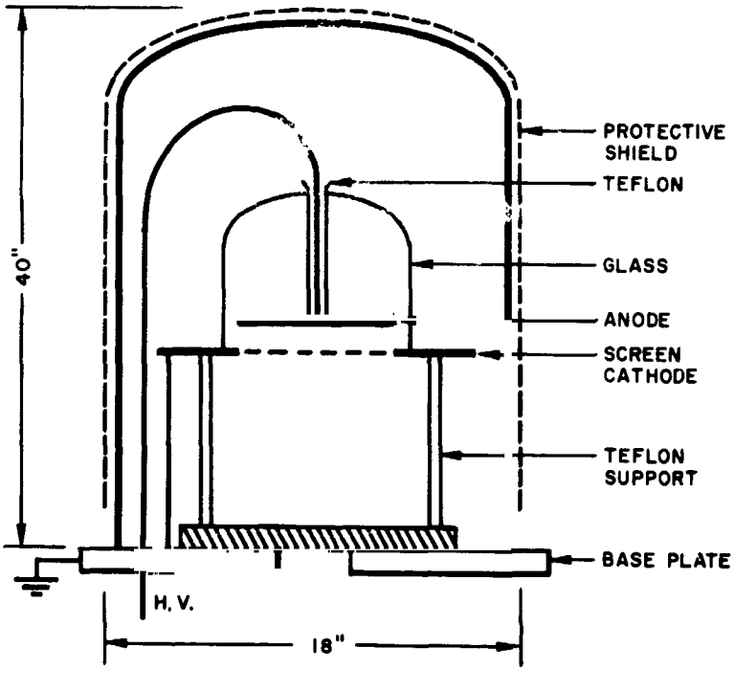


Figure D-10–Bell jar and screen cathode.

A circular section of stainless steel screening welded to a ring of the same material was located as the upper surface of a cylinder supported by three teflon legs anchored to a base ring of the same diameter. Above the screen electrode a solid disk employed as the anode was fixed in place by a glass support structure.

The voltage for the discharge was provided by a 2 kv variable dc supply through a 10,000 ohm series resistor. A discharge was established at a pressure of 1 mm Hg. with an applied voltage of 700 and a current of 30 milliamperes. Under these conditions a purple glow appeared about the anode and a small pink section formed in a region just below the cathode. At a pressure of 500 microns the discharge spread out more uniformly over the electrodes maintaining the same general characteristics. With a further reduction in pressure to 300 microns, the discharge changes. The pink glow formerly beneath the screen fills the teflon tube supporting the anode while the region beneath the screen takes on a blue-purple hue which also begins to appear on the surface of the base plate separated from it by a dark space a few millimeters in thickness. The discharge is unstable in these circumstances and waves of a reddish color bubble up through the teflon tube supporting the anode.

At a pressure of 50 microns the glowing regions spread out over a larger volume, decrease in intensity, and eventually fill most of the bell jar with a diffuse bluish light.

In general the discharge is quite susceptible to any geometrical misalignment of the electrodes themselves or with respect to the axis of the bell jar. Occasionally, glowing streamers flash out in the direction of the walls and removal of the bell jar's protective screen confirmed the opinion that it had a direct influence tending to constrict the discharge.

In an attempt to produce a more stable discharge the glass supporting the anode was removed and the spacing between anode and cathode maintained by teflon blocks 8 millimeters in thickness. A pink glow now appeared on the upper surface of the anode while the cathode and base plate had a violet blue glow approximately one inch in thickness. The thin layer of pink on the upper side of the anode was assumed to be part of a positive column and the violet blue negative glow. The thickness of the negative glow increased with decreasing pressure until it began to pervade the whole vacuum enclosure at a pressure of 5 microns. The dark region between the screen cathode and the base plate might be assumed to be a Faraday dark space. Because of the complicated nature of the discharge due to the effect of the base plate, the production of the discharge was modified. The anode assembly was removed and the screen biased at a high negative potential with respect to the grounded base plate. Positive ions formed in the region between the base plate and screen are now accelerated upwards. A pink glow covers part of the base plate and a violet-blue glow extends for two inches or more above and below the screen. This configuration produced a very stable discharge and measurements were begun using single and double probe systems as indicated in Figure D-11.

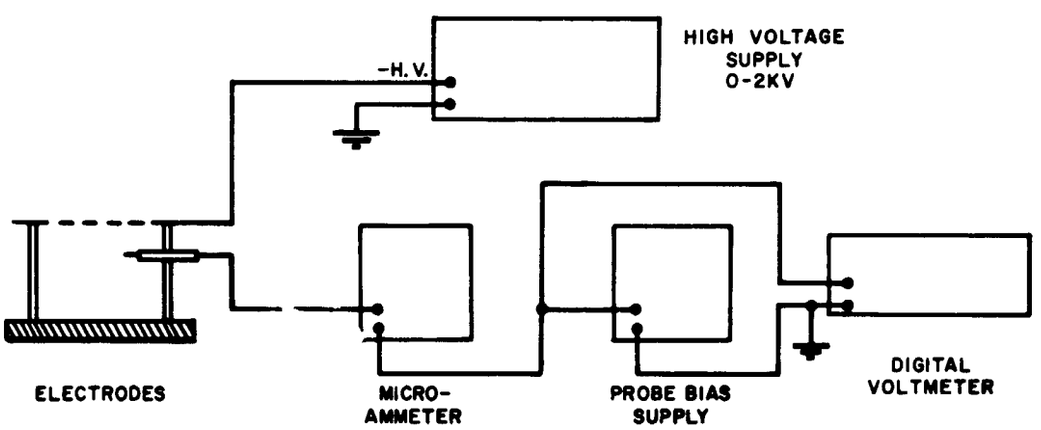


Figure D-11–Discharge and probe system.

The electronic components used were a 2 kc, 500 ma variable high voltage supply, an electronic milli-micrometer, a variable dc probe bias supply and a digital voltmeter to measure the probe bias. The double probe system followed that of Johnson and Malter (1).

Brush Cathode

The next series of experiments used a cathode consisting of many hundreds of steel needles densely packed over the surface of a two and one half inch square steel plate. This cathode assembly somewhat resembled one used at the National Bureau of Standards. The needle assembly was placed in one end of a 12" long rectangular lucite box. At the opposite end a ring electrode was inserted. When simply placed in the bell jar the discharge often chose the path over the outer surface rather than within the electrode enclosure. Many geometrical and high voltage arrangements were explored. The most

satisfactory and stable mode of discharge was achieved when the brush electrode was in contact with the grounded bell jar base plate and the ring placed at a high negative potential. A discharge was obtained at a pressure of 200 microns at 450 volts carrying a current of 1 to 2 milliamperes. The brush was covered with a thin pink glow followed by a dark region of one inch thickness, followed by three separate pink-orange glow regions one and one half inches in thickness forming a striated column as shown in Figure D-12(a). A rather large dark space preceded the blue negative glow surrounding the ring electrode. As the discharge current was increased the layers of the positive column contracted into two, and then one layer and ultimately adhered to the needle surface in the form of a small bubble whose location moved occasionally from place to place. Meanwhile the negative glow expanded and increased in density.

A decrease in pressure to 50 microns resulted in the formation of a clearly visible beam structure formed in the center of the ring electrode as shown in Figure D-12 (b). At a still lower pressure of 10 microns the diffuse negative glow filled the whole interior of the bell jar and a beam an inch or so in diameter could be seen emerging from the electrode assembly and proceeding straight towards the top of the bell jar without any discernible change in diameter, Figure D-12 (c).

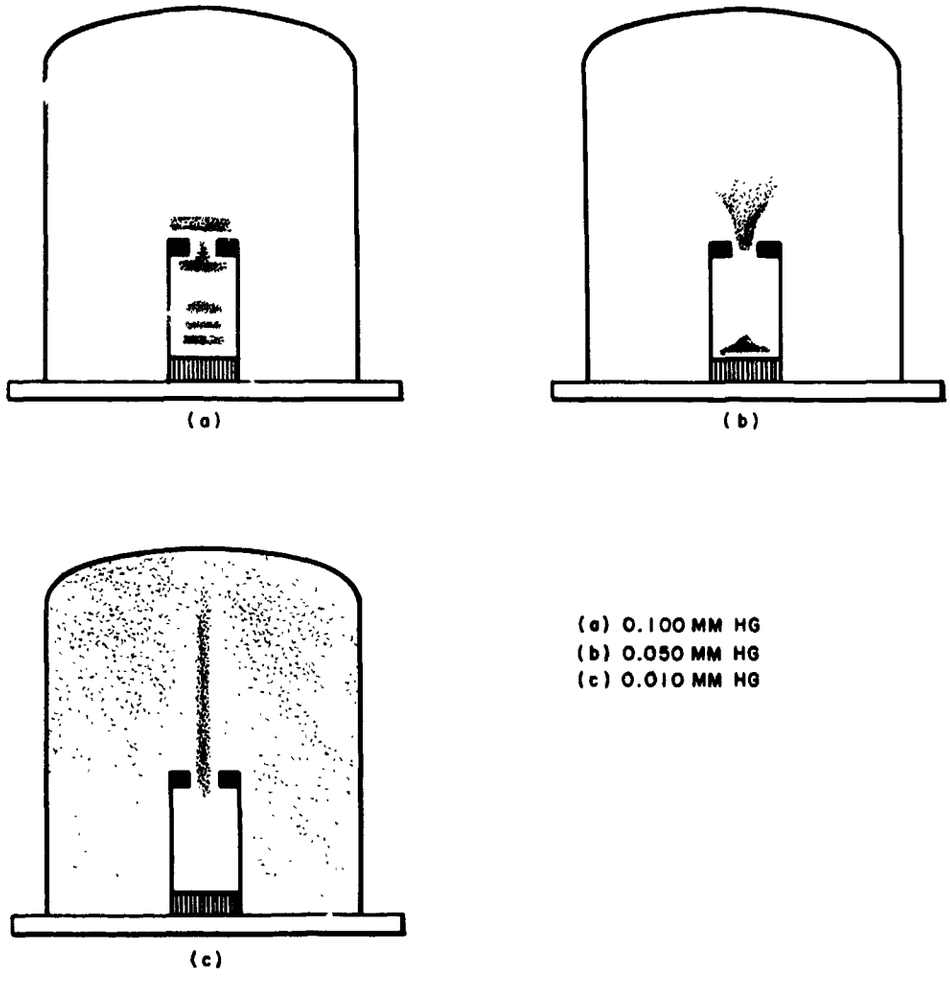


Figure D-12–Brush discharges at various pressures.

The Radio Frequency Discharge

Encouraged by the ability of direct current discharges to fill the whole of the bell jar, at least at favorable pressures, a small air core coil designed to resonate at 21 megacycles and connected to a coaxial cable as an auto-transformer, was mounted on a teflon rod inside the bell jar. RF power for the coil was supplied by a radio transmitter operating in the 20-30 megacycle region. A discharge was easily obtained at a pressure of 500 microns. An intense pink glow formed within the coil and diffused outwards on all sides. The discharge was maintained without difficulty down to a pressure of 5 microns. A decrease in pressure resulted in a more uniform glow density throughout the whole volume which gradually changed from a concentrated pink to a rather deep blue diffuse glow.

LARGE CHAMBER EXPERIMENTS

Radio Frequency Discharge

Introduction—To determine the practicality of producing a large volume discharge under radio frequency excitation, a coil was placed in a 4' × 5' vacuum chamber. The chamber easily reduced to pressures of a few microns with the forepump alone and was capable of attaining pressures as low as 10^{-6} mm Hg. A coil designed to resonate at 21 megacycles was fabricated of 1/8" copper tubing having a diameter of approximately 2", a pitch of 1/8" and a length of 3". Its measured inductance was 3.29 millihenries. The coil was mounted on a teflon support rod in the chamber as shown in Figure D-13. Excitation was accomplished by operating the coil as an auto-transformer with a one turn primary. A 50 ohm coaxial cable lead from the coil through the vacuum wall to a radio transmitter of one kilowatt capacity.

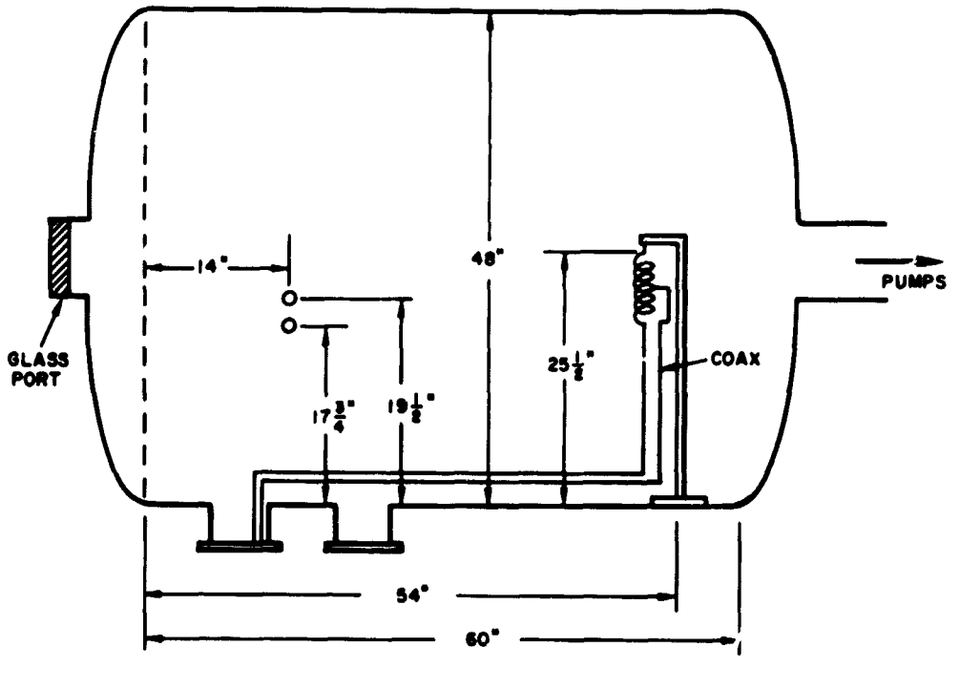


Figure D-13–Large chamber, R.F. coil and double probe.

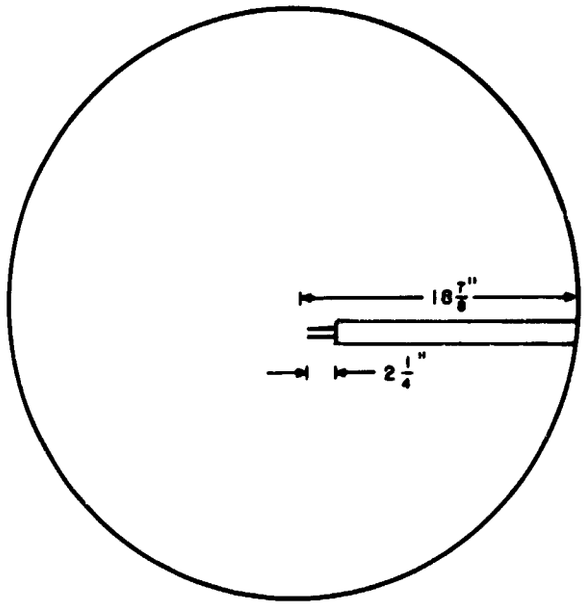


Figure D-14–Chamber cross-section showing double probe.

A double probe system was installed in the chamber as shown in Figures D-13 and D-14. The coil was excited at a frequency of 21.5 megacycles at a pressure of approximately 1 mm Hg.; this produced a luminous glow about a few top turns of the coil. With a decrease in pressure to 200 microns an intense pink glow appeared in the center of the coil. The effect of the auto-transformer connection was clearly visible as two dark regions formed along the exterior surface of the coil followed by diffuse luminous areas whose shape followed the flux lines of the coil as shown in Figure D-15. With a continued stepwise decrease in pressure to 100, 5, 1.0 and 0.1 microns Hg, the glow first extended itself by covering the coaxial feed cable with a luminous sheath and later spread out over a section of the chamber floor. The whole chamber now appeared to have acquired a faint luminosity throughout its whole volume.

A further decrease in pressure resulted in a less intense glow within the coil and corresponding increase in the intensity in the rest of the chamber. The color underwent a transition from a strong pink at high pressures to a deep blue at lower pressures. At the lowest pressure achieved, 10^{-6} mm Hg., a faint blue glow filled the whole chamber appearing to be very uniform in intensity so that even the region inside the coil itself had no marked luminosity.

The character of the discharge was observed as a function of the input power level at pressures in the 50 micron range. Generally an increase in input power resulted in an increase in light intensity; however, if the power further increased, discharge instabilities occurred in the form of large globs of distinctly higher light intensity on the chamber walls. More than a dozen of such spots have been observed around an inner circumference of the chamber at one time, and these may be made to rotate in one direction or the other depending on the input power. Under this excessive stimulation the discharge is very bright and a beam of considerable light intensity passes out through the viewing port.

Measurements—Using the experience gained in the bell jar studies, a program of measurements to determine the characteristics of the discharge was undertaken. To avoid excessive cycling of the chamber an axial and radial array of single probes was placed in the chamber as shown in Figure D-16. Supported by a glass composition rod, the probes

consisted of a series of individual wires strung along the rod each terminating at a specific location. At its location the wire probe extended four inches below the rod ending in an exposed portion one inch in length. Exterior access to each individual probe was provided by a 32 pin vacuum feed through connector. Attempts to interpret individual probe data according to Langmuir's method were not completely satisfactory. Measurements were taken of the floating potential. This is the potential a probe in a plasma must assume if it is to collect equal numbers of ions and electrons. In the external circuitry this is observed as that potential to which a probe must be biased to draw zero current. A measurement of the floating potential enables one to determine readily the uniformity of the generated plasma through the chamber volume; furthermore, it is the potential a test object placed in the plasma would assume. The fact that the floating potential varies little in the discharge indicates that the potential should be nearly uniform over the interior of the chamber. Axial and radial potential profiles for the discharge, using the probe locations of Figure D-16 were obtained at various pressures and degrees of excitation using nitrogen gas, (see Figures D-17, D-18, and D-19). The uniformity of these curves, excepting the potential rise in the vicinity of the coil, in both the axial and radial directions appears to indicate that a large uniform volume, (approximately 40 ft^3), of plasma has been generated.

In order to explore the temperature and density properties of the plasma generated, probes F and G of the probe array were connected as a double probe circuit. The plots of Figures D-20 and D-21 are representative of the results obtained. These curves exhibit many of the characteristics of double probe curves and are reasonably symmetric; they differ from the normal double probe characteristic curve in the vicinity of zero voltage. The slope through this zero point is essential in the usual method of temperature calculation, and the probe current is not a simple function of bias voltage.

The radio-frequency field which penetrates the plasma may also be a contributing factor. A simple radio-frequency filter of two series chokes and a capacitor in parallel with the probe output leads produced no appreciable change in the results. It seems possible that it might be necessary to shield the probes from the radio-frequency field.

Discharge Frequency Spectrum—It is well known that discharges produce noise in the microwave region and several empirical laws analogous to Wien's displacement law have been observed. Under the assumption that the discharge radiates as a black body at a noise temperature, τ_N, a value for the electron temperature, τ_e, may be calculated. The noise power radiated in a given frequency interval Δf appears to be accounted for by the fluctuations in the electron current density in the plasma. Though no microwave noise measurements were made on the plasma an investigation of the noise spectrum was made in the 30-240 megacycle frequency range. One of the probes in the chamber was used as an antenna and the relative signal strength was recorded as shown by a radio-frequency signal level meter on the receiver. Two different frequencies 21.2 and 27.0 megacycles were used to excite the discharge at a pressure of 3.6 microns. The results are recorded in Tables D-1 and D-2.

In most cases the strength of the harmonic signals was larger when the plasma was present. There does not seem to be any simple correlation between the signals received at non-harmonic frequencies in the two cases.

Direct Current Discharge

Introduction—The investigation of the properties of the direct current discharge in the 4' × 5' vacuum chamber utilized the screen cathode assembly as an ionization source. This discharge, for the pressure range in which it is useful, is largely diffusion controlled, hence it is not critically dependent upon source geometry or its exact location. The source

a 100 Hg

b 5 Hg

c

d 0.1 Hg

[illegible] 3. [illegible] the pressure [illegible]
[illegible] the pressure [illegible]

(e) 250 Hg

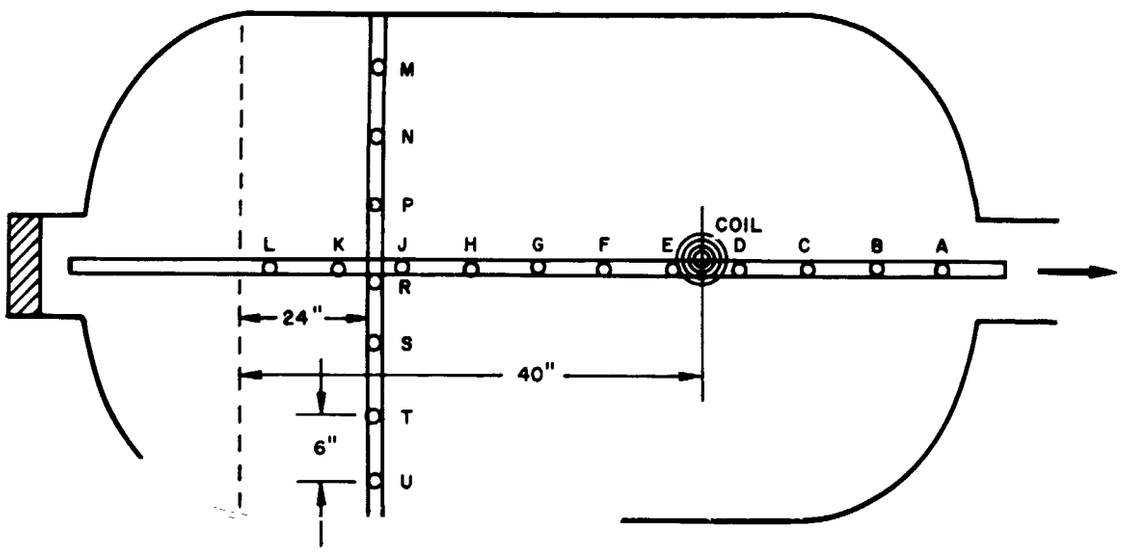
(f) 100 Hg

(g) 30 Hg

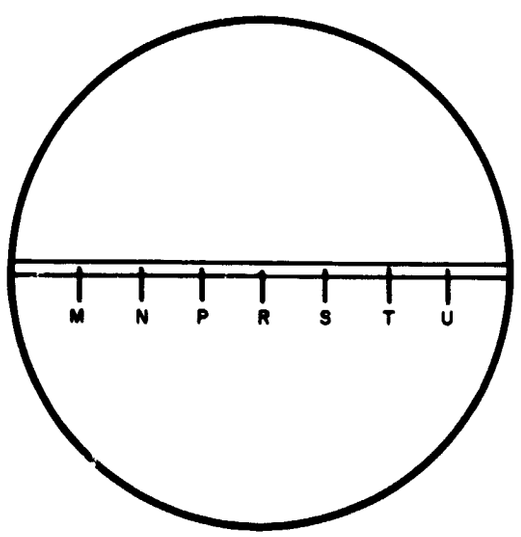
(h) 10 Hg

Figure D-15 (Continued) Glow discharges in air. Pictures (a) through (d) depict RF discharges as the pressure decreases. Pictures (e) through (h) depict screen cathode discharges as the pressure decreases.

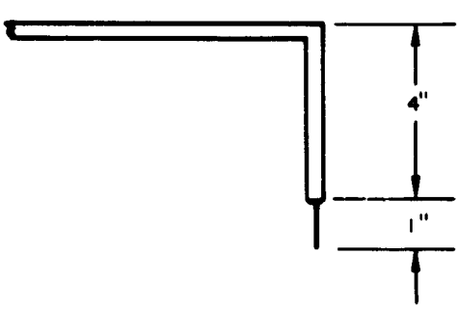
BLANK PAGE



(a) CHAMBER PROBE ARRAY
TOP VIEW

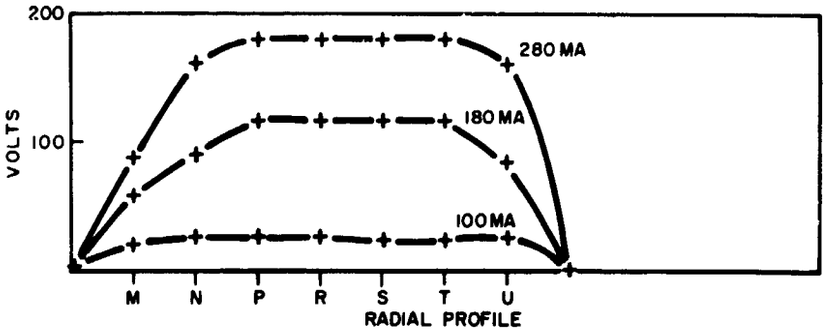


(b) CHAMBER PROBE
ARRAY
CROSS SECTION



(c) SINGLE PROBE DETAIL

Figure D-16–Chamber probe array



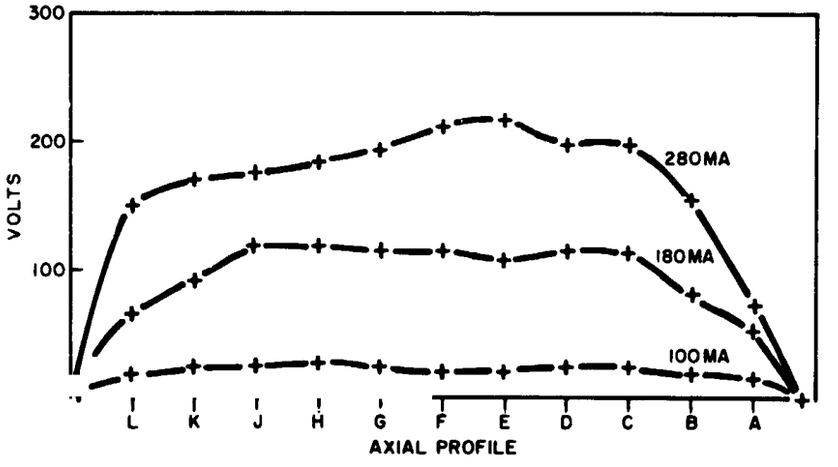
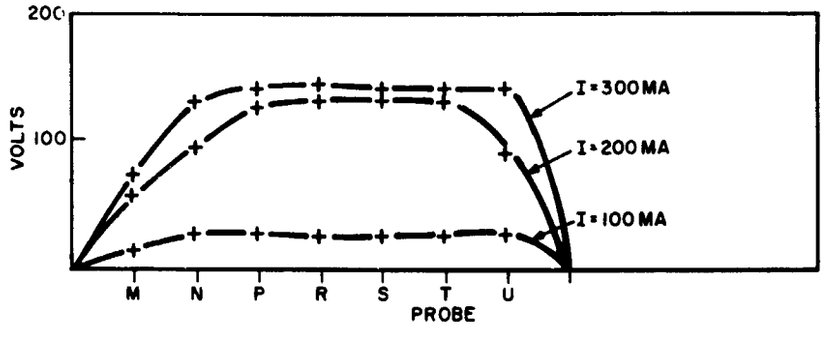


Figure D-17–N2-gas, pressure 150μ, frequency 28.5 Mc.



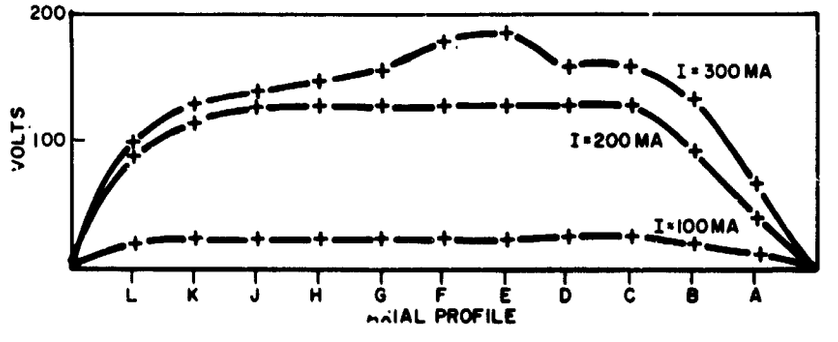


Figure D-18–N2-gas pressure 50μ, frequency 28.5 Mc.

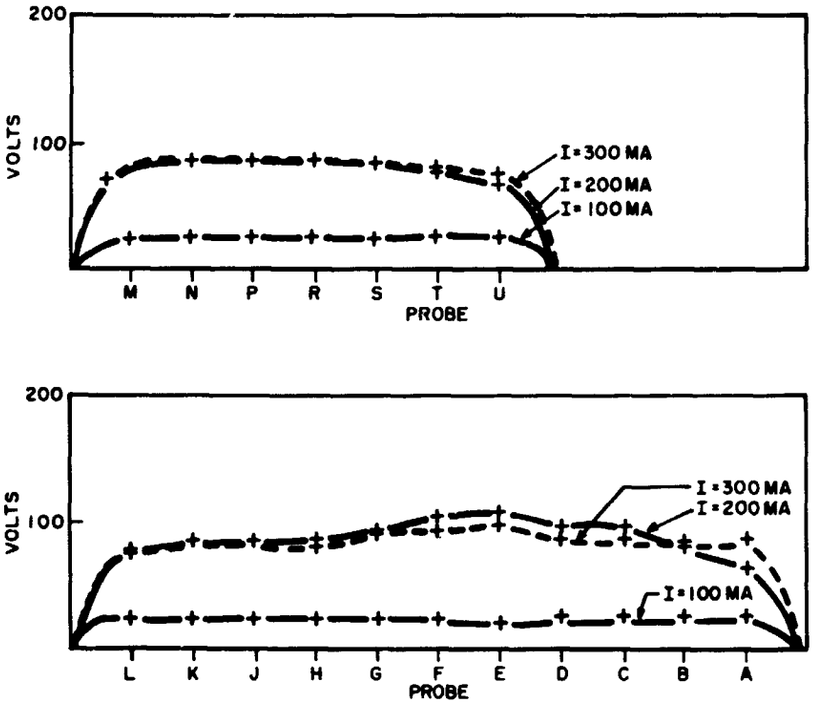


Figure D-19–N2 gas, pressure 10μ, frequency 28.5 Mc.

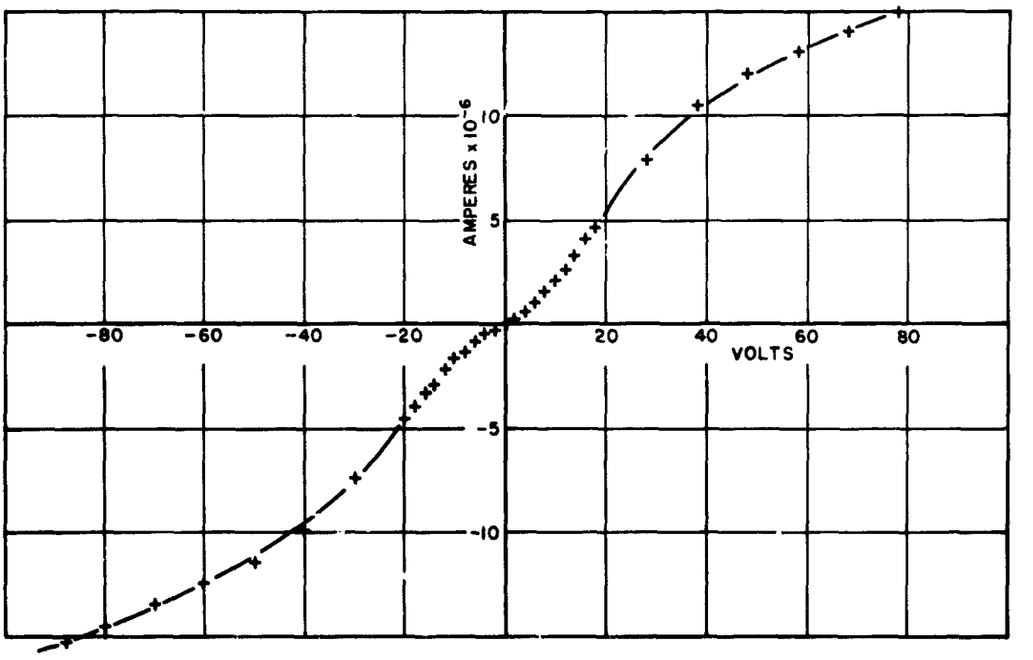


Figure D-20–Double probe, pressure .05 mm Hg, frequency 21.9 Mc, current 150 ma.

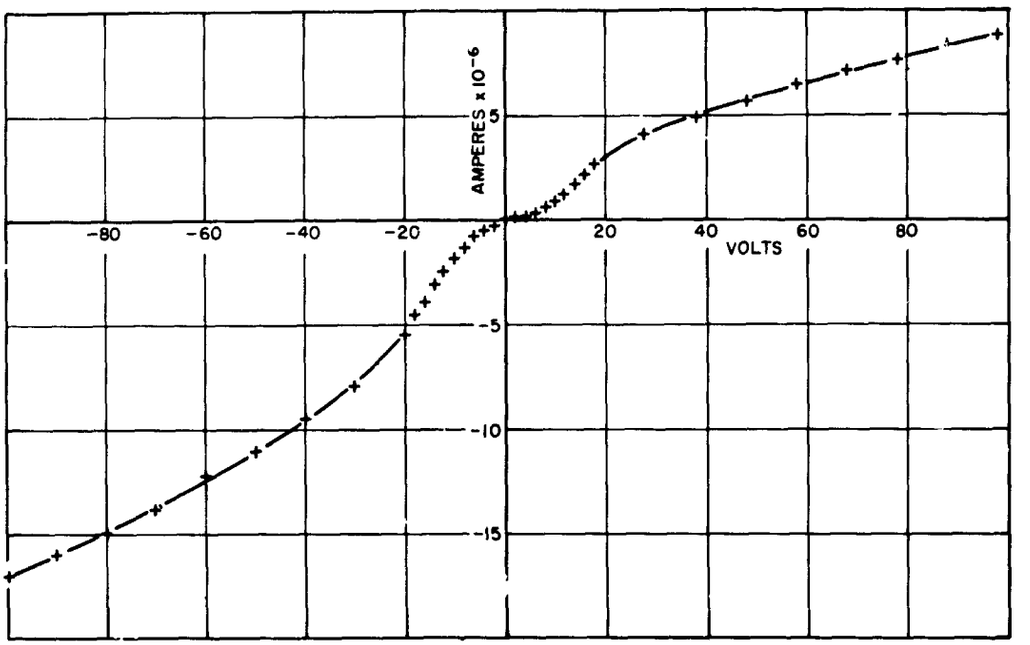


Figure D-21–Double probe, pressure .010 mm Hg, frequency 21.9 Mc, current 50 ma.

Table D-1
Relative Intensity of Observed Signals
Frequency 27.0 Mc

Frequency Mc	Plasma Present	Plasma Absent	Harmonic	Frequency Mc	Plasma Present	Plasma Absent	Harmonic
54.0	8.0	3.0	2	99.4	0.0	4.0	
61.9	6.0			108	10.0	4.0	4
63.8	9.0			131	5.8		
66.0	8.0			134.8	10.0		5
69.0	8.5			161.5	5.0		
71.2	8.0	2.0		163.5	10.0		6
76.0	6.5			179.5	0.0	6.0	
79.2	7.5			188.0	10.0		7
80.8	6.5			198.0	0.0	5.5	
81.6	6.0	7.0	3	216.0	10.0	4.0	8
86.0	6.0	1.5		239.0	5.0		
91.1	7.5			243.0	6.0		
93.1	6.5			244.0	10.0		9
98.0	7.0	3.0		265.0	5.0		

Table D-2
Relative Intensity of Observed Signals
Frequency 21.2 Mc

Frequency Mc	Plasma Present	Plasma Absent	Harmonic	Frequency Mc	Plasma Present	Plasma Absent	Harmonic
63.6	10.0	1.0		148.5	9.0	1.0	7
67.3	9.0	1.5		169.1	4.0		
84.6	10.0	2.0	4	170.0	9.0	0.0	8
95.5	5.0			190.0	9.0	0.0	9
102.8	3.0			212.0	9.0	5.0	10
105.0	5.0			231.0	3.0		
106.0	9.0	1.0	5	[illegible]4.0	8.0	0.0	11
115.0	10.0			254.0	5.0	0.0	12
127.2	9.0	1.0	6	255.0	6.0		
147.0	3.0						

was placed on the floor of the chamber 40" from one end occupying a position near to the location of the coil used for the radio frequency discharge. The positive terminal of the direct high voltage source was grounded and connected to the chamber wall which served as one electrode, the other, the screen, was biased to a high negative potential with respect to ground. Visual observations were made of the discharge as the pressure decreased. At 500 microns a deep purple glow appeared about the screen partially covering it.

A stable discharge was obtained at a pressure of 200 microns, a current of 17 milliamperes and a voltage of 550 V. The discharge was very stable. As expected the discharge spreads out over a larger volume as the pressure falls, the glow on either side of the screen thickens to several inches and a reddish film covers a portion of the chamber beneath the screen. In the low micron pressure range, the whole chamber takes on a pale bluish hue of constant density. Successively higher voltages are required as the pressure drops below 20 microns to sustain the discharge. A potential of 1000 volts is required at 10 microns.

Floating Potential Measurements—The probe array constructed for the radio-frequency discharge was left in the chamber and a measure of the floating potential at each probe location was taken as a test of the uniformity of the discharge. The results of these measurements for pressures of 10, 50 and 200 microns are displayed in Figure D-22.

Note that the floating potential increases with pressure, a sign which is a manifestation of the particle energy. For any given pressure the potential is very uniform over a large volume of the chamber. Because of these characteristics a study was made to determine how the floating potential varied with source voltage at constant pressure. Taken at a pressure of 200 microns under slightly different circumstances than those of Figure D-22. Table D-3 tabulates values of the floating potential as the high voltage is changed.

The current through the discharge increases while the floating potential remains substantially constant. This leads one to conjecture that the particle energy has a tendency to remain constant.

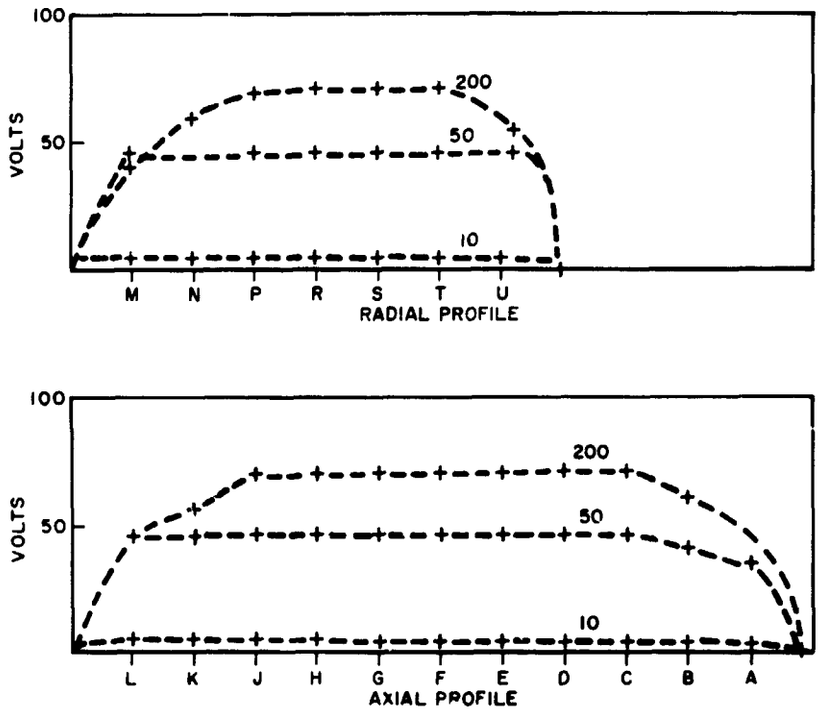


Figure D-22–Direct current potential distribution as a function of pressure (in microns).

Table D-3
Floating Potential Versus High Voltage

Floating Potential (volts)	High Voltage (volts)	Discharge Current (milliamperes)	Floating Potential (volts)	High Voltage (volts)	Discharge Current (milliamperes)
73.0	550	17.0	74.8	640	23.0
73.4	560	18.0	74.6	650	25.0
74.5	580	19.0	74.5	660	26.0
74.6	590	20.0	74.5	670	27.0
74.7	600	20.2	74.3	680	28.0
74.8	610	21.0	74.9	690	29.0
74.8	630	22.5	74.4	700	29.2

Langmuir Probe Measurements–The determination of the temperature and electron density in the discharge under various conditions was accomplished by using probe H of the array as a single probe. It was connected in the manner shown in Figure D-11. Illustrative of this method are the graphs shown in Figures D-23 and D-24. The first is a plot of the measured probe current as the bias voltage on the probe is increased from negative through positive values. When biased strongly negative the probe collects only positive ion current exhibiting a saturation characteristic. For smaller values of negative bias the probe collects ions and some fast electrons which have sufficient energy to

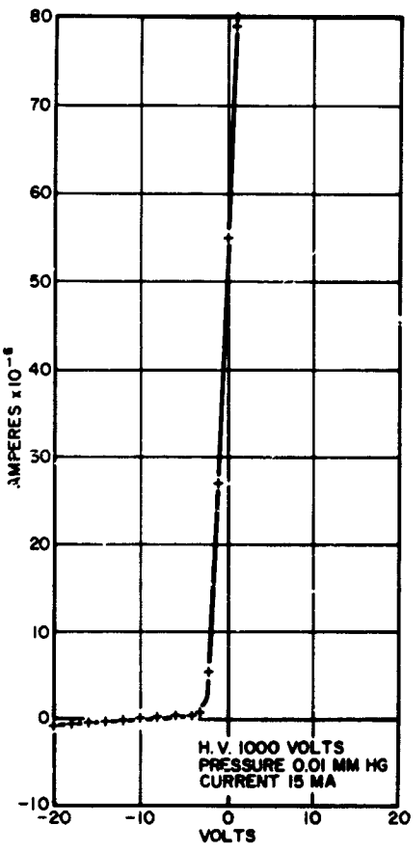


Figure D-23–Measured probe current as the bias voltage on the probe is increased from negative through positive.

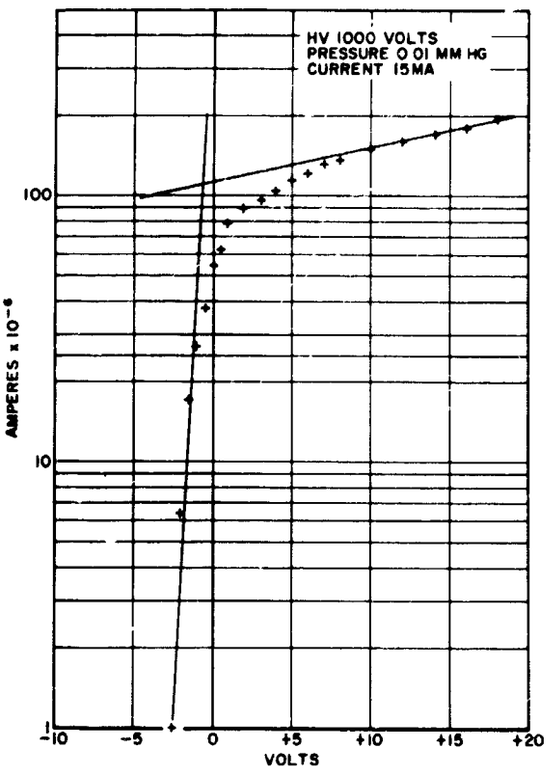


Figure D-24–Semi-logarithmic plot of current vs. voltage.

overcome the repulsive potential so that the total collected current begins to decrease. For probe bias values near zero and for positive values, the current rises abruptly as electrons are collected from the discharge.

Figure D-24 is a semi-logarithmic plot which includes the very large values of electron current collected when the probe is biased positively. From a determination of the slope of the vertical portion of the curve of the semi-log plot of current versus voltage, the temperature may be calculated.

$$\text{slope} = \frac{\ln 27.0 - \ln 1.0}{1.0}$$

$$\tau = \frac{11{,}600}{5.59} = 2072^{\circ}\text{K}$$

$$n_e = \frac{i}{Ae}\sqrt{\frac{2m}{kT}} = \frac{1.0 \times 10^{-6}}{\sqrt{2072}}\,(4.43 \times 10^{19})$$

$$n_e = 9.74 \times 10^{5}/\text{cm}^3$$

The current, i, in the above equation for the electron density, n_e, is that attributable to positive ions alone and is derived from Figure D-23. A is the probe area, m, the electron mass, e, its charge and k the Boltzmann constant. A more detailed description of the application of probe techniques to gaseous discharges may be found elsewhere in this volume.

In Table D-4 are listed the results obtained for different discharges.

Table D-4
Results of Different Discharges

Pressure Microns	Discharge Current Milliamperes	Temperature °K	Electron Density n/cm^3
200	17	7919	1.25×10^4
50	20	7918	6.23×10^4
20	20	4595	7.96×10^5
10	15	2072	9.74×10^5

It is apparent from these values that the electron concentration increased as the pressure decreased. In the pressure range from 200 to 50 microns, the temperature remained constant after which it begins to fall rapidly. A re-examination of the potential profiles of Figure D-22 also seems to substantiate this trend.

Constant Current Discharges—The circuit arrangements for a discharge require that an external current limiting resistor be used in series with the high voltage lead since the discharge itself is highly conducting. It is possible to vary the value of this external resistor and by adjustment of the high voltage keep the same current flowing through the discharge. This procedure was followed in an effort to determine the effect of a varying potential applied to the discharge electrodes themselves under constant current conditions. The results are shown in Table D-5 with the electrode potential V_{po}

Table D-5
Results of a Varying Potential Applied to the Discharge Electrodes Under Constant Current Conditions

Pressure Microns	High Voltage Volts	V_p Volts	Resistor Kilohms	Current Milliamperes
200	510	433	3.3	20
200	520	420	5.0	20
200	610	410	10.0	20
50	585	518	3.3	20
50	602	502	5.0	20
50	690	490	10.0	20

CONCLUSIONS

The electron concentration attained by the direct current discharge method of ionosphere simulation varied over the range 10^4 to 10^6 electrons/cm^3. A comparison of these quantities and the normal daytime values for the ionosphere itself reveal that a considerable portion of the D, E and F regions fall within this range.

Figure D-25 shows three single probe characteristic curves taken under these conditions. The coincidence of these curves implies that the total number of primary electrons generating the plasma remains invariant while their energy changes and neither the electron temperature nor the concentration changes.

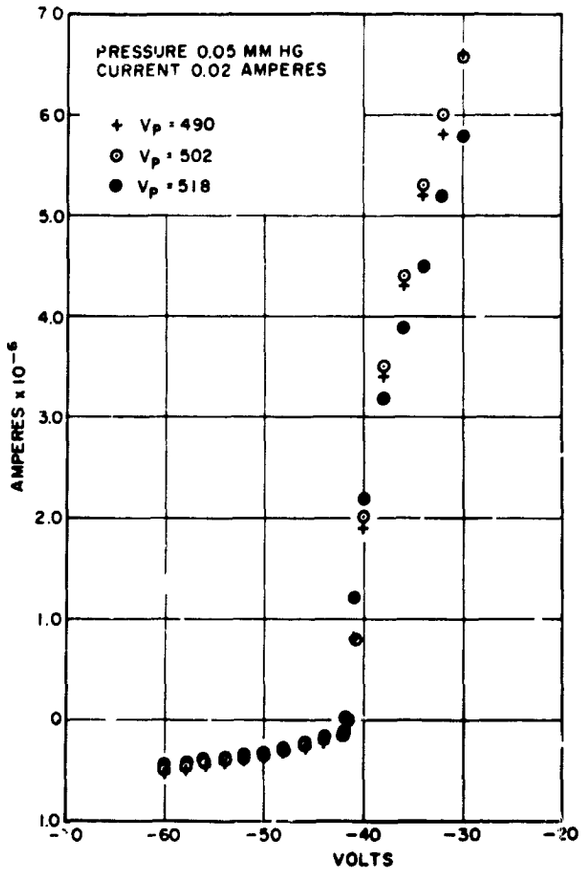


Figure D-25–Single probe characteristics under constant discharge current conditions.

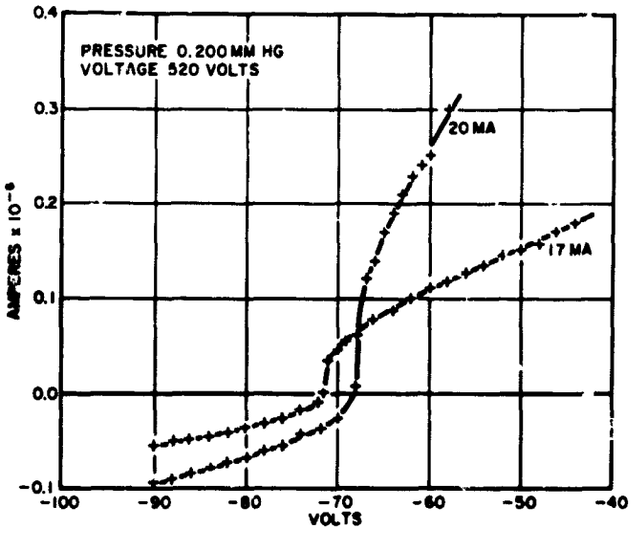


Figure D-26–Single probe characteristics under constant discharge voltage conditions.

Variable Current Discharges—The results of the previous section pose a further question. What is the effect of varying a current through a discharge? Figure D-26 shows the result. The temperature which is related to the logarithm of the slope of the vertical portion of the curve remains unchanged while the concentration which is dependent on the collected positive ion current increases with increasing discharge current. Control over the charged particle energy can be exercised by a variation of the environmental pressure and control over particle concentration by a change in the discharge current.

Because of the correlation between the floating potential and the temperature, which is a property of the plasma itself rather than its mode of excitation, and the fact that this floating potential is generally higher than the corresponding value in the direct current discharge, it is probable that the temperature in the radio frequency discharge is higher. The total power input used was larger than in the direct current discharge and accordingly the charged particle density will be larger. The useful pressure range for the radio frequency discharge is 1000 times larger than that of the direct current discharge.

The use of both the direct and radio frequency discharges for simulation purposes will be enhanced if the plasma is allowed to diffuse through a grid or screen into the test area. This ensures that the test area is a field free space and adds a second means of control over particle energies.

It is quite probable that no difficulty will be experienced in producing an even larger volume of plasma than that used in this study.

In large volume excitation the geometrical relations of the source to the containing vessel are of minimal importance compared to their relative effect in a small discharge.

ACKNOWLEDGMENT

The authors wish to express their appreciation to Mr. George Cope for fabricating the coils and to Dr. Donald Waidelich for his continued suggestions and encouragement.

REFERENCES

1. Johnson and Malter, Phys. Rev. Volume 80, page 58, (1950)

2. "Satellite Environment Handbook," page 28, Stanford University Press (1961)

Appendix A

IONOSPHERE DATA*

Altitude km	Temperature °K	Particle Speed m/sec	Free Path m	Collision Frequency sec^{-1}	Concentration n/m^3	Pressure mm Hg.	Ionosphere Region
50	283	454	7.5 (-5)	6.06 (6)	2.25 (22)	0.66	D 50-85 km
80	166	348	3.80 (-3)	9.10 (4)	4.42 (20)	7.56 (-3)	
90	166	348	2.90 (-2)	1.22 (4)	5.90 (19)	1.02 (-3)	E 85-140 km
140	850	795	26.3	30.0	6.40 (16)	5.63 (-6)	
200	1404	1062	201	5.30	8.40 (15)	1.22 (-6)	
400	1480	1266	6,098	0.21	2.77 (14)	4.25 (-8)	
600	1691	1429	74,270	0.019	2.28 (13)	3.98 (-9)	

*ARDC Model atmosphere

Appendix B

KENETIC THEORY DATA

Pressure mm Hg.	Concentration n/m^3	Ion Free Path cm	Electron Free Path cm
750	3×10^{19}	10^{-5}	5.6×10^{-5}
0.50	2×10^{16}	1.8×10^{-2}	10.2×10^{-2}
0.40	1.6×10^{16}	2.25×10^{-2}	12.7×10^{-2}
0.30	1.2×10^{16}	3.0×10^{-2}	16.9×10^{-2}
0.20	8×10^{15}	4.5×10^{-2}	25.5×10^{-2}
0.10	4×10^{15}	9.0×10^{-2}	50.9×10^{-2}
0.05	2×10^{15}	1.8×10^{-1}	10.2×10^{-1}
0.02	8×10^{14}	4.5×10^{-1}	25.5×10^{-1}
0.01	4×10^{14}	9.0×10^{-1}	50.9×10^{-1}

BLANK PAGE

MEASUREMENTS IN A PLASMA

J. Boright

INTRODUCTION

In ionospheric research the important parameters are electron temperature (T_e), ion temperature (T_i), electron density (N_e), ion density (N_i), velocity distribution $f(v)$, percent composition, and charged-to-neutral density ratio. In a laboratory plasma, the space potential (V_s) will also be of interest. The concept of temperature is of limited application: When there is a Maxwellian distribution of velocities, τ is defined by

$$f(v) \propto e^{-\frac{mv^2}{2kt}} \qquad (1)$$

where $f(v)$ is the number of particles with energy E. T_e and T_i are usually unequal in a plasma, T_e being higher because of greater electron mobility. Also, it has been observed that distinct populations of electrons with different T_e's can coexist in a plasma. The ionospheric T_e's are on the order of 1500°K or less. The percent composition refers to the relative amounts of molecular and atomic gases, e.g., below 100 km the ionosphere is about 20% O_2 and 80% N_2. The ion composition may differ from the neutral composition, i.e., at 100 km most ions are NO^+ (1). A brief investigation was made of the methods of making measurements in such a laboratory plasma and the results are reported here. Also some experimental work on various methods of measurement was done and a few typical results are presented. Atmosphere electron densities vs altitude are shown in Figure D-27.

MEASUREMENT METHODS

Some possible measuring techniques are spectroscopy (2), microwaves (3), mass spectrometry (4, 5), ion beams (6), emission of noise in the EM spectrum (7), and Langmuir probes (8).

Microwave experiments are based on the observation of the effect of a plasma on the passage of an electric field. The plasma acts as a dielectric with constant

$$K = 1 - \frac{4\pi N_e e^2}{m\omega^2} \qquad (1)$$

where ω is the electric field frequency. From this N_e can be found. This method draws no current and involves no internal apparatus.

In a mass spectrometer probe, ions are drawn out of the plasma and their e/m ratio is found from their path in a magnetic field. Few measuring methods give the composition as this one does. Ion traps on satellites also give ion composition results (9). Noise measurement, with the assumption that power radiated is given by $KT_e \Delta f$, gives values for T_e. The following discussion will treat only the methods used in our experimental work, i.e., Langmuir probes and their extensions. A general problem with the other methods are the low densities involved in ionospheric simulation.

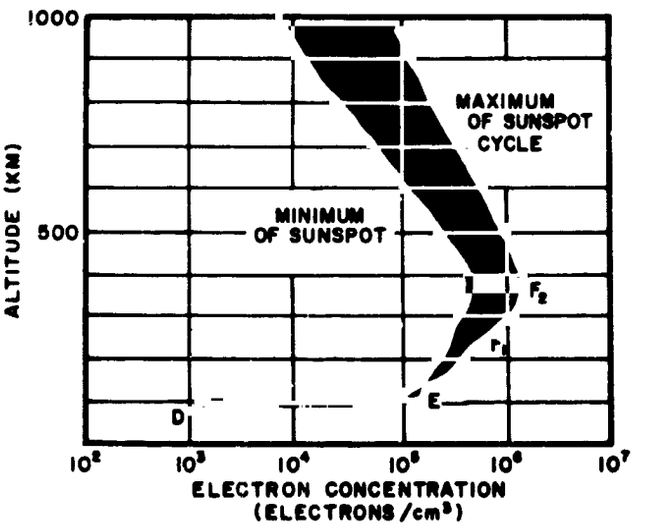


Figure D-27–Electron concentration (electrons/cm^3).

PROBES

The classical probe method is the single probe of Langmuir (8), abbreviated SPM in the following. A probe is a small electrode placed in a plasma, e.g., a few centimeters of wire. The circuit is shown in Figure D-28.

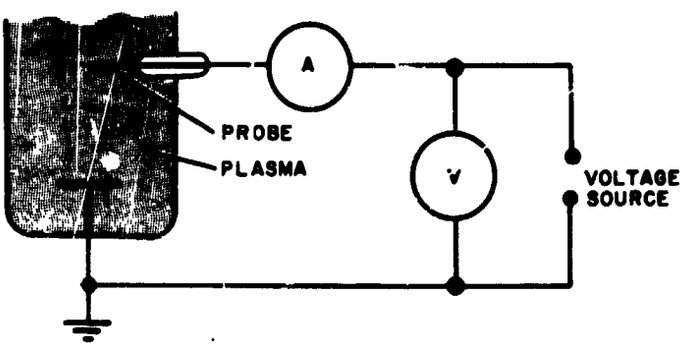


Figure D-28–Circuit diagram for single probe.

The probe characteristic is a plot of probe current vs. probe voltage. The general form, and a semilog plot of i_e, are shown in Figure D-29. The length A B represents ion current, having little variation with voltage. In B C the higher energy electrons penetrate the sheath, an area around the probe which contains few electrons. It is assumed that all the potential change from probe to plasma takes place across the sheath. With increasing voltage, more electrons get through. If Equation (1) hold, i.e., there is a Maxwellian distribution, then the i_e value in the BD part of the curve will have an exponential rise, and the log plot in Figure D-29(b) will yield T_e. The equation is

$$\ln i_e = \left(\frac{e}{KT_e}\right) V_p + \ln Aj_0 - \left(\frac{eV_e}{KT_e}\right) . \qquad (2)$$

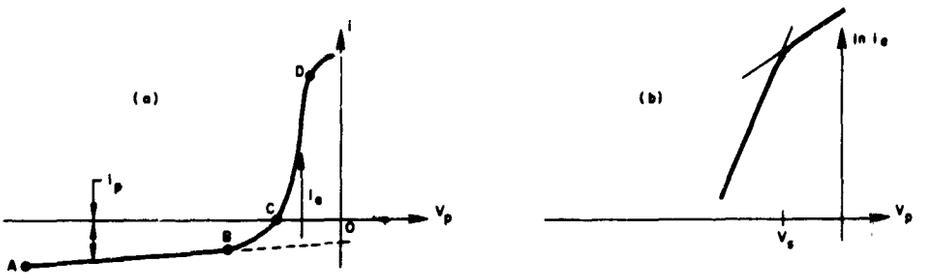


Figure D-29–Single probe i-v plot and semilog plot.

The slope of the log plot is e/KT_e. The point D of Figure D-29(a), corresponding to V_s in Figure D-29(b), is identified as the plasma potential. Here, with no potential difference from probe to plasma, the sheath disappears. The point C of Figure D-29(a), with no current, is the "floating" or wall potential, V_f.

Since the electron current density is

$$j_e = eN_e \sqrt{\frac{KT_e}{2m}} \cdot e^{-eV/KT_e}$$

N_e can be found from the value of i_e at V_s, since here V, the plasma to probe potential, is zero:

$$N_e = \frac{i_e}{Ae} \sqrt{\frac{2m}{KT_e}} \tag{3}$$

where A is the sheath area. Thus from the single probe characteristic can be found N_e, T_e, V_s, and V_f. If the pressure and charge neutrality are known, N_i and the charge-to neutral ratio can be inferred. Also, Druyvestein (10) devised a method for finding an unknown velocity distribution from single probe current readings:

$$f(eV) = \frac{m^2}{2\pi e^3 N_e} \frac{d^2}{dV^2} J_e \tag{4}$$

So a single probe characteristic gives a large amount of information.

As a practical method, the single probe method (SPM) draws rather large currents which at V_s and above can become comparable to the discharge current in a dc discharge. This changes the configuration, in effect introducing a new electrode. In particular, the N_e determination (see Equation (3)) may be impossible. Also, the single probe is attached to ground or an electrode, and is therefore not applicable to an electrodeless or a changing plasma. The probe changes the plasma potential. Decaying plasmas and high frequency discharges are examples of changing plasmas. It was found that Langmuir's assumption that no field extends beyond the sheath was incorrect (11). The ions, at lower energy, are affected by a weak field outside the sheath. Thus the region AB of Figure D-29(a) will not be a straight line. Ion collection theory has been developed, e.g., by Allen, et al (12). It is found that ion current is not constant, and should depend on T_e, not on T_i. Several complex methods for finding electron density from ion current data have been proposed. The ion portion of the curve will interest us, as discussed below.

The large currents drawn in the SPM led to the proposal of a double probe method (DPM) by Johnson and Matter (13). The circuit is connected as in Figure D-30, and the i-v characteristic resulting has the form of Figure D-31. The system as a whole floats, i.e., the total current of ions and electrons to the probes is zero. As with the single probe, V_f is negative with respect to the plasma due to the higher velocities of the electrons. Varying the voltage then varies the percent of the (constant) total number of electrons collected by the probes. At B all electrons go to probe 1, and at D all go to probe 2, if v is defined to be the potential of probe 2 with respect to probe 1. The voltage at point C is the potential difference between the plasma areas surrounding the two probes. The ion currents are assumed to be constant, independent of v, in the simplified theory. The discussion of sections AB and DE is the same as the ion collection portion of the SPM.

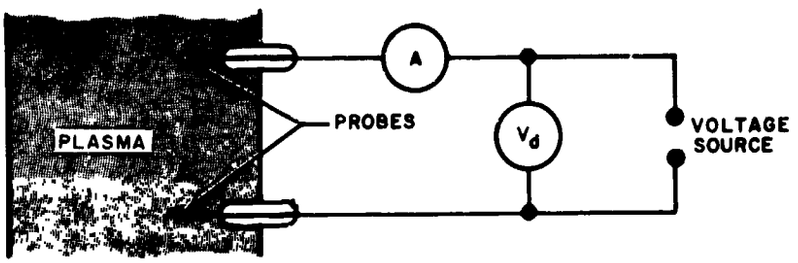


Figure D-30–Circuit diagram for double probe.

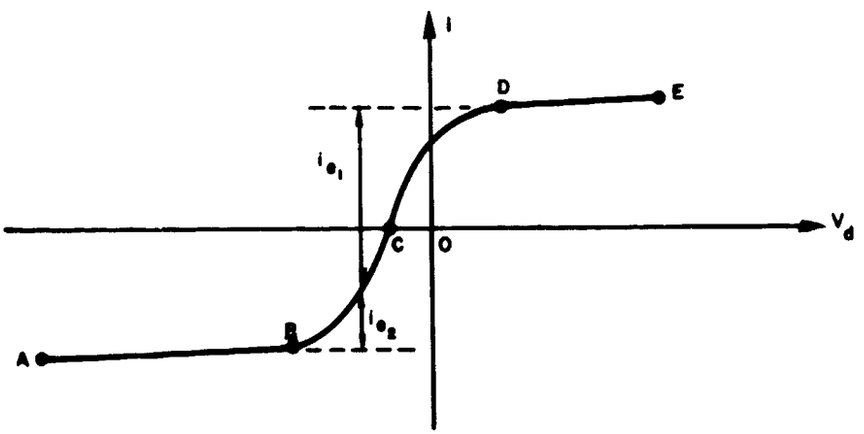


Figure D-31–Double probe i-v characteristic.

Three methods are described for deriving T_e from the characteristic, the simplest of which is similar to that of the SPM. From Kirchhoff's laws and assuming the Maxwellian distribution can be derived

$$\ln [(\Sigma ip/i_{e_2}) - 1] = -\phi V_d + \ln \sigma \qquad (5)$$

where

$$\sigma = \left(\frac{A_1}{A_2}\frac{j0_1}{j0_2}\right) e^{\phi V_c}, \quad \phi = \frac{e}{kT_e}$$

and V_c is the contact potential in the system. Then a semi-log plot gives T_e. Note that σ does not include V_s as does the corresponding term in the SPM. Yamamoto and Okuda (14) give the criterion that the log plot method is only applicable when

$$S << 1/R_0 \tag{6}$$

where S is the slope of section AB of Figure D-31 and $1/R_0$ is the slope at C. A second method, corrected for $s \neq 0$ gives

$$T_e = \frac{e}{k}(G - G^2)\,[\Sigma i_p]_{V_d=0}\left(\frac{1}{1/R_0 - S/2}\right) \tag{7}$$

where

$$G = [i_{e_2}/\Sigma i_p]_{V_d=0}$$

A sample calculation will be carried out later on for Figure D-32. The determination of N_e suggested in (13) involves: use of Langmuir's ion current theory, the assumption that T_i equals T of the neutral gas, and calculation of sheath size. N_e, then, is often difficult to determine.

The DPM gives no V_s or V_f data. DPM theory assumes a Maxwellian distribution of velocity distribution. The single probe characteristic is more useful, where it can be obtained, then the double probe characteristic. The problem, then, is to obtain the equivalent of the single probe characteristic under changing conditions. The simplest approach is to make one probe of a double probe much larger than the other. Then even the electron current to the smaller probe at V_s does not affect the larger, which remains at V_f. The characteristic then becomes the single probe curve of the smaller. Kagan and Perel (15) obtain the size criterion

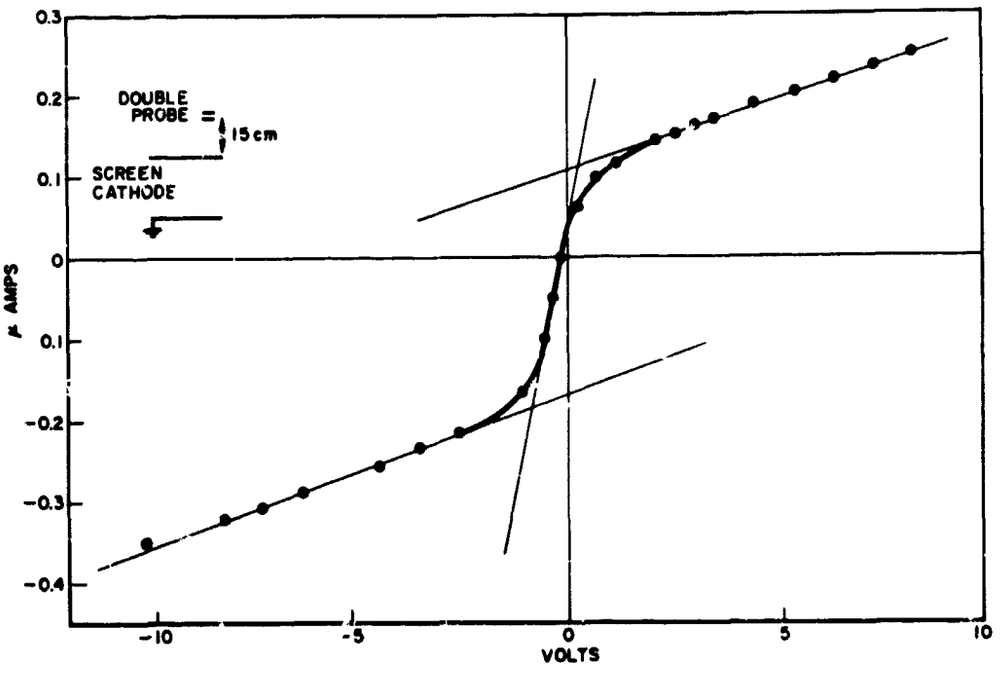


Figure D-32–Double probe screen cathode.

$$\frac{S_1}{S_2} >> \sqrt{\frac{M}{m}} \tag{8}$$

where S_1 is the area of the counterprobe, S_2 the area of the smaller.

In order to have a reference potential in an electrodeless discharge, a third probe can be introduced (14). This is always kept at V_f, by adjusting, for each V, so that the current in the third probe is zero. The circuit is shown in Figure D-33. This arrangement can give a single probe curve for all voltages up to slightly less than V_f. A modification of the triple probe is which probe 1 is larger than probe 2 (16) can give the single probe curve over the entire range up to V_s. Kagan and Perel (15) obtain for the size criterion here,

$$\frac{S_1}{S_2} \geq \sqrt{\frac{M}{m}}, \tag{9}$$

which is much more easily achieved than that in Equation (8). The counter probe in either a double or triple probe arrangement has the disadvantage of introducing a rather large electrode into the discharge and to some extent altering the configuration. In small tubes this is especially important. The effect of counterprobes is studied in (17).

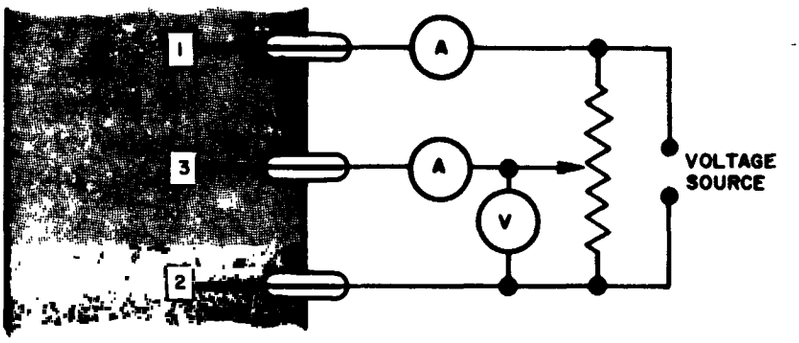


Figure D-33–Circuit diagram for triple probe.

Various aspects of probe measurements in high frequency discharges are discussed in (15), (17), (18), (19). A filter, consisting of a capacitor between the probe leads and chokes in the circuit, not shown in Figures D-30, D-33, keeps the ac current component out of the circuit. One extension of the use of the Langmuir probe involves the application of a periodic waveform to the probe (20), (21), (22). This method has the large advantage of giving density measurements in a non-Maxwellian plasma. All of the results of the SPM and DPM depend on a Maxwellian velocity distribution.

EXPERIMENTAL WORK

To illustrate probe interpretation and problems and to indicate some experimental results, some typical data will be considered. The equipment used is illustrated in Figure D-34. Pressures down to about 5 microns can be obtained in the jar. The probes can be raised or lowered above and below the screen. The following configurations were used:

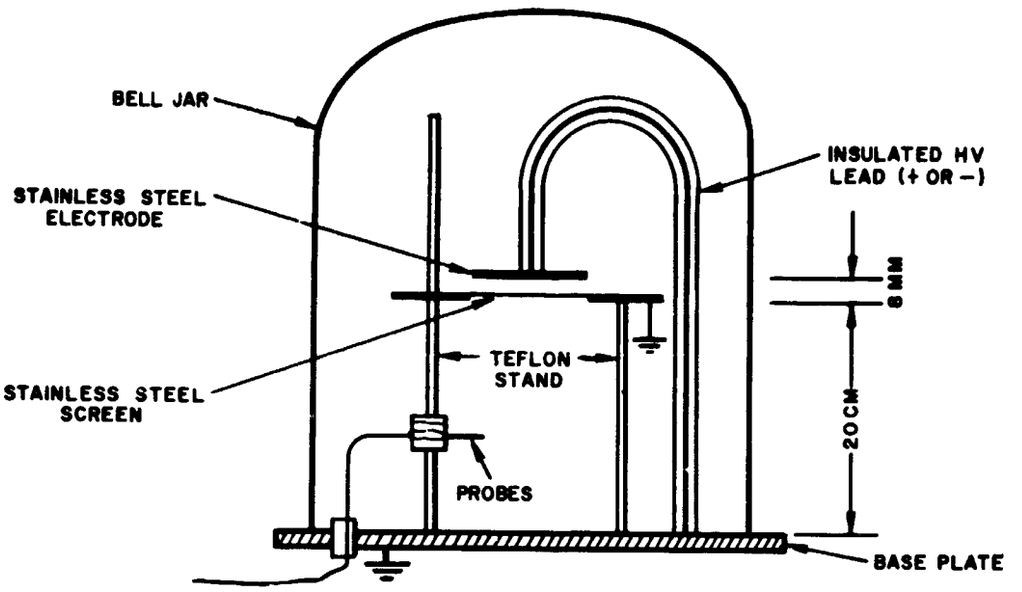


Figure D-34–Bell jar.

(a) as drawn, with the electrode + or −;

(b) with the electrode removed, and the high voltage connected to the screen (screen disconnected from ground).

In configuration (a), both the screen and the base are at ground. Electrons ionize the gas (air) between the screen and the anode, and ions are drawn to the screen, some passing through. We might expect a surplus of positive charges just below the screen.

The results of a run with a discharge voltage of 650 volts are shown in Figure D-35. The following observations might be made:

1. There is an excess of positive ions, since at $V_p = 0$ there is positive ion current. Also note that the V_p necessary to make the current zero is 90 volts.
2. The electron current, instead of rising exponentially, rises slowly.
3. The most interesting part of these results is the positive ion current at large negative voltage. We recall that Langmuir found a nearly constant ion current. However, our V_p is larger than that usually used.

Later, Bohm, Burhop, and Massey (11) derived the expression

$$I_i = .57\, n_e\, e(kT_e/M_i)^{1/2}\, A \qquad (10)$$

Here the ion current depends on T_e and the mass of ions, but not on V_p. This is because the sheath thickness was assumed negligible. No assumption that $N_e = N_i$ was made for Equation (3) and thus Equation (3) should be applicable to case at hand. Other treatments of ion collection, e.g. Allen, Boyd, Reynolds (12) derive ion characteristics, assuming $N_i = N_e$. Their theoretical curves do not resemble Figure D-35. Figure D-36 is the data of Figure D-35 as a semilog plot. The straight line resulting is very good, indicating an exponential dependence on V_p. The expression in (15) for ion current is

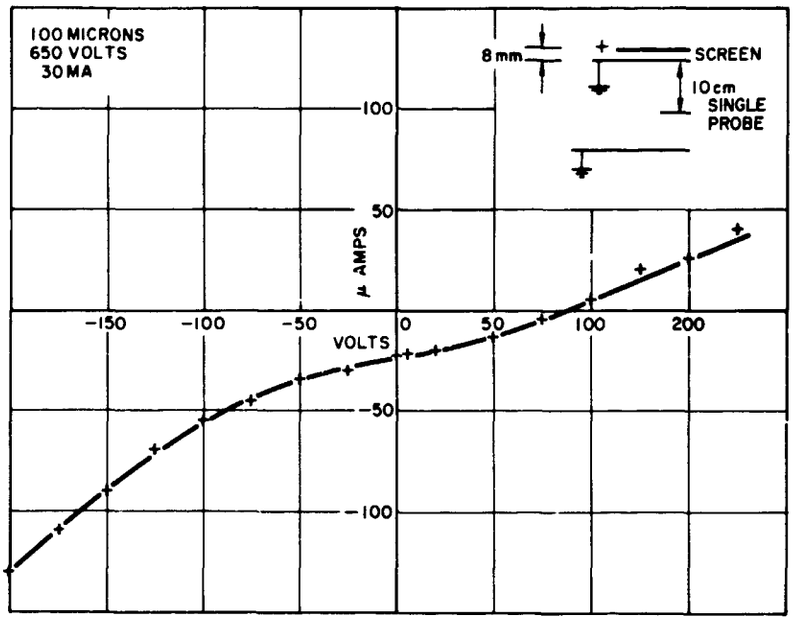


Figure D-35–Single probe characteristic.

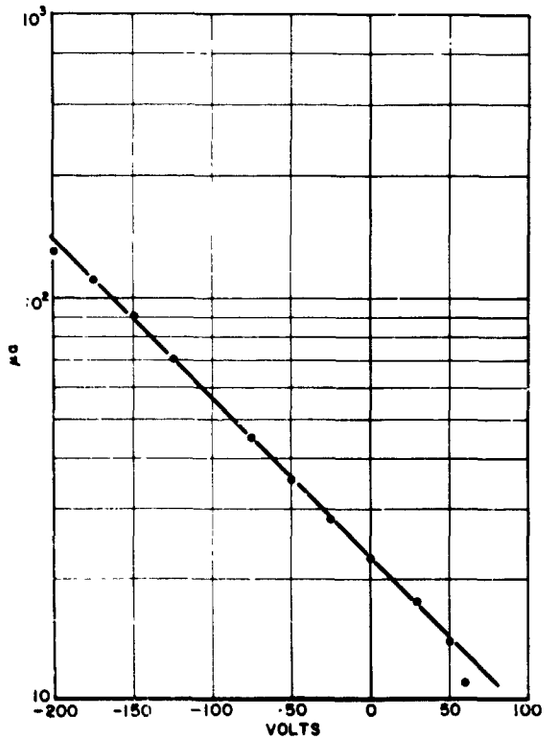


Figure D-36–Semilog plot of data from Figure D-35.

$$i_p = cN_e\, e^{-eV_s/kT_e} \sqrt{\frac{2\,eV_s}{M}}\; S_s,$$

where V_s is the plasma potential, and S_s is the sheath area. Assuming V_s is of the order of kT_e gives the expression

$$i_p = cN_e\, e \sqrt{\frac{2kT_e}{M}}\; S_s$$

This assumption has been made in other treatments, though not always stated. Since we use a large V_p, that assumption is not warranted here. Now $V_s = V_p + V$ where V is the potential across the sheath. Thus there is an exponential factor in V_p, but we still have

$$\sqrt{V_s}\, e^{-eV/kT_e}$$

in the expression. Values were determined from the log plots for the corresponding T_e, and the values resulting are unreasonably large in comparison to other T_e determinations. It seems apparent then that more work on ion collection and theory is needed.

The single probe data taken between the screen and anode of configuration (a) had the form of Figure D-29, with no obvious saturation, and the log plot is shown in Figure D-37.

T_e determined from line 1 of Figure D-9 is

$$T_{e_1} = 11{,}600 \times 13 = 1.51 \times 10^5 \text{ °K}$$

The sample double probe results are shown in Figure D-32. The asymmetry along the i axis can be caused by instrumentation problems or by the differences in probe dimensions. From this graph, T_e is determined by Equation (7)

$$G = \frac{0.17}{0.275} = 0.617, \quad G - G^2 = 0.236,$$

$$R_0 = \frac{1}{0.25} = 4\frac{V}{\mu}$$

$$T_e = \frac{(11{,}600)\ (.236)\ (.275)}{.25 - .009} = 3125 \text{ °K}$$

An estimate of N_e, using probe area instead of sheath area, gives

$$N_e \cong 2 \times 10^7 \text{ cm}^{-3}$$

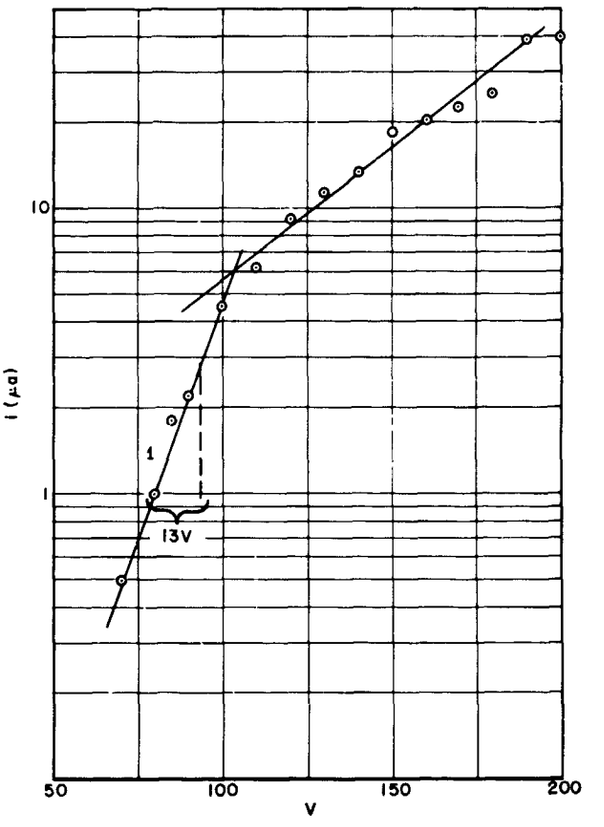


Figure D-37–Semilog plot of single probe data.

The asymmetry of this graph illustrates the importance of the location of $V_d = 0$. The factor $[G - G^2]$ can vary from about .25 to .05. DPM data was taken for probes located between the cathode screen and plate anode of configuration (b). The probes were placed side by side, in equivalent positions in the plasma and the resulting characteristics were quite symmetrical.

Sample results were

$$\text{log plot:} \quad T_e = 16{,}500° \text{ K}$$

$$\text{eq. resis.:} \quad T_e = 15{,}900° \text{ K}$$

which is, by our present standards, very good agreement. Here S = .017, R_0 = .25, so from Equation (6) the theory above, and in particular, the log plot method, is valid.

Figure D-38 illustrates some of the problems with double probe data. The conditions here are lower pressure, 15 microns, and fairly high discharge voltage, 900 v. The double probe characteristic is still quite recognizable, but the asymmetry and slope S have increased. The problem can be seen in our T_e values:

$$\text{equiv. R method:} \quad T_e = 15{,}000° \text{ K}$$

$$\text{log plot method:} \quad T_e = 34{,}800° \text{ K}$$

But here S = .32, R_0 = .95, so the log plot result is invalid.

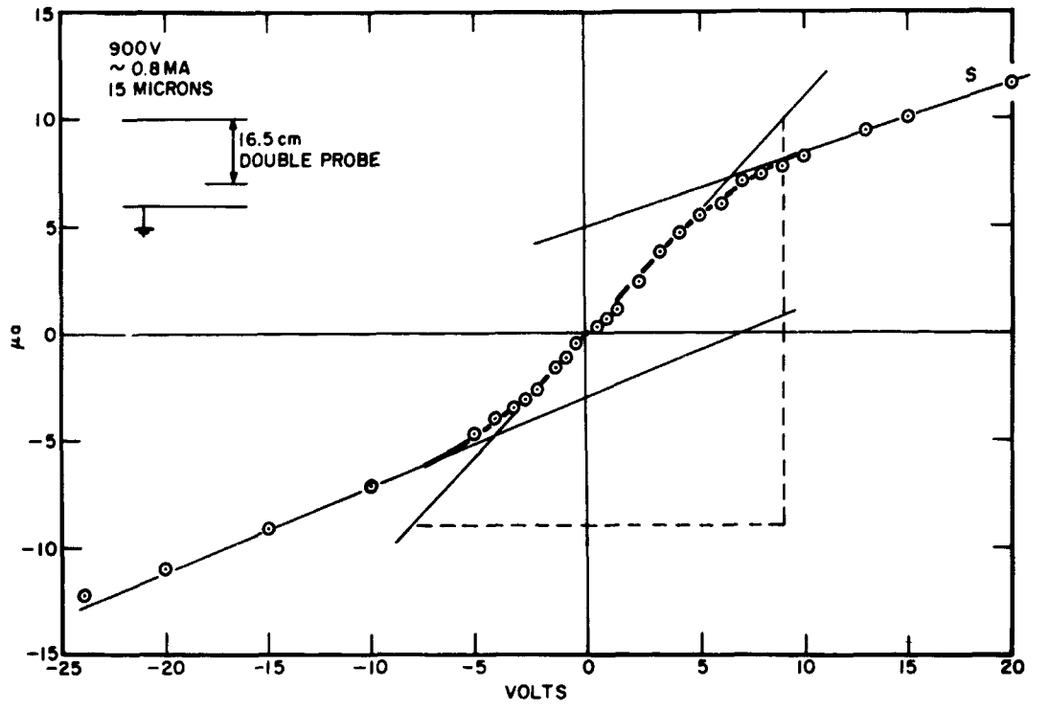


Figure D-38–Double probe in low pressire dc discharge.

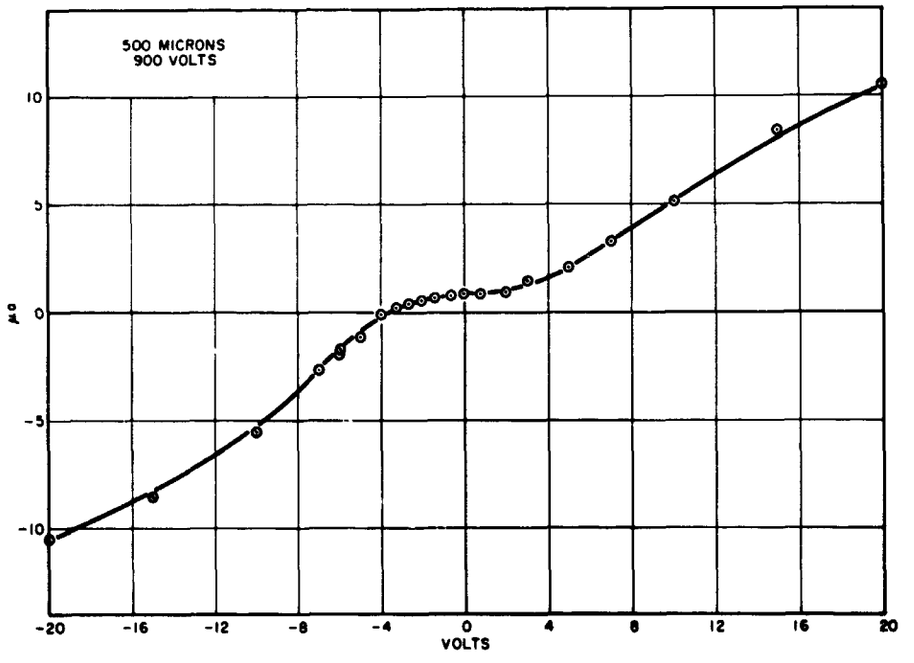


Figure D-39–Double probe in high pressure dc discharge.

As mentioned earlier, virtually all probe theory is based on the assumption of Maxwellian velocity distribution. The positive column of a glow discharge is one region with such a distribution. However, in experimental work one would often like to find plasma parameters in other, or less well-defined, regions. Probe theory is needed for regions with a large acceleration or other distorting influence. A few examples of distorted probe results are given in Figures D-39, D-40 and D-41. Figure D-39 is a double

probe characteristic taken in configuration (b) of Figure D-34, while Figure D-40 shows triple probe results at low pressure in the dc discharge.

Figure D-41 shows double probe results in the glowing region immediately surrounding an RF discharge coil. These curves appear similar to the curves in Figure D-40, and have the form of the deviation from linearity noted on the middle section of Figure D-38.

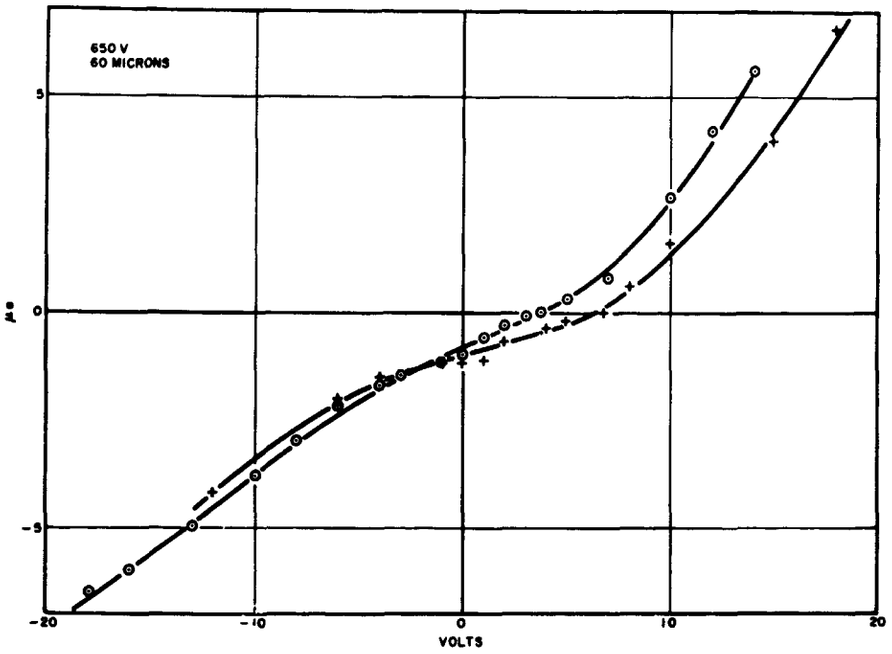


Figure D-40–Triple probe in dc discharger.

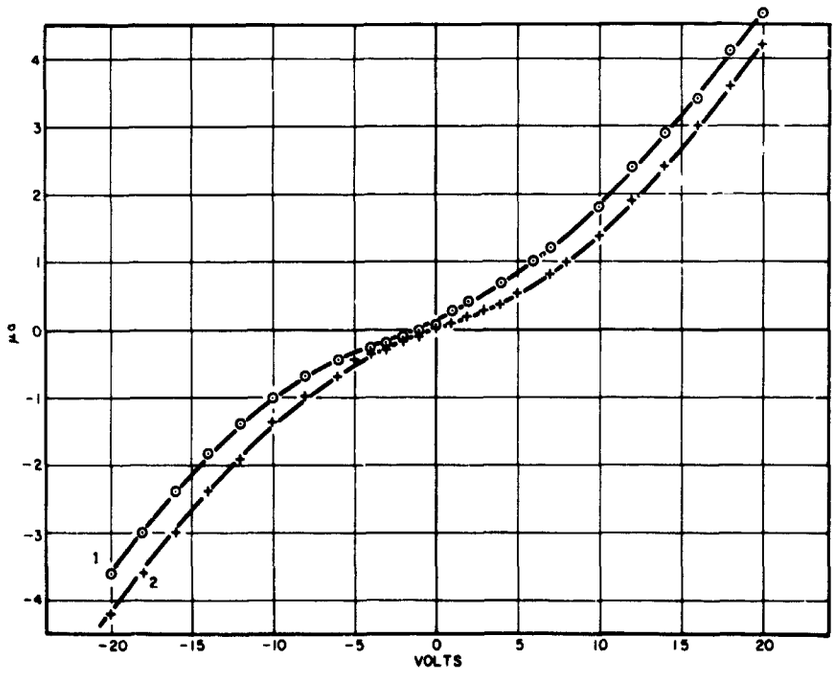


Figure D-41–Double probe in the glow region of coil.

All of these curves have the form of the distorted characteristics found by Yamamoto and Okuda (14) in the negative glow and cathode glow regions of a dc glow discharge.

SUGGESTIONS

It is suggested that for this type of research, a well-defined glow discharge, measured as precisely as possible by means of several methods, be maintained as a reference to check probe behavior before their use in an unknown plasma. Also, the regions other than the positive column might be explored to become familiar with "distorted" characteristics, and to begin a theoretical treatment of these, which would, as mentioned above, be extremely valuable. A velocity distribution as given by Equation (4) might be a logical starting point.

REFERENCES

1. Bourdeau, R. E., Science, 148, page 585, April 1965
2. Griem, H. R., Plasma Spectroscopy, McGraw-Hill Book Co., New York, (1964)
3. Hasted, J. B., Physics of Atomic Collisions, Butterworths, Washington (1964)
4. Haigh, C. and T. Smith, Phil. Mag. Volume 41, page 557 (1950)
5. Boyd, R. L. F., Nature, Volume 165, page 142 (1950)
6. Gabor, et al., Nature, Volume 176, page 961 (1955)
7. Vth Conf. Ionization Phenomena Gases, Munich, N. Holland Publishing Co., (1961)
8. Langmuir, I., and H. Mott-Smith, Phys. Rev., Volume 28, page 727 (1926)
9. Bourdeau, R. E., and J. L. Donley, Proc. Royal Soc., page 281 (1964)
10. Druyvestein, M., Zs. Phys., Volume 64, page 781 (1930)
11. Bohm, D., Burhop, E. H. S., and Massey, H. S. W., "The Characteristics of Electrical Discharges in Magnetic Fields, eds Guthrie, A., and Wakerling, R. K., McGraw-Hill, New York, (1949)
12. Allen, J. E., Boyd, R. L. F., and Reynolds, P., Proc. Phys. Soc., LXX 3-B
13. Johnson, E. O., and Malter, L., Phys. Rev., Volume 80, No. 1 (1950)
14. Yamamoto, K., and Okuda, T., J. Phys. Soc., Japan, Volume 11, page 57 (1956)
15. Kagan, Y. M. and Perel, V. I., Soviet Phys Uspekhi, page 767 (May-June 1964)
16. Okuda, T., and Yamamoto, K., J. Appl. Phys., Volume 31, page 158, (1960)
17. Dzherpetov, K. A., and Pateiuk, G. M., Soviet Physics JETP, Volume 1, page 326 (1955)
18. Vagner, S. D., Zudov, A. I., and Khakhaev, A. D., Soviet Physics-Technical Physics, Volume 31, page 240 (1961)
19. Kojima, S., Takauama, K., and Shimachi, A., J. Phys. Soc., Japan, Volume 8(1), page 55 (1953)
20. Crawford, F. W., and Mlodnosky, R. F., J. Geophys. Res. Volume 69, page 2765 (1964)
21. Harp, R. S., and Crawford, F. W., J. Appl. Phys., Volume 35, page 3436 (1964)
22. Crawford, F. W., and Harp, R. S., J. Geophys Res., Volume 70, page 587 (1965)

BLANK PAGE

MAGNETIC PROPERTIES OF MATERIALS

A. V. Dralle and J. Moore

INTRODUCTION

Considerable interest has developed in recent years in the measurement of the magnetic properties of the "non-magnetic" alloys, in magnetic fields of the magnitude encountered near the earth (0.5 oersteds) and over a wide range of temperatures. This has arisen largely from the increased application of these alloys in instruments for use in the space environment near the earth, where the interaction of components with the earth's field could give rise to undesirable torques and induced fields. Of particular interest has been recent work at the Naval Ordnance Laboratory, where it was found that several alloys, suitable for non-magnetic application at room temperature, assume strong magnetic characteristics at low (−200°C) temperatures (1).

A search was made for references that would give the magnetic properties of non-magnetic materials at low temperatures. Several of these were found although in almost every instance the measurements were made at relatively high values of magnetic field. Pasnak and Gordon (2) performed measurements on a variety of materials at temperatures ranging from room temperature down to −30°C. The magnetic field intensities used were of the order of about 0.5 oersteds. The temperature dependence of the magnetic properties of silver-zinc (3), palladium-rhodium (4) and vanadium (5) alloys was investigated. Develey (6) presents the variation of the magnetic susceptibility in the neighborhood of the Curie points, and Clogston (7) discusses the properties of a variety of alloys along with a brief explanation of the theory of the magnetic properties.

This report describes work on three methods for measuring magnetic permeability of alloy samples. The work follows closely that of E. Levy in the 1963 Goddard Summer Workshop (8); the equipment described by Levy was developed further with particular emphasis on the capability of making low temperature measurements. The work was performed at the Magnetic Test Facility of the Test and Evaluation Division at GSFC, with close cooperation from Fabrication Division. Measurements were made on approximately 30 alloys of the type commonly used in situations requiring non-magnetic materials. The samples were those tested by NOL in the work mentioned above. All samples were cylindrical rods of 6-inch length and 1-inch diameter.

TEST METHODS

dc Induction Method

The dc Induction method was used at NOL in the recent low temperature work and by Gordon and Pasnak in an early project in this field (2). It is based on the familiar permeability measurement method (9) using toroidal samples.

The circuit shown in Figure D-42 was constructed. The mutual inductor M is adjusted to balance the EMF produced upon a change in the magnetic field in the search coil without a specimen. When the search coil contains a specimen, the current in the solenoid is changed by means of the switch S, altering the magnetic field H. This produces a galvanometer deflection arising exclusively from the presence of the sample and proportional to the sample's susceptibility. As shown by Gordon and Pasnak, the permeability of the sample is then given by

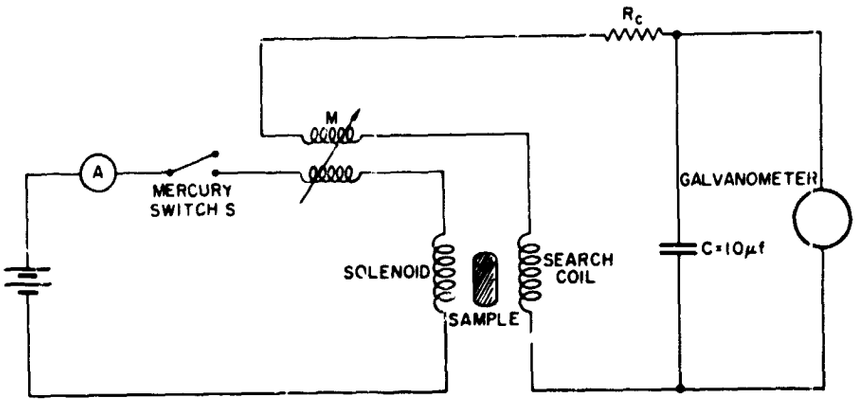


Figure D-42—Schematic diagram of dc induction method.

$$\mu = 1 + \frac{Kd}{2HNA} \tag{1}$$

where

μ = permeability (cgs)

K = galvanometer sensitivity (maxwell-turns/mm)

d = galvanometer deflection (mm)

H = magnetic field strength (oersteds)

N = number of turns on search coil

A = cross section area of sample (cm^2)

The solenoid windings, of density 80 turns/cm, were on a 2.25-inch diameter form 20.75-inch in length. The search coil was supplied by the Fabrication Division and consisted of 2000 turns wound on a 6-inch region of a 12-inch long phenolic tube designed to fit inside the solenoid and to contain a dewar for low temperature work. Figures D-43 and D-44 show the dewar, coil, and solenoid arrangement.

The galvanometer used was a Leeds and Northrup Model 2285D, which is a ballistic galvanometer having a lamp-mirror-scale arrangement for making measurements. The resistance R_c of Figure D-42 was added to bring the total secondary circuit resistance to 60,000 Ω, the critical damping resistance of the galvanometer. Capacitor C, 10 microfarads, was present to eliminate eddy current effects caused by a changing magnetic field in the metallic sample. The mutual inductor M was a General Radio Co. variable inductor; the system was found to balance for M = 33.5 millihenries, this then being the mutual inductance of the solenoid and search coil. Switch S was a single pole mercury switch, used to eliminate oscillations in the pulse to the galvanometer. Figure D-45 shows the complete experimental setup.

The system of Figure D-42 was calibrated and found to have a constant K = 3990 maxwell-turns/mm (deflection of reflected galvanometer cross hair at one meter). Measurements were made to 0.2 mm, so that a permeability of 1.075 or greater could be detected.

All measurements were made with the axes of the coils oriented vertically. This was mandatory in the low temperature work because of the need to contain liquid nitrogen

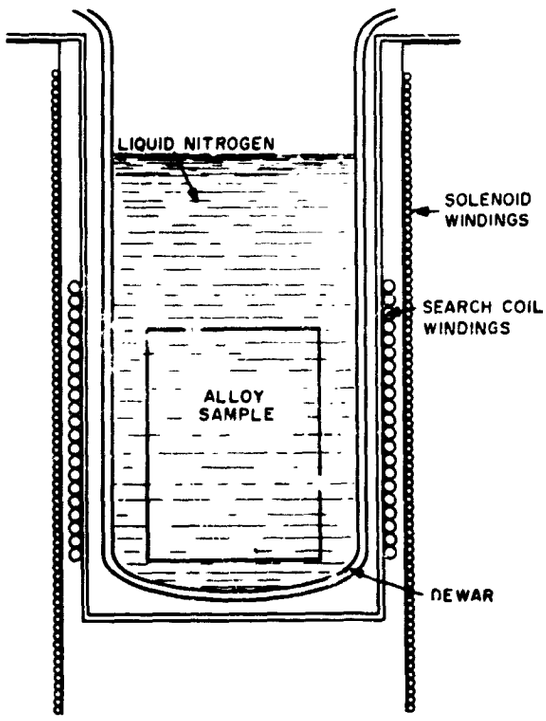


Figure D-43–Dewar and coil system for dc induction method.

Figure D-44–Search coil, dewar, and alloy sample (attached to insert rod) for dc induction low temperature measurements.

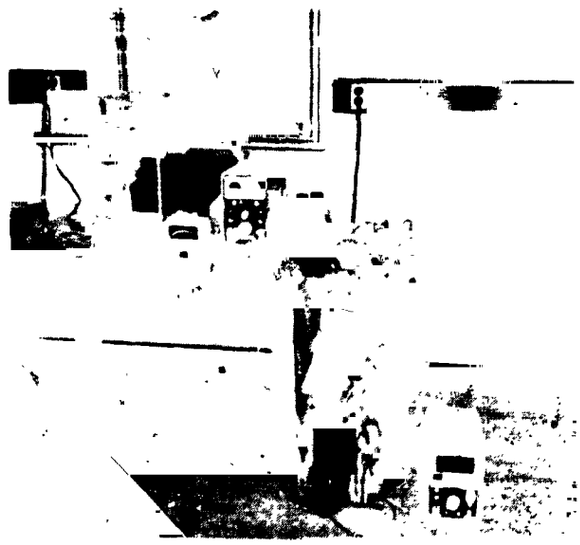


Figure D-45–dc induction experimental apparatus.

in the dewar. In addition, this orientation permitted benefit to be made of the vertical component H_a of the earth's field (H_a = −0.53 oersteds): the solenoid current was adjusted so that the net field was +0.53 oersteds; turning off the current effectively reversed the field through the sample, providing a flux-turns change equal to twice that obtained from a sample and coil system mounted perpendicular to the earth's field.

Samples were inserted into the dewar by means of a threaded rod screwed into a small hole tapped at the end of each sample, as shown in Figure D-44.

The following experimental procedure was employed for low temperature measurements:

1. Field in solenoid was set at +0.53 oersteds, using a Hewlett-Packard Model 3529A magnetometer.

2. Sample was attached to insertion rod and placed in a large liquid nitrogen container for cooling.

3. Sample was placed in dewar which had already been half-filled with liquid nitrogen. After sample was in place, dewar was completely filled with liquid nitrogen.

4. Current in solenoid was turned off, creating flux-turns change in sample and deflection of galvanometer.

Room temperature measurements were made with the sample centered in the search coil, and following steps 1 and 4 of the low temperature procedure.

The dc induction method is most desirable for its speed and ease of measurement. 10 samples can easily be measured in one hour, the limiting factor being the time required for the samples to cool to liquid nitrogen temperature. The greatest difficulty is the low sensitivity of the 2285D galvanometer. A 2285E galvanometer was available at one time, yielding about ten times the sensitivity of the 2285D and permitting permeabilities as small as 1.006 to be measured; this instrument was damaged before data could be taken.

Further increase in sensitivity might be gained by means of a wideband amplifier, but care should be exercised in assuring that it be suitably matched to the system. A gain in sensitivity of about 3 could also be obtained by employing an undamped galvanometer.

Of considerable importance is the prevention of unwanted vibrations of the galvanometer. This was achieved by mounting the instrument on a window-sill, where it was subject only to vibrations of the entire building.

ac Induction Method

The ac Induction method has been used, with various modifications, by Norder (10), by Broersma (11), and by Collinson, Molyneux, and Stone (12), and is capable of very sensitive susceptibility measurements. An indication of its sensitivity is its wide use in the measurement of the very weak magnetic properties of the diamagnetic salts.

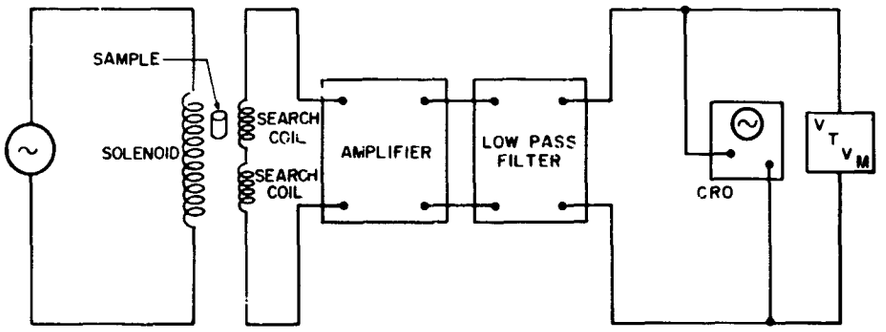


Figure D-46–Schematic diagram of ac induction method.

Figure D-46 shows the circuit used. The apparatus generates an ac field of the desired magnitude in a large diameter solenoid containing two identical, oppositely wound, search coils. The search coils generate zero total EMF until a sample is placed within one of them, the resulting imbalance voltage being a measure of the susceptibility of the sample. Susceptibilities are best measured with the use of calibration standards of the same geometry as the unknown samples.

The primary coil consisted of 280 turns of #16 magnet wire wound on a 5.5-inch outside diameter, 15-inch long tube. The secondaries each contained 3 250-turn layers of #27 magnet wire wound over the middle 4-inch section of a 1-inch inside diameter, 20-inch long coil form.

Although the secondary coils were nearly identical, a small imbalance voltage existed between them. This was largely eliminated by the insertion between the coils in Figure D-46 of a small, handwound, third coil. Even greater balance was obtained by the careful positioning within one of the secondary coils of a small iron rod. The best "null" system obtained had an imbalance voltage of about 400 nanovolts, compared with a 1 nanovolt balance obtained by Broersma.

The amplifier used was an Electro Instruments Inc. Model A15, with a gain of 100. The electronic filter, used largely to eliminate disturbances from the National Bureau of Standards Time Signal Station WWV, was an SKL Model 308A. The filter was the origin of the noise limiting further balance of the system.

The ac signal was chosen with an amplitude corresponding to a solenoid field of 1 oersted. The frequency selected was 10 cps, based on estimates that eddy-current effects at this low frequency would be negligible. Eddy currents would alter the permeability to an extent dependent mainly on the frequency of the ac signal and on the conductivity of the sample.

It was concluded that this method is not a reliable means of measuring the permeabilities of solid metallic bars. Experimentation showed that current effects arising from finite conductivity of some samples dominated the output from the search coils at as low a frequency as 5 cps, and that in any case it would be extremely difficult to distinguish between eddy current and magnetic effects. In particular, a sample of pure (non-magnetic) copper produced a much larger output signal than slightly magnetic Ni-O-Nel 825. Work on this apparatus was consequently abandoned.

Astatic Magnetometer

Figure D-47 shows the basic features of an astatic magnetometer, described by Johnson and Steiner (13) as a means for measuring magnetic susceptibility at low field strengths. Its operation is based on the distortion of the non-uniform field of two small bar magnets originally in rotational equilibrium with respect to one another; the introduction of a sample gives rise to a new equilibrium position, and the consequent rotation of a mirror mounted between the magnets. For small angles the amount of rotation measured with a lamp, mirror, and scale arrangement is proportional to the susceptibility of the sample.

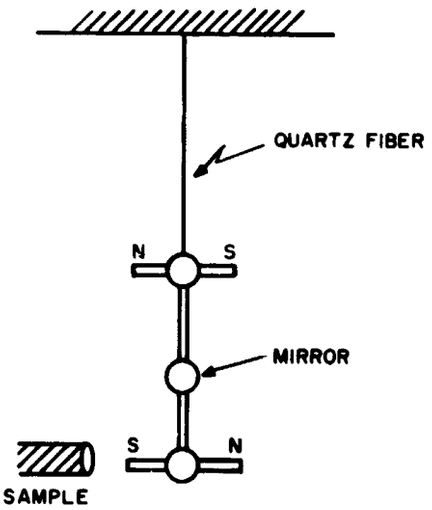


Figure D-47–Schematic diagram of astatic magnetometer.

The instrument is a very highly sensitive means for indicating the presence of magnetic materials. It was found to be extremely sensitive to position and orientation of the sample, so much as to make quantitative data largely irreproducible except under highly controlled conditions.

Measurements were made by placing samples in a position about 10 centimeters from the axis formed by the quartz fiber, and observing the deflection of the light beam. Efforts were made to demagnetize the samples before measurement; these succeeded only partially for it was found, especially for the "hard" magnetic specimens, the different samples of the same substance yielded vastly different magnetometer deflections, apparently depending on the magnetic history of the individual sample. Figure D-48 shows a sample in position for measurement.

Figure D-48–Astatic magnetometer with sample in position for measurement.

Calibration by means of known standards is obviously indicated in working with this instrument. Since no standards were obtained, the measurements made with the astatic magnetometer should be interpreted only as qualitative indications of the magnetic nature of the samples.

Cryogenics

The dewar used for low temperature work was supplied through Fabrication Division. It is pictured in Figure D-44.

The inside diameter of the dewar was 1-1/8 inches, large enough to accommodate a 1-inch diameter sample and a jacket of liquid nitrogen around the sample. The outside diameter of the dewar was 1-1/2 inches, large enough to fit securely into the search coil. A small Teflon insert fits at the bottom of the dewar as a rest for the frequently heavy metallic sample. The glass surfaces of the dewar were not silvered, as a guarantee against the introduction of magnetic impurities into the dewar-coil system. Nonetheless, there was observed to be no significant loss of liquid nitrogen in the few minutes required to make a permeability measurement.

CALIBRATION STANDARDS

Considerable effort was made in attempting to acquire suitable standard specimens, enabling a direct calibration of the permeability measuring equipment. Such standards would need to have the following characteristics:

1. The standard and unknown samples should be geometrically identical.
2. The unknown should have magnetic properties of magnitude sufficient to be measured by the equipment to be calibrated.
3. The unknown should not have properties (such as extreme deliquesence) which could subject its magnetic properties to change over a period of time.

Many examples of salts possibly meeting these qualifications were considered. Particular emphasis was directed to finding standards suitable for calibrating the dc induction apparatus, which could measure permeabilities no smaller than 1.075.

It was found that no readily available salts met the requirements completely. Various ferrous and nickel salts were investigated and found to have permeabilities in the 1.003 or smaller range. These were rejected as being too weak to be detectable by the present apparatus. The salts of the rare earth elements, such as holmium and gadolinium, appear to be very promising for this work. They are not, however, readily available in sufficiently pure form. Sanford and Cooter (14) offer other suggestions as to possible standards.

RESULTS OF MEASUREMENTS

Table D-6 is a list of different alloys tested with the chemical composition of each. Several alloys were available in different stages of thermal or mechanical treatment, and these are identified in Table D-7. The sample numbers appearing in Table D-7 are used in the tables listing the results of measurements.

Table D-6
Alloys Tested and their Chemical Composition

Alloy Type	Designation	Chemical Composition (Weight, %)											
		Al	C	Co	Cr	Cu	Fe	Mg	Mn	Ni	Si	Zn	Other
[1]Aluminum	356	Bal.				.014	0.16	0.21	< 0.02		6.8	< 0.03	Ti-020
[2]Aluminum	5086	Bal.			.05/.25	0.10	0.50	3.5/4.5	.2/.7		0.40	0.25	Ti-0.15
[2]Aluminum	6061	Bal.			.15/.35	.15/.40	0.70	.8/1.2	.15		.40/.80	0.25	Ti-0.15
[2]Aluminum	7075	Bal.			.18/.40	1.2/2.0	0.70	2.1/2.9	.30		0.50	5.1/6.1	Ti-0.20
[3]Cobalt	Elgiloy		0.15	40.0	20.0		16.0		2.0	15.0			Be-0.04 Mo-7.0
[1]Cobalt	Haynes 25		0.08	Bal.	19.97		2.41		1.32	9.70	0.56		W-15.22
[2]Copper	Beryllium Copper			.18/.30		Bal.							Be-1.80 to 2.05
[1]Copper	Cartridge Brass					69.56						30.38	
[3]Copper	Alloy 720					60.0			20.0	20.0			
[1]Iron	AISI 310		0.06		25.12	0.09	Bal.		1.49	20.58	0.61		Mo-0.10
[3]Magnesium	AZ31B	3.0						Bal.				1.0	
[3]Magnesium	AZ92	9.0						Bal.				2.0	
[3]Magnesium	ZK60A							Bal.				5.7	Zr-0.55
[3]Nickel	Inconel X-750	0.80	0.04		15.0	0.05	6.75		0.70	73.0	0.30		Cb-0.85 Ti-2.50
[3]Nickel	Monel K-500	2.80	0.15			29.5	1.0		0.60	65.0	0.15		Ti-0.50
[3]Nickel	Ni-O-Nel 825	0.15	0.03		21.5	1.80	30.0		0.65	41.8	0.35		Mo-3.0 Ti-0.90

[1]Actual Analysis
[2]Composition Limits
[3]Nominal Composition

Table D-7
Thermal and Mechanical Treatment of Samples

Sample No.	Base Metal	Alloy Designation	Condition or Thermal Treatment
1	Aluminum	7075	T6
2	Aluminum	5086	0
3	Aluminum	5086	H32
4	Copper	Beryllium-Copper	1/2 Hard
4B	Copper	Beryllium-Copper	1/2 Hard + aged (600°F–2 hrs–air cool)
5	Nickel	K-500 Monel	Cold drawn + aged
5B	Nickel	K-500 Monel	Solution heat treated (1600°F–1 hr–water quench)
5C	Nickel	K-500 Monel	Solution heat treated (1600°F–1 hr–water quench) + aged (1000°F–8 hrs–slow cool to 900°F–air cool)
5D	Nickel	K-500 Monel	Cold drawn + aged + scale removed
6	Copper	Alloy 720	Hot rolled + scale removed
6B	Copper	Alloy 720	Solution heat treated (1050°F–1 hr–air cool) + 33% cold reduction + aged (700°F–24 hrs–air cool)
7	Magnesium	ZK60A	T5
8	Magnesium	AZ31B	H24
9	Aluminum	6061	T6
10	Aluminum	6061	0
11	Iron	AISI 310	Solution heat treated
11B	Iron	AISI 310	Solution heat treated + 50% cold reduction
12	Cobalt	Haynes 25	Solution heat treated
12B	Cobalt	Haynes 25	Solution heat treated + 15% cold reduction
12C	Cobalt	Haynes 25	Solution heat treated + 20% cold reduction
13	Copper	Cartridge Brass	Cold drawn
13B	Copper	Cartridge Brass	Annealed (1000°F–1 hr–air cool)
13C	Copper	Cartridge Brass	Annealed (1000°F–1 hr–air cool) + 50% cold reduction
14	Magnesium	AZ92	T6
15	Aluminum	356	T6
16	Nickel	Inconel X-750	Cold drawn + aged
16B	Nickel	Inconel X-750	Solution heat treated (2100°F–4 hrs–air cool)
16C	Nickel	Inconel X-750	Solution heat treated (2100°F–4 hrs–air cool) + 30% cold reduction
17	Nickel	NI-O-NEL 825	Annealed
17C	Nickel	NI-O-NEL 825	Annealed + scale removed

Table D-8 lists the results of measurements using the dc induction apparatus and the astatic magnetometer. Note that in the dc induction results, actual permeabilities are listed, while the magnetometer results are to be interpreted as proportionate measures of sample susceptibility. Permeability measurements of the same samples at NOL are also included.

Table D-8
Magnetometer and dc Induction Measurements of Samples Compared With NOL Measurements

	Astatic Magnetometer (Relative Susceptibility)		dc Induction Permeability (GSFC)		dc Induction Permeability (NOL)	
	20°C	−196°C	20°C	−196°C	20°C	−196°C
1	0	0	< 1.075	< 1.075	< 1.004	< 1.004
2	0	0	< 1.075	< 1.075	< 1.004	< 1.004
3	0	0	< 1.075	< 1.075	< 1.004	< 1.004
4	0	4.5			< 1.004	< 1.004
4B	0	5.0	< 1.075	< 1.075	< 1.004	< 1.004
5	8.5	VL	< 1.075	8.98	< 1.004	8.83
5B	1.0	VL	< 1.075	17.7	< 1.004	16.5
5C	1.0	VL	< 1.075	15.8	< 1.004	14.7
5D	1.0	VL	< 1.075	8.90	< 1.004	8.36
6	1.5	4.0	< 1.075	< 1.075	< 1.004	1.099
6B	0	4.5	< 1.075	< 1.075	< 1.004	< 1.004
7	0	12.0	< 1.075	< 1.075	< 1.004	< 1.004
8	0	0	< 1.075	< 1.075	< 1.004	< 1.004
9	0	2.0	< 1.075	< 1.075	< 1.004	< 1.004
10	0	0	< 1.075	< 1.075	< 1.004	< 1.004
11	0	46.	< 1.075	< 1.075	< 1.004	1.011
11B	0	32.	< 1.075	< 1.075	< 1.004	1.011
12*					< 1.004	< 1.004
12B*					< 1.004	< 1.004
12C*					< 1.004	< 1.004
13	0	0	< 1.075	< 1.075	< 1.004	< 1.004
13B	0	0	< 1.075	< 1.075	< 1.004	< 1.004
13C	0	0	< 1.075	< 1.075	< 1.004	< 1.004
14	0	0	< 1.075	< 1.075	< 1.004	< 1.004
15	0	0	< 1.075	< 1.075	< 1.004	< 1.004
16	1	VL	< 1.075	1.44	< 1.004	1.40
16B	0.5	45.	< 1.075	1.075	< 1.004	1.073
16C	0.5	VL	< 1.075	1.495	< 1.004	1.036
17	2.5	VL	< 1.075	1.69	< 1.008	1.61
17C	1.5	VL	< 1.075	1.57	< 1.006	1.58

VL: Very Large
*The Haynes 25 Samples tested by NOL were not available for the work at GSFC.

GALVANOMETER CALIBRATION

Figure D-49 shows the circuit used for calibrating the dc Induction system. Note that it may be obtained from the measuring circuit of Figure D-42 simply by setting the mutual inductor in Figure D-42 to zero. In doing this it is essential that the resistance of the secondary circuit of Figure D-49 be equal to that of the secondary circuit of Figure D-42. Measurements are made of the galvanometer deflection caused by turning off current I.

The graph (Figure D-50) shows the galvanometer deflection d as a function of solenoid current I. From the slope $\Delta d/\Delta I$ of this curve is obtained the calibration constant K :

$$K = \frac{N\Phi}{d} = M\frac{\Delta I}{\Delta d}$$

In this circuit, $M = 33.5 \times 10^{-3}$ henries, and $\Delta I/\Delta d = 1/840$ amp/mm. Hence, $K = 3.99 \times 10^{-5}$ weber-turns/mm = 3990 maxwell-turns/mm.

Equation (1) is used to obtain a relation for permeability. $H = 0.53$ oersteds, $N = 2000$ turns, $A = 5.08\ cm^2$, so that $\mu = 1 + 0.371\ d$. This relationship was used for permeability determinations.

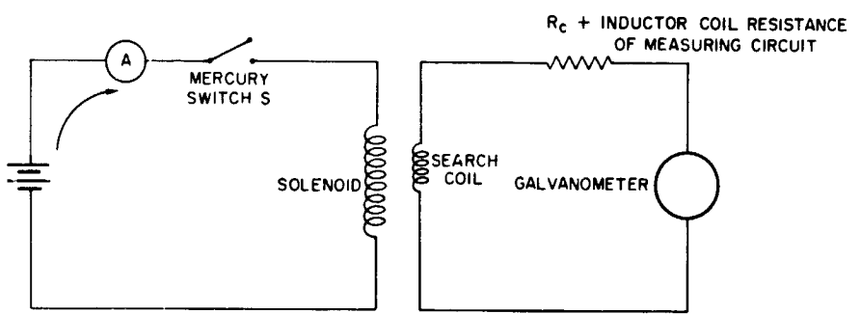


Figure D-49–dc induction method calibration circuit.

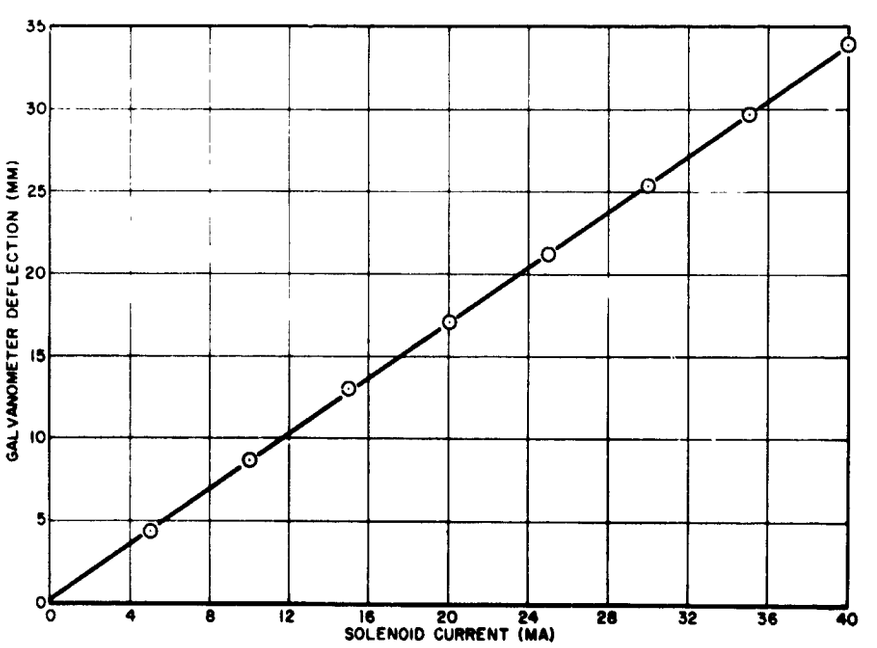


Figure D-50–Calibration curve for dc induction apparatus.

CONCLUSIONS

The dc induction apparatus developed is a reliable and convenient means for measuring magnetic permeabilities. The measurements made agree well with those of NOL, and even better agreement might be expected if the system were calibrated by means of known standards. The present sensitivity of the system precludes the use of any immediately available standards. A 10-fold increase in sensitivity will be obtained upon replacement of the 2285D galvanometer by Type 2285E.

As explained above, the ac induction system was found to be unsatisfactory for measuring the permeabilities of conducting samples.

The astatic magnetometer was found to be extremely sensitive, and hence it serves well as a qualitative indicator of the presence of magnetic materials. It appears that quantitative measurements may be made only with the aid of magnetic standards geometrically identical to the unknown samples, and it is recommended that further investigations be made in this area.

REFERENCES

1. "Feebly Magnetic (Non-Magnetic) Materials," GSFC Fabrication Division, private communication.
2. Pasnak, N., and Gordon, D. I., "The Measurement of Magnetic Characteristics of 'Non-Magnetic' Materials," DDC AD156450, (May 1952)
3. Meyer, L. and Weiner, D., "Temperature Dependence of the Magnetic Susceptibility of α-Silver-Zinc Alloys," Physical Review, Volume 108, pages 1426-1427, (Dec. 15, 1957)
4. Manuel, A. J. and St. Quinton, J. M. P., "The Magnetic Susceptibilities of Palladium and Palladium-Rhodium Alloys from 1.85 to 293°K.," Royal Society Proceedings (London), Series A, Volume 273, pages 412-426, (May 21, 1963)
5. Childs, B. G., Gardner, W. E. and Penfold, J., "The Magnetic Susceptibilities of Vanadium-Based Solid Solutions Containing Titanium, Manganese, Iron, Cobolt and Nickel," Philosophical Magazine, Volume 8, pages 419-433, (March 1963)
6. Develey, G., "Variation de la Susceptibilite Magnetique du Fer, du Gadolinium et du Nickel au Voisinage du Point de Curie," Comptes Rendus de l'Academie des Sciences (Paris), Volume 260, pages 4951-4953, (May 10, 1965)
7. Clogston, A. M., "Localized Magnetic Moments," Journal of Metals, Volume 17, pages 728-734, (July 1965)
8. Levy, E. K., "Testing of Non-Magnetic Materials," Final Report of the Goddard Summer Workshop Program in Measurement and Simulation of the Space Environment, GSFC Document X-320-63-264, (1963)
9. Golding and Widdis, Electrical Measurements and Measuring Instruments, London Pitman Press, page 382 (1963)
10. Norder, W., "Induction Apparatus for the Continuous Measurement of Magnetic Susceptibility," Rev. Sci. Instr., Volume 31(8), pages 849-851 (August 1960)
11. Broersma, S., "Induction Apparatus for Diamagnetic Measurements," Rev. Sci. Instr., Volume 20(9), pages 660-663 (September 1949)
12. Collinson, D. W., Molyneux, L., and Stone, D. B., "A Total and Anisotropic Magnetic Susceptibility Meter," J. Sci. Instr., Volume 40, pages 310-312 (1963)

13. Johnson, E. A. and Steiner, W. F., "An Astatic Magnetometer for Measuring Susceptibility," Rev. Sci. Instr., Volume 8, pages 236-238 (July 1937)

14. Sanford, R. L. and Cooter, I. L., "Basic Magnetic Quantities and the Measurement of the Magnetic Properties of Materials," National Bureau of Standards Monograph No. 47, 1962, 36 pages

BLANK PAGE

X-100-65-407(S)

SUPPLEMENT TO FINAL REPORT OF THE GODDARD SUMMER WORKSHOP PROGRAM IN ANALYSIS OF SPACE DATA AND MEASUREMENT OF SPACE ENVIRONMENTS

June 15 to September 15, 1965

"... Men in small groups, collaborating, can solve problems in a few weeks or months or even days that one man working alone and in isolation could never have solved had he lived a thousand years."

Pierre Teilhard de Chardin

Goddard Space Flight Center
Greenbelt, Maryland

BLANK PAGE

PREFACE

There has been considerable need for a handbook on the production of uniform magnetic fields. A start on assembling such a handbook was made this summer (1965) by assembling information on the application, history and methods of producing uniform magnetic fields. An annotated bibliography was also compiled. Even though this work is not complete it was decided to gather the material finished so far and present it as part of the 1965 Summer Workshop Report so that it will be available for use by interested persons.

BLANK PAGE

CONTENTS

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Continued)

LIST OF TABLES

COIL SYSTEMS FOR UNIFORM MAGNETIC FIELDS

R. E. Gebhardt and D. L. Waidelich

INTRODUCTION

During the three past Goddard Summer Workshops a great deal of information about uniform magnetic fields has been developed and collected by the members of the Workshops. It was felt that this information should be gathered together into a report for the following reasons:

1. To have in one place a list of most of the references on uniform magnetic fields
2. To indicate where information on a particular problem might be found
3. To provide information about the installations of coil systems, past and present
4. To present information that might be helpful in the design of future coil systems
5. To record some information that might be quite inaccessible or impossible to obtain in the future.

This work will be concerned with the production of relatively weak uniform magnetic fields in large volumes and thus will deal with air-core coil systems in almost every case. The designs for these coil systems have covered a wide range from small solenoidal windings to coil systems as large as forty feet in diameter for the testing of spacecraft. Some of the systems are designed to produce a very homogeneous field, 0.001 percent or better, over a relatively small volume, while other systems were designed to give a relatively large volume with a homogeneity of one to five percent.

The effort so far has been to make a detailed outline of the work, to assemble a brief history of the methods used, to bring together a short account of the principal ways of producing uniform fields and to provide an annotated bibliography. The work on the bibliography has not been completed but it was decided to present it in its unfinished state so that the various references would be available for use.

APPLICATIONS

Generally the applications employing uniform magnetic fields may be divided into two groups, first, that of relatively intense fields in small volumes for which coils with either air or ferromagnetic cores or coils of superconducting materials are used. These applications are concerned primarily with obtaining the highest intensity of the field and only secondarily with the uniformity of the field. Second would be the case of relatively weak fields usually in larger volumes for which coils with air cores are used in almost every case. The coils might be made of superconducting materials. The primary consideration would be the uniformity of the field intensity with the magnitude of the intensity as a secondary requirement. Considerable overlaps exist between the groups. This report will be concerned almost entirely with the second case.

Applications in which these coil systems are employed include:

1. The testing of magnetic components, instruments, satellites and spacecraft in the simulated uniform magnetic fields of space (27), (84), (93), (94), (95), (98), (109), (110), (111), (112), (113), (114), (130), (131), (132), (133), (134).

2. The testing of magnetic instruments used in geomagnetic work, (1), (17), (52), (85), (97), (99), (116), (135).

3. The production of uniform fields for electron or ion focusing, (4), (5), (34), (53), (54), (115), (143), (144), (147).

4. The construction of beta-ray spectrometers, (18), (26), (33), (46), (55), (61).

5. The measurement of the magnetic properties of materials, (28), (56), (62), (63), (67), (69), (70), (77), (86), (109), (119), (120), (121), (122).

6. The development of mass spectrometers, (18), (57), (71).

7. The construction of cloud chambers, (35), (64), (68), (87), (117), (140), (148).

8. The measurement of nuclear magnetic resonance, (43), (58), (96), (118), (145).

9. The use of superconductors, (44), (72).

10. The application to ordnance work, (100), (101), (136).

11. The use in a variety of other applications such as masers (102), magnetic demonstrations (73), electroacoustics (137), and plasmas (103).

The above list is not intended to be complete but shows a representative group of the applications.

WORK DONE IN THE NINETEENTH CENTURY

The single circular coil of Ampère (74) dating to about 1820 had a small volume of uniform field about its center. To increase the volume of uniformity, two coils were placed in parallel planes with the same axis and were at such a distance that the second order term of the field along the axis was zero. All odd order terms were zero because of the symmetry. The discovery of this two-coil system is usually credited to Helmholtz (123) in 1849 although there is some reason to credit either Gaugain with the discovery in 1853 or possibly Neumann (74), (75).

A solution for the four-coil system seems to have been given first by Neumann about 1856 (75). A complete solution of the six-coil system and a partial solution of the eight-coil system were given by Neumann shortly thereafter (74). Riecke in 1872 (88) showed that a spherical or ellipsoidal coil could be made to have a very homogeneous field inside and later made such coils.

About 1873 Maxwell (76) presented one solution of the three-coil system (74) as well as another solution of the four-coil system. In addition, he showed that a perfectly uniform field would exist inside a sphere if the windings on the outside of the sphere are such that each unit of length along the axis contains the same number of turns.

In 1885 Ellis, Clarkson and Rainy (104) made an experimental study of the field of a Helmholtz pair of coils.

Some references from the Nineteenth Century are presented in Appendix A, but no special effort has been made to make this a complete list.

WORK DONE IN THE FIRST HALF OF THE TWENTIETH CENTURY

In 1902 Watson (138) constructed Helmholtz coils accurate to one part in 10^4. Gehrke and von Wogan (45) in 1909 used conically wound coils in the

Helmholtz position and Rosa (107) used a two coil system with a spacing somewhat larger than that of the optimum Helmholtz coil-pair. In 1911 Bestelmeyer (7) calculated the field both on and off the axis of a Helmholtz system and compared the results with experimental measurements. Several very precise Helmholtz coil systems were constructed during the early 1920's (8), (116), (139). Nagaoka (89) in 1921 gave formulas for the field of a Helmholtz coil system both on and off the axis along with curves showing the variation of the field from that at the origin. Additional equations for the field of a Helmholtz coil system were given by Bromwich (9). Ruark and Peters (104) in 1926 considered approximately the effect of finite cross sectional area on the design of a Helmholtz coil system. In 1929 Bock (10) considered the size of the volume of homogeneity of the Helmholtz coil system. In the same year Fanselau (36) presented a theory of coil systems with any number of coil pairs such that each coil carried the same current and that for the four-coil system for example the fourth order term was made zero (74), (94), (131), (123). This solution was the same as that found many years earlier by Neumann.

The year 1931 brought forth a two coil system in whcih the coils had a rectangular shape (4) and also another method (33) of designing a two circular coil system with a finite rectangular cross section for each coil.

In 1933 Fanselau (37) developed a four-coil system in which both coils on one side of the center of the system lay in the same plane (94), (131). In 1934 Braumbek (6) gave a four-coil solution in which each coil carried the same current and the sixth order term was made zero (74), (94), (131). He also showed that for a six-coil system the tenth order term could be made zero. The Braunbek four-coil system is the pattern for a number of coil systems built recently. In 1934 additional two coil systems with rectangular coils were developed (53), (54). In the same year two coil systems with coils of other shapes such as triangular, square and octagonal were considered (99).

McKeehan in 1936 (74) presented an additional four-coil system which had all coils lying on the surface of a sphere and which made the sixth order term zero. Unfortunately the ampere-turns ratio was not a ratio of small integers as would be necessary if the same current were used in all coils.

The use of a two-coil system with coils of finite cross section was presented by Bacon in 1936 (11) and Crane in 1937 (20).

Clark in 1938 (19) constructed a nearly spherical coil for producing a uniform magnetic field. During the same year Shaw (115), (124) produced a uniform field in an annular region by reducing the spacing of two coils from that used in the Helmholtz configuration. In 1939 Hipple (57) constructed a spherical

coil with a uniformity of 0.3 percent, while Helmholtz coils were used for susceptibility measurements (86). In the same year Lyddane and Ruark (68) made an analysis of a pair of rectangular coils.

In the 1940's further work was done on uniform annular magnetic fields (34), and an extensive review and bibliography of uniform magnetic fields (88) was assembled by Neumann. Glaser (47) suggested a solution for making a uniform field by the use of integral equations. Sauerwein (117) designed and constructed a Helmholtz coil pair with rectangular cross section. Rössiger (105) used Helmholtz coils in a system to compensate automatically for variations in the earth's field. Sauter and Sauter (123) in 1944 presented a graphical solution of the general eighth-order, four-coil system. In the same year Crittenden and McDaniel (21) presented a method of obtaining a uniform field using two magnetic dipoles. A two coil system of finite cross section was designed by von Zeipel (148) in 1944. In 1945 Nageotte (87) described two large rectangular coils with bent sides to increase the uniformity and Rubens (101) gave two different square coil systems consisting of five coils on the surface of a cube. Ramsay and Yost in 1947 (100) described some of the coil systems used at the Kensington Magnetic Laboratory and Craig (18) discussed various volumes of uniform magnetic fields using two coils on the same axis. In 1948 Foss (35) presented a design for two coils with rectangular cross section, Garrett (49) indicated that systems producing uniform central fields act as dipoles at great distances and McKeehan (79) recalculated some roots of Legendre polynomials that are useful in coil system theory. Naumann (90) in 1949 used Helmholtz coil-pairs in an automatic system to reduce the earth's field and in another reference (22) a great many constructional details were given for Helmholtz coil pairs. In the same year Barker (12) presented calculated results for new three- and four-coil systems.

RECENT WORKS

In the last decade and a half the amount of works on coil systems for producing uniform magnetic fields has increased markedly. In 1950 a three-coil system with finite cross section (13) was developed, a two-coil system with large cross section (64) was reported and a three-coil system (137) was used in electroacoustic work. Garrett (48) reported on his theory of coil systems, and three long rectangular coils (55) were used in a beta-ray spectrometer in 1951. Two rectangular coils (125) were used in 1952 to cancel the vertical component of the earth's field. A description of a two-coil system of finite cross section (67) was presented in 1953. The next year a Helmholtz coil-pair was used as the rotating element of an astatic magnetometer (28). In 1955 Warburton (136) summarized the work on coil systems at the U.S. Naval Ordnance Laboratory

and further work on both the two-circular and two-square coil systems (52) was reported. Details on two systems each using two coils of finite cross section (96), (140) were presented, and the theory of such a system (50) was given. A Helmholtz coil pair was used in a shield box (118) to produce an eighth-order field. In 1956 further work on a two-coil system with finite cross section (15) was given, followed by additional papers on the same subject (40), (58) in 1957. In the same year Scott (119) described his four-coil system, and the details of a number of coil systems (26) for a beta-ray spectrometer were reported. A magnetically shielded chamber with a Helmholtz coil pair (2) was reported, and the use of one ellipsoidal winding within a second ellipsoidal winding (129) was presented. In 1958 Krisyuk and Latyshev (65) gave the theory for various three- and four-coil systems plus tables of the axial and radial field components of several systems. In a study of astatic magnetometers, four- and six-coil systems (56), and the effects of tolerances were discussed. A brief review of some of the work on two- and four-coil systems is given in (142) and in (51). In very large electromagnets it is suggested that two coils be used (66) to improve the uniformity. A Helmholtz coil pair was used in an apparatus (14) to stabilize a magnetic field. In 1959 a description of the two- and four-coil systems at the Fredericksburg Magnetic Observatory (91) was presented.

In 1960 a Helmholtz coil pair with conical windings was used to calibrate a magnetometer for flight in a satellite (109), and additional computations were made for the field of such a coil pair (43). Williams and Cain (141) developed a theory for multiple coil systems lying on the surface of a sphere and Scott (120) presented further information on his four-coil system and on Kettering Magnetic Laboratory. Some work was done using series other than power series (61) in developing larger volumes of homogeneous magnetic fields, and the use of two circular or two rectangular coils was discussed (69). The next year Franzen (38) reviewed Garrett's theory (48), some useful information on solenoids and current loops (80) was compiled, and the properties of an infinite set of current loops (92) was discussed. The use of a four-coil system on a satellite (60) was mentioned, and the measurement of the magnetic dipole moment of a satellite in a large spherical coil system (130) was discussed. Two square coils were used in magnetic measurements on rocks (62), and Helmholtz coil pairs were used to reduce alternating magnetic fields (77). In 1962 Franzen (39) considered the effect of the finite cross section on a two- and a four-coil system. Further information on the Fredericksburg Magnetic Observatory (85) was given, and some results of measurements using the Scott coil system (121) were presented. Two coils along with shielding (71) were used to reduce the effect of the earth's field, while a method of producing a uniform field with conductors on the surface of a cylinder and in the axial direction (41) was presented. Waidelich and Pittman (131) developed a computer solution of the eighth-order, four-coil problem. The following year the application of coil systems to the testing of satellites (93)

and the use of a four-coil system on a satellite (59) was mentioned. A large two coil system with finite cross section (103) was constructed, a system using three Helmholtz coil-pairs on mutually perpendicular axes (102) was used in maser research, and the possibilities of homogeneity in intense fields (81) were reviewed. Further work on the four-circular coil system (132) was done, work was started on a system of four-square coils (146), (110), the six-circular-coil system (111) was solved, and the effects of finite cross section on a four-coil system (98) was assessed. Conductors in the axial direction on the surface of a cylinder were used (3), (16) to obtain a uniform magnetic field, a superconducting Helmholtz coil pair (30) was designed, and a method of obtaining homogeneous circular magnetic fields (31) was described.

In 1964 a complete solution of the four-circular coil system (94) and of the four-square coil system (133) appeared. Parameters and contour curves for four-circular coil systems whose turns are the ratios of integers (112) were given, and a complete solution of the four-octagonal-coil system (113) was presented. The approximation problem for two coils (114), (23) was considered, and ways of providing very large magnetic environment (95), (134) were suggested. A magnetic laboratory was established at Oakland University (122) and various coil systems were suggested as useful in controlling the magnetic environment (135). Superconducting coils were used to produce very uniform fields (32), (82), (145), and a Helmholtz coil pair was used inside a magnetic shield (118). The application of various coil systems to space problems (84) and to the testing of ferromagnetic materials (63) was presented. A computer program for the six-circular-coil problem (78) was given a "half-Helmholtz" coil system (73) was described, the use of a Helmholtz coil pair with two solenoids (83) was suggested and a Helmholtz coil-pair was used at high frequencies (97). This year (1965) Rinard (106) used Tschebyscheff polynomials in determining the parameters of a number of four-circular-coil systems; studies were made of the four-octagonal-coil and of dipole systems (24), (25), and of two circular coils with finite cross section (108). Two rectangular coils with finite cross section (44) were considered, and some experimental work with two superconducting solenoids (127) was done. Helmholtz coils were used in magnetic measurements (1) and in radio frequency work (42), while a four-coil system was employed in reproducing the magnetic fields of space (27).

THEORY

The basis for all of the work on circular coil systems starts with the works of Ampère (74) and Maxwell (76). The scaler potential in Gaussian units is given by the Legendre polynomial:

$$V_p = \frac{2\pi i_p}{c}\left[1 + u_p - (1-u_p)^2 \sum_{n=1}^{\infty} \frac{1}{n}\left(\frac{r}{r_p}\right)^2 p_n'(u_p)\, p_n(u)\right] \qquad (1)$$

where:

a_p = Radius of p^{th} loop

i_p = current in p^{th} loop

c = velocity of light

r = $(x^2 + z^2)^{1/2}$

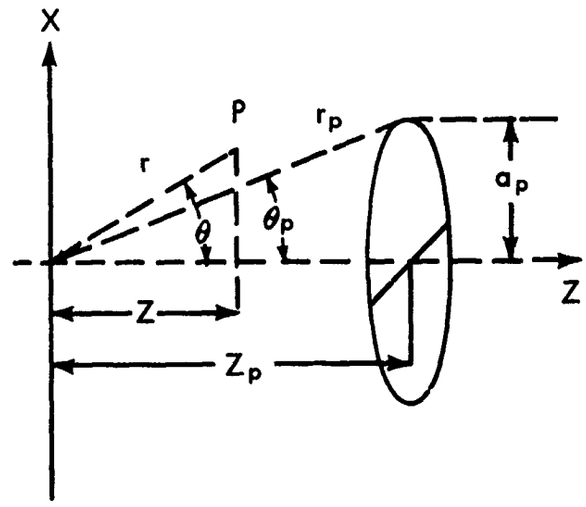
r_p = $(a_p^2 + z_p^2)^{1/2}$

u = $\cos\theta = \frac{z}{r}$

u_p = $\cos\theta_p = \frac{z_p}{r_p}$

$p_n(x)$ = n^{th} Legendre polynomial

$p_n'(x)$ = First derivative of the n^{th} polynomial



The axial and radial components of the field from a single current loop can also be expressed in elliptic functions, which offer some advantage if the field at a number of points is to be computed by means of high speed digital computers. Symthe (128) has shown that the axial and radial components of the field can be expressed in elliptic functions as:

$$B_x = \frac{\mu NI}{2\pi} \left\{ \frac{z}{x\,[(a+x)^2 + z^2]^{1/2}} \right\} \left\{ -\mathbf{K} + \frac{a^2 + x^2 + z^2}{(a-x)^2 + z^2}\,\mathbf{E} \right\}$$

$$B_z = \frac{\mu NI}{2\pi} \left\{ \frac{1}{[(a+x)^2 + z^2]^{1/2}} \right\} \left\{ \mathbf{K} + \frac{a^2 - x^2 - z^2}{(a-x)^2 + z^2}\,\mathbf{E} \right\}$$

where:

μ = permeability

I = current (amperes)

N = number of turns

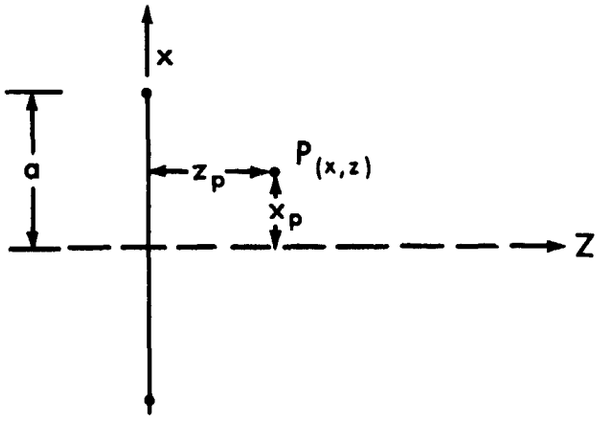
x = radial distance from coil axis (meters)

z = axial distance from plane of coil (meters)

a = radius of coil (meters)

K = complete elliptic integral of the first kind

E = complete elliptic integral of the second kind



For the computer program to evaluate the axial and radial components, the **K** and **E** functions were evaluated using Hastings[1] expressions for the fourth power terms in m.

$$\mathbf{K} = (a_0 + a_1 m + a_2 m^2 + a_3 m^3 + a_4 m^4) - (b_0 + b_1 m + b_2 m^2 + b_3 m^3 + b_4 m^4) \ln \frac{1}{m}$$

$$\mathbf{E} = (1 + c_1 m + c_2 m^2 + c_3 m^3 + c_4 m^4) - (d_1 m + d_2 m^2 + d_3 m^3 + d_4 m^4) \ln \frac{1}{m}.$$

where $m = 1 - k^2$ and $k^2 = \dfrac{4 a_x}{(a+x)^2 + z^2}$

a_0 =	1.3862	9436 112	b_0 =	0.5	
a_1 =	0.0966	6344 259	b_1 =	0.1249	8593 597
a_2 =	0.0359	0092 383	b_2 =	0.0688	0248 576
a_3 =	0.0374	2563 713	b_3 =	0.0332	8355 346
a_4 =	0.0145	1196 212	b_4 =	0.0044	1787 012
c_1 =	0.4432	5141 463	d_1 =	0.2499	8368 310
c_2 =	0.0626	0601 220	d_2 =	0.0920	0180 037
c_3 =	0.0475	7383 546	d_3 =	0.0406	9697 526
c_4 =	0.0173	6506 451	d_4 =	0.0052	6449 639

Figures 1 and 2 are plots of the axial and radial components of the field from a single coil (Ampère's Coil). The field contours are plotted as percent departure from the value of the field at the center. The constants for the coil plotted are:

Radius = 1.0 meter

NI = 1 ampere-turn

B_0 = 628.3185 gammas/meter radius/ampere turn

In the case of a combination of loops, the potential function can be written from Equation (1) as:

---

[1]Hastings, C., Jr., "Approximations for Digital Computers," Princeton University Press, Princeton, N.J., pages 170 to 175, (1955).

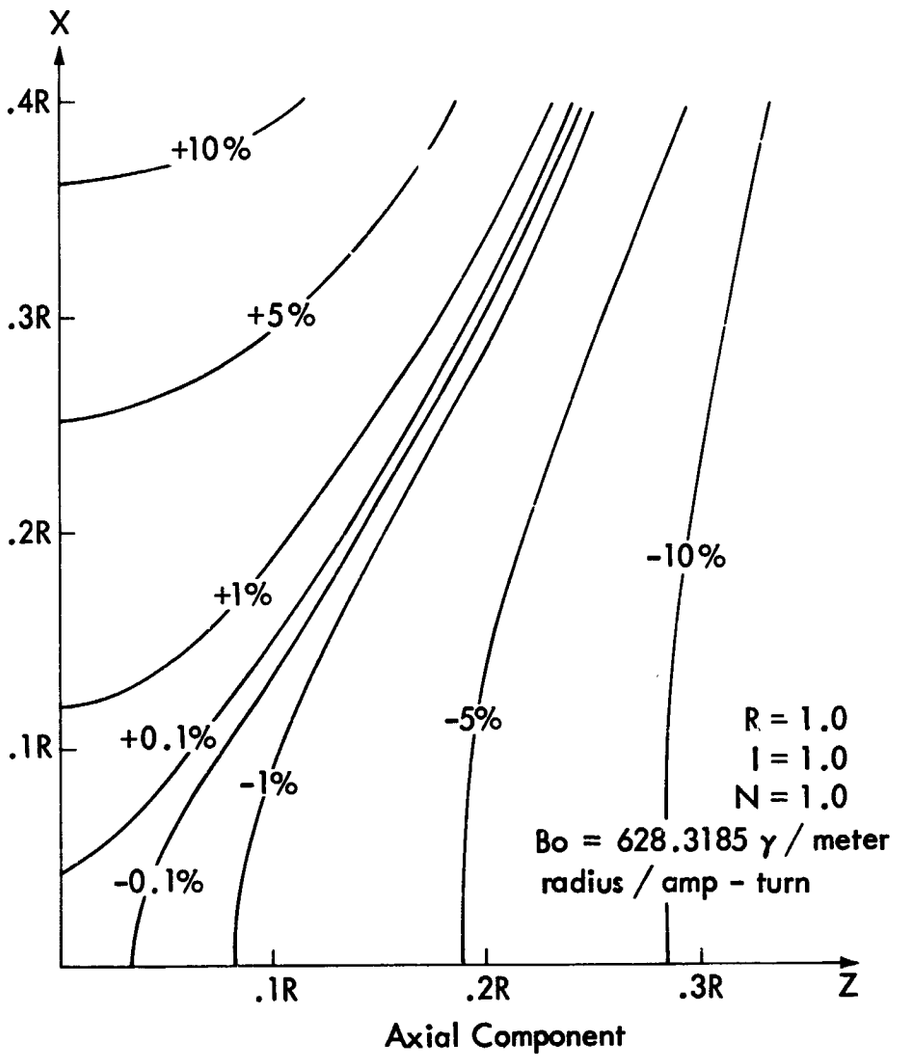


Figure 1. Ampere's coil (axial field plot).

$$\mathbf{V} = \sum \mathbf{V}_p = \frac{2\pi}{c} \left\{ \sum i_p + \sum i_p u_p + \sum_{n=1}^{n=p} \frac{1}{n} \sum \left[i_p (1-u_p^2)\, r_p^{-n}\, P_n'(u_p) \right] \; r^n\, P_n(u) \right\}$$

If two loops are placed symmetrically about the origin and carry equal currents, that is if $i_2 = i_1$; $r_2 = r_1$ and $u_2 = -u_1$; in the above equation for the potential V all terms with n = even vanish due to symmetry.

The solution by Helmholtz (89) places the current loops such that $u_1 = 5^{-1/2}$ and the term for n = 3 is equal to zero. The field then is uniform to fourth-order terms in r/r_p.

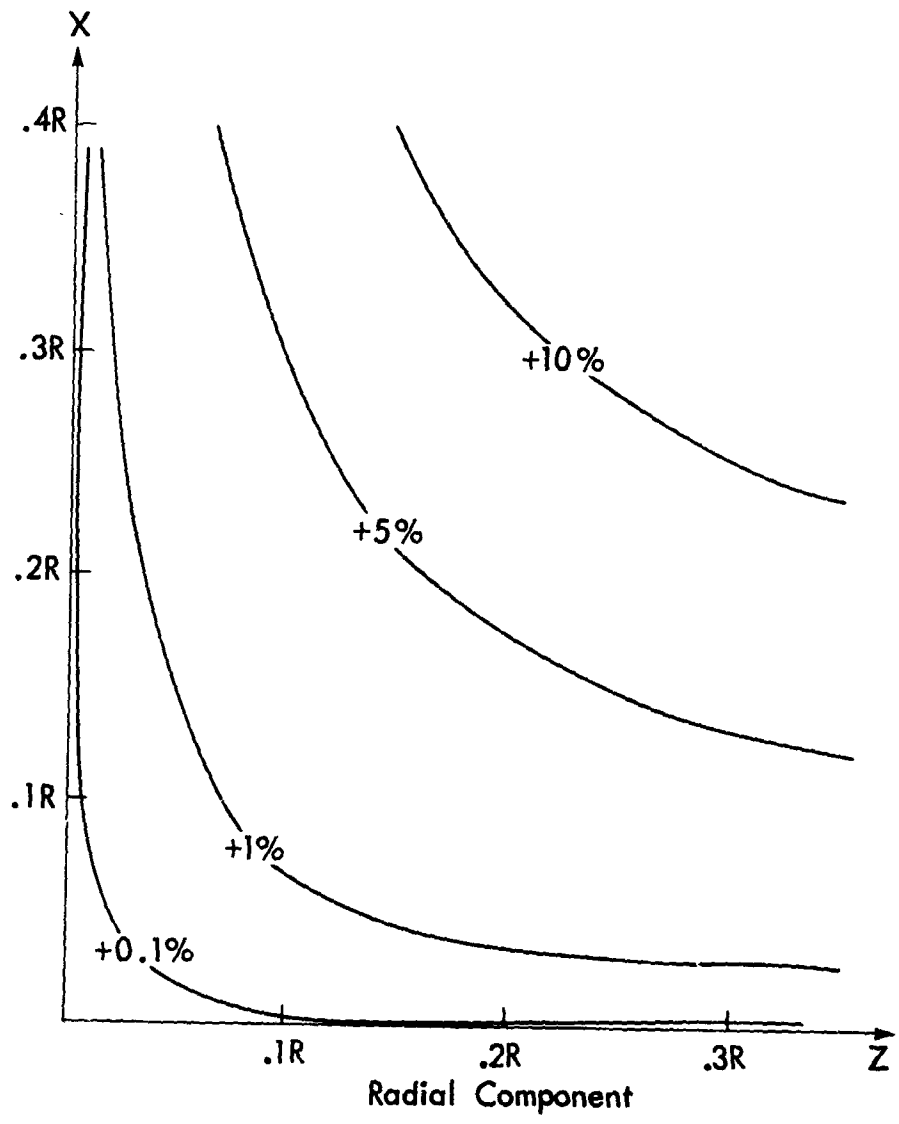


Figure 2. Ampere's coil (radial field plot).

Figures 3 and 4 are plots of the axial and radial components of the field from the Helmholtz coil configuration. The constants for the coil plotted are:

Radius = 1 meter
Spacing = 1 meter ($a_1 = \pm 0.5$ m)
B_0 = 899.176 gammas/meter radius/amp-turns

The discussion above has been based on an idealized current filament of negligible cross sectional area. For a winding of finite rectangular cross section in a two loop coil (Helmholtz), Maxwell (76) has shown that an important correction can be reduced to zero if the ratio of the breadth of the winding to its depth is $\sqrt{31/36}$ or approximately 5.57:6. Franzen (39) independently has also shown

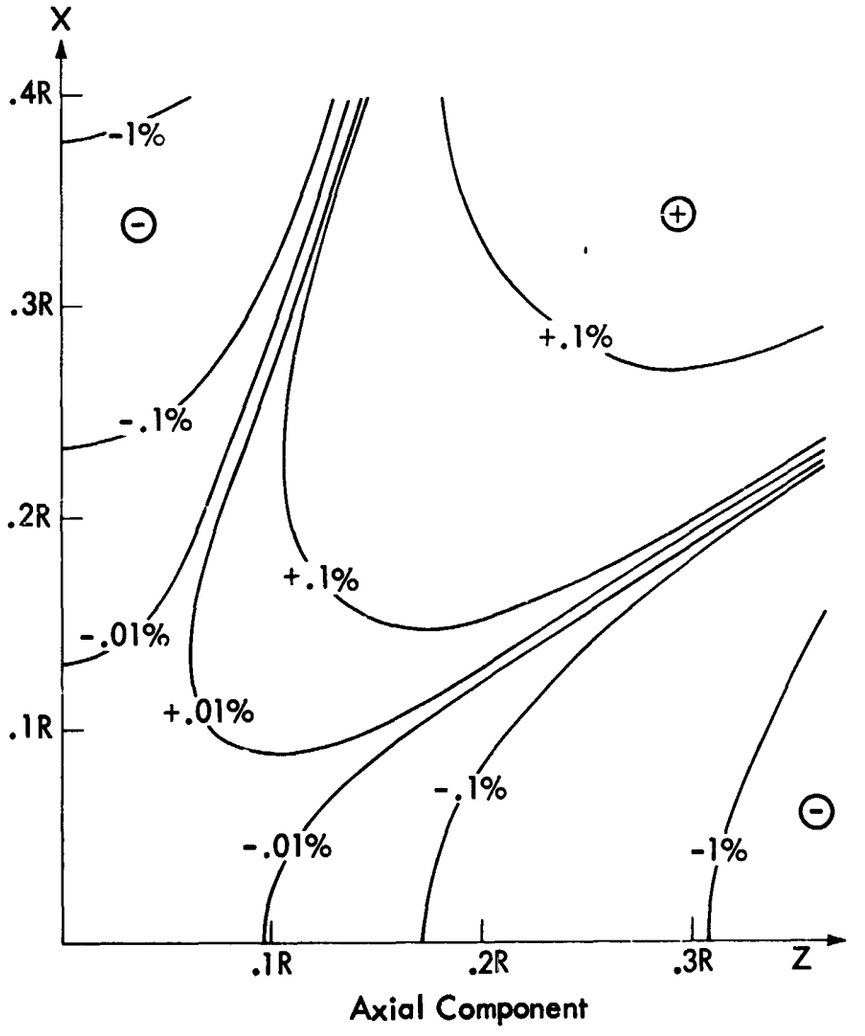


Figure 3. Helmholtz coil (axial field plot).

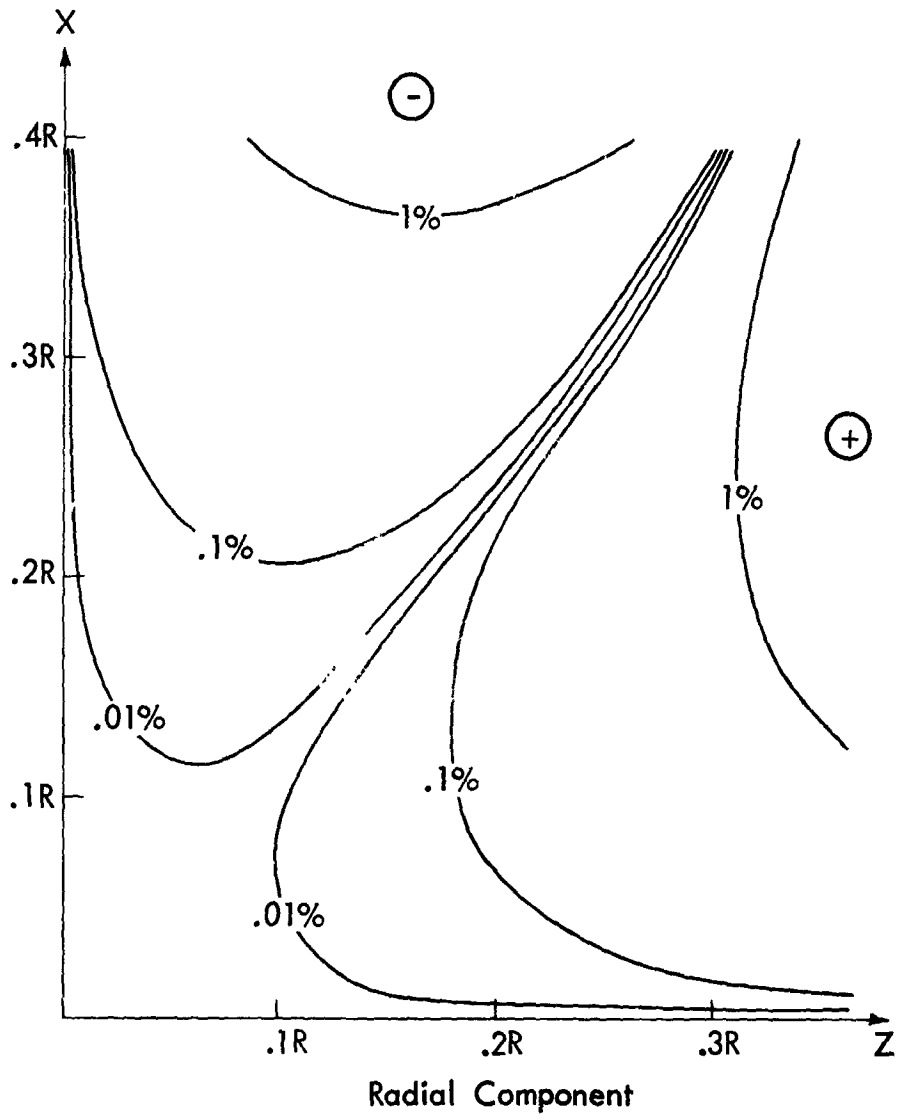


Figure 4. Helmholtz coil (radial field plot).

that this ratio can be determined and is:

$$\text{breadth/depth} = \left[1 - \left(\frac{1}{9}\sin^2\theta\right)\right]^{1/2} = 0.928 \text{ or } 5.57{:}6$$

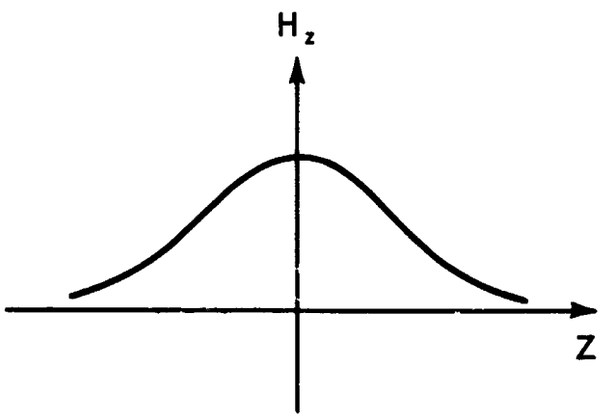
The uniformity of the field produced will be substantially identical with the field produced by a coil pair of infinitesimal cross section. In Franzen's work several minor restraints were imposed, first that the cross section be rectangular and secondly that the dimensions for the breadth and depth be small compared to the distance between the coils (coil spacing).

The work of Helmholtz, Gaugain (74) and others made the second derivative of the field at the coil center equal to zero. In following this technique, the axial

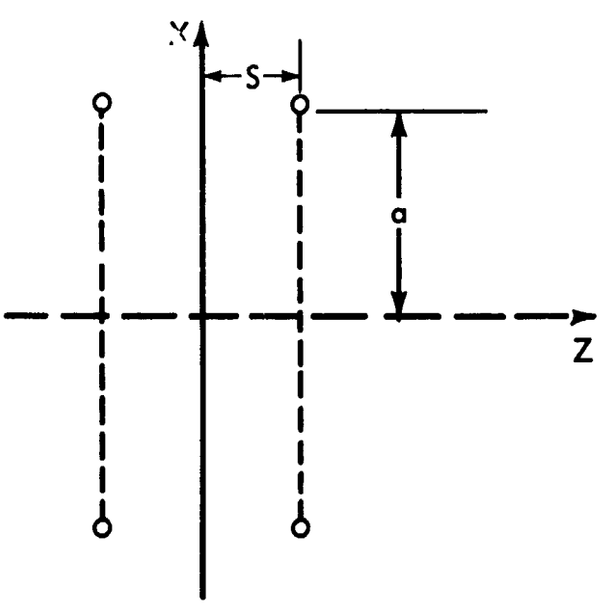
component of the field is uniform and flat through this central region and falls off as one moves along the coil axis.

In many applications one is interested not in a small field with a homogeneity of 0.001%, but rather in a large volume with a homogeneity of a few percent (3% to 5%), for example in deperming spacecraft or components, a magnetic field of modest homogeneity is satisfactory.

This problem has been investigated by Saperstone (114) for a circular coil system with a single pair of coils. If the axial-field on the axis is plotted against the distance from the center, the shape of the curve is as shown below.



It has been suggested that the volume of homogeneity can be increased by increasing the coil spacing as indicated below. Heller (52) noted that in the case



where the spacing (2S) is equal to the radius of coil, there is the maximum shown on page 15 as the spacing is increased the axial field yields two maxima as indicated in the figure below. For maximum homogeneities Table 1 gives a comparison of the distance along the axis that can be obtained for different spacing of the coils.

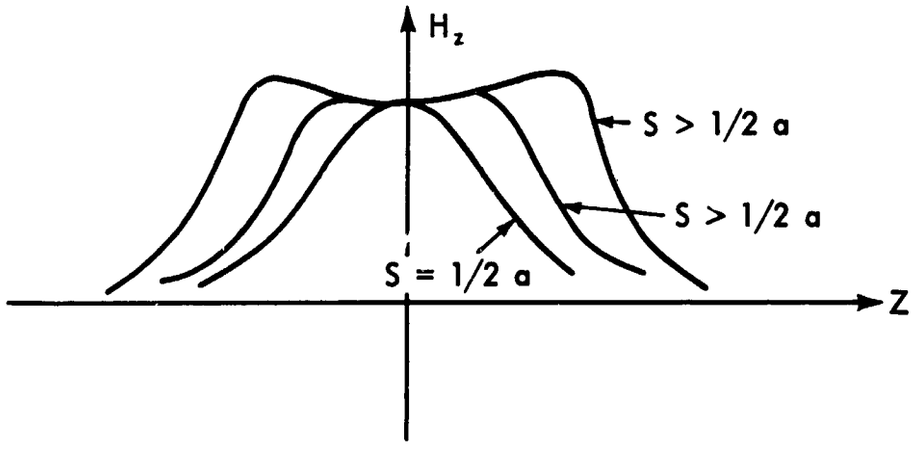


Table 1
Comparison of Axial Distance Obtained for
Various Spacing of Coils

% Homogeneity	K	Δz for S = Ka	Δz for S = 1/2 a
0.1	.518	.247	.085
0.5	.541	.376	.131
1.0	.558	.451	.157
2.0	.583	.546	.188
3.0	.603	.613	.211
4.0	.621	.669	.238
5.0	.636	.714	.245
6.0	.650	.754	.254

A vast amount of work has been done on circular coil systems having more than a single pair of coils. Some of the early work of Neumann (88) and Maxwell (76) made various assumptions to reduce the number of variables of their coil systems. In general for a four-coil system there are the following variables:

$N_1 I_1$ = ampere turns of 1st coil pair

$N_2 I_2$ = ampere turns of 2nd coil pair

$\pm z_1$ = distance of 1st coil pair from origin

$\pm z_2$ = distance of 2nd coil pair from origin

a_1 = radius of 1st coil pair

a_2 = radius of 2nd coil pair.

With the above, the following other relations exist:

$$r_1 = \left(z_1^2 + a_1^2\right)^{1/2} \qquad r_2 = \left(z_2^2 + a_2^2\right)^{1/2}$$

$$u_1 = \cos \theta_1 = \frac{z_1}{r_1} \qquad u_2 = \cos \theta_2 = \frac{z_2}{r_2}$$

Neumann chose to make $N_1 I_1 = N_2 I_2$ and solved for the other variables. Maxwell on the other hand placed the coils on the surface of a sphere so that $r_1 = r_2$. Pittman and Waidelich (94) allowed the ratio of the ampere turns of the pairs of coils ($N_2 I_2 / N_1 I_1$) to vary in small increments and obtained numerous solutions, starting with the coils lying in the same plane perpendicular to the axes of the system and extending to the case where the coils have equal radii (Barkers solution (12)). Included in the solutions reported are Braunbek's (6) arrangement with equal currents (ampere-turns) in the two pair of coils, as well as McKechan's (74) solution where the coils lie on the surface of a sphere.

Plots have been made of a number of coil configurations. Probably of most interest is the Braunbek coil system, with equal ampere turns on each of the coil pairs. Figures 5 and 6 are plots of the axial and radial components of the field to show the homogeneous volume obtainable.

The parameters for the idealized Braunbek coil system are:

$N_1 I_1 = N_2 I_2$

$a_1 = 1.0000000$ $\qquad a_2 = 0.76389941$

$d_1 = 0.27802813$ $\qquad d_2 = 0.84566402$

Coil constant = 1619.3190 gamma/meter radius/ampere turn

In the course of computing the data for these figures, it was apparent that very little information was available on the field in the immediate vicinity of the

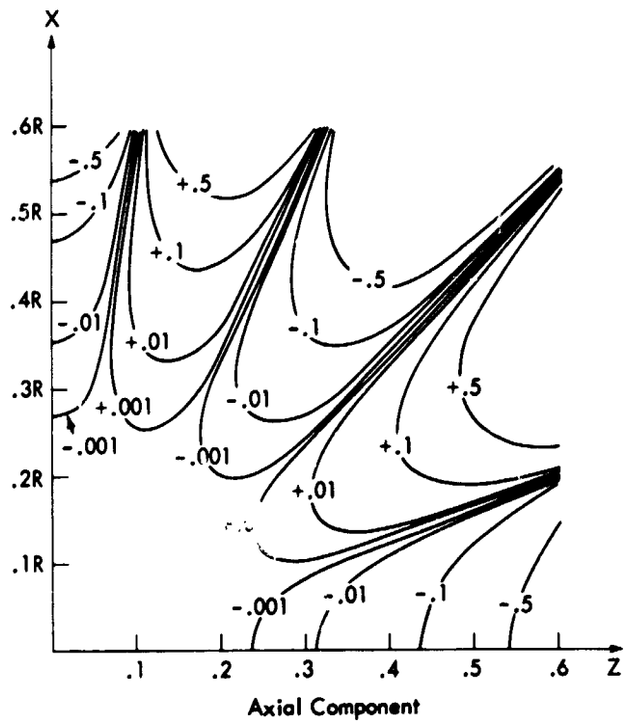


Figure 5. Braunbek coil (axial field plot).

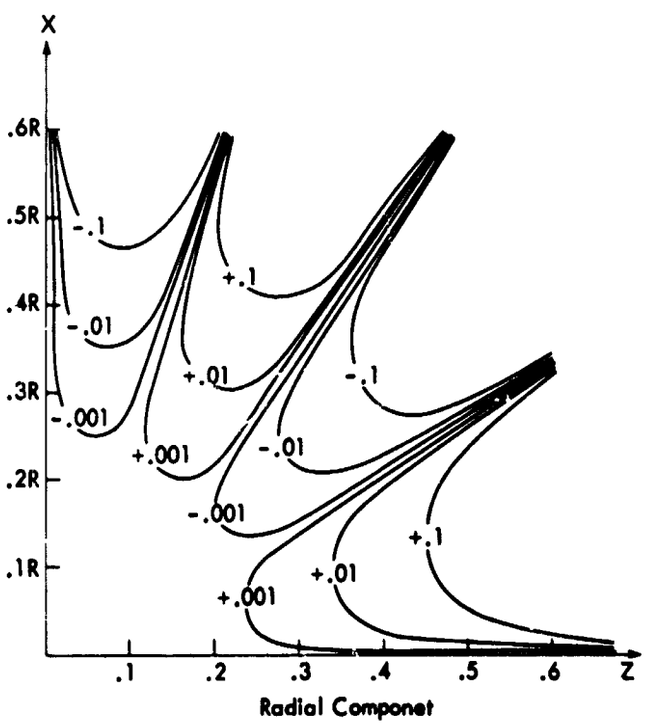


Figure 6. Braunbek coil (radial field plot).

conductors and in the space just outside the coil system. Figures 7 and 8 are plotted to a reduced scale to show the nature of the field of the two components (axial and radial) in the region near the conductors and immediately outside the coil system.

In multiple coil systems the misalignment or improper spacing of the coil to an extent smaller than the tightest possible tolerances, will introduce linear gradients. These gradients can be eliminated by mechanically adjusting the coils, a rather time consuming process, or by shimming the coils electrically, using special gradient cancelling coils.

All linear gradients can be expressed in terms of a basic set of five orthogonal gradient terms just as a field vector **H** can be expressed in terms of three orthogonal component vectors: $\mathbf{H}_x$, $\mathbf{H}_y$ and $\mathbf{H}_z$.

There are nine linear gradient terms, but the relationship that exists, and inherent in Maxwell's equations, reduces the number of independent terms to five. The type of gradient coils designed for the Magnetic Fields Component Test Facility at the Goddard Space Flight Center are easily installed on almost any orthogonal coil system. In this particular coil arrangement the following orthogonal set of gradients are present. (The constant numerical factors are omitted.)

$$\mathbf{A} = \frac{\partial H_z}{\partial z} = -\left(\frac{\partial H_x}{\partial x} + \frac{\partial H_y}{\partial y}\right)$$

$$\mathbf{B} = \frac{\partial H_x}{\partial x} - \frac{\partial H_y}{\partial y}$$

$$\mathbf{C} = \frac{\partial H_x}{\partial y} = -\frac{\partial H_y}{\partial x}$$

$$\mathbf{D} = \frac{\partial H_x}{\partial z} = -\frac{\partial H_z}{\partial x}$$

$$\mathbf{E} = \frac{\partial H_y}{\partial z} = -\frac{\partial H_z}{\partial y}$$

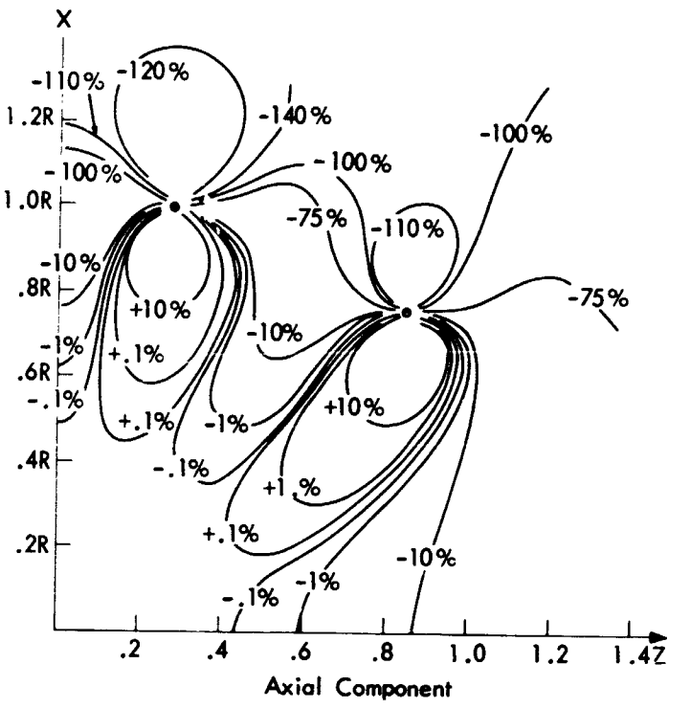


Figure 7. Braunbek coil (axial field plotted to reduced scale).

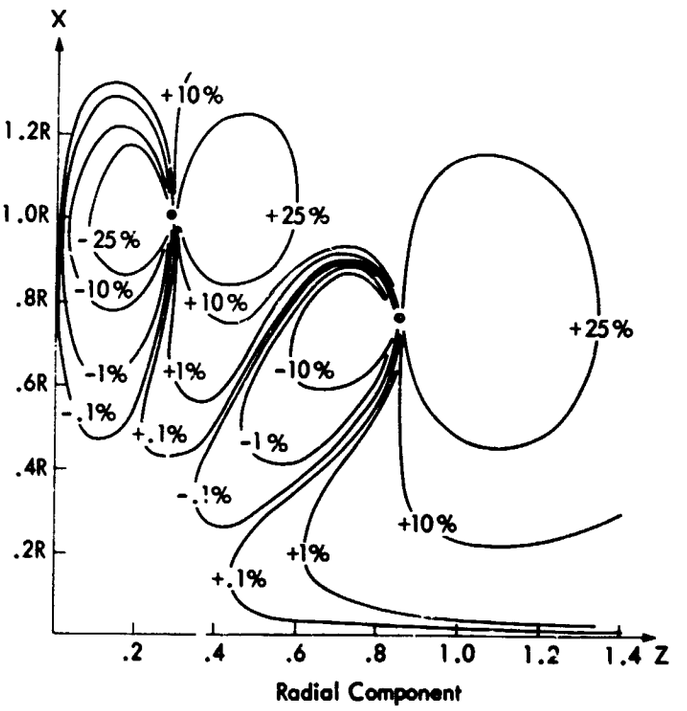


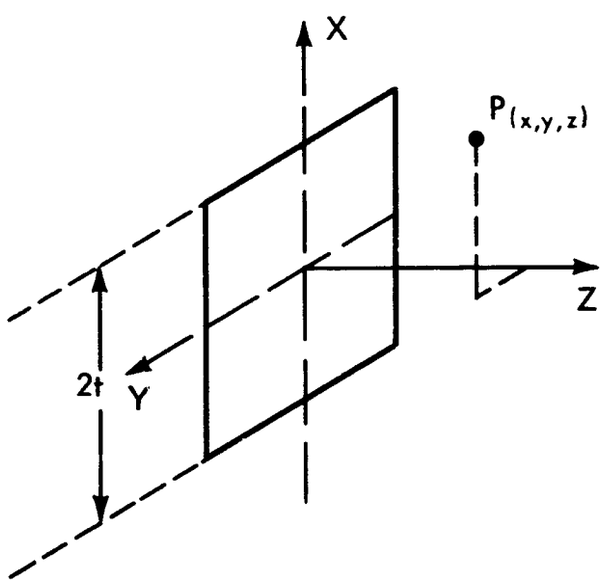
Figure 8. Braunbek coil (radial field plotted to reduced scale).

In choosing the coordinate system on which to mount the necessary coils, it is expected that the direction in which a gradient most certainly will be present is in the vertical direction. Consequently by choosing the z axis to be vertical, gradients of the Case A type can be corrected by one pair of coils. The North-South and East-West directions are then chosen as x and y.

These gradient coils are most useful when the resulting field at the center of the coil system is very small or zero. The gradient coils in use at GSFC are wound directly on the coil forms in a manner similar to those shown in Figure 9.

Square coils have been used by many investigators because of the ease with which they can be constructed. Some of the first theoretical work was published by E. L. Kaplan (136) as a Naval Ordnance Laboratory report and has received little notice.

The geometry of a single square coil is shown below.



The field produced by this single square coil is:

$$H_x = \frac{NI}{t} \left[f(1-x,\ z,\ 1+y) + f(1-x,\ z,\ 1-y) - f(1+x,\ z,\ 1+y) - f(1+x,\ z,\ 1-y) \right]$$

$$H_y = \frac{NI}{t} \left[f(1-y,\ z,\ 1+x) + f(1-y,\ z,\ 1-x) - f(1+y,\ z,\ 1+x) - f(1+x,\ z,\ 1-y) \right]$$

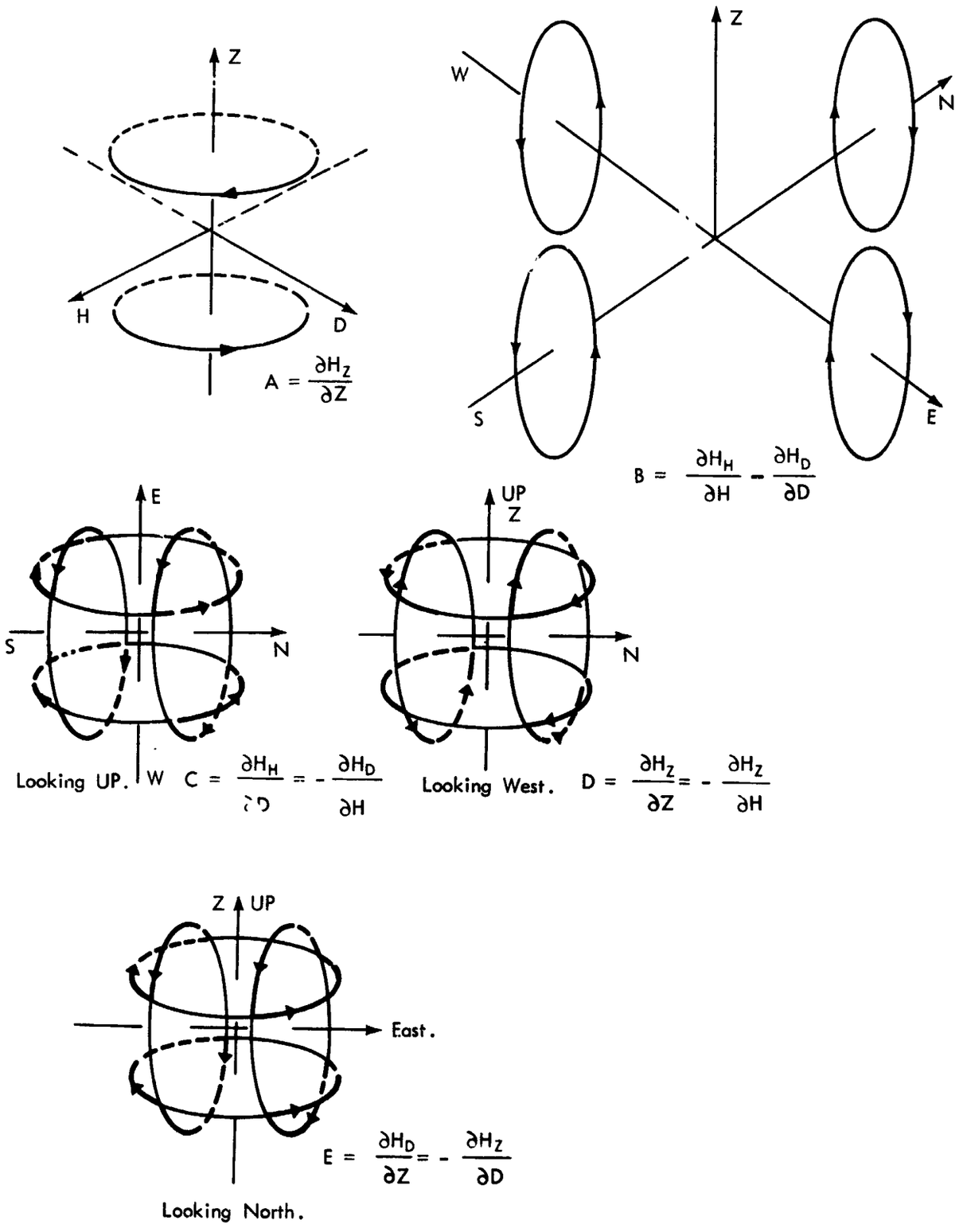


Figure 9. Schematic arrangement of the gradient coils.

$$H_z = \frac{NI}{t} [f(z, 1+x, 1+y) + f(z, 1+x, 1-y) + f(z, 1-x, 1+y) + f(z, 1-x, 1-y)$$

$$+ f(z, 1+y, 1+x) + f(z, 1+y, 1-x) + f(z, 1-y, 1+x) + f(z, 1-y, 1-x)]$$

where $f(u, v, w)$ represents the expression

$$v\,w\,(u^2 + v^2)^{-1}\,(u^2 + v^2 + w^2)^{-1/2}$$

on the axis this reduces to

$$H_z = \frac{8NI}{t}\left[\frac{1}{(1+z^2)\,(\sqrt{2+z^2})}\right] \text{ and } H_x = H_y = 0$$

The contour curves for the axial component of the field from a single square coil are shown in Figure 10.

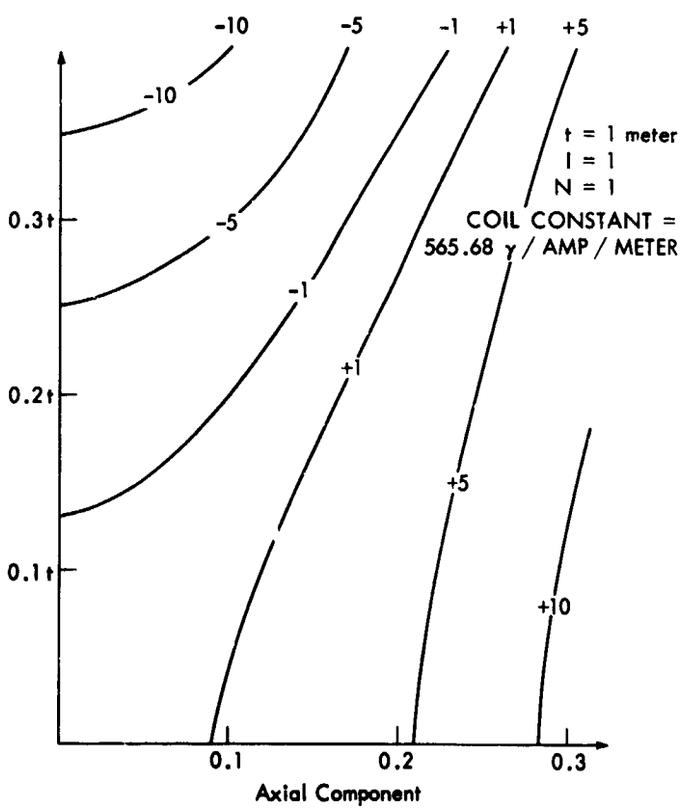


Figure 10. Single square coil (axial field plot).

The solution for a two loop square coil system, corresponding to the Helmholtz circular coil system, has been developed by Warburton (136) and Heller (52) by proceeding in the same manner as for the circular coil system. The spacing is chosen so that the second derivative of the axial component is zero at the center of the system. This distance from the center to each coil turns out to be $\pm 0.5445\,t$ (where $t = 1/2$ side length). The curves for the field components are given in Figures 11 and 12.

In the case of the four loop square coil systems, Waidelich et al (133) have shown that the axial field can be expressed as:

$$H_z = \frac{2N_1I_1}{\pi t_1}\left[f\left(\frac{d_1-z}{t_1}\right) + f\left(\frac{d_1+z}{t_1}\right)\right] + \frac{2N_2I_2}{\pi t_2}\left[f\left(\frac{d_2-z}{t_2}\right) + f\left(\frac{d_2+z}{t_2}\right)\right]$$

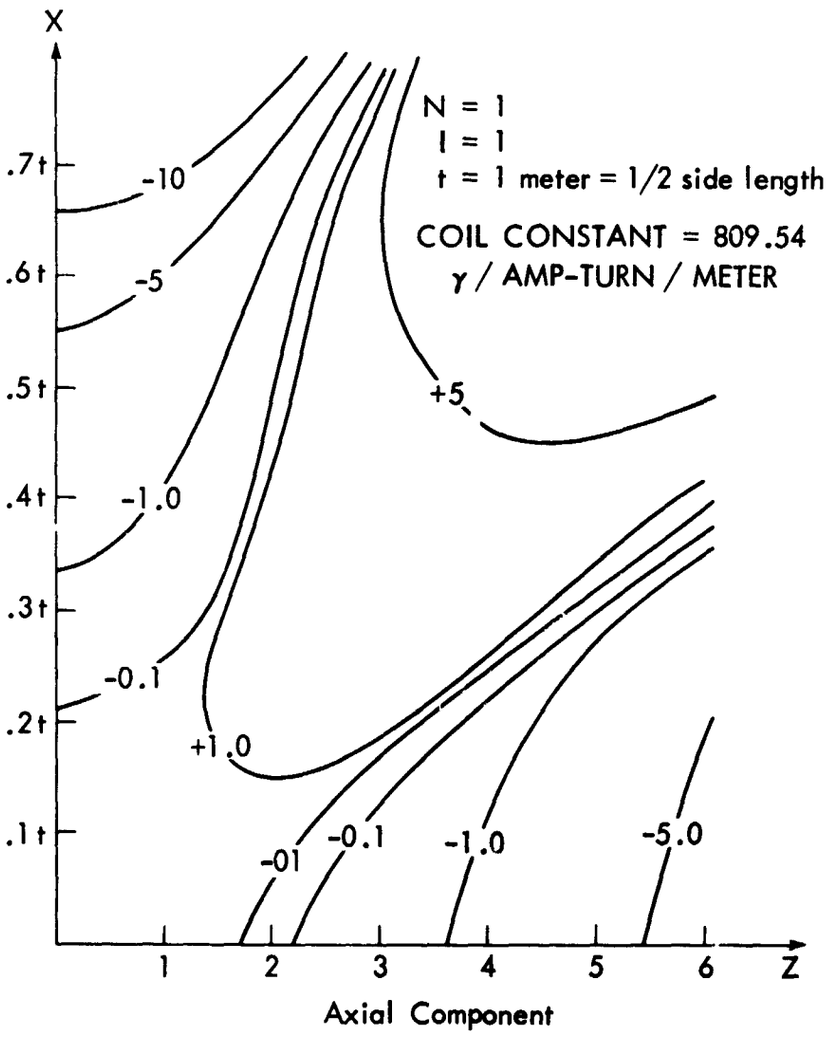


Figure 11. Square Helmholtz coil (axial field plot).

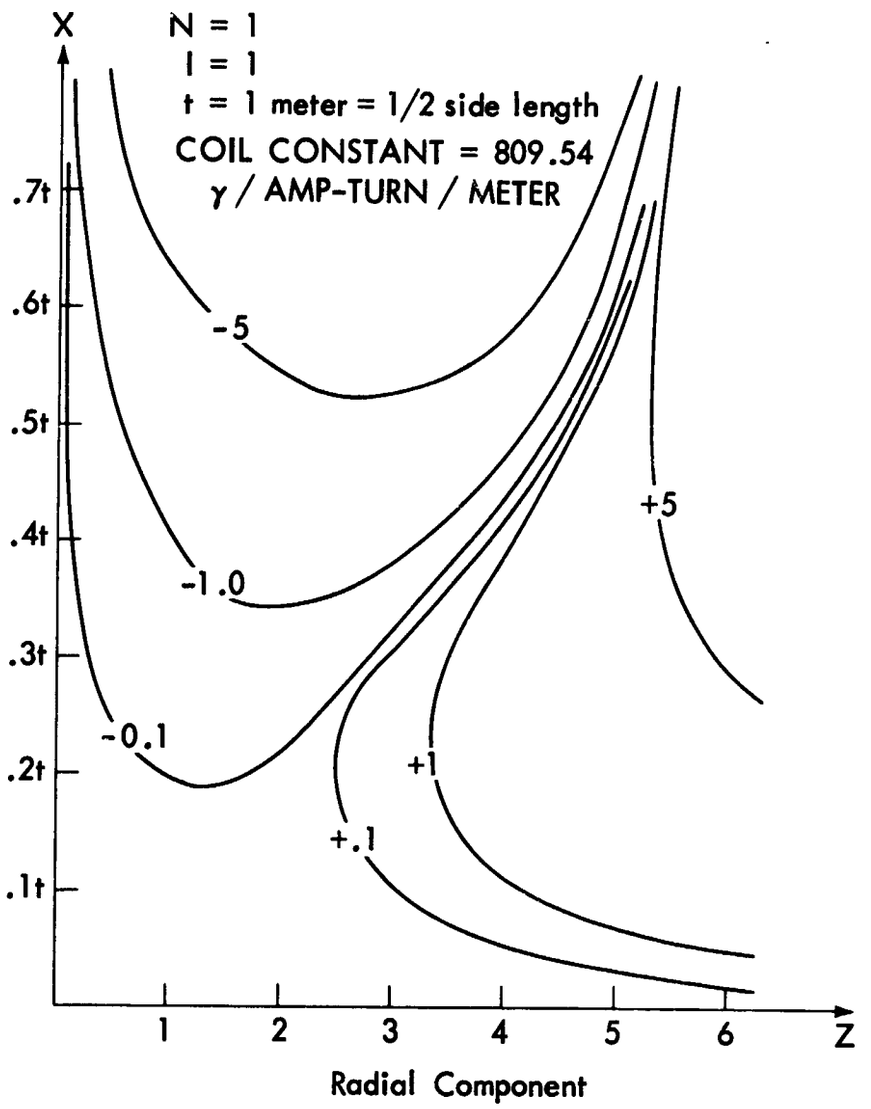


Figure 12. Square Helmholtz coil (radial field plot).

The elevation and end views of the four-square-coil system are shown in Figure 13. This work included a determination of the parameters of systems having unequal ampere-turns in the two pairs of coils, but having an integral turn ratio so that the coils could be connected in series. Table 2 gives the results for some of the solutions.

As the turns-ratio (N_2/N_1) is changed from 7:8 to 7:3, the center coils start to move toward the center while the outer coils move away from the center and increase in size.

Figures 14, 15, and 16 are plots of the axial component of the field in three different planes for the case with the turns-ratio equal to 1:1 (see Figure 13 (b)).

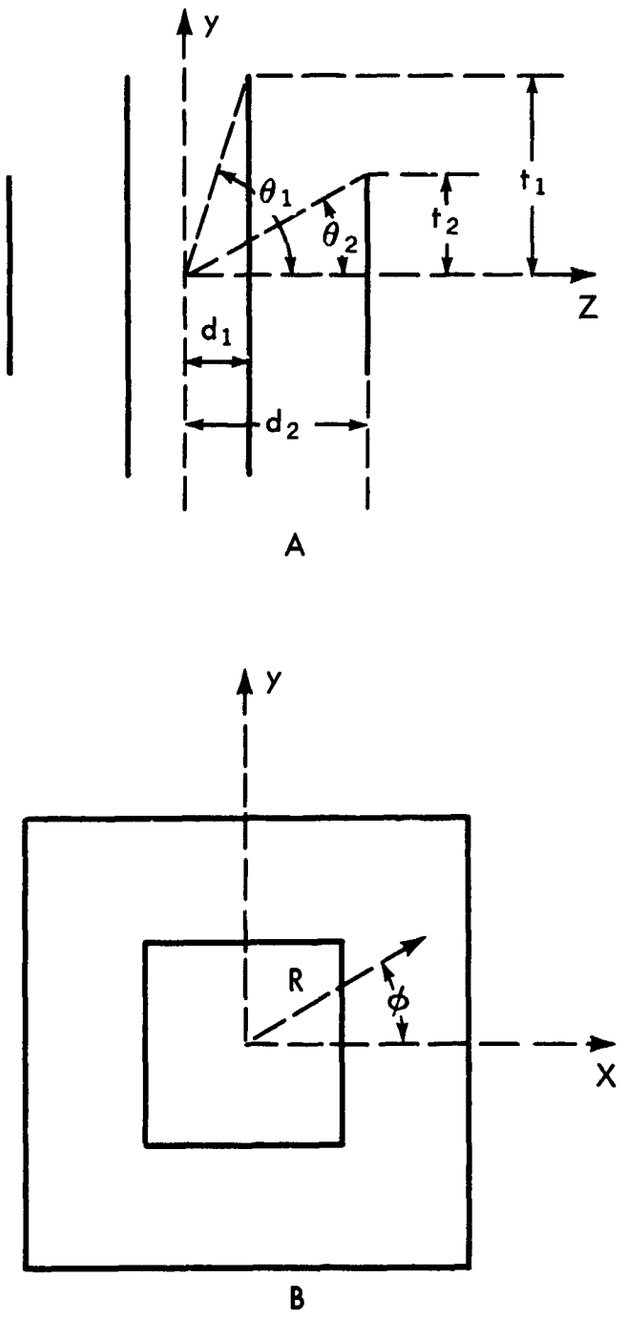


Figure 13. Elevation and end view of the four-square-coil system.

Table 2.
Parameters for Square Coil Systems with Integral Turn-Ratios

N_2/N_1	x_1	x_2	t_2	d_1	d_2	B_0	u
7/8	0.3012243	1.2296968	0.7205466	0.3012243	0.8860538	1427.195	0.2471588
1/1	0.2941177	1.1988007	0.7536688	0.2941177	0.9034988	1489.335	0.2506530
5/3	0.2697345	1.0837782	0.8936690	0.2697345	0.9685390	1806.113	0.2645352
2/1	0.2623200	1.0449411	0.9483309	0.2623200	0.9909499	1958.130	0.2696621
7/3	0.2566294	1.0133642	0.9962491	0.2566294	1.0095631	2107.070	0.2740560

t_1 = 1.00 for all coil systems

$x_1 = \frac{d_1}{t_1} = \cot \theta_1$; $x_2 = \frac{d_2}{t_2} = \cot \theta_2$

B_0 = Flux constant in gammas per ampere-turn per meter

u = Uniformity constant. Distance along axis from center of coil system in which uniformity is equal or better than .001%.

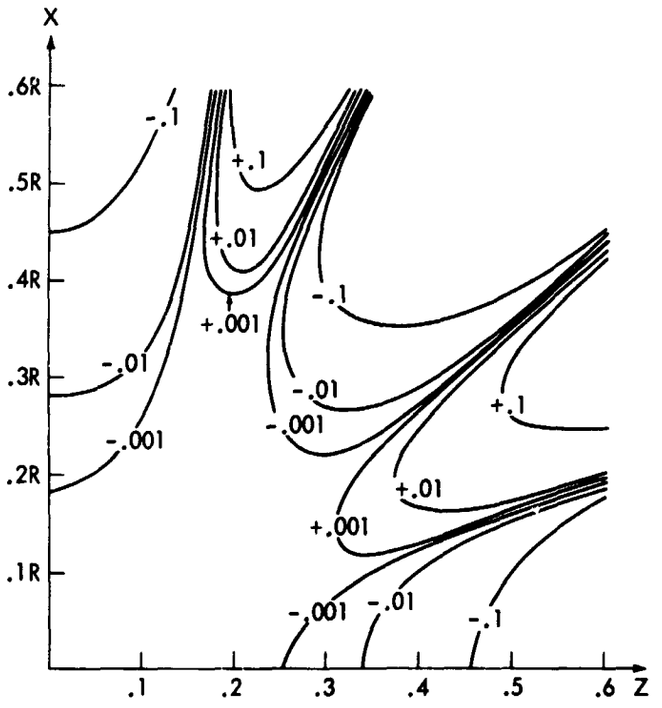


Figure 14. Square coil-turn ratio 1/1-θ = 0°
(axial field plot).

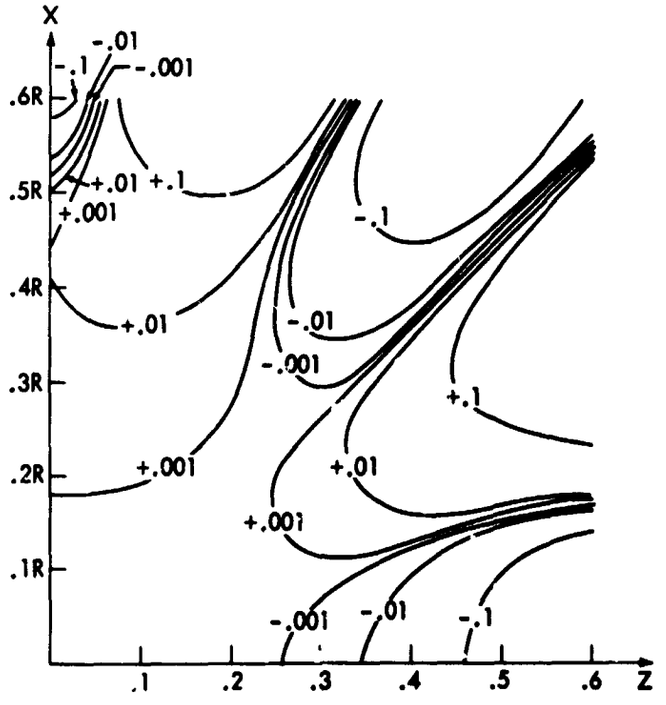


Figure 15. Square coil-turn ratio 1/1-θ = 30°
(axial field plot).

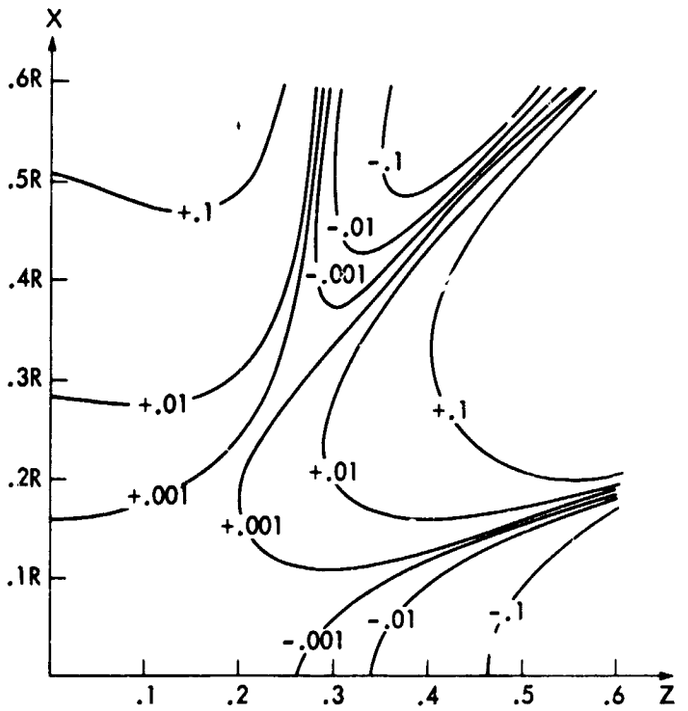


Figure 16. Square coil-turn ratio 1/1-θ = 45° (axial field plot).

It should be noted that the effect of the corners is very pronounced and in the plane θ = 45° the volume with a homogeneity better than 0.001% has decreased considerably.

In Figures 17 and 18 for the 7:3 turns ratio, the effect of the corners in the plane θ = 45° is not as pronounced as with a turns-ratio of 1:1.

The preceding discussion of various square coil systems has concerned those which yield a small volume at the center with a high degree of homogeneity (.001% or better). For some special types of work, particularly with ordnance items, the interest has beer directed toward large volumes with a modest homogeneity (0.1% to 5.0%).

Some of the first work on these multiple square coil systems is that of S. M. Rubens (101) and M. S. Raff (136) of the Naval Ordnance Laboratory.

Ruben's coil system consisted of five equally spaced square coils, forming the surface of a cube. The ampere-turns ratio were found by assuming first that t' e center coil was a single ampere-turn. If $H_0(z)$ is the axial field from this center coil at a distance z along the axis of the coil (a line perpendicular to the plane of the coil), and $H_1(z)$ and $H_2(z)$ be the axial field from the inner and

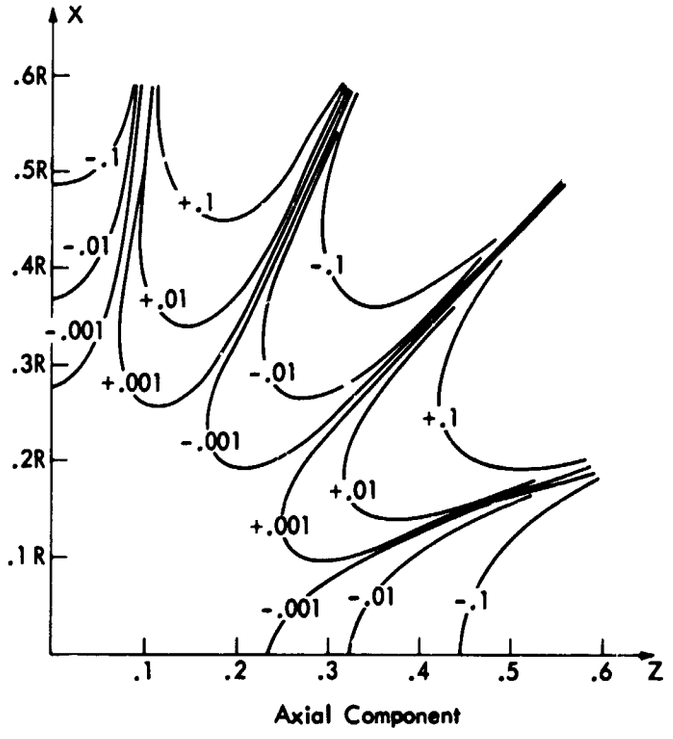


Figure 17. Square coil-turn ratio 7/3-$\theta = 0°$ (axial field plot).

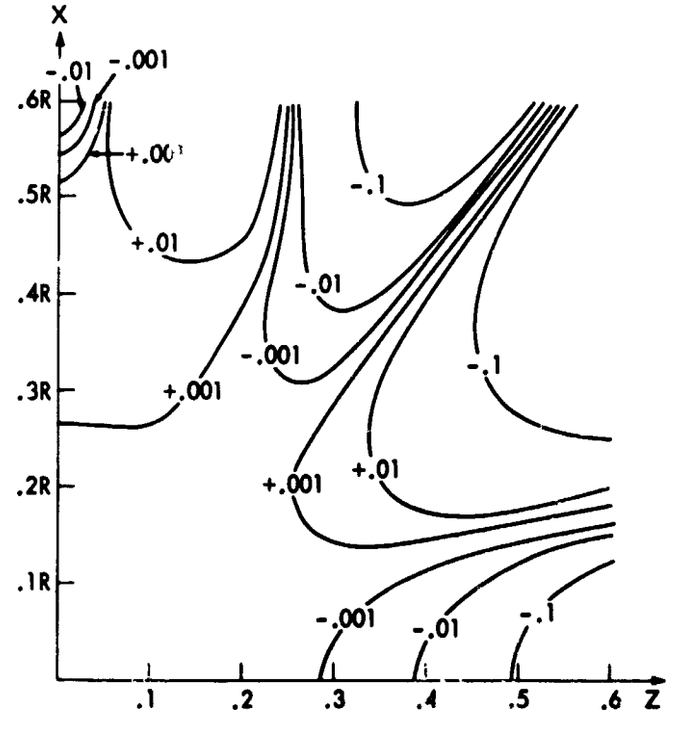


Figure 18. Square coil-turn ratio 7/3-$\theta = 45°$ (axial field plot).

outer pairs of coils. Then the field at z from the coil system will be

$$H_t(z) = H_0(z) + AH_1(z) + BH_2(z)$$

where A and B are the ratios of currents in the inner and outer coil pairs to the current in the center coil, i.e.,

$$A = \frac{N_1 I_1}{N_0 I_0}; \quad B = \frac{N_2 I_2}{N_0 I_0}.$$

By choosing three values of z where the values of H_z are to be identical, two simultaneous equations result and thus a solution for values of A and B may be made. In Ruben's work, he selected values of z at 0, 0.30 t and 0.50 t (where t = 1/2 the side length of the coils) to give a large volume with acceptable homogeneity. He obtained solutions of A = 0.405 and B = 1.92. This gives 71, 15, 37, 15, 71 for a set of integers proportional to the ampere-turns if the five coils are connected in series. Rubens found that by using A = 0.400 and B = 1.900, the homogeneity along the axis was substantially the same but that greater uniformity resulted off the axis.

If the number of turns on the system are exactly 19, 4, 10, 4, 19 then the axial field at the center is:

$$\text{Coil Constant} = \frac{35.69}{2t} \text{ oersteds per ampere}$$

where 2 t is the length of one side of the coil in centimeters. This coil system is shown in Figure 19.

With the turns ratio shown in Figure 19 the homogeneous volume available is:

% Homogeneity	Cylinder length	Cylinder diameter
0.1	0.90 t	0.40 t
1.0	1.00 t	1.00 t
5.0	1.50 t	1.00 t

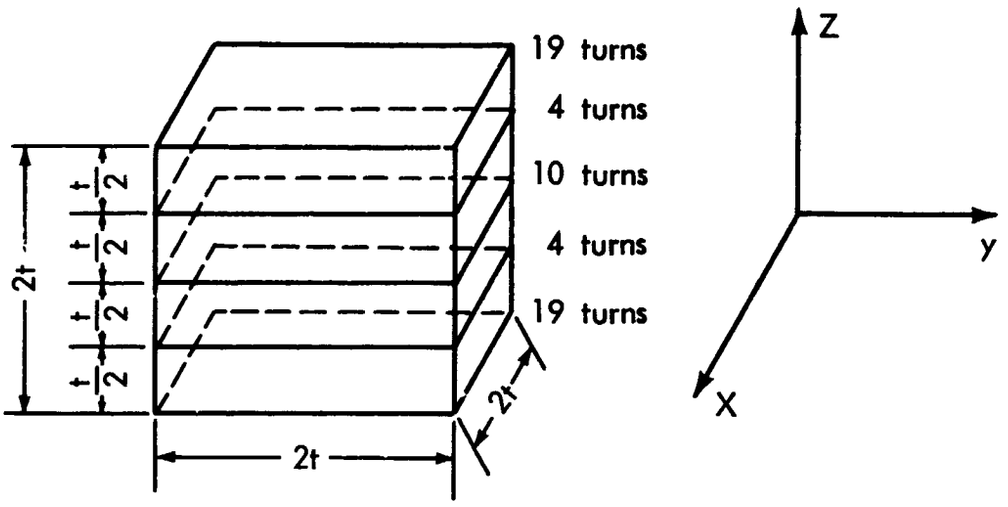


Figure 19. Ruben's five-coil system.

M. S. Raff (136) recomputed the original Ruben's design by setting up equations involving the turn-ratios as unknowns and selecting values that satisfied the equations and made as many of the lower order coefficients in the expansion vanish. The five coil system, computed on this basis, retained the equal spacing of the coils on the surface of a cube and was designed to give a volume with a homogeneity of 0.05% over a sphere with a radius equal to 0.3t (t is 1/2 the side length). The solution obtained uses turn ratios of 134, 44, 66, 44, 134.

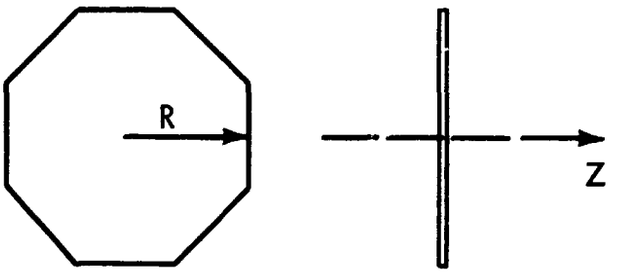
For many types of work, the three and five coil systems are inconvenient because the center coil is directly on the center-line of the assembly.

Raff (136) also designed a four coil system to give a homogeneity of 0.1% in a sphere with a radius equal to 0.2 t (where t is 1/2 the side length). The geometry of the coil system was specified and turn-ratios computed. The design selected by Raff to meet these requirements is shown in Figure 20.

There has been some interest in octagonal coil systems because of the ease with which they can be constructed and installed. The homogenity of the system falls somewhere between the circular coil systems and the square coil systems previously discussed.

Saperstone (114) and Chiang (25) have published some information on octagonal coils.

The magnetic field intensity on the axis of a single octagonal coil with N turns is:



$$H_z = \frac{4NIR^2\left(\tan\frac{\pi}{8}\right)}{\pi(R^2+Z^2)\sqrt{R^2\left(\tan\frac{2\pi}{8}+1\right)+Z^2}}$$

by substituting:

$$V = \frac{Z}{R}$$

$$B = \sqrt{1+\tan\frac{\pi}{8}} = 1.0823922$$

$$K = \frac{\pi}{4\left(\tan\frac{\pi}{8}\right)}$$

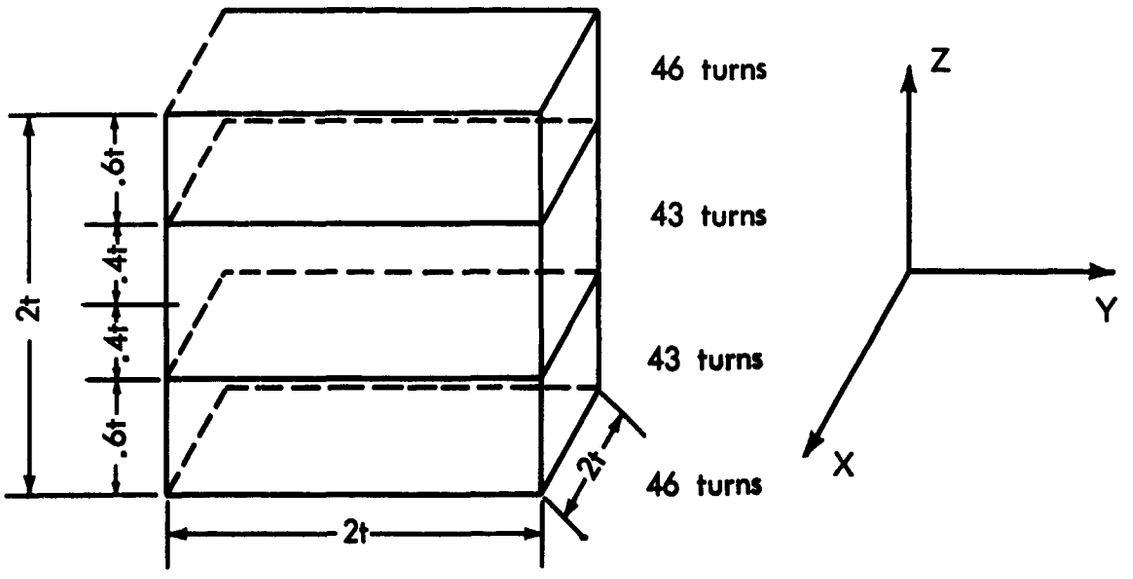


Figure 20. Ruben's four-coil system.

the above equation can be rewritten as:

$$H_z = \frac{NI}{RK}\left[\frac{1}{(1+V^2\sqrt{B^2+V^2}}\right]$$

For a four coil system, Figure 21, the field intensity along the axis can be expressed as:

$$H_z = \frac{N_1 I_1}{R_1 K}\left[f(d_1-z) + f(d_1+z)\right] + \frac{N_2 I_2}{R_2 K}\left[f(d_2-z) + f(d_2+z)\right]$$

Where R_1 and R_2 are the radii of the inscribed circles of the octagonal coils. If we substitute $x_1 = d_1/R_1$ and $x_2 = d_2/R_2$, we can then expand the above expression as a Taylor series about $z = 0$ hence:

$$H_z = \sum_{n=0}^{\infty} a_{2n} z^{2n} \quad \text{where} \quad a_n = \frac{1}{n!}\frac{d^n H}{dZ^n}\Bigg|_{z=0}$$

Suitable values of x_1 and x_2 can be determined, such that the first three terms of the expansion are equal to zero. Saperstone computed the parameters

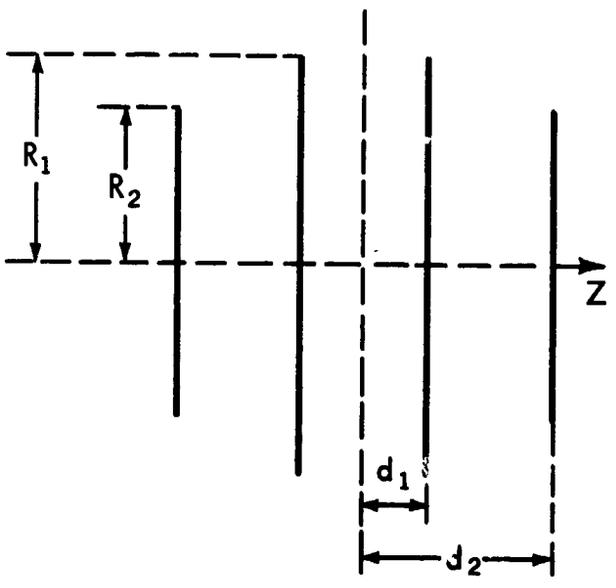


Figure 21. Four coil octagonal system.

of several coil configurations having integral turns ratios. His results are tabulated in Table 3.

Table 3
Octagonal Coil Systems with Integral Turns-Ratios

$I = \frac{N_2 I_2}{N_1 I_1}$	d_1	d_2	R_2	R_1
16/16	.28435432	.86560663	.76317254	1.0
15/16	.28764020	.85751313	.74677648	1.0
14/16	.29120383	.84881337	.72964492	1.0
13/16	.29508305	.83942090	.71170181	1.0
12/16	.29932539	.82923067	.69285597	1.0
11/16	.30398098	.81809756	.67299845	1.0
10/16	.30913377	.80589842	.65200038	1.0
9/16	.31485706	.79237031	.62968647	1.0
8/16	.32126987	.77724546	.60584340	1.0

Figure 22 is a plot of the axial field of an octagonal coil system with a 1:1 turn ratio, (similar to the Braunbek coil configuration) a close-comparison of the axial field of this octagonal coil and the circular Braunbek (Figure 7) shows

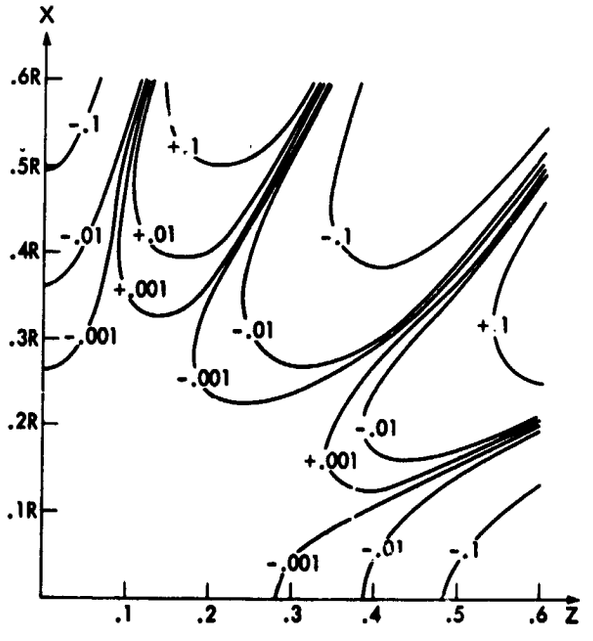


Figure 22. Octagonal coil (axial field plot).

very little difference in the plane perpendicular to the side of the coil. As this plane is moved to the corner of the system, an effect similar to that experienced with the square coil becomes apparent. It is evident, as is to be expected, that the octagonal coil is between the circular and square coil systems in volume of homogenity available.

REFERENCES

1. Albach, W., "Methoden zur genauen Messung von Spulenkonstanten, Feldverteilungen und Windungsflachen," Zeitschrift für Instrumentenkunde, Volume 73, pages 7 to 10, (Janua.· 1965)

   A method is given for measuring the axial field of a solenoid by the use of a Helmholtz coil system.

2. Albach, W. and Voss, G. A., "Ein durch vorerregte Dynamobleche magnetisch abgeschirmter Messraum," Zeitschrift für angewandte Physik, Volume 9, pages 111 to 115, (March 1957).

3. Azatyan, A. A., Badalyan, G. V., and Eritsyan, G. N., "Obtaining Steady Magnetic Fields of a Prescribed Shape," Instruments and Experimental Techniques, Volume 3, pages 514 to 518, (May-June 1963).

4. Beyerle, K., "Ein Beitrag zur Entwicklung des Kathodenoszillographen mit kalter Kathode," Archiv für Elektrotechnik, Volume 25, pages 267 to 276, (1931).

   Two coils were used in a Helmholtz system to counteract the earth's field for use in a cathode ray oscilloscope. Two types of coils were employed, one was rectangular and the other was rectangular but with rounded ends. Equations for the axial field are also given.

5. Blewett, J. P., "Magnetic Field Configurations due to Air Core Coils," Journal of Applied Physics, Volume 18, pages 968 to 976, (November 1947).

   The production of a uniform field perpendicular to the axis of and inside an infinitely long cylinder is discussed. Also the equations are given for ellipsoidal coils for uniform fields along with the special cases of a long solenoid and a sphere.

6. Braunbek, W., "Die Erzengung weitgehend homogener Magnetfelder durch Kreisströme," Zeitschrift für Physik, Volume 88, pages 399 to 402, (1934).

   The method of solution and parameters of the eighth-order coil system with two coil pairs in which the currents are the same in all of the coils are given.

7. Bestelmeyer, A., "Berechnung, Herstellung und Messung eines homogenen Magnetfeldes," Physikalische Zeitschrift, Volume 12, pages 1107 to 1111, (1911).

One large solenoid and two smaller solenoids were used to produce a very uniform field. Calculated and measured curves are given of the axially-directed field measured both on and off the axis of the solenoids. The results for two different spacings of the small solenoids from the large solenoid are given.

8. Barnett, S. J., Researches of the Department of Terrestrial Magnetism, Volume 4, Publication No. 175 of the Carnegie Institution of Washington, (1921).

9. Bromwich, T. J. I., "Note on the Magnetic Field Produced by Circular Currents," Philosophical Magazine, Volume 46, pages 108 to 112, (1923).

The field of a Helmholtz coil-pair is expanded along the axis and integrated to obtain the scalar potential. Legendre functions are then inserted to obtain the potential and both components of the field at any point on or off the axis.

10. Bock R., "Über die Homogenität des magnetischen Feldes in der Helmholtz-Gaugainschen Doppelkreisanordnung," Zeitschrift für Physik, Volume 54, pages 257 to 259, (1929).

The absolute magnitude of the field of two circular coils in a Helmholtz system is calculated. The results are given as a table and as contour curves.

11. Bacon, R. H., "Method of Producing Uniform Magnetic Field," Review of Scientific Instruments, Volume 7, pages 423 to 425, (November 1936).

Two coils of finite cross section are used. The cross section of each coil is divided into three parts and the fields at a point on the axis, and at a point off the axis are calculated. By choosing different currents in the three parts, the variation of the field at the two points may be minimized.

12. Barker, J. R., "New Coil Systems for the Production of Uniform Magnetic Fields," Journal of Scientific Instruments, Volume 26, pages 273 to 275, (August 1949).

Three- and four-coil systems are presented in which all of the coils have the same diameter but carry different currents. The three-coil system is a sixth order system while the four-coil system is an eighth order one. Corrections for the finite cross section of the windings are given. Various practical problems in making the coils and obtaining the necessary current ratios are discussed.

13. Barker, J. R., "An Improved Three Coil System for Producing a Uniform Magnetic Field," Journal of Scientific Instruments, Volume 27, pages 197 to 199, (July 1950).

This paper presents a three-coil system with rectangular coil cross section in which all coils carry the same current, but the middle coil has one half the number of complete layers that the two outer coils have. The solution is that of a sixth-order system and uses the superposition of three solenoids. A table of design data is presented and some design considerations are given along with a numerical example.

14. Brunelli, B. Ye., Nizyayev, D. A. and Kanonidi, Kh. D., "A Magnetic Field Stabilizer," Bulletin of the Academy of Sciences of the USSR, Geophysics Series, No. 7, pages 523 to 525 (July 1958).

The mirror in a magnetometer moves a light spot off a photocell which controls through an amplifier the current in a Helmholtz coil system. Some construction details of the coils and the rest of the apparatus are given. A mean variation of 60 to 100 gammas was reduced to 2 to 3 gammas.

15. Berger, W. and Butterweck, H. J.. "Die Berechnung von Spulen Zur Erzengung homogener Magnetfelder und Konstanter Feldgradienten," Archiv für Elektrotechnik, Volume 42, pages 216 to 222, (1956).

16. Badalyan, G. V., "Production of Constant Magnetic Fields of Prescribed Forms by Means of a Magnetic System without Manifest Poles," Soviet Physics -- Technical Physics, Volume 8, pages 253 to 256, (September 1963).

Presents a method of producing a very uniform field inside a hollow cylinder of ferromagnetic material by having windings parallel to the axis of the cylinder and on the inside surface of the cylinder.

17. Chapman, S. and Bartels, J., "Geomagnetism" Oxford University Press, Volume 1, pages 83 to 85, (1940).

Discusses briefly the Helmholtz-Gaugain, Fanselau and Braunbek coil systems. Gives dimensions of a Fanselau system useful in testing magnetometers.

18. Craig, H., "The Production of a Uniform Magnetic Field over a Specific Volume by Means of Twin Conducting Circular Coils," Proceedings of the Physical Society, Volume 59, pages 804 to 814, (September 1947).

Curves are given for the optimum coil separation for two coils when the axially directed field is considered off the axis for a circular area and for an annulus. Contour curves of the axial field for one case are presented and the variation of the radial field is also shown. Also considered is the effect of finite cross section of the coils.

19. Clark, J. W., "A New Method for Obtaining a Uniform Magnetic Field," Review of Scientific Instruments, Volume 9, pages 320 to 322, (October 1938).

Construction of a spherical coil system to produce a uniform magnetic field is shown. The uniform field was about 100 oersteds in a ten-inch diameter sphere with a current of 4.35 amperes and a power of 1090 watts. The measured field uniformity was better than one percent.

20. Crane, H. R., "Cloud Chamber for Nuclear Disintegration Studies," Review of Scientific Instruments, Volume 8, pages 440 to 444, (November 1937).

Constructional details of a two-coil system with finite cross section. The experimental variation of the axial field with radial distance is shown.

21. Crittenden, E. C., Jr., and McDaniel, B. D., "Production of Uniform Magnetic Field by a Pair of Magnetic Dipoles," Review of Scientific Instruments, Volume 15, page 270, (October 1944).

In the field between two magnetic dipoles, there is a straight line along which the field is nearly uniform for quite some distance. The dipoles were approximated by two coils, and a table of measured distances for one and two percent uniformity is presented.

22. Coonrod, L. G., Cecconi, W. E., Gedraitis, C., Harper, K. R. and Lewis, R. M., "Magnet Coils," Review of Scientific Instruments, Volume 20, pages 408 to 410, (June 1949).

This paper presents some constructional details of Helmholtz coils to produce large fields.

23. Brownfield, F. R., Jr., "Optimum Spacing of Coil Pairs," Review of Scientific Instruments, Volume 35, pages 240 to 241, February 1964.

This paper shows that if the spacing of a coil-pair is made more than that of the Helmholtz condition, a longer volume of uniformity in the axial direction is possible.

24. Chan, S. B., "The Homogeneity of Coil-Dipole Systems," Research Report, University of Missouri Engineering Library, Columbia, Mo., (June 1965).

25. Chiang, B. A., "Four-Octagonal-Coil Systems and Coil-Dipole Systems," Research Report, University of Missouri Engineering Library, Columbia, Mo., (January 1965).

This paper presents a complete solution of the four-octagonal-coil problem with the results given as curves and tables. Also considered is the variation of the axial field component as the axial distance is varied for a coil pair with various spacings between the coils. The same study is then made for one large coil and two dipoles on the axis of the coil.

26. DuMond, J. W. M., "An Iron-Free Magnetic Beta-Ray Spectrometer of High Luminosity, Resolving Power, and Precision for the Study of Decay Schemes of the Heavy Isotopes," Annals of Physics, Volume 2, pages 283 to 371, (1957).

This instrument required an ellipsoidal winding producing a field of 665 gauss and using a current of 30.8 amperes and 3.4 kilowatts power. The uniformity was 3 parts in 10^4. Also two coaxial pairs of Helmholtz coils connected in opposition were used so that the external field varied as the inverse seventh power of the distance and yet the internal field remained quite homogeneous. Two sets of large square Helmholtz coils were used to neutralize the earth's field. One set, 124 inches on a side, was used for the vertical component and the other set, 65-3/4 inches on a side was used for the horizontal component.

27. Devane, J. F., "Research Directed Toward the Collection, Reduction and Evaluation of Geomagnetic Field Data and Electric Field Phenomena Atmospheric and Terrestrial," NASA STAR N65-20888.

A coil system was constructed to reproduce the magnetic fields encountered by space vehicles. The coil system has four coils and is of the Braunbek design. The larger coils have a diameter of 90 inches and the axis of the system can be oriented with respect to the inclination of the earth's magnetic field.

28. Dürschner, H., "Ein Magnetometer zur Bestimmung det magnetischen Eigenschaften von Gesteinen," Annales de Géophysique, Volume 10, pages 152 to 156, (1954).

An astatic magnetometer is constructed using a Helmholtz coil-pair suspended by a gold fiber. The advantage of the use of the Helmholtz pair is that the measurements depend less upon the shape of the magnetic sample.

29. Dye, D. W., "A Magnetometer for the Measurement of the Earth's Vertical Magnetic Intensity in C. G. S. Measure," Proceedings of the Royal Society of London, Volume 117, pages 434 to 458, (1928).

The same Helmholtz coil pair was used as that used by Smith (116). Within a sphere two centimeters in diameter the maximum difference in axial intensity from that at the center is less than four parts in 10^5. Gives the coil constant with a correction for the temperature change.

30. Day, J. D. A., "Superconducting Magnet Design," Journal of Scientific Instruments, Volume 40, pages 503 to 505, (December 1963).

One coil-pair with finite cross section is used and the system is made to be a fourth-order system. The winding space is minimized, and a table of dimensionless results is given. A design of a particular coil system is presented.

31. Daniel, H., "Einige Aufwendungsmöglichkeiten des homogenen magnetischen Kreisfeldes, $H_r = H_z = 0$, H_θ = Constant," Nuclear Instruments and Methods, Volume 24, pages 181 to 184, (August 1963).

This paper discusses a method of producing a field with H_θ = constant using a number of coils with approximately constant current density along the radius and each coil placed in plane through the axis.

32. Day, J. D. A. and Orton, J. W., "The Design of a Superconducting Magnet for a Travelling Wave Maser Using a Helmholtz Pair," IAA A64-23453.

33. d'Espine, J., "Contributions à l'Etude des Rayons β par La Méthode des Spectres Magnétiques," Annales de Physique, Series 10, Volume 16, pages 5 to 70, (1931).

This article is concerned with the axial field in the plane between two Helmholtz coils of rectangular cross section. The coils are 44 centimeters in radius, have 624 turns, carry a current of 14 amperes and produce a field of 179 gauss. The field has a uniformity of 0.2 percent in a 16 centimeter diameter circle and 0.8 percent in one 34 centimeters in diameter. Sauerwein (J17), this section points out that if a value of b/c = 0.928 had been used, a more homogeneous field would have resulted.

34. Ference, M., Shaw, A. E. and Stephenson, R. J., "The Production of Annular Magnetic Fields of Great Uniformity," Review of Scientific Instruments, Volume 11, pages 57 to 62, (February 1940).

Two coils of rectangular cross section and about 0.6 meter in diameter had the axial magnetic field in the center transverse plane calculated and also measured for the Helmholtz spacing and for a slightly smaller spacing. The uniformity for the smaller spacing was better in the annular region considered.

35. Foss, M. H., "Design of Cloud Chamber Coils," Contract N7onr 303 Task Order 1 NR 025-035, Technical Report No. 2, (1948) Carnegie Institute of Technology Library.

A method is given for two coils with rectangular cross section to make the second term in the series expansion of the field zero. Two numerical examples are presented.

36. Fanselau, G., "Die Erzengung weitgehend homogener Magnetfelder durch Kreisströme," Zeitschrift für Physik, Volume 54, pages 260 to 269, (1929).

This paper gives a sixth-order solution of the two coil-pair system with equal currents and using the roots of the Legendre functions. Also uses the same method to obtain an eighth order solution for three coil-pairs. Shows the contour curves for a two coil-pair system.

37. Fanselau, G., "Über die Homogenität des Magnetfeldes bei symmetrischer Spulenanordnung," Zeitschrift tür Geophysik, Volume 9, pages 236 to 237, (1933).

This paper gives a sixth order solution of a two-pair circular coil system in which the two coils on one side of the center of the system lie in the same plane perpendicular to the axis. The winding ratio could be approximated by 85/3 or 113/4.

38. Franzen, W., "Reinterpretation of Garrett's Theory of Axial Magnetic Fields," (May 5, 1961). Not published but a copy was obtained from Professor W. Franzen, Department of Physics, Boston University, 700 Commonwealth Ave., Boston, Mass. 02215.

This paper reviews the work of Garrett, see reference (48) and shows that a sufficient condition for producing a uniform field is to make as many derivatives of the field equal to zero as is possible. The conditions are derived for a Helmholtz coil pair and for the eighth-order four-coil system first given by McKeehan (74), this section.

39. Franzen, W., "Generation of Uniform Magnetic Fields by Means of Air-Core Coils, "Review of Scientific Instruments, Volume 33, pages 933 to 938, (September 1962).

This is a continuation of the earlier paper of Franzen but it is concerned mainly with coils of rectangular cross section. The two-coil system with each coil of rectangular cross section is considered first and the axial field is derived as a power series in the axial distance. Franzen finds that the breadth to depth ratio of the coil cross section should be 0.928 which agrees with Maxwell (76) and Sauter (123). For the eighth-order four- coil system of McKeehan (74), it is possible to make a sixth-order system when finite rectangular coil cross sections are assumed.

40. Foner, S. and Kolm, H. H., "Coils for the Production of High-Intensity Pulsed Magnetic Fields," Review of Scientific Instruments, Volume 28, pages 799 to 807, (October 1957).

This paper presents curves showing the variation along the axis of a coil-pair with rectangular cross-section. Details of construction are given for a coil which has a peak field at the center of 350 kilogauss.

41. Fiks, I. G., "Selecting the Criteria of a Winding Which Forms a Homogeneous Magnetic Field," Elektrichestro (USSR) No. 5, pages 43 to 49, (May 1962) (in Russian).

A short abstract in English is in Electric Technology USSR, Volume 2, pages 331, (1962).

42. Franzen, W., "Moebius Strip Series Connection for RF Helmholtz Coils," Review of Scientific Instruments, Volume 36, page 1062, (July 1965).

43. Granzow, K. D., "Homogeneity of the Magnetic Field of a Helmholtz Coil," Sandia Corporation Monograph SCR-193, (July 1960).

The axial and radial components of the field of a Helmholtz coil pair are calculated and plotted. Contour curves for both components are also presented.

44. Grant, W. J. C. and Strandberg, M. W. P., "Magnetic Field of Noncylindrical Coils," Review of Scientific Instruments, Volume 36, pages 343 to 345, (March 1965).

This paper considers two rectangular shaped coils of rectangular cross section. The field components of many different configurations were computed. Some empirical equations for the error curves are also given.

45. Gehrcke, E. and von Wogan, M., "Magnetische Messungen," Verhandlungen der Deutschen Physikalichen Gesellschaft, Volume 11, pages 646 to 681, (1909).

Two conically-wound coils about 70 centimeters in diameter were used in a Helmholtz system to produce a large volume of uniform magnetic field. Construction details are presented.

46. Gardner, M. E., Jungerman, J. A., Lichtenstein, P. G. and Patten, C. G., "Production of a Uniform Magnetic Field by Means of an End-Corrected Solenoid," Review of Scientific Instruments, Volume 31, pages 929 to 934, (September 1960).

This paper gives calculations and the construction details of a thick solenoid with 0.01 percent uniformity in the axial region comprising 63 percent of the solenoid length. The conditions used were to make the magnitude of the field at the center and at a given point along the axis equal, together making the derivative of the field at the given point zero. A four coil system of Maxwell (76) was used to cancel the field component of the earth perpendicular to the axis of the solenoid.

47. Glaser, W., "Über die zu einem vorgegebenen Magnetfeld gehörende Windungs dichte einer Kreisspule," Zeitschrift für Physik, Volume 118, pages 264 to 268 (1941).

This paper expresses the magnetic field along the axis of a system as an integral equation in terms of the unknown function of the ampere-turns per unit axial length. Indicates that this same problem would arise in producing a homogeneous electric field. Through the use of Fourier transforms the function of the ampere-turns per unit length is given as an integral which has to be evaluated.

48. Garrett, M. W., "Axially Symmetric Systems for Generating and Measuring Magnetic Fields. Part I," Journal of Applied Physics, Volume 22, pages 1091 to 1107, (September 1951).

The properties of Lengendre polynomials as used in axially symmetric fields are discussed. The error terms and associated contour curves are given. For use in coils of finite cross section the form of the series and associated tables are presented. Various applications are discussed with reference to the theory presented including uniform and gradient field systems, thin and thick solenoids for fourth, sixth and eighth order systems.

49. Garrett, M. W., "Dipole, Quadrupole and Octupole Field and Search Coils Using Loops and Solenoids," Physical Review, Volume 74, page 1231, (1948).

This paper mentions that systems producing uniform magnetic fields produce nearly ideal dipole fields at distant points and as search coils they tend to respond only to the axial component of the central field. Flux changes in these systems measure the dipole moments of the magnet or coil suddenly removed from the central region.

50. Gandal, P., "Sur la Theorie des Bobines de Hélmholtz," Revue Générale de l'Eléctricité, Volume 64, pages 199 to 202, (April 1955).

The expressions for the axial and radial field components are derived for two circular coils with rectangular cross sections. The Helmholtz condition is obtained and also a method to make the axial field component have a minimum variation at a given radial distance is presented.

51. Gaume, F., "Bobines sans Fer pour la Production de Chamos Magnétiques Constant Très Intenses Calcul et Réalisation, I, Classification et Usage des Bobines sans Fer Alimentation. Calcul des Enroulements," Journal des Recherches du Centre National de la Recherche Scientifique, Volume 9, No. 43, pages 93 to 115, (June 1958).

A discussion is presented of the various methods of producing a uniform magnetic field. The paper has 54 references, many of which are concerned with the production of uniform fields.

52. Heller, C., "Über die Erzengung gross räumiger homogener Magnetfelder zum Studium des Verhaltens von Magnetkompassen und Kompensiermitteln auf verschiedenen magnetischen Breiten," Deutschen Hydrographischen Zeitschrift, Volume 8, pages 157 to 164, (1955).

This paper states that two rectangular coils have a 20 percent greater volume of uniformity than two circular coils. The variation of the axial field along the axis is presented for two rectangular and for two circular coils with spacings both smaller and larger than the optimum values. Curves for a one percent variation in the axial field are also given along with the maximum values. Curves of the variation of the axial field along the plane of symmetry transverse to the axis are also shown.

53. Harris, F. K., "A New Cathode-Ray Oscillograph and its Application to the Study of Power Loss in Dielectric Materials," National Bureau of Standards Journal of Research, Volume 12, pages 87 to 102, (1934).

Two pairs of long narrow coils are used to compensate for the horizontal components of the earth's field. The long sides of the coils are parallel to the axis of a cylinder and the rounded ends of the coils lie in the cylindrical surface. Over 0.8 of the length of the cylinder and for distances of not more than 0.1 the radius from the axis, the field is uniform to within 5 percent.

54. Harris, F. K., "Coil Arrangements for Producing a Uniform Magnetic Field," National Bureau of Standards Journal of Research, Volume 13, pages 391 to 410, (1934).

An analysis of the field of the coils in (53) was made for the three cases of straight-head, circular-head and bent-head coils. Both theoretical and experimental field variation and contour curves are shown for all three field components. These curves show that the bent-head coils are superior to both the straight-head and circular-head coils with the circular-head type being somewhat better than the straight head.

55. Haynes, S. K. and Wedding, J. W., "The Use of Three Long Rectangular Coils for Neutralization of the Earth's Magnetic Field in a Lens-Type Beta-Ray Spectrometer," Review of Scientific Instruments, Volume 22, pages 97 to 101, (1951). See abstract in Physical Review, Volume 78, pages 641 to 642, (1950).

Three long rectangular coils located symmetrically on the surface of a circular cylinder were used. Dimensions were such that only sixth order terms were of importance, and the proper values of the currents were determined experimentally. One set of three coils was used for the vertical component of the earth's field and the second set for the horizontal component.

56. Hellbardt, G., "Über die Grenzempfindlichkeit des astatischen Spulenmagnetometers," Zeitschrift für Geophysik, Volume 24, pages 34 to 52, (1958).

Presents the theory of a coil system within coaxial pairs of coils following Sauter (123). In particular he gives the four coil solution of McKeehan (74) and the six coil solution of Neumann (74) which is approximated by using turns ratios of 17:11:22. The effect of a small change in geometry or current on the homogeneity of the coil system is also calculated.

57. Hipple, J. A., Jr., "A Spherical Coil for a Mass Spectrometer," Physical Review, Volume 55, pages 597, (1939).

A spherical coil was constructed with a uniformity of 0.3 percent throughout the volume. Constructional details are given and the field inside was 2500 gauss when the coil was cooled by circulating water.

58. Hara, K., "The Magnetic Field Near the Center of a Coil Pair like a Helmholtz Coil with Helical Windings of Finite Cross Section," Researches of the Electrotechnical Laboratory, Tokyo, No. 560, (May 1957).

The field produced by a coil of finite cross section with a helical winding along with the effects of the connecting leads is derived. Then the field produced by a pair of such coils is calculated using power series. A numerical example is presented.

59. Heppner, J. P., Ness, N. F., Skillman, T. L. and Scearce, C. S., "Explorer X Magnetic Field Results," in Space Research III, Proceedings of the Third International Space Science Symposium (COSPAR), Washington, D. C., May 2-8, 1962, pages 553 to 601, (1963).

The McKeehan (74) four-coil system was used to provide a bias field for the rubidium-vapor magnetometer (141) flown on the satellite.

60. Hibbs, A. R., Eimer, M. and Neugebauer, M., "Early Ranger Experiments," Astronautics, Volume 6, pages 26 to 27, (September 1961).

The rubidium-vapor magnetometer cell on board the spacecraft was enclosed within a 13-inch diameter fiberglass spherical shell on which a McKeehan (74) four-coil system was wound (141).

61. Ingarden. R. S. and Michalczyk, J., "On Optimum Field Homogeneity of High Energy Coil Magnets," Bulletin de l'Académie Polonaise des Sciences, Série des Science Mathematique Astronomie et Physique, Volume 8, No. 5, pages 319 to 324, (1960).

This paper discusses the case of two circular coils of rectangular cross section. The field is expanded in a Maclaurin series and also in a Gegenbauer series. A fourth or sixth order field is specified in the Maclaurin series to produce the best uniformity in the vicinity of the origin for small objects. A similar specification in the Gegenbauer series is useful where the objects are large but small oscillations of the field are not harmful. A numerical example is given.

62. Irving, E., Stott, P. M. and Ward, M. A., "Demagnetization of Igneous Rocks by Alternating Magnetic Fields," Philosophical Magazine, Volume 6, no. 62, pages 225 to 241, (1961).

The construction details are given for two pairs of square coils, one pair of which is used to annul the vertical component of the earth's field and the other pair the horizontal component.

63. Kifer, I. I., "Testing of Ferromagnetic Materials," NASA STAR N64-31801.

This paper presents a discussion of ellipsoidal coils, cylindrical solenoids with several end windings, Helmholtz coils and the five square coil system of Rubens (101).

64. Koch, H. W., "Magnetic Field Calculations for Large Cross-Section Cloud-Chamber Coils," Journal of Applied Physics, Volume 21, pages 387 to 395, (May 1950).

This paper is concerned with two circular coils on the same axis with rectangular coil cross sections. The magnetic scalar potential and the field are obtained as integrals and expressed as series expansions. Curves of the axial field component on the coil axis and at various radii are presented. An experimental check with some coils that were constructed previously is also included.

65. Krisyuk, E. M. and Latyshev, G. D., "Compensation of the Earth's Magnetic Field," Bulletin of the Academy of Science of the USSR, Physics Series, Volume 22, pages 966 to 974, 1958 (English translation).

This paper presents the theory of coil systems and derives the condition for the Helmholtz system. Tables of the axial and radial components for a Helmholtz system at various points in space are presented. The conditions for the three coil systems of Barker (12) and Maxwell (76) are presented and tables of the axial and radial components of Barker's system are given. The four coil systems of Scott (119), Barker (12) and McKeehan (74) are discussed and tables of the axial and radial components of Barker's four-coil system are presented. A final table summarizes the results for all of the systems.

66. Kurti, N., "High Energy Coil Magnets," Physica, Volume 24, pages s123 to s124, (1958).

67. Lin, S. T. and Kaufmann, A. R., "Helmholtz Coils for Production of Powerful and Uniform Fields and Gradients," Review of Modern Physics, Volume 25, pages 182 to 190, (January 1953).

A Helmholtz coil of rectangular cross section was designed and constructed to produce both a high field intensity and good uniformity. Two methods of determining the separation were used, the first employed, interpolations and the second used Garrett's method (48). The effect of heating and the power requirement were investigated. A two coil magnetic gradient system was designed to increase the uniformity. The construction details of the coil system are presented along with test results of the uniformity. The result is a system with a field of 35,000 oersteds and having a uniformity of 10 parts per million, along an axial distance of 13 millimeters. The power consumption is 1700 kilowatts.

68. Lyddane, R. H. and Raurk, A. E., "Coils for the Production of a Uniform Magnetic Field," Review of Scientific Instruments, Volume 10, pages 253 to 257, (September 1939).

Two rectangular coils of rectangular cross section are used but the analysis is made on the basis that the coils extend to infinity in the direction perpendicular to the plane in which the field is calculated. Conditions are given for a uniform field and the effects of the coil ends are calculated. Bent-head coil ends (4), (53), (54) are considered and a set of contour curves is presented.

69. Murrmann, H. and Schwink, C., "Homogene Magnetfelder für optische Untersuchungen ferromagnetischer Proben," Zeitschrift für angewandte Physik, Volume 12, pages 155 to 157, (1960).

This paper discusses the case of two circular single layer coils on the same axis and determines the field variation along the axis. Similar results are obtained for rectangular coils along with the variation perpendicular to the axis. Also, a more homogeneous field is obtained by the use of two additional coils placed close to the gap between the two main coils. A homogeneity of better than one percent was obtained in a volume $20 \times 16 \times 4$ millimeters.

70. McKeehan, L. W., "Pendulum Magnetometer for Crystal Ferromagnetism," Review of Scientific Instruments, Volume 5, pages 265 to 268, (August 1934).

Gives the relationships for a Helmholtz coil pair with the optimum thickness and breadth relationship for a finite coil size (76). Also gives the same type of relationships when the coil pair is used as a gradient system.

71. Marmet, P., Morrison, J. D. and Swingler, D. L., "Neutralization of Fringing Magnetic Fields in the Ion Source of a Mass Spectrometer," Review of Scientific Instruments, Volume 33, pages 239 to 240, (February 1962).

Presents the use of two coils along with shielding by Mumetal to reduce the effect of the earth's field to less than 0.01 gauss.

72. Marshall, H. L. and Weaver, H. E., "Application of the Garrett Method to Calculation of Coil Geometries for Generating Homogeneous Magnetic Fields in Superconducting Solenoids," Journal of Applied Physics, Volume 34, pages 3175 to 3178, (November 1963).

Garrett's method (48) is used to calculate a sixth order solenoid with end-correcting coils. Over a volume 2 millimeters in diameter and 5 millimeters long a homogeneity of one part in a million has been achieved. Construction details are given.

73. Moore, A. D., "Half-Helmholtz Coils and a New Sectoral Coil Scheme," Institute of Electrical and Electronics Engineers Transactions, Volume 83, pages 314 to 317, (May 1964).

This paper describes a four-coil system. The upper two coils are in a gradient configuration and the lower two coils produce a field in the same direction as that of the uppermost of the two gradient coils. A uniform field is produced in a thin disk-shaped region just below the uppermost coil. No mathematical theory is given, but details of construction and test results are presented.

74. McKeehan, L. W., "Combinations of Circular Currents for Producing Uniform Magnetic Fields," Review of Scientific Instruments, Volume 7, pages 150 to 153, (March 1936).

This paper reviews the history of the various coil systems using combinations of current loops and lists a number of systems using three four, six and eight loops. It also introduces an eighth-order system with all coils lying on the surface of a sphere and with an approximate coil winding ratio of 22 to 15. A discussion of the effect of the finite dimensions of the cross-section of the coils is also presented.

75. McKeehan, L. W., "Gaugain-Helmholtz Coils for Uniform Magnetic Fields," Nature, Volume 133, pages 832 to 833, (June 2, 1934).

This paper provides information and references on the early history of coil systems during the Nineteenth Century.

76. Maxwell, J. C., "A Treatise on Electricity and Magnetism," Oxford University Press, Third Edition, Volume 2, page 306 and pages 356 to 359, (1892).

This paper discusses the use of a winding on a sphere to produce a uniform field inside the sphere. The Helmholtz coil-pair condition is obtained and the breadth to depth ratio of $\sqrt{31/36} = 0.928$ for a coil of rectangular cross section is specified so that a fourth order system is obtained. It is suggested that a rectangular cross-sectional shape is perferable to winding the coil upon a conical surface. A sixth order system of four coils is presented. For one special case the two center coils combine into one making a sixth-order three-coil system with a winding ratio of 64 to 49.

77. Marzetta, L. A., "Use of an Operational Amplifier with Helmholtz Coils for Reducing AC Induced Magnetic Fields," Review of Scientific Instruments, Volume 32, pages 1192 to 1195, (November 1961).

A solenoid is used to produce a steady field in the E-W direction and a Helmholtz coil with the same axis is driven by AC to reduce the AC field existing in the solenoid. Two other Helmholtz pairs are used to cancel the vertical and horizontal N-S components of the earth's field.

78. Miller, H., "IBM 7094 program for the Six-Coil Problem," NASA STAR N65-21653.

This paper discusses the solution of the six-coil system as given by Speiser and Waidelich (111). Includes the flow chart and the Fortran listing of the program. A numerical example is presented.

79. McKeehan, L. W., "Approximations to a Uniform Magnetic Field: Roots of Legendre Polynomials of Even Order," Review of Scientific Instruments, Volume 19, page 475, (1948).

The roots of some Legendre polynomials are recalculated and listed.

80. Montgomery, D. B. and Terrell, J., "Some Useful Information for the Design of Air-Core Solenoids," Report AFOSR-1525, National Magnet Laboratory, Massachusetts Institute of Technology, Cambridge, Mass., (1961).

81. Montgomery, D. B., "The Generation of High Magnetic Fields," Reports on Progress in Physics, Institute of Physics and the Physical Society, London, Volume,26, pages 69 to 104, (1963).

82. Montgomery, D. B., "Superconducting Magnets," Spectrum, Volume 1, pages 103 to 113, (February 1964).

83. McIntyre, D. E., "A Method for Generating a Cylindrical Homogeneous Magnetic Field Volume," NASA STAR N64-33834.

Two coil systems were considered, the first consisting of two circular coils with rectangular cross sections while the second was the same as the first with the addition of two more coils of zero cross section. The axial field is expressed as an integral equation which was solved numerically. The results are presented in tabular form.

84. Ness, N. F., "Magnetic Fields: Reasons for Simulation and Methods Available," NASA STAR N65-21654.

This paper derives the condition for uniformity of a Helmholtz coil-pair and discusses the percent homogeneity. It has a table listing some details of the Helmholtz, Fanselau (36), Braunbek (6), McKeehan (74), two square-coil (99) and Rubens (101), (136) coil systems. Some details of the twenty-foot diameter coil system facility at the Goddard Space Flight Center are presented.

85. Nelson, J. H., Hurwitz, L. and Knapp, D. G. "Magnetism of the Earth," pages 4 and 56 to 58, Publication No. 140-1, U. S. Coast and Geodetic Survey, (1962).

This article mentions the use of Helmholtz, Fanselau and Braunbek coil systems in a magnetometer and in a magnetic observatory. It also discusses the coil systems at the Fredericksburg Observatory.

86. Nettleton, H. R. and Sugden, S., "The Magnetic Susceptibility of Nickel Chloride," Proceedings of the Royal Society, Series A, Volume 173, pages 313 to 323, (1939).

Some constructional details are given for a Helmholtz coil pair used in susceptibility measurements.

87. Nageotte, E., "Sur un Système de Bobines Produisant un Champ Magnétique Uniforme pour une Chambre de Étroite," Comptes Rendus Academie des Sciences Paris, Volume 220, pages 557 to 559, (April 16, 1945).

This paper discusses two rectangular coils of rectangular cross section with the long sides bent to increase the uniformity in a somewhat similar manner to the bent-head coils of Harris (53), (54). Constructional details are presented along with some measurements of the uniformity of the field.

88. Neumann, H., "Herstellung räumlich und zeitlich konstanter Magnet felder für Messzwecke," Archiv für Technisches Messen, No. 113, Z 60-1, pages T128 to T129, (November 1940). No. 114, Z 60-2, pages T138 to T139, (December 1940). No. 115, Z 60-3, pages T14 to T15, (January 1941). No. 116, Z 60-4, pages T28 to T29, (February 1941).

Extensive review and bibliography of coil systems and solenoids for the production of uniform magnetic fields. The bibliography has more than 200 items although many items are concerned with iron cores or with very strong rather than uniform fields.

89. Nagaoka, H., "Magnetic Field of Circular Currents," Philosophical Magazine, Volume 41, pages 377 to 388, (1921).

This paper takes expressions for the axial a. . radial field components of a single circular coil in terms of elliptic functions and applies them to a Helmholtz coil-pair. Series expansions of the elliptic functions are used for computation purposes. Contour curves of the second, fourth and sixth order terms for both axial and radial field components of the single coil are shown. Similarly, contour curves of the fourth order terms for both axial and radial field components of the Helmholtz coil system are plotted.

90. Naumann, H., "Über eine Einrichtune zur Konstanthaltung magnetischer Felder," Zeitschrift für angewandte Physik, Volume 1, pages 260 to 264, (1949).

Helmholtz coil pairs are used to automatically reduce the earth's field to near zero.

91. Nelson, J. H., Gebhardt, R. E., and Bottum, J. L., "The Constant-Field Coil House at the Fredericksburg Magnetic Observatory," U. S. Department of Commerce, Coast and Geodetic Survey, Washington, D. C., (1959).

This paper describes the coil systems used at the observatory, in particular the Braunbek four-coil systems (6) for the vertical and magnetic meridian directions. A Helmholtz coil pair was used for the direction perpendicular to the magnetic meridian. Some details of the coil control circuits are given.

92. North, G. R., "Some Properties of Infinite Lumped Solenoids," Report ORNL-2975, Oak Ridge National Laboratory, Oak Ridge, Tennessee, (September 1961).

An infinite set of similar circular loops with a common axis and with equal axial spacing is analyzed by using Fourier series methods. The analysis is then extended to coils of rectangular cross section, and a numerical example is presented.

93. Peterson, M. C., "The Orbiting Geophysical Observatory Test Program," Institute of Electrical and Electronics Engineers Transactions, Volume AS-1, pages 362 to 373, (August 1963).

This paper indicates that the magnetic tests will be performed at the Space Technology Laboratories facility at Malibu. One figure shows an earlier spacecraft being tested in a coil system composed of two Helmholtz coil-pairs, one pair with a vertical and the other with a horizontal axis.

94. Pittman, M. E. and Waidelich, D. L., "Three and Four Coil Systems for Homogeneous Magnetic Fields," NASA Technical Note D-2095, January 1964. Also Institute of Electrical and Electronics Engineers Transactions on Aerospace, Volume AS-2, pages 36 to 45 (February 1964).

The complete range of solutions for the sixth-order, three-coil systems is given in the form of curves and tabulated results. This includes the Maxwell (76) and Barker (12) systems. A similar range of solutions for the eighth-order, four-coil systems is presented including the Braunbek (6), McKeehan (74) and Barker (12) systems. Information on other systems is given including the known six- and eight-coil systems.

95. Piziali, R. and Waidelich, D. L., "Large Magnetic Environments," pages C-21 to C-33 of "Final Report of the Goddard Summer Workshop Program in Measurement and Simulation of the Space Environments, Publication X-320-64-341, NASA Goddard Space Flight Center, Greenbelt, Maryland (1964).

Various coil systems were considered for use in producing a very large volume of uniform magnetic field. Four-dipole and four-circular-coil systems were tried first, and then two-square-coil and two-rectangular-coil systems were also looked at. Next, arrays of dipoles were considered. The system which is probably the best, appeared to be one composed of two dipoles and one very large coil, and a possible numerical design is presented.

96. Pontier, J., "Réalisation des Champs Magnétiques tres Homogènes pour l'Étude des Structures Fines en Résonance Paramagnétique Nucléaire," L'Onde Électrique, Volume 35, pages 481 to 484, (May 1955).

Two circular coils of rectangular cross section are used. The field is calculated by considering the cross section to be divided into a number of squares, each of whose field is calculated separately and then added. Constructional details are given and in a 0.5 cubic centimeter volume at the center of the system the homogeneity is approximately 10^{-8}.

97. Phillips, R. C. and Turner, E. B., "Construction and Calibration Techniques of High Frequency Magnetic Probes," NASA STAR N64-12006.

Helmholtz coil systems are used in the frequency range from one to fifty megacycles per second to provide a uniform field in which small magnetic probes can be calibrated.

98. Rinard, G., "Dimensions of Coil Systems Necessary to Produce Uniform Magnetic Field," NASA STAR N64-28216.

An analysis is made of an eighth-order four-coil system in which each coil is considered to be a solenoid with finite breadth and zero depth. This was done for both circular and square coils. Indications were found that a tenth-order system could be obtained by making the solenoids long enough. Calculated results for several eighth-order systems with solenoids of finite breadth and for both circular and square coils are given in a table.

99. Rankine, A. O., "The Measurement of Magnetic Field Distortion," Proceedings of the Physical Society, Volume 46, Part I, pages 1 to 15, (1934).

This paper gives the constants for fourth-order two-coils systems for coils of the following shapes: circular, octagonal, hexagonal, pentagonal, square, triangular and two infinitely long wires.

100. Ramsay, B. P. and Yost, F. L., "Facilities Provided by the Kensington Magmetic Laboratory," Journal of Geophysical Research, Volume 52, pages 357-to 367, (1947).

This article provides some details of the construction of a magnetic laboratory and of more than twenty large solenoids with auxiliary booster coils at the ends to increase the volume of uniformity. Three square Helmholtz coil-pairs on three mutually perpendicular axes were used to control each component of the field. The spacing was made slightly larger than optimum to increase the volume of uniformity; also contour curves of the field variation are shown. Construction details are also presented, and the design of a gradient coil system is given. The generators to drive the coil-systems and the spacings needed between coil systems are discussed.

101. Rubens, S. M., "Cube-Surface Coil for Producing a Uniform Magnetic Field," Review of Scientific Instruments, Volume 16, pages 243 to 245, (September 1945).

Five equally-spaced square coils located on the surface of a cube are used to produce a large uniform volume of field. The spacing between loops is one-quarter of the side of a loop and the turns vary as 19, 4, 10, 4, 19. To determine these values three points along the axis were chosen at which the fields were made equal. Another choice of the three points gave 71, 15, 37, 15, 71 as the turns values. Variation of the field both on and off the axis is presented as curves and a table. A table is also given of the dimensions in which the field varies less than 0.1, 1 and 5 percent and the comparison with a circular Helmholtz coil-pair is also provided.

102. Ramsey, N. F., "The Atomic Hydrogen Maser," Microwave Journal, Volume 6, pages 89 to 94, (March 1963).

A Helmholtz circular coil-pair on each of three mutually perpendicular axes was used to reduce the effect of the earth's field.

103. Rudzki, J. E. and Holt, E. H., "A 10 Kilogauss, Air Cooled Magnet for Plasma Research," NASA STAR N63-13070.

Two coils of shaped copper sheets soldered together were constructed and placed slightly farther apart than the optimum spacing. Along the axis

the field variation was 3 percent higher in the center of each coil than at the midway point. The radial variation was also measured and presented.

104. Raurk, A. E. and Peters, M. F., "Helmholtz Coils for Producing Uniform Magnetic Fields," Journal of the Optical Society of America and the Review of Scientific Instruments, Volume 13, pages 205 to 212, (August 1926).

Has approximate expressions for the axial and radial field components for two circular coils with a rectangular cross section. The breadth to depth ratio of the cross section is taken as 0.928 (76). Discusses the errors introduced by the finite size of the wire and a slight departure from the Helmholtz condition. Construction details of a Helmholtz coil-pair is given.

105. Rössiger, M., "Die Entstörung magnetischer Beobachtungsräume und erdmagnetischer Observatorien von Gleich strom-Magnetfeldern der elektrischen Bahnen," Die Naturwissenschaftern, Volume 30, pages 753 to 755, (1942). Also Volume 31, page 211, (1943).

A system containing a Helmholtz coil-pair was used to compensate automatically for variations in the earth's field caused by an electrified railroad.

106. Rinard, G. A., "The Use of Tschebyscheff Polynomials in Producing Uniform Magnetic Fields," Ph. D. Thesis, University of Missouri, Columbia, Mo., (August 1965).

Gives a process by which the axial field component of a coil system can be expressed as a series in a set of general polynomials. For a system of L coil-pairs then, the first 2L-1 coefficients are made zero to approximate a uniform field. Tschebyscheff polynomials are selected and applied to the particular case of a four-coil system (L = 2). A numerical method is used to solve the equations and computed parameters are presented in the tables for a number of possible systems. Error contour curves are given for volumes of uniform fields in the forms of a prolate spheroid and also an oblate spheroid.

107. Rosa, E. B., "A New Method for the Absolute Measurement of Resistance," Bulletin of the National Bureau of Standards, Volume 5, pages 499 to 509, (1909).

Two circular coils were used with a spacing slightly greater than the optimum spacing for a Helmholtz coil-pair. The shape of the flux lines is pictured.

108. Rao, D. M., "Two-Coil Systems for the Production of a Uniform Magnetic Field," Research Report, University of Missouri Engineering Library, Columbia, Mo., (June 1965).

Two circular coils of solenoidal and spiral shapes along with coils with rectangular cross section are used. The maximum field and the distance along the axis for a given percentage of field uniformity are obtained. The results are presented as curves and computed tables.

109. Shapiro, I. R., Stolarik, J. D. and Heppner, J. P., "The Vector Field Proton Magnetometer for IGY Satellite Ground Stations," NASA Technical Note D-358, (October 1960).

Two Helmholtz coil-pairs with perpendicular axes are used. The turns are wound on the surface of a cone to produce a fourth-order coil system. Some details of the construction are given.

110. Speiser, M. and Waidelich, D. L., "Square-Coil System for Production of Uniform Magnetic Field," NASA STAR N64-28214.

A computer solution of the general four-square-coil system was made for the eighth-order case. Tables are presented for systems with currents in the same direction and in opposite directions. A table of systems with integral turns ratio is given and field contours for a one-to-one turns ratio are given. Expressions for the field intensity at any point and approximate values of the inductance of the system are presented.

111. Speiser, M. and Waidelich, D. L., "Mathematical Analysis of Six Circular-Coil System for Production of Uniform Magnetic Field," NASA STAR N64-28215.

A computer solution of the general twelfth-order, six-circular-coil system was made. The computer program is given in the paper by Miller (78) and four solutions are presented in a table. The regions in which solutions were found are indicated by a graph.

112. Saperstone, S. H., "Four-Circular-Coil Systems," Final Report of the Goddard Summer Workship Program in Measurement and Simulation of the Space Environments, pages C-3 to C-6. Publication X-320-64-341, NASA Goddard Space Flight Center, Greenbelt, Maryland (1964).

The parameters for a number of eighth-order, four-circular-coil systems with the turns given as ratios of integers were obtained using a computer. In addition the field contours for the one-to-one and one-to-two turns ratio systems were plotted by the computer.

113. Saperstone, S. H., "Four-Octagonal-Coil Systems," Final Report of the Goddard Summer Workshop Program in Measurement and Simulation of the Space Environments, pages C-7 to C-16, Publication X-320-64-341, NASA Goddard Space Flight Center, Greenbelt, Maryland (1964).

The parameters for a number of eighth-order, four-octagonal-coil systems with the turns given as ratios of integers were obtained using a computer. The equations for the field intensity at any point are presented.

114. Saperstone, S. H., "Approximation Methods for Obtaining Uniform Magnetic Fields," Final Report of the Goddard Summer Workshop Program in Measurement and Simulation of the space Environment, pages C-17 to C-20. Publication X-320-64-341, NASA Goddard Space Flight Center, Greenbelt, Maryland, (1964).

Spacings greater than the usual Helmholtz spacing were investigated for two circular coils and a table of the results is given. The methods of least squares and of a Tchebyshev approximation were used in an attempt to extend this idea to multiple coil systems.

115. Shaw, A. E., "A New Precision Method for the Determination of e/m for Electrons," Physical Review, Volume 54, pages 193 to 209, (August 1, 1938).

Two circular coils of rectangular cross section were used in producing a uniform field in an annular region by reducing the spacing of the coils from that used in the usual Helmholtz configuration. The necessary spacing is calculated by making a gradient term approximately zero. The variation of the field is compared with that obtained from the Helmholtz spacing.

116. Smith, F. E., "On An Electromagnetic Method for tne Measurement of the Horizontal Intensity of the Earth's Magnetic Field," Philosophical Transactions of the Royal Society of London, Series A, Volume 223, pages 175 to 200, (1923).

Constructional details of a Helmholtz coil system with the wires wound on a marble cylinder are given. Great care was taken in making the coils and in determining the constants. Contour curves are shown for the axial intensity and for the ratio of transverse to axial intensities.

117. Sauerwein, K., "Ein Magnetfeld grosser Homogenität für die Wilsonsche Nebelkammer," Die Naturwissenschaften, Volume 30, pages 494 to 495, (1942).

A Helmholtz coil pair of finite cross section is considered for use in producing a uniform magnetic field for a Wilson cloud chamber. Expressions for the two field components are given and the relationship of the breadth to depth for a rectangular cross section of the coil is given as 1.097. Sauter (123) regards this value as a typographical error. Equipotential curves and calculated field variation curves are also shown. Construction details are given.

118. Sauzade, M. and Sallé, F., "Influence d'un Blindage Magnétique sur l'Homogénie d'un Champ Magnétique Créé par une Paire de Bobines en Position de Helmholtz. Applicàtion à un Spectrographe à Résonance Nucléaire à Champ Faible," Comptes Rendns, Academie des Sciences, Paris, Volume 259, pages 73 to 76, (July 6, 1964).

Presents the radial magnetic field of a Helmholtz coil-pair within a spherical shield. In the actual experiment the shielding is composed of two concentric cubes of thin mumetal which do not affect the homogeneity very much. The shielding does reduce the effect of external magnetic fields.

119. Scott, G. G., "Compensation of the Earth's Magnetic Field," Review of Scientific Instruments, Volume 28, pages 270 to 273, (1957).

The details of the sixth-order coil systems used by Scott (120) are presented. Constructional details are also given.

120. Scott, G. G., "The Kettering Magnetics Laboratory," Report GMR-291 General Motors Corporation Research Laboratories, (August 1960).

Contains a complete description of the laboratory. The axial and radial field components of a Helmholtz coil system are given in terms of polar coordinates about the center of the system. A second Helmholtz system coaxial with the first has its turns and radii adjusted along with a current in the opposite direction to that of the first system so a sixth order system results. Experimental measurements confirmed that a considerable improvement in homogeneity resulted over that of a single Helmholtz system. Two of the sixth-order systems were used to cancel the vertical and the north-south components of the earth's field, while a Helmholtz system was used to cancel the east-west component. Construction details are given.

121. Scott, G. G., "Review of Gyromagnetic Ratio Measurements," Review of Modern Physics, Volume 34, pages 102 to 109, (January 1962).

A brief description is given of the coil systems presented in more detail in (120).

122. Scott, G. G., "Kettering Magnetics Laboratory Installed at Oakland University," General Motors Engineering Journal, Volume 11, pages 66 to 69, Third Quarter (1964).

Description of equipment installed after the transfer of the laboratory to Oakland University. The coil system consists of three Helmholtz systems on three mutually perpendicular axes.

123. Sauter, A. and Sauter, F., "Die Erzengung von möglichst homogenen Magnetfeldern durch Stromsysteme," Zeitschrift für Physik, Volume 122, pages 120 to 136, (1944).

This paper reviews the theory of multiple coil systems and gives the expressions for the Helmholtz two-coil, the Fanselau, four-coil (36), the Maxwell three-coil (76) and the McKeehan four-coil (74) systems. This paper indicates that with four coils the best possible system is an eighth-order system. Outlines the general solution graphically for the four-coil system, although no calculated values are given. The case of four coils with equal radius solved later by Barker (12) and that of two coils in the same plane solved later by Waidelich and Pittman (131) are mentioned. Also given is a method of obtaining a fourth-order system for a circular coil of rectangular cross section. This paper shows that the Helmholtz condition should be satisfied and that the depth to breadth ratio should be $\sqrt{36/31} = 1.0776$ which agrees with Maxwell (76). The possibility of extending this result to a four-coil system is suggested.

124. Stephenson, R. J., Ference, M. and Shaw, A. E., "Production of a Uniform Magnetic Field," Physical Review, Volume 53, pages 208, (1938).

The uniformity of the field off the axis of two coils is increased by decreasing the spacing from the Helmholtz spacing. The coils constructed had a mean radius of 30 centimeters and a square cross section of 6 × 6 centimete

125. Schmidt, F. H., "A Uniform Solenoidal Magnetic Field Nuclear Spectrometer," Review of Scientific Instruments, Volume 23, pages 361 to 366, (1952).

126. Schlosser, E. G., "Ein abgeschirmtes homogenes Magnetfeld," Zeitschrift für Angewandte Physik, Volume 7, pages 59 to 61, (1955).

127. Sampson, W. B. and Kruger, P. G., "Superconducting Pseudo-Helmholtz Coil," Review of Scientific Instruments, Volume 36, pages 1081 to 1083, (August 1965).

Two solenoids wound with superconducting ribbon are placed close together to make a two-coil system. Curves of the axial field variation are presented for various spacings between the coils. Similar curves are given for the two coils connected in opposition to produce a gradient field.

128. Smyth, W. R., "Static and Dynamic Electricity," McGraw-Hill Book Co., New York, N. Y., pages 270 to 271, (1950).

129. Tseitlin, L. A., "On a Possible Method of Obtaining a Homogeneous Magnetic Field," Soviet Physics – Technical Physics, Volume 2, pages 2592, (1957).

Two ellipsoidal windings are used, one enclosing the other, to produce a very uniform magnetic field within the innermost winding. The outer winding makes it possible to reduce the external field practically to zero.

130. Usiskin, C. and Wilkes, R., "Measuring the Magnetic Dipole of a Satellite," Ballistic Missile and Aerospace Technology, Volume 2, pages 207 to 220, Academic Press, (1961).

Coils were wound on a 100-inch diameter sphere to cancel the earth's field inside the sphere. The sphere could be opened in the middle for access to the inside volume. Construction details are given.

131. Waidelich, D. L. and Pittman, M. E., "Magnetic Field Simulation," Final Report of the Goddard Summer Workshop Program in Simulation of Space Environment, pages 5-3 to 5-15, Publication X-320-62-193, NASA Goddard Space Flight Center, Greenbelt, Maryland, (1962).

Presents the theory of coil systems and a table giving data on the known coil systems including those having six- and eight-circular coils. A complete solution for the three-circular coil, sixth-order and four-circular coil eighth-order systems was made. A six-coil solution was attempted but not completed. The effect of tolerances in the system dimensions was discussed.

132. Waidelich, D. L. and Speiser, M., "Computer Analysis of Problems of Four Circular-Coil Systems," NASA STAR N64-28213.

The parameters of the Braunbek (6) four-circular coil system were calculated for greater accuracy. A table of the calculated parameters for eighth-order four-circular-coil systems with opposing currents is presented. Formulas for the field components at any point of a four-circular-coil system and a computed contour curve for the axial component of the field are given. A method along with an example of calculating the inductance of such a system is presented.

133. Waidelich, D. L., Yildirim, G., Speiser, M. and Gebhardt, R. E., "Four-Square-Coil Systems for the Magnetic Testing of Spacecraft," Institute of Electrical and Electronics Engineers Transactions on Aerospace, Volume AS-2, pages 1181 to 1186, (October 1964).

The complete solution of the general four-square-coil system is given including the cases of currents in the same and in opposing directions. The computed parameters are given in tables and curves and tables of parameters for the systems having integral turns-ratios are also presented. Curves of the flux contours for the system with unity turns-ratio are shown.

134. Wright, F., "Contour Curves for the Axial Magnetic Field of a Four-Dipole System," Final Report of the Goddard Summer Workshop Program in Measurement and Simulation of the Space Environments, page C-35, Publication X-320-64-341, NASA Goddard Space Flight Center, Greenbelt, Maryland, (1964).

Shows one contour curve of the axial field component of a four-dipole system with the theory given by Piziali and Waidelich (95).

135. Wolff, N., "The Application of Magnetometers to the Dynamic Simulation and Control of the Geomagnetic Environment," Institute of Electrical and Electronics Engineers International Convention Record, Volume 12, part 8, pages 149 to 155, (1964).

This paper shows the diagram of an apparatus to cancel the earth's field using a Helmholtz coil-pair. Also for a homogeneous magnetic field, the use of a solenoid, a single circular coil, a Helmholtz coil-pair and a five square coil system by Rubens (101) are discussed. Construction details and a picture of a coil system using three Helmholtz coil-pairs on mutually perpendicular axes are given.

136. Warburton, F. W., "Design of Coil Systems for Magnetic Field Control at the Naval Ordnance Laboratory," NavOrd 3768, U. S. Naval Ordnance Laboratory, (June 18, 1955) DDC AD-105-974.

This is a summary of the work on coil systems during and after World War II at the U. S. Naval Ordnance Laboratory. A bibliography of 44 items is included some of which are reports and memoranda of the Naval Ordnance Laboratory. Both circular and square coil-pairs have been used at the optimum and also wider and narrower distances. Homogeneous fields in annular regions (115), (34) are mentioned and some information on three coil systems (13), (74) is given. Spherical windings are discussed (19), (57) and Garrett's work (48) is mentioned.

A system of four square coils on the surface of a cube and with windings of 46, 43, 43 and 46 turns is described, and two systems of five square coils are given, one of which is that of Rubens (101) and the other has the turns ratios of 134, 44, 66, 44 and 134. A number of solenoid systems are presented and the bent-head coils are mentioned (4) (53) (54) (68). The characteristics of various gradient systems and the use of booster coils are also given. A discussion on the impedance of a coil system and a table indicating the volume of uniformity of a number of coil systems are added.

137. Weber, K. H. R., "Ein magnetischer Störfeldmesser und ein magnetisches Messtörfeld für elektroakustische Zwecke," Elektrotechnik, Volume 4, pages 216 to 218, (June 1950).

138. Watson, W., "A Determination of the Value of the Earth's Magnetic Field in International Units, and a Comparison of the Results with the Values given by the Kew Observatory Standard Instruments," Philosophical Transactions of the Royal Society of London, Series A, Volume 198, pages 431 to 462, (1902).

139. Watanabe, N., "Portable Electric Magnetometer," Proceedings of the Physics and Mathematics Society of Japan, Volume 2, pages 210 to 223, (November 1920).

140. Wind, G. and Schérer, M., "Bobines Productrices de Champ Magnetique pour Chambre de Wilson," Le Journal de Physique et le Radium, Volume 16, pages 92 to 96, (February 1955).

Two rectangular shaped coils of rectangular cross section are considered and the axial component of the field is computed for various points, both on and off the axis. A uniformity of one percent in a volume $60 \times 40 \times 12$ centimeters is obtained with a field intensity of 1953 oersteds and using a power of 53 kilowatts. Construction details and a set of contour curves are presented.

141. Williams, V. L. and Cain, J. C., "Homogeneous Magnetic Fields Produced by Circular Current Loops," Unpublished memorandum (1960) NASA Goddard Space Flight Center, Greenbelt, Maryland.

Discusses the eighth-order, four coil system of McKeehan (74) later solved again in a much more compact manner by Garrett (48). The coils of this system lie on the surface of a sphere. A general method for obtaining the parameters of a system with n loops lying on the surface of a sphere was developed. The positions of the loops are given by the roots of $P'_{n+1}(u) = 0$ and the weighting factor for the currents is $W = 2/\{[n+2]\,[n+1]\,[P_{n+1}(u_p)]^2\}$. With the aid of these formulas the twelfth-order six-coil and the sixteenth-order, eight-coil systems were calculated (131) (94). Tables giving parameters

for a number of coil systems are included along with curves showing the variation of the field both on and off the axis for a number of different coil systems.

142. Wendt, G., "Statische Felder und stationare Ströme" Handbuch der Physik, Edited by S. Glugge, Volume 16, Elektrische Felder und Wellen, pages 127 to 128, (1958), Springer Verlag, Berlin.

143. Yavor, S. Ya., Shpak, E. V. and Minina, R. M., "Cylindrical Magnetic Lenses with a Plane of Symmetry" Soviet Physics – Technical Physics, Volume 5, pages 369 to 377, (1960).

Two rectangular solenoids are used with their axes at right angles to compensate for the vertical and horizontal components of the earth's magnetic field. Some extra turns are placed on the end of one of the solenoids to decrease the inhomogeneity of the field. An expression is given for the field along the axis of a rectangular solenoid.

144. Yavor, S. Ya and Siladi, M., "Creation of a Homogeneous Magnetic Field with a Rectangular Solenoid of Finite Length," Instruments and Experimental Techniques, No. 5, pages 151 to 153, (January 1961).

The expressions for the field components of a rectangular solenoid are presented, and the variation of the field along the axis is given as a curve and in a table. The effect of putting end turns on the solenoid to decrease the inhomogeneity is also shown.

145. Younse, J. M. and Einspruch, N. G., "A Superconducting Helmholtz Pair," Institute of Electrical and Electronics Engineers Proceedings, Volume 52, pages 1238 to 1239, (October 1964).

The dimensions and construction of two circular coils of superconducting wire and with a rectangular cross section are given.

146. Yildirim, G., "Production of Uniform Magnetic Field by Using Square Coils," Research Report. University of Missouri Engineering Library, Columbia, Mo., (June 1963).

This paper presents the solution of the two and three square-coil systems as curves and tables. An analysis of the four square-coil system is given but no calculated values are included.

147. Zacharov, B., "A Method for the Design of Axially Symmetric Magnetic Fields for Image Tube Systems," Nuclear Instruments and Methods, Volume 7, pages 132 to 136, (1962).

The outline is given for a method of obtaining the currents in a series of equal coils with a common axis to produce any axial field variation desired such as a homogeneous field. An example is given in which the current densities in 19 coils were calculated to produce a field uniform to better than 0.1 percent over an axial distance of 44 centimeters. The measured value was about one percent and was caused by differences in the individual coils and misalignments.

148. von Zeipel, E., "An Automatic Wilson Cloud Chamber with New Illumination and a New Type of Coil for a Homogeneous Magnetic Field," Arkiv för Matematik, Astronomi och Fysik, Volume 30A, No. 2, pages 1 to 27, (1944).

The potential and field components of two coils with finite cross section were obtained. Two methods of increasing the uniformity of the field are given, one of which appears to lead to a fourth-order system and the other leads to an eighth order system. The constants for the system are worked out, and construction details are given. Expressions and tables of the field components are presented and a comparison is made with the usual Helmholtz system. For one percent deviation in the Helmholtz system, the axial distance is 0.38 of the radius, while in the present system, it is 0.71 of the average radius. A table of measured field strengths is given.

BLANK PAGE

Appendix A

NINETEENTH CENTURY REFERENCES

1. Gaugain, J.M., "Boussole des Tangentes Établie sur un Principe Nouveau d'Electrodynamique," Comptes Rendus, Volume 36, pages 191 to 193 (1853).

2. Bravais, A., "Zusatz su den elektro-dynamischen Versuchen des Hrn. Gaugain," Annalen der Physik und Chemie, Volume 88, pages 446 to 451 (1853).

3. Gaugain, J.M., "Tangentenbussole nach einem neuen elektrodynamischen Prinzip," Annalen der Physik und Chemie, Volume 88, pages 442 to 446, (1853).

4. Gaugain, J.M., "Memoire sur une Nouvelle Boussole des Tangentes," Annales de Chimie et de Physique, Volume 41, pages 66 to 71, (1854).

5. Wiedemann, G. H., "Die Lehre von der Elektricität," Braunschwieg, Volume 3, page 275 (1895).

6. "Die Fortschritte der Physik ion Jahre 1849," der physikalischen Gesellschaft zu Berlin, Verlag Georg Reimer, Berlin, Volume 5, page vii, (1853).

7. Mascart, E. and Joubert, J., "Electricité et Magnetisme," Volume I, pages 390 to 405; Volume II, pages 119 to 125, Paris, 1896-1897.

8. Maxwell, J. C., "Treatise on Electricity and Magnetism," Oxford University Press, Third Edition, Volume II, page 306 and pages 356 to 359, (1892).

9. Ellis, R., Clarkson, R. D., and Rainy, H., Proceedings of the Royal Society of Edinburg, Volume 3, page 523 (1885).

10. Houllevigue, L., Journal de Physique, Volume 7, pages 466 to 468, (1898).

11. Gray, A., "Absolute Measurements in Electricity and Magnetism," Macmillan, First edition 1893, Second edition, 1921, pages 210 to 221.

12. Weber, W., "Elektrodynamische Massbestimmungen," Leipzig, Hirzel, First edition 1852, Second edition 1867, Volume 1, page 548.

13. Ampère, A.M., Annales de Chimie et de Physique, Series II, Volume 15, pages 59 to 76, (1820).

14. Ampère, A.M., Annales de Chimie et de Physique, Series II, Volume 15, pages 170 to 218, (1820).

15. Ampère, A. M., Annales de Physique, Volume 67, pages 113 to 163, (1821).

16. Ampère, A.M., Annales de Physique, Volume 67, pages 225 to 258, (1821).

17. Bravais, A., Comptes Rendus, Volume 36, pages 194 to 197, (1853).

18. Riecke, E., "Über die Ersetzbarkeit geschlossener galvanischer Ströme durch magnetische Doppelflächen insbesondere über die Ersetzung eines eine beliebige Oberfläche spiralförming umziehenden Stromes durch eine räumliche Verteilung magnetischer Massen," Pogg. Ann., Volume 145, pages 218 to 234, (1872).

19. Stefan, J., "Zur Theorie der magnetischen Kräfte," Wien. Ber., Volume 69, (1874).

20. Riecke, E., "Über einen Tangentenmultiplicator," Wied. Ann., Volume 3, pages 36 to 53, (1878).

21. Riecke, E., "Nachtrag zu dem Aufsatz: Über einen Tangentenmultiplicator," Wied. Ann., Volume 4, pages 226 to 232, (1878).

22. "Collection de Mémoires Relatifs a la Physique," published by the Société Francaise de Physique, Volume 2, pages 7 to 53, (1885).

23. Neumann, F. E., "Vorlesungen über elektrische Ströme," B. G. Teubner, Leipzig, (1884), page 197.

24. Wild, H., Vierteljahrsschr. d. naturf. Ges. in Zürich, Volume 2, page 239, (1857).

25. Wiedemann, G. H., "Galvanismus," Second Edition, Volume 2, page 225.

26. Mascart, E., and Joubert, J., "Legons sur l'Electr. et le Magn.," Paris, Volume 1, page 546, (1882).

Appendix B

ADDITIONAL REFERENCES FROM THE TWENTIETH CENTURY

1. Auerbach, F., "Strommessung" in A. Winkelmann's "Handbuch der Physik," Second Edition, Volume 4, pages 263 to 264, Verlag J. A. Barth, Leipzig, (1905).

   This articles gives the expression for the axial field of a Helmholtz system and also a figure showing the flux lines and equipotentials of the system. Indicates that for a rectangular cross section, the breadth of the depth ratio should be 1.097. Sauter (123) seems to regard this figure as a typographical error. The work of Maxwell (76) on three and four coil systems and that of Riecke (88) on ellipsoidal coils are mentioned.

2. Bestelmeyer, A., "Berechnung, Herstellung und Messung eines homogenen Magnetfeldes," Physikalische Zeitschrift, Volume 12, pages 1107 to 1111, (1911).

   Experimental measurements on a 13-coil system are given.

3. Blackett, P. M. S., "A Negative Experiment Relating to Magnetism and the Earth's Rotation," Philosophical Transactions of the Royal Society of London, Series A, Volume 245, pages 309 to 370, (1952).

   Three Helmholtz systems with mutually orthogonal axes are used to compensate for the earth's field. Subsidiary coils could be used to compensate automatically for small changes in the earth's field.

4. Beller, W. S., "Attitude Control Testing Advanced by Sophisticated Magnetic Facility," Missiles and Rockets, Volume 17, No. 11, pages 22 to 23, (September 13, 1965).

   Information is presented on both the 20 foot and 40 foot coil systems at NASA's Goddard Space Flight Center (84) as well as the coil systems at the Fredericksburg Magnetic Observatory (91), at Malibu Beach, California (93), at the Naval Ordnance Laboratory (136) and at RCA, Princeton, New Jersey (130).

5. Chan, S. B., "Coil and Dipole Systems for a Large Volume of Uniform Magnetic Field," Research Report, University of Missouri Engineering Library, Columbia, Mo., (January 1965).

This paper considers the variation of the axial field component as the radial distance is varied for a coil pair with various spacings between the coils. The same study is then made for one large coil and two dipoles on the axis of the coil.

6. Goldstein, K. S., and Gray, G. R., "Handbook of Definitions, Terms, Formulas, Conversion Tables and Measurement Methods for Magnetic Phenomena," NASA STAR N64-28892.

   Some basic information is given on the following coil systems: Helmholtz, Fanselau (36) and Braunbek (6).

7. Morris, R. E., "Four-Cell Rocketborne Magnetometer," NASA STAR N63-16776.

   For testing the magnetometer a four-coil system was constructed using the McKeehan (74) design. The larger coils were about 79 inches in diameter and their axis could be aligned along the direction of the earth's magnetic field. Constructional details are given. The system was mounted in an aluminum trailer at the Kirtland Air Force Base, New Mexico. Also information about the current and regulator supply is presented. A separate Helmholtz coil system with coils eight feet in diameter was used to provide transient magnetic fields.

8. Taylor, D. B., "Synthesis of Arbitrary-Axial-Magnetic Fields with Standard Coils," DDC AD-616690.

   The field along the axis of a system containing an odd number of coils is treated by the method of least squares where the variables are the axial positions of the coils. A number of solutions are presented along with the calculated field variation both on and off the axis.