

## **General Disclaimer**

### **One or more of the Following Statements may affect this Document**

- This document has been reproduced from the best copy furnished by the organizational source. It is being released in the interest of making available as much information as possible.
- This document may contain data, which exceeds the sheet parameters. It was furnished in this condition by the organizational source and is the best copy available.
- This document may contain tone-on-tone or color graphs, charts and/or pictures, which have been reproduced in black and white.
- This document is paginated as submitted by the original source.
- Portions of this document are not fully legible due to the historical nature of some of the material. However, it is the best reproduction available from the original submission.

# ***Space Programs Summary No. 37-27, Volume IV***

**for the period April 1, 1964 to May 31, 1964**

**Supporting Research and Advanced Development**



**JET PROPULSION LABORATORY  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
PASADENA, CALIFORNIA**

**June 30, 1964**

**N67-36467**  
(ACCESSION NUMBER)  
**244**  
(PAGES)  
**CR-88106**  
(NASA CR OR TMX OR AD NUMBER)

(THRU)  
1  
(CODE)  
**34**  
(CATEGORY)



***Space Programs Summary No. 37-27, Volume IV***  
***for the period April 1, 1964 to May 31, 1964***

***Supporting Research and Advanced Development***

**JET PROPULSION LABORATORY  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
PASADENA, CALIFORNIA**

**June 30, 1964**

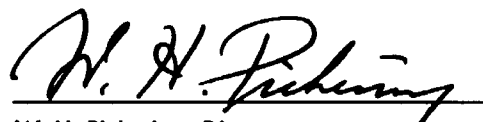
## Preface

The *Space Programs Summary* is a six volume, bimonthly publication designed to report on JPL space exploration programs, and related supporting research and advanced development projects. The subtitles of all volumes of the *Space Programs Summary* are:

- Vol. I. The Lunar Program (Confidential)
- Vol. II. The Planetary-Interplanetary Program (Confidential)
- Vol. III. The Deep Space Network (Unclassified)
- Vol. IV. Supporting Research and Advanced Development (Unclassified)
- Vol. V. Supporting Research and Advanced Development (Confidential)
- Vol. VI. Space Exploration Programs and Space Sciences (Unclassified)

The *Space Programs Summary*, Volume VI consists of an unclassified digest of appropriate material from Volumes I, II, and III and a reprint of the space science instrumentation studies of Volumes I and II.

Beginning with the *Space Programs Summary* 37-27 series, the Space Flight Operations Facility development progress, previously reported in Volume VI, will be reported in Volume III. The scope of Volume III is being expanded to incorporate the activities of the Deep Space Network.



W. H. Pickering, Director  
Jet Propulsion Laboratory

### Space Programs Summary No. 37-27, Volume IV

Copyright © 1964, Jet Propulsion Laboratory, California Institute of Technology  
Prepared under Contract No. NAS 7-100, National Aeronautics & Space Administration

## Contents

### SYSTEMS DIVISION

<b>I. Systems Analysis</b>	1
A. A Geometrical Interpretation of an Optimal Trajectory, Task No. 329-00401-1-3127 (129-04-01-01), C. G. Pfeiffer	1
B. Error Analysis of Multiple Planet Trajectories, Task No. 324-00602-2-3124 (124-04-01-02), F. M. Sturms, Jr.	6
References	14

### GUIDANCE AND CONTROL DIVISION

<b>II. Spacecraft Electrical Power</b>	15
A. Power Sources, Task No. 323-03302-2-3420 (123-33-02-01), P. Rouklove and D. W. Ritchie	15
Reference	19
<b>III. Guidance and Control Analysis and Integration</b>	20
A. Satellite Orbit Determination Accuracy Using On-Board Instruments, Task No. 325-01703-1-3430 (125-17-05-01), R. A. Virzi and R. V. Morris	20
<b>IV. Guidance and Control Research</b>	25
A. Optical Physics Research, Task No. 329-00201-1-3450 (129-02-05-01), T. Tomotsu and A. R. Johnston	25
References	28

### ENGINEERING MECHANICS DIVISION

<b>V. Materials</b>	29
A. Effect of Plastic Deformation on Preferred Orientation of Synthetic Carbons, Task No. 329-00310-1-3513 (129-03-15-04), D. B. Fischbech	29
B. Growth and Perfection of Crystals, Task No. 329-00310-1-3511 (129-03-15-04), P. J. Shlichta	31
C. Graphite, Task No. 329-00311-1-3513 (129-03-04-01), W. V. Kottlensky	32
D. Properties of Tungsten at High Temperatures, Task No. 329-00301-1-3511 (129-03-02-01), J. L. Taylor	34
References	34
<b>VI. Applied Mechanics</b>	36
A. Thermochemistry and Normal Shock Program, Task No. 324-00714-2-3540 (124-07-01-01), T. Horton	36
B. Thermal Scale Modeling: Arthur D. Little Contract, Task No. 324-00905-2-3540 (124-09-05-02), J. M. F. Vickers	38

## Contents (Cont'd)

C. Development of Radio Frequency Transparent Energy Absorbing Structural Elements,	
Task No. 384-10801-2-3540 (186-08-00-01), R. K. McFarland . . . . .	39
References . . . . .	40

## ENGINEERING FACILITIES DIVISION

VII. Technical Facilities . . . . .	43
A. An Intense Magnetic Field Facility,	
Task No. 324-00901-1-3280 (124-09-01-01), R. Rust . . . . .	43

## PROPULSION DIVISION

VIII. Solid Propellant Engineering . . . . .	49
A. Low-Pressure Combustion Studies,	
Task No. 328-03211-1-3810 (128-32-06-01), L. D. Strand . . . . .	49
B. Silicone Propellant Development for Heat Sterilization,	
Task No. 543-55819-2-3810 (807-58-00-19), L. C. Montgomery . . . . .	52
References . . . . .	53
IX. Polymer Research . . . . .	54
A. Compatibility of Polyethylene Oxide and Poly- $\beta$ -Vinyl-naphthalene,	
Task No. 329-00304-1-3820 (129-03-11-03), J. Moacanin and E. F. Cuddihy . . . . .	54
B. Kinetics of the Anionic Polymerization of Acenaphthylene,	
Task No. 329-00304-1-3820 (129-03-11-03), J. Moacanin and A. Rembaum . . . . .	59
C. Kinetics and Mechanism of the Cyanoethylation of 1- and 2-Propanols,	
Task Nos. 328-03204-1-3820 (128-32-99-03) and 329-00307-1-3260 (129-03-15-01), S. H. Kalfayan . . . . .	61
D. Polymer Chemistry: C <sup>14</sup> Distribution in Labeled 2, 4-Toluene Diisocyanate (TDI),	
Task Nos. 328-03204-1-3820 (128-32-99-03) and 329-00307-1-3820 (129-03-15-01), E. F. Kopka and J. D. Ingham . . . . .	64
References . . . . .	68
X. Liquid Propulsion . . . . .	70
A. Advanced Liquid Propulsion System,	
Task No. 328-03101-2-3841 (128-31-01-01), R. N. Porter, H. B. Stanford, and L. K. Jones;	
Task No. 328-03103-2-3841 (128-31-02-02), L. H. Peebles and W. H. Tyler;	
Task No. 328-03102-2-3841 (128-31-02-01), W. F. MacGlashan . . . . .	70
References . . . . .	74

## Contents (Cont'd)

<b>XI. Advanced Propulsion Engineering . . . . .</b>	<b>75</b>
A. Liquid MHD Power Conversion, Task No. 320-02704-2-3850 (120-27-06-03), D. Elliott, D. Carini, D. O'Connor, and E. Weinberg . . . . .	75
References . . . . .	82

## SPACE SCIENCES DIVISION

<b>XII. Space Instruments Development . . . . .</b>	<b>85</b>
A. A Gas Chromatograph for the Analysis of the Martian Atmosphere, Task No. 384-26202-2-3222 (185-62-00-02), C. Josias, L. Bowman, and H. Mertz . . . . .	85
References . . . . .	91
<b>XIII. Space Instruments Systems . . . . .</b>	<b>92</b>
A. Carrier Collection and Spectral Response of Radiation Detectors, Task No. 384-16213-2-3242 (186-62-00-13), R. Y. Wong . . . . .	92
B. A New Head for Recording Twenty-Four Tracks on Quarter-Inch Magnetic Tape, Task No. 325-02402-2-3245 (125-24-01-02), W. G. Clement . . . . .	97
C. A Multi-Instrument Data Conditioning, Storage and Retrieval System for Planetary Spacecraft, Task No. 325-02402-2-3245 (125-24-01-02), D. W. Slaughter . . . . .	101
References . . . . .	107
<b>XIV. Chemistry . . . . .</b>	<b>109</b>
A. Effects of Ultrahigh Vacuum on Three Types of Microorganisms, Task No. 386-15502-1-3260 (189-55-00-02), P. J. Geiger, F. A. Morelli, and H. P. Conrow . . . . .	109
B. Fluorometric Measurements of Growth: III. The Interference of Soil with the Fluorescence of Proteins in a Neutral Aqueous Solution, Task No. 386-15502-1-3260 (189-55-00-02), J. H. Rho . . . . .	116
C. Many-Electron-Theory Ab-Initio Calculation for the Be Atom: I. Numerical Results and Discussion, Task No. 329-00208-1-3260 (129-02-07-01), M. Geller, H. S. Taylor, and H. B. Levine . . . . .	122
References . . . . .	125
<b>XV. Fluid Physics . . . . .</b>	<b>127</b>
A. The Structure of a Supersonic, Low-Density, Free Jet, Task No. 329-00105-1-3270 (129-01-10-01), H. I. Ashkenas . . . . .	127

## Contents (Cont'd)

B. The Flow Between a Rotating Disk and a Coaxial, Stationary Disk, <i>Task No. 329-00108-1-3270 (129-01-05-02), T. Maxworthy</i>	129
C. The Stability of the Compressible Laminar Boundary Layer According to a Direct Numerical Solution, Part V, <i>Task No. 329-00102-1-3270 (129-01-09-01), L. M. Mack</i>	130
References	133

<b>XVI. Physics</b>	135
A. Rocket Measurement of Nitric Oxide in the Upper Atmosphere, <i>Task No. 385-04701-1-3280 (188-47-00-01), C. A. Barth</i>	135
B. Improved Efficiency Calculations for the Fission-Electric Cell, <i>Task No. 322-02801-1-3280 (122-28-02-01), J. L. Shapiro</i>	138
C. Local Measurement of Gravitational Fields, <i>Task No. 329-00209-1-3280 (129-02-07-02), H. Wahlquist and F. Estabrook</i>	139
D. Derivation of the Master Equation with the Wigner- Weisskopf Formalism for Radiation Damping, <i>Task No. 329-00209-1-3280 (129-02-07-02), O. von Roos</i>	140
E. Internal Automorphisms of a Two-Particle System, <i>Task No. 329-00209-1-3280 (129-02-07-02), J. S. Zmuidzinas</i>	141
F. Two Variational Principles for London's Equation, <i>Task No. 329-00209-1-3280 (129-02-07-02), M. M. Saffren</i>	142
References	145

## TELECOMMUNICATIONS DIVISION

<b>XVII. Communications Elements Research</b>	147
A. Low Noise Amplifiers, <i>Task No. 350-22207-1-3332 (150-22-02-23), C. T. Stelzried, W. V. T. Rusch, R. Brantner, and R. Clauss</i>	147
B. Optical Communications Components, <i>Task No. 325-02201-1-3335 (125-22-02-01), J. Siddoway and J. Maserjian</i>	152
C. Antennas for Space Communications, <i>Task No. 350-22206-1-3331 (150-22-02-21), P. Potter</i>	155
D. RF Techniques, <i>Task No. 350-22211-1-3336 (150-22-02-31), T. Otoshi and D. White</i>	159
References	167
<b>XVIII. Communication System Development</b>	169
A. Signal-to-Noise Monitoring, <i>Task No. 388-60613-2-3340 (684-06-02-01), C. E. Gilchrist</i>	169
References	184

## Contents (Cont'd)

<b>XIX. Communications Systems Research</b>	<b>185</b>
<b>A. Coding Theory,</b>	
Task No. 350-22204-2-3310 (150-22-02-07), R. C. Tittsworth, G. Solomon, and E. C. Posner	185
<b>B. Detection and Filter Theory,</b>	
Task No. 350-22204-2-3310 (150-22-02-07), W. B. Kendall, W. C. Lindsey, J. J. Stiffler, R. C. Tittsworth, and E. A. Yerman	195
<b>C. Information Processing,</b>	
Task No. 350-22204-2-3310 (150-22-02-07), E. C. Posner and I. Eisenberger	222
<b>References</b>	<b>234</b>

## SYSTEMS DIVISION

### I. Systems Analysis

#### A. A Geometrical Interpretation of an Optimal Trajectory

C. G. Pfeiffer

Recent work in the calculus of variations has been directed toward the optimal final-value control problem, which can be described as the task of controlling the trajectory of a point mass so as to achieve a desired final state while at the same time minimizing (or maximizing) a given function of the final state variables. In Ref. 1 some properties of an optimal trajectory were discussed, leading to the development of necessary and sufficient conditions which must be satisfied if the trajectory is to be minimizing. In this paper we shall augment that analysis by describing necessary and sufficient conditions from a geometrical (functional analysis) point of view, and discuss the concept of abnormality.

The problem analyzed here will be somewhat simplified in order to facilitate the exposition. Thus, it will be assumed that the final time is fixed, that there is only one control variable to consider, and that the control

variable is not bounded. In the perturbation analysis we shall limit ourselves to considering only small deviations in the control variable. By this restriction we are dealing with what is called a "weak" variation in the terminology of the classical calculus of variations.

The notation used is as follows: The independent variable is  $t$ , which may be thought of as time;  $T$  is the (fixed) final time; other capital letters are matrices; column vectors are denoted by a bar ( $\bar{\phantom{x}}$ ) over a small letter; and the transpose of a vector or matrix is indicated by the superscript  $'$ . The notation  $(t)$  will occasionally be omitted in equations in order to simplify the notation.

##### 1. Formulation of the Problem

Let the rate of change of the state of the system be given by

$$\frac{d}{dt} \bar{x} = \bar{f}(\bar{x}, y, t) \quad (1)$$

where

$$\bar{x}' = (x_1, x_2, \dots, x_n)$$



is the state vector of the system (such as the position and velocity coordinates of a rocket vehicle);  $y(t)$  is the continuous control function which is to be determined (such as the steering angle of the rocket vehicle); and  $\bar{f}(\bar{x}, y, t)$  is a known vector function of  $\bar{x}$  and  $y$ , and is differentiable up to third order with respect to these quantities. Given some initial condition  $\bar{x}(0)$ , the problem is to choose the control variable  $y(t)$  in such a way as to minimize at some fixed final time  $T$  the function  $\beta_0(\bar{x})$ , subject to the  $r$  distinct constraints  $\beta_i(\bar{x}) = 0$ . The  $\beta_i(\bar{x})$  will be called the "boundary functions," where  $i = 0, 1, \dots, r$ . Without loss of generality it can be assumed that the  $\beta_i(\bar{x})$  are linear functions of the final state  $\bar{x}$ . This is the well-known Mayer formulation of the optimization problem in the calculus of variations, which includes the case of minimizing a functional of  $\bar{x}(t)$  and  $y(t)$  and/or of applying integral constraints.

A geometrical interpretation of the optimization problem is afforded by constructing an  $r+1$  dimensional boundary function space (Fig. 1), where the  $i$ th coordinate of a point in the space is given by

$$\delta\beta_i(\bar{x}) = \beta_i(\bar{x}) - \beta_i(\bar{x})_{\text{nominal}} \\ \text{for } i = 0, 1, \dots, r$$

The  $\beta_i(\bar{x})_{\text{nominal}}$  is the value of the  $i$ th boundary function on a nominal (reference) trajectory, attained by applying the nominal control function. The  $\delta\beta_i$  is the variation in this boundary function resulting from perturbing the control by some arbitrary  $\delta y(t)$ , where

$$\delta y(t) = y(t) - y(t)_{\text{nominal}}$$

Thus a point described by  $\delta\bar{\beta}$  represents the mapping of the function  $\delta y(t)$  to the boundary function space, and the coordinates  $\delta\beta_i$  are functionals of  $\delta y(t)$ . The analysis of optimal trajectories from the geometrical point

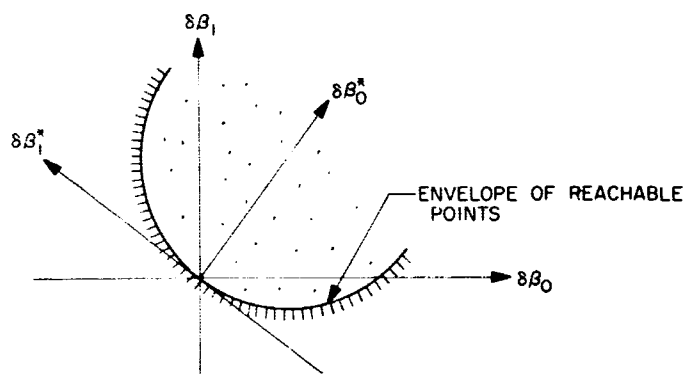


Fig. 1. The boundary function space

of view consists of studying the envelope of points reachable by applying arbitrary control variations, as will be described below.

## 2. The Variation of the Boundary Functions

We suppose that a nominal trajectory (not necessarily optimal) exists, and construct a Taylor series expansion of Eq. (1) in terms of the  $\delta x_i$  and  $\delta y$ . It is shown in Ref. 1 that variation in the  $i$ th boundary function can be written

$$\delta\beta_i = \int_0^T \eta_i(t) \delta y(t) dt + \frac{1}{2} \int_0^T \xi_i(t) \delta y^2(t) dt \\ + \frac{1}{2} \int_0^T \int_0^T k_i(t, \tau) \delta y(t) \delta y(\tau) dt d\tau \\ + \text{higher order terms} \quad \text{for } i = 0, 1, \dots, r$$

where  $\eta_i(t)$  is called the influence (Green's) function,  $\xi_i(t)$  is called the weighting function, and  $k_i(t, \tau)$  is a real symmetric kernel. The term containing  $\delta y(t)$  is said to be the first variation of  $\beta_i$ , and the terms containing  $\delta y^2(t)$  (with the factor  $\frac{1}{2}$  deleted) are said to be the second variation of  $\delta\beta_i$ . If we consider only small (weak) variations  $\delta y(t)$ , the higher order terms in Eq. (2) may be dropped and the resulting expression is then a second-order functional expansion of  $\beta_i$ .

## 3. The First Necessary Condition

Suppose the nominal trajectory pretends to be an optimal trajectory, and we seek to test it by applying the special control perturbation

$$\delta y(t) = \sum_{i=0}^r \eta_i(t) \epsilon_i = \bar{\eta}'(t) \bar{\epsilon} \quad (3)$$

where the  $\epsilon_i$  are arbitrary numbers. We take the  $\epsilon_i$  to be very small, and consider

$$\delta\bar{\beta} \text{ (first variation)} = \int_0^T [\bar{\eta}(t) \bar{\eta}'(t)] \bar{\epsilon} dt = P \bar{\epsilon} \quad (4)$$

where  $P$  is an  $r+1$  by  $r+1$  matrix with elements

$$p_{ij} = \int_0^T \eta_i(t) \eta_j(t) dt \quad (5)$$

If the trajectory is optimal (minimizing) it is necessary that the  $P$  matrix be singular, for otherwise it would be possible to invert  $P$  and find an  $\bar{\epsilon}$  which would attain a

smaller value of  $\beta_0$  while preserving the conditions  $\delta\beta_i = 0$  for  $i = 1, 2, \dots, r$ . Thus we have

**The First Necessary Condition for optimality.** The influence functions must be linearly dependent, that is,

$$\sum_{i=0}^r \eta_i(t) v_i = 0 \quad (6)$$

where the  $v_i$  are constants not all equal to zero.

This statement is verified by recognizing that  $P$  is a real symmetric matrix, and hence there exists an orthonormal  $r+1$  by  $r+1$  matrix  $L$  such that

$$LPL' = P^* = \text{diagonal} \quad (7)$$

where the elements of  $P^*$  are given by

$$p_{ij}^* = \int_0^T \eta_i^*(t) \eta_j^*(t) dt \quad \begin{cases} = 0 & \text{if } i \neq j \\ \geq 0 & \text{if } i = j \end{cases} \quad (8)$$

and

$$\bar{\eta}^*(t) = L\bar{\eta}(t) \quad (9)$$

If  $P$  is singular then

$$|P| = |P^*| = (p_{00}^*) (p_{11}^*) \cdots (p_{rr}^*) = 0 \quad (10)$$

where  $|\cdots|$  indicates the determinant of a matrix. Thus at least one  $p_{ii}^* = 0$ , and from Eq. (8) we conclude that this is true if and only if the corresponding  $\eta_i^*(t) = 0$  over the interval  $(0, T)$ . Eq. (6) is now verified, where the  $v_i$  are elements of a row of the  $L$  matrix.

It is interesting to note that the obvious statement

$$|P| = |P^*| \geq 0 \quad (11)$$

is a generalization of the well-known Schwarz inequality, as may be verified by carrying out the expression for  $|P|$  for the case of only two influence functions.

#### 4. Abnormality and Uncontrollability

The first necessary condition states that the rank of the  $r+1$  by  $r+1$  matrix  $P$  must be at most equal to  $r$ . If

the rank is less than  $r$  the trajectory is said to be abnormal (Ref. 2, p. 210), where

**Definition:** A trajectory is said to be "abnormal of order  $q$ " if there are  $q+1$  elements  $p_{ii}^*$  of Eq. (10) equal to zero.<sup>1</sup>

The abnormal case presents difficulty for the analysis to follow, so we shall henceforth assume  $q = 0$ .

Abnormality is related to the concept of uncontrollability introduced by Kalman in Ref. 3. Treating only the linear perturbation equations (the first variation) and with the final time fixed, we say that an initial state deviation  $\delta\bar{x}(0)$  is controllable if there exists a control signal  $\delta y(t)$  defined over the interval  $(0, T)$  such that  $\delta\bar{x}(T) = 0$ . Consider a  $\delta\bar{x}(0)$  and some  $\beta_i^*$ , where

$$[\text{grad } \beta_i^*] \left[ \frac{\partial \bar{x}(T)}{\partial \bar{x}(0)} \right] \delta\bar{x}(0) \neq 0 \quad (12)$$

The  $\delta\bar{x}(0)$  is uncontrollable if  $\eta_i^*(t) = 0$ , for then the control  $\delta y(t)$  has no (first order) effect on the projection of  $\delta\bar{x}(T)$  onto the  $\beta_i^*$  direction, and the final state cannot be driven to zero. Thus it can be said that an optimal trajectory is uncontrollable of order  $q+1$  if it is abnormal of order  $q$ .

#### 5. The Second Necessary Condition

Let us rotate the boundary function coordinate system (Fig. 1) by the transformation  $L$  [Eq. (7)], to get

$$\delta\bar{\beta}^* = L\delta\bar{\beta} \quad (13)$$

We rearrange the functions  $\delta\beta_i^*$  such that (only)  $\delta\beta_0^*$  has an influence function  $\eta_0^*(t) = 0$  (recall that we are assuming the trajectory to be normal). Thus

$$\begin{aligned} \delta\beta_0^* &= \frac{1}{2} \int_0^T \xi_i^*(t) \delta y^2(t) dt \\ &+ \frac{1}{2} \int_0^T \int_0^T k_0^*(t, \tau) \delta y(t) \delta y(\tau) dt d\tau \\ &+ \text{higher order terms} \end{aligned} \quad (14)$$

<sup>1</sup>It has recently come to our attention that the relationship between the rank of the  $P$  matrix and the abnormality of the trajectory has also been noted by W. G. Melbourne and F. D. Faulkner. The notion arose during Melbourne's review of an unpublished paper by Faulkner dealing with techniques for computing optimal trajectories (presented at UCLA Conference on Computing Methods in Optimization Problems, January 30-31, 1964).

$$\delta\beta_i^* = \int_0^T \eta_i^*(t) \delta y(t) dt \quad i = 1, \dots, r$$

+ higher order terms (15)

Assuming that the projection of the  $\beta_0^*$  axis on the  $\beta_0$  axis is positive (otherwise we choose  $-\beta_0^*$ ), we have

*The Second Necessary Condition for optimality (minimality).* The zeroth weighting function must be greater than or equal to zero, that is

$$\xi_0^*(t) \geq 0 \quad (16)$$

This is the classical Legendre condition. The statement is verified heuristically by imagining a positive impulse of control at the time  $t$ , immediately followed by an identical negative impulse. In this case the first term on the right-hand side of Eq. (14) dominates all other terms in Eqs. (14) and (15), and the second necessary condition is established.

For the following discussion it will be convenient to consider only "strictly nonsingular" extremals, where

*Definition:* A strictly nonsingular extremal is characterized by  $\xi_0^*(t) > 0$  everywhere in  $(0, T)$ .

The more general case suggested by Eq. (16) will be discussed in a future work.

## 6. The Reachable Envelope for Strictly Nonsingular Extremals

Let us simplify the subsequent analysis by assuming that there is only one boundary condition to be met (the more general case is developed in Ref. 4), and introduce the control variation

$$\delta \hat{y}(t) = [\xi_0^*(t)]^{-1/2} \delta y(t) \quad (17)$$

The corresponding influence function becomes

$$\hat{\eta}(t) = [\xi_0^*(t)]^{-1/2} \eta_1^*(t) \quad (18)$$

and the symmetric kernel becomes

$$\hat{k}(t, \tau) = [\xi_0^*(t) \xi_0^*(\tau)]^{-1/2} k_0^*(t, \tau) \quad (19)$$

The  $\Lambda$  quantities are now substituted into Eqs. (14) and (15) and the first term on the right-hand side of Eq. (14) is simplified by having  $\hat{\xi}_0(t) = 1$ .

Seeking an equation for the envelope of points reachable for small variations  $\delta \hat{y}(t)$  (Fig. 1), we delete the higher order terms in Eqs. (14) and (15) and find the extremal value of  $\delta\beta_0^*$ , subject to the condition  $\delta\beta_1^* = b$ . The  $b$  is an arbitrary parameter which will be varied to trace out the envelope. Thus we extremalize with respect to  $\delta \hat{y}(t)$  the functional

$$p = \delta\beta_0^* + \nu(\delta\beta_1^* - b) \quad (20)$$

where the  $\nu$  is a Lagrange multiplier. It follows that

$$0 = \delta \hat{y}(t) + \sum_{i=1}^{\infty} \frac{c_i \phi_i(t)}{\omega_i} + \nu \hat{\eta}(t) \quad (21)$$

where  $\omega_i$  are characteristic constants and  $\phi_i(t)$  are the characteristic functions (Refs. 5, 6) of the symmetric kernel  $\hat{k}(t, \tau)$ , and

$$c_i = \int_0^T \phi_i(t) \delta \hat{y}(t) dt \quad (22)$$

Multiplying Eq. (21) by  $\delta \hat{y}(t)$  and integrating, we have

$$0 = 2\delta\beta_0^* + \nu\delta\beta_1^* \quad (23)$$

Multiplying Eq. (21) by  $\phi_i(t)$  and integrating, we have

$$0 = c_i(1 + 1/\omega_i) + \nu d_i \quad (24)$$

where

$$d_i = \int_0^T \phi_i(t) \hat{\eta}(t) dt \quad (25)$$

Multiplying Eq. (21) by  $\hat{\eta}(t)$  and integrating, we have

$$0 = \delta\beta_1^* + \sum_{i=1}^{\infty} \frac{c_i d_i}{\omega_i} + \nu \int_0^T \hat{\eta}^2(t) dt \quad (26)$$

If there are no  $\omega_i = -1$ , Eqs. (23) to (26) can be combined to yield

$$\delta\beta_0^* (\text{extremum}) = \frac{1}{2} \left[ \frac{(\delta\beta_1^*)^2}{\rho} \right] \quad (27)$$

where  $\delta\beta_1^* = b$  is the specified value, and

$$\rho = \int_0^T \hat{\eta}^2(t) dt - \sum_{i=1}^{\infty} \frac{d_i^2}{1 + \omega_i} \quad (28)$$

Considering  $\delta\beta_1^*$  as a parameter which takes on all real values, Eq. (27) defines the parabola which either envelopes the reachable points or represents the locus of inflection points. The  $\rho$  is the radius of curvature at the origin.

We assert without proof the following

**Theorem:** Suppose there are no  $\omega_i = -1$ , then a necessary and sufficient condition that the optimal trajectory be minimizing is that Eq. (27) define a lower bound of the reachable points. Thus

$$\delta\beta_0[\text{arbitrary } \delta\hat{y}(t)] - \delta\beta_0[\text{corresponding extremal value}]$$

$$= \frac{1}{2} \left[ \int_0^T \delta\hat{y}^2(t) dt + \int_0^T \int_0^T \hat{m}(t, \tau) \delta\hat{y}(t) \delta\hat{y}(\tau) dt d\tau \right]$$

$$= \frac{1}{2} \left[ \int_0^T \delta\hat{y}^2(t) dt + \sum_{i=1}^{\infty} \hat{c}_i^2 (\hat{\omega}_i)^{-1} \right] \geq 0 \quad (29)$$

where the modified kernel is

$$\hat{m}(t, \tau) = \hat{k}(t, \tau) - \frac{1}{\rho} [\hat{\eta}(t) \hat{\eta}(\tau)] \quad (30)$$

The  $\hat{\omega}_i$  are the characteristic constants of this kernel, and  $\hat{c}_i$  is defined analogously to Eq. (22). From Eq. (29) it can be shown that  $\hat{m}(t, \tau)$  can have no characteristic constants between 0 and -1.

The sufficiency statement of the theorem follows immediately, but the necessary condition is not obvious. We shall discuss this point more fully in work to be reported in this journal.

## 7. Geodesics on a Sphere

To apply the above results, let us examine the curve of minimum arc length on the surface of a sphere which connects the point (0,0) with the point (T,0), where the coordinates are, respectively, longitude and latitude (Fig. 2). Letting the coordinate  $t$  be the independent variable, the derivative of arc length is

$$\frac{dx_0}{dt} = (y^2 + \cos^2 x_1)^{1/2} = f_0(x_1, y) \quad (31)$$

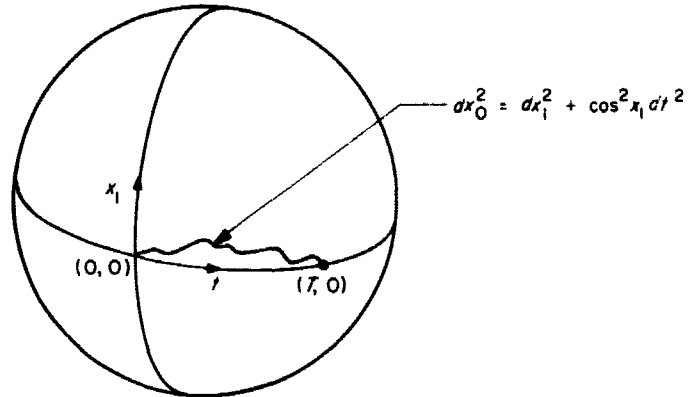


Fig. 2. Motion on the unit sphere

where the control function is

$$\frac{dx_1}{dt} = y = f_1(y) \quad (32)$$

The nominal trajectory is obviously the great circle arc between the given initial and final points.

Applying the theory described above we find that  $\hat{\eta}(t) = 1$  and the symmetric kernel is

$$\hat{k}(t, \tau) = \begin{cases} (t - T) & \text{for } t \geq \tau \\ (\tau - T) & \text{for } \tau > t \end{cases} \quad (33)$$

The characteristic constants are

$$\omega_i = -\frac{\pi^2}{4T^2} (2i - 1)^2 \quad (34)$$

and the characteristic functions are

$$\phi_i(t) = \left(\frac{2}{T}\right)^{1/2} \cos(-\omega_i)^{1/2} t \quad (35)$$

Thus (see Titchmarsh, *The Theory of Functions*)

$$\begin{aligned} \rho(T) &= T \left\{ 1 - \frac{8}{\pi^2} \sum_{i=1}^{\infty} [2i - 1]^{-2} \left[ 1 - \left(\frac{\pi^2}{4T^2}\right) (2i - 1)^2 \right]^{-1} \right\} \\ &= \tan T \end{aligned} \quad (36)$$

The radius of curvature  $\rho(T)$  is plotted versus  $T$  in Fig. 3. The modified kernel is

$$\hat{m}(t, \tau) = \hat{k}(t, \tau) - \frac{1}{\rho} \quad (37)$$

which has characteristic constants  $\hat{\omega}_i$  given by

$$(-\hat{\omega}_i)^{1/2} \tan(-\hat{\omega}_i)^{1/2} T = \rho(T) \quad (38)$$

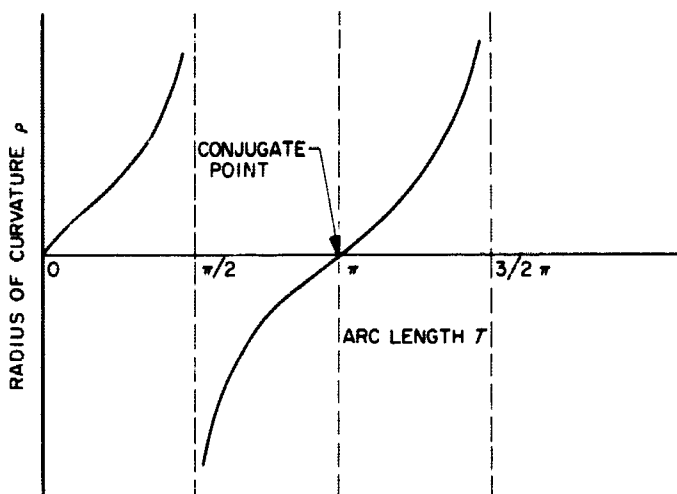


Fig. 3. Radius of curvature versus arc length

Upon applying the theorem we conclude that the trajectory is minimizing if and only if  $T$  is less than  $\pi$ . Notice that  $\rho(\pi) = 0$  corresponds to the "conjugate point" (Ref. 2).

### 8. Discussion

We have presented here a new approach to the analysis of optimal trajectories which extends and augments the material of Ref. 1. Still to be done is a rigorous establishment of the theorem stated in Sect. 6, and a treatment of abnormal points and a more general class of extremals. Such work will be reported in a future volume of this journal.

## B. Error Analysis of Multiple Planet Trajectories

F. M. Sturms, Jr.

It has been proposed (Refs. 7, 8) that trajectories using the gravitational perturbations of intermediate planets be used in order to achieve exploration of the solar system with launch velocity requirements significantly lower than are needed for direct trajectories. Typical uses of these lower velocity requirements might be to travel to remote planets with smaller launch vehicles or to extend the launch period available for the closer planets. This report gives the results of an initial feasibility study, and derives equations for examining the

effects of errors associated with missions to a target planet along a trajectory passing near an intermediate planet. Numerical results are presented for two typical missions in 1970: an Earth-Venus-Mercury trajectory and an Earth-Venus-Mars trajectory. Both of the selected trajectories use the 1970 Venus launch opportunity.

### 1. Intermediate Planet Hyperbola

Using the heliocentric conic trajectory program (HECON), described in Ref. 9, the Earth-intermediate planet, and the intermediate planet-target planet trajectories are chosen such that the asymptotic arrival and departure speeds at the intermediate planet are equal.

$$V_{h_2} = V_{h_3} \quad (1)$$

The right ascension and declination of the arrival and departure asymptotes are then used to determine the components of unit vectors along the asymptotes.

$$\hat{S} = (\cos \phi \cos \Theta, \cos \phi \sin \Theta, \sin \phi) \quad (2)$$

At this point, a rotation may be performed to express the components in an ecliptic referenced system

$$\begin{aligned} \hat{S}' &= (\cos \beta \cos \lambda, \cos \beta \sin \lambda, \sin \beta) \\ &= \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \epsilon & \sin \epsilon \\ 0 & -\sin \epsilon & \cos \epsilon \end{pmatrix} \hat{S} \end{aligned} \quad (3)$$

and wherever  $\hat{S}$  appears in the following equations, either or both the equatorial or ecliptic referenced vector may be used, depending on the requirements of the user. The semimajor axis of the encounter hyperbola is given by

$$a = \frac{-\mu}{V_h^2} \quad (4)$$

A unit vector normal to the encounter hyperbola trajectory plane is given by

$$\hat{W} = \frac{\hat{S}_2 \times \hat{S}_3}{|\hat{S}_2 \times \hat{S}_3|} \quad (5)$$

The inclination to the reference plane is then found from

$$\cos i = W_z \quad (6)$$

and the longitude of the ascending node from

$$\tan \Omega = \frac{W_x}{-W_y} \quad (7)$$

The turning angle between the arrival and departure asymptotes is found from

$$\sin \psi = |\mathbf{S}_2 \times \mathbf{S}_3| \quad (8)$$

Eccentricity is found from

$$\frac{1}{e} = \sin \frac{\psi}{2} \quad (9)$$

and the impact parameter from

$$b = -a(e^2 - 1)^{1/2} \quad (10)$$

Fig. 4 illustrates the encounter hyperbola in the vicinity of the intermediate planet. The arrival impact parameter vector,  $\mathbf{B}_2$ , is shown in a plane perpendicular to  $\hat{\mathbf{S}}_2$ . A reference axis is defined in this plane parallel to the reference plane (either equator or ecliptic) and a unit vector  $\hat{\mathbf{T}}$ , is defined along the reference axis. To complete the right-handed system,

$$\hat{\mathbf{R}} = \hat{\mathbf{S}} \times \hat{\mathbf{T}} \quad (11)$$

The angle measured from  $\hat{\mathbf{T}}$  to  $\mathbf{B}$  may be found from

$$\tan \theta_2 = \frac{\sin i \cos (\Theta_2 - \Omega)}{\cos i \cos \phi_2} \quad (12)$$

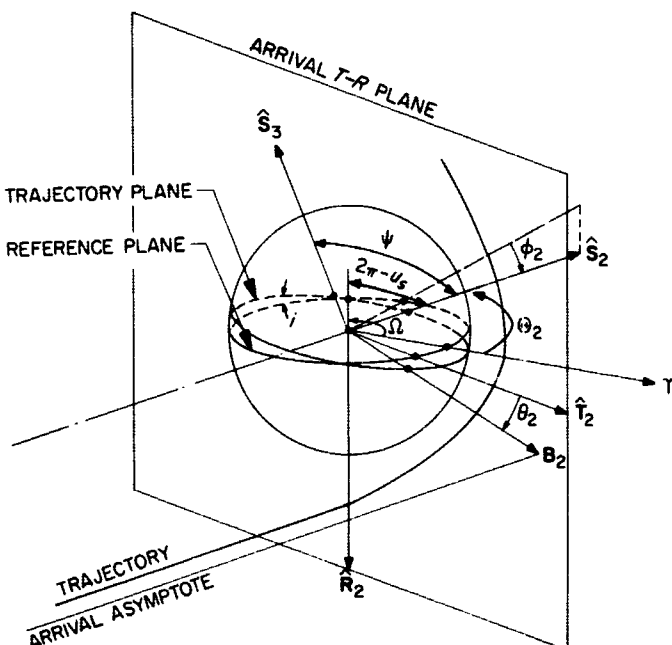


Fig. 4. Intermediate planet hyperbola

and similarly with a subscript 3 for the departure asymptote. [Actually,  $\phi$  and  $\Theta$  should be replaced in Eq. (12) by  $\beta$  and  $\lambda$ , respectively, if the ecliptic reference is being used.] The components of  $\mathbf{B}$  in the  $T$ - $R$  plane are often used in conjunction with the ecliptic reference.

$$\left. \begin{aligned} \mathbf{B} \cdot \mathbf{T} &= b \cos \theta \\ \mathbf{B} \cdot \mathbf{R} &= b \sin \theta \end{aligned} \right\} \quad (13)$$

Table 1 presents a summary for the two sample trajectories of parameters taken from HECON results and computed from the above equations. The Mercury trajectory was chosen to illustrate the reduced launch velocity requirement. The minimum launch hyperbolic excess speed required for a direct Mercury trajectory

Table 1. Nominal conic trajectory parameters for two typical missions

Parameter	Earth-Venus-Mercury		Earth-Venus-Mars	
	Equatorial	Ecliptic	Equatorial	Ecliptic
Launch date, $T_1$	July 29, 1970	0 <sup>h</sup> U.T.	Aug. 12, 1970	0 <sup>h</sup> U.T.
$V_{h_1}$ , km/sec	3.724		3.309	
$\phi_1$ , deg	-22.94	—	16.83	—
$\Theta_1$ , deg	242.01	—	240.84	—
$\Delta T_{12}$ , days	120.0		130.0	
Arrival at Venus	Nov. 26, 1970	0 <sup>h</sup> U.T.	Dec. 20, 1970	0 <sup>h</sup> U.T.
$V_{h_2} = V_{h_3}$ , km/sec	7.168		5.526	
$\phi_2, \beta_2$ , deg	-35.20	-15.72	-66.77	-50.62
$\Theta_2, \lambda_2$ , deg	232.48	238.87	219.70	245.21
$\phi_2, \beta_2$ , deg	-25.18	-3.64	-11.01	1.59
$\Theta_2, \lambda_2$ , deg	294.79	292.35	211.31	212.97
$a$ , km	-6320.9		-10639.2	
$e$	2.2056350		2.0306745	
$b$ , km	12426.4		18803.6	
$\psi$ , deg	53.92		59.00	
$i$ , deg	35.89	17.21	93.31	113.24
$\Omega$ , deg	335.32	304.20	210.66	213.65
$\theta_2$ , deg	352.52	7.10	99.62	128.47
$\mathbf{B} \cdot \mathbf{T}_2$ , km	—	12331.2	—	11696.8
$\mathbf{B} \cdot \mathbf{R}_2$ , km	—	1535.1	—	14722.8
$\theta_2$ , deg	26.46	16.83	93.37	113.25
$\mathbf{B} \cdot \mathbf{T}_2$ , km	—	11894.0	—	7422.4
$\mathbf{B} \cdot \mathbf{R}_2$ , km	—	3598.3	—	17276.7
$\Delta T_{24}$	57.0		175.5	
Arrival at target	Jan. 22, 1971	0 <sup>h</sup> U.T.	June 13, 1971	12 <sup>h</sup> U.T.
$V_{h_4}$ , km/sec	13.935		7.052	
$\phi_4$ , deg	4.96	—	-24.10	—
$\Theta_4$ , deg	343.48	—	230.20	—
Venus gravitational constant $\mu = 324769.5 \text{ km}^3/\text{sec}^2$				

in 1970 (a particularly bad year) is 6.85 km/sec as opposed to the typical (not minimum) value of 3.724 km/sec shown in Table 1. (These correspond to injection energies,  $C_s$ , of 46.9 km<sup>2</sup>/sec<sup>2</sup> and 13.9 km<sup>2</sup>/sec<sup>2</sup>, respectively.) The Mars trajectory was chosen to illustrate the use of the Venus 1970 launch opportunity in order to go to Mars at a time other than the normal opportunities for direct Mars flights occurring in 1969 and 1971.

## 2. Partial Derivatives for Encounter Hyperbola Errors

It is convenient to describe the errors in the arrival and departure conditions at the intermediate planet in terms of the following six parameters, arranged in a column matrix, where  $T$  is the time of closest approach.

$$Y = \begin{pmatrix} V_h \\ \phi \\ \Theta \\ b \\ \theta \\ T \end{pmatrix} \quad (14)$$

Errors in the arrival conditions are mapped to first order into errors in the departure conditions by the  $6 \times 6$  matrix  $B$

$$\delta Y_s = B \delta Y_2 \quad (15)$$

where the elements of  $B$  are the partial derivatives of the departure parameters with respect to the arrival parameters.

$$B_{ij} = \frac{\partial y_{i_s}}{\partial y_{j_2}} \quad \begin{pmatrix} i = 1, \dots, 6 \\ j = 1, \dots, 6 \end{pmatrix} \quad (16)$$

It is immediately seen that, since  $V_h$ ,  $b$ , and  $T$  are the same for both arrival and departure,

$$\frac{\partial V_{h_3}}{\partial V_{h_2}} = \frac{\partial b_3}{\partial b_2} = \frac{\partial T_3}{\partial T_2} = 1 \quad (17)$$

and

$$\frac{\partial y_{i_3}}{\partial y_{j_2}} = 0 \quad \begin{pmatrix} i = 1, 4, 6 \\ j = 1, \dots, 6 \\ j \neq i \end{pmatrix} \quad (18)$$

The remaining partial derivatives are most easily expressed in terms of intermediate partials of the param-

eters  $u$ ,  $i$ ,  $\Omega$ , and  $\Psi$ . From Fig. 4 and spherical trigonometry, we have

$$\tan u_s = \frac{\tan \phi_2}{\sin \theta_2} \quad (19)$$

$$\cos i = \cos \phi_2 \cos \theta_2 \quad (20)$$

$$\tan(\Theta_2 - \Omega) = \frac{\sin \phi_2}{\tan \theta_2} \quad (21)$$

and from Eqs. (4), (9), and (10)

$$\cos \Psi = 1 - \frac{2}{e^2} = \frac{\frac{b^2 V_h^4}{\mu^2} - 1}{\frac{b^2 V_h^4}{\mu^2} + 1} \quad (22)$$

Therefore, from Eq. (19),

$$\left. \begin{aligned} \frac{\partial u_s}{\partial V_{h_2}} &= \frac{\partial u_s}{\partial \Theta_2} = \frac{\partial u_s}{\partial b_2} = \frac{\partial u_s}{\partial T_2} = 0 \\ \frac{\partial u_s}{\partial \phi_2} &= \frac{\sin \theta_2 (1 + \tan^2 \phi_2)}{\sin^2 \theta_2 + \tan^2 \phi_2} \\ \frac{\partial u_s}{\partial \theta_2} &= \frac{-\cos \theta_2 \tan \phi_2}{\sin^2 \theta_2 + \tan^2 \phi_2} \end{aligned} \right\} \quad (23)$$

From Eq. (20),

$$\left. \begin{aligned} \frac{\partial i}{\partial V_{h_2}} &= \frac{\partial i}{\partial \Theta_2} = \frac{\partial i}{\partial b_2} = \frac{\partial i}{\partial T_2} = 0 \\ \frac{\partial i}{\partial \phi_2} &= \frac{\tan \phi_2}{\tan i} \\ \frac{\partial i}{\partial \theta_2} &= \frac{\tan \theta_2}{\tan i} \end{aligned} \right\} \quad (24)$$

From Eq. (21),

$$\left. \begin{aligned} \frac{\partial \Omega}{\partial V_{h_2}} &= \frac{\partial \Omega}{\partial b_2} = \frac{\partial \Omega}{\partial T_2} = 0 \\ \frac{\partial \Omega}{\partial \Theta_2} &= 1 \\ \frac{\partial \Omega}{\partial \phi_2} &= -\frac{\sin 2(\Theta_2 - \Omega)}{2 \tan \phi_2} \\ \frac{\partial \Omega}{\partial \theta_2} &= \frac{\sin 2(\Theta_2 - \Omega)}{\sin 2\theta_2} \end{aligned} \right\} \quad (25)$$

From Eq. (22),

$$\left. \begin{aligned} \frac{\partial \Psi}{\partial \phi_2} &= \frac{\partial \Psi}{\partial \Theta_2} = \frac{\partial \Psi}{\partial \theta_2} = \frac{\partial \Psi}{\partial T_2} = 0 \\ \frac{\partial \Psi}{\partial V_{h_2}} &= \frac{-4(e^2 - 1)^{1/2}}{V_h^2 e^2} \\ \frac{\partial \Psi}{\partial b_2} &= \frac{-2(e^2 - 1)^{1/2}}{b e^2} \end{aligned} \right\} \quad (26)$$

The remaining partial derivatives of the  $B$  matrix may now be formulated. From Fig. 4 and spherical trigonometry, we have

$$\sin \phi_3 = \sin i \sin (u_2 + \Psi) \quad (27)$$

$$\sin (\Theta_3 - \Omega) = \frac{\tan \phi_3}{\tan i} \quad (28)$$

$$\tan \theta_3 = \tan i \cos (u_2 + \Psi) \quad (29)$$

Then, for  $j = 1, \dots, 6$ , the remaining partials are given by,

$$\frac{\partial \phi_3}{\partial y_{j_2}} = \frac{\tan \phi_3}{\tan (u_2 + \Psi)} \left[ \frac{\partial u_2}{\partial y_{j_2}} + \frac{\partial \Psi}{\partial y_{j_2}} \right] + \frac{\tan \phi_3}{\tan i} \left[ \frac{\partial i}{\partial y_{j_2}} \right] \quad (30)$$

$$\begin{aligned} \frac{\partial \Theta_3}{\partial y_{j_2}} &= \left[ \frac{\partial \Omega}{\partial y_{j_2}} \right] + \frac{2 \tan (\Theta_3 - \Omega)}{\sin 2\phi_3} \left[ \frac{\partial \phi_3}{\partial y_{j_2}} \right] \\ &\quad - \frac{2 \tan (\Theta_3 - \Omega)}{\sin 2i} \left[ \frac{\partial i}{\partial y_{j_2}} \right] \end{aligned} \quad (31)$$

$$\begin{aligned} \frac{\partial \theta_3}{\partial y_{j_2}} &= \cos^2 \theta_3 \left\{ \sec^2 i \cos (u_2 + \Psi) \left[ \frac{\partial i}{\partial y_{j_2}} \right] \right. \\ &\quad \left. - \tan i \sin (u_2 + \Psi) \left[ \frac{\partial u_2}{\partial y_{j_2}} + \frac{\partial \Psi}{\partial y_{j_2}} \right] \right\} \end{aligned} \quad (32)$$

The previous Eqs. (14) to (32) are given for the equatorial reference; however, the form is identical for the ecliptic reference, and may be obtained by merely replac-

ing  $\phi$  and  $\Theta$  by  $\beta$  and  $\lambda$  respectively, and properly interpreting the remaining parameters with respect to the ecliptic rather than the equatorial plane. Table 2 gives the elements of the  $B$  matrix for both equatorial and ecliptic reference planes for each of the two sample trajectories specified in the previous section.

### 3. Effects of Errors at Intermediate Planet Encounter

The errors in the arrival conditions at the intermediate planet will propagate along the encounter hyperbola into errors in the departure conditions as described by Eq. (15) of the previous section. The effects of these errors on the postencounter trajectory may be examined in two respects. First, the departure errors will, if uncorrected, result in sizable errors at the target planet. Second, the departure errors may be used to represent the size of a postencounter midcourse maneuver.

The second consideration may be easily seen if we consider the first postencounter midcourse maneuver to be that which nulls the errors in the departure asymptotic velocity. The errors in the departure velocity are expressed in polar form by the upper three elements of the vector matrix,  $\delta Y_3$  [Eqs. (14) to (15)]. The velocity errors may be obtained in Cartesian components by the following matrix multiplication.

$$\begin{pmatrix} \delta V_{A_x} \\ \delta V_{A_y} \\ \delta V_{A_z} \end{pmatrix} = \begin{pmatrix} \cos \phi \cos \Theta, & -V_A \sin \phi \cos \Theta, & -V_A \cos \phi \sin \Theta \\ \cos \phi \sin \Theta, & -V_A \sin \phi \sin \Theta, & V_A \cos \phi \cos \Theta \\ \sin \phi, & V_A \cos \phi, & 0 \end{pmatrix} \begin{pmatrix} \delta V_A \\ \delta \phi \\ \delta \Theta \end{pmatrix} \quad (33)$$

Then the midcourse velocity is

$$\delta V_{mc} = (\delta V_{A_x}^2 + \delta V_{A_y}^2 + \delta V_{A_z}^2)^{1/2} \quad (34)$$

The size of the postencounter midcourse maneuver may be used to compare the sensitivity of various trajectories to arrival errors. This may also be done from a statistical viewpoint, as discussed in the next section.

Let us now examine the effect at the target of departure errors. The output of the heliocentric conic trajectory program (Ref. 9) includes "differential corrections" which, among other quantities, list the values of the partial derivatives of  $B \cdot T$ ,  $B \cdot R$  and  $T$  at the target



Table 2. Partial derivatives (B matrix) of departure parameters with respect to arrival parameters

Earth-Venus-Mercury (equatorial reference)						
	$V_{h_2}$	$\phi_2$	$\Theta_2$	$b_2$	$\theta_2$	$T_2$
$V_{h_2}$	1.0	0	0	0	0	0
$\phi_2$	-0.100475	0.464687	0	$-2.89788 \times 10^{-5}$	0.723561	0
$\Theta_2$	-0.223085	-0.416295	1.0	$-6.43419 \times 10^{-5}$	0.397913	0
$b_2$	0	0	0	1.0	0	0
$\theta_2$	-0.0949145	-0.978451	0	$-2.73751 \times 10^{-5}$	0.419588	0
$T_2$	0	0	0	0	0	1.0
Earth-Venus-Mercury (ecliptic reference)						
	$V_{h_2}$	$\beta_2$	$\lambda_2$	$b_2$	$\theta_2$	$T_2$
$V_{h_2}$	1.0	0	0	0	0	0
$\beta_2$	-0.0653160	0.595106	0	$-1.88384 \times 10^{-5}$	0.773573	0
$\lambda_2$	-0.216277	0.0511208	1.0	$-6.23784 \times 10^{-5}$	-0.234564	0
$b_2$	0	0	0	1.0	0	0
$\theta_2$	-0.0137300	-0.805270	0	$-3.96000 \times 10^{-6}$	0.573994	0
$T_2$	0	0	0	0	0	1.0
Earth-Venus-Mars (equatorial reference)						
	$V_{h_2}$	$\phi_2$	$\Theta_2$	$b_2$	$\theta_2$	$T_2$
$V_{h_2}$	1.0	0	0	0	0	0
$\phi_2$	-0.309756	0.989290	0	$-4.55073 \times 10^{-5}$	-0.0504538	0
$\Theta_2$	0.0186065	0.0283888	1.0	$2.73354 \times 10^{-6}$	-0.871755	0
$b_2$	0	0	0	1.0	0	0
$\theta_2$	$3.55347 \times 10^{-11}$	0.148647	0	$5.22053 \times 10^{-7}$	0.348504	0
$T_2$	0	0	0	0	0	1.0
Earth-Venus-Mars (ecliptic reference)						
	$V_{h_2}$	$\beta_2$	$\lambda_2$	$b_2$	$\theta_2$	$T_2$
$V_{h_2}$	1.0	0	0	0	0	0
$\beta_2$	-0.285094	0.845856	0	$-4.18841 \times 10^{-5}$	-0.338421	0
$\lambda_2$	0.122551	-0.0147945	1.0	$1.80044 \times 10^{-5}$	-0.787867	0
$b_2$	0	0	0	1.0	0	0
$\theta_2$	$-3.39770 \times 10^{-11}$	0.533629	0	$-4.99168 \times 10^{-7}$	0.536835	0
$T_2$	0	0	0	0	0	1.0

Units: km, rad, sec

with respect to  $C_2$ ,  $\phi$ , and  $\Theta$  at the departure planet. The quantities  $b$  and  $\theta$  at departure are not pertinent in the conic approximation, but the time of departure is. Therefore, it is planned to add to the multiplanet version of HECON, the partials of  $B \cdot T$ ,  $B \cdot R$  and  $T$  at the target with respect to departure time. With the above described partials, then, we can construct the matrix

$$K = \begin{pmatrix} \frac{\partial B \cdot T_4}{\partial V_{h_3}} & \frac{\partial B \cdot T_4}{\partial \phi_3} & \frac{\partial B \cdot T_4}{\partial \Theta_3} & \frac{\partial B \cdot T_4}{\partial T_3} \\ \frac{\partial B \cdot R_4}{\partial V_{h_3}} & \frac{\partial B \cdot R_4}{\partial \phi_3} & \frac{\partial B \cdot R_4}{\partial \Theta_3} & \frac{\partial B \cdot R_4}{\partial T_3} \\ \frac{\partial T_4}{\partial V_{h_3}} & \frac{\partial T_4}{\partial \phi_3} & \frac{\partial T_4}{\partial \Theta_3} & \frac{\partial T_4}{\partial T_3} \end{pmatrix} \quad (35)$$

where the first column is obtained from the change of variable

$$\frac{\partial}{\partial V_A} = 2V_A \frac{\partial}{\partial C_3} \quad (36)$$

Now define the miss vector at the target planet

$$\delta M = \begin{pmatrix} \delta B \cdot T_1 \\ \delta B \cdot R_1 \\ \delta T_1 \end{pmatrix} \quad (37)$$

and we have

$$\delta M = (KB') \delta Y_2 \quad (38)$$

where  $B'$  consists of the first, second, third and sixth rows of the  $B$  matrix. Eq. (38), then, maps the errors in the six arrival parameters into the three miss coordinates at the target planet. The elements of the  $3 \times 6 (KB')$  matrix indicate the sensitivity of target errors to errors in the arrival parameters at the intermediate planet, and would therefore be useful for comparing selected trajectories.

Before giving numerical results, it is useful to simplify the problem by isolating the more important parameters. Experience has shown that for expected injection and midcourse errors, the errors in  $V_{A_2}$ ,  $\phi_2$ ,  $\Theta_2$ , and  $T_2$  are small in comparison to errors in  $b_2$  and  $\theta_2$ . Preliminary numerical results show that the target errors resulting from expected errors in  $b_2$  and  $\theta_2$  are from 10 to 100 times larger than from expected errors in  $V_{A_2}$ ,  $\phi_2$  and  $\Theta_2$ . Let us then examine only the errors in the impact parameter components. Define

$$\left. \begin{aligned} \delta M' &= \begin{pmatrix} \delta B \cdot T_1 \\ \delta B \cdot R_1 \end{pmatrix} \\ \delta Y_2' &= \begin{pmatrix} \delta b_2 \\ \delta \theta_2 \end{pmatrix} \\ \delta Y_2'' &= \begin{pmatrix} \delta B \cdot T_2 \\ \delta B \cdot R_2 \end{pmatrix} \end{aligned} \right\} \quad (39)$$

Then, from Eq. (13),

$$\begin{aligned} \delta Y_2' &= \begin{pmatrix} \cos \theta_2 & \sin \theta_2 \\ -\sin \theta_2 & \cos \theta_2 \\ b_2 & b_2 \end{pmatrix} \delta Y_2'' \\ &= R' \delta Y_2'' \end{aligned} \quad (40)$$

Then, defining  $K'$  as the appropriate  $2 \times 2$  portion of  $(KB')$ , we have

$$\delta M' = (K' R') \delta Y_2'' = R \delta Y_2'' \quad (41)$$

where

$$R = \begin{pmatrix} \frac{\partial B \cdot T_1}{\partial B \cdot T_2} & \frac{\partial B \cdot T_1}{\partial B \cdot R_2} \\ \frac{\partial B \cdot R_1}{\partial B \cdot T_2} & \frac{\partial B \cdot R_1}{\partial B \cdot R_2} \end{pmatrix} \quad (42)$$

The numerical values of the elements of  $R$  are given in Table 3 for the two sample trajectories to Mercury and Mars.

Table 3. Impact parameter error sensitivities ( $R$  matrix)

Earth-Venus-Mercury (ecliptic reference)		
	$B \cdot T_1$	$B \cdot R_1$
$B \cdot T_1$	528.8	348.5
$B \cdot R_1$	-161.6	-760.5
Earth-Venus-Mars (ecliptic reference)		
	$B \cdot T_1$	$B \cdot R_1$
$B \cdot T_1$	-1236	-2607
$B \cdot R_1$	-4046	3904

The  $R$  matrix is useful for targeting of multiplanet trajectories on the Space Trajectories Program. The procedure is as follows. For the initial search, select the aiming point values of  $B \cdot T_2$  and  $B \cdot R_2$  that are obtained from the conic approximation. The resulting miss at the target planet may then be used with the inverse of the  $R$  matrix to generate increments in  $B \cdot T_2$  and  $B \cdot R_2$  that will remove most of the target error. Whether or not this process converges for successive trials with the same  $R$  matrix has not been determined. However, the first iteration does yield trajectories which pass very close to the target planet, as indicated by the results in Table 4 for the Earth-Venus-Mars sample trajectory.

#### 4. Statistics of Encounter Errors

Since actual values of arrival errors are not known in advance, it is useful to examine the statistical nature of the errors, based on the expected distribution of arrival errors. Let  $\Delta_{Y_2}$  be the  $6 \times 6$  covariance matrix of expected

Table 4. Earth-Venus-Mars trajectory on space trajectories program

Parameter	HECON	1st Search	Corrections from Eq. (41) for $B_1 = 0$	2nd Search
<b>Launch</b>				
$V_A$ , km/sec	3.309	3.2959626		3.2959626
$C_A$ , km <sup>2</sup> /sec <sup>2</sup>	10.949	10.863419		10.863419
$\phi_A$ , deg	16.83	16.484351		16.484953
$\theta_A$ , deg	240.84	241.25110		241.25091
$\Sigma_A$ , deg	90	90		90
$R_A$ , km	6567.4	6567.4		6567.4
$\gamma_A$ , deg	2.0	2.0		2.0
$T_A$	8/12/70 0 <sup>h</sup> U.T.	8/12/70 0 <sup>h</sup> U.T.		8/12/70 0 <sup>h</sup> U.T.
$\Delta T_{12}$ , days	130.0	129.598464		129.598841
<b>Venus</b>				
$V_A$ , km/sec	5.526	5.5250332		5.5250049
$\phi_A$ , deg	-69.77	-69.541426		-69.542074
$\theta_A$ , deg	-219.70	220.13314		220.13184
$B \cdot T_A$ , km	-11697	-11698.334	+ 87.89	-11603.263
$B \cdot R_A$ , km	14723	14724.085	+ 114.45	14837.774
$\phi_A$ , deg	-11.01	-10.809301		-10.865752
$\theta_A$ , deg	211.31	211.41984		211.80882
$\Delta T_{21}$ , days	175.5	175.734447		176.941396
<b>Mars</b>				
$V_A$ , km/sec	7.052	7.0359599		6.9600939
$\phi_A$ , deg	-24.10	-24.173786		-24.159954
$\theta_A$ , deg	230.20	230.11903		229.68421
$B \cdot T_A$ , km	0	406936.32		-11062.911
$B \cdot R_A$ , km	0	-91264.984		12517.591
$B_A$ , km	0	417044.92		16705.630

arrival errors, which may be determined from a variety of analyses.

$$\Lambda_{Y_2} = (\delta Y_2)(\delta Y_2)^T \quad (43)$$

Then from Eq. (15),

$$\Lambda_{Y_3} = B \Lambda_{Y_2} B^T \quad (44)$$

The covariance matrix of the departure asymptotic velocity components may be obtained from the upper-left  $3 \times 3$  part of  $\Lambda_{Y_3}$  and the  $3 \times 3$  matrix of Eq. (33), which we shall denote by the symbol,  $A$ . Then

$$\Lambda'_{Y_3} = A \Lambda_{Y_3} A^T \quad (45)$$

and the one-sigma value of the rms postencounter midcourse velocity is the square root of the trace of  $\Lambda'_{Y_3}$ .

A number of different sets of  $\Lambda_{Y_3}$  matrices could be generated based on a variety of assumptions concerning the number and accuracy of pre-encounter maneuvers and the associated orbit determination. An analysis has been conducted with an assumed approach guidance system for the Earth-Venus-Mars trajectory. A complete description of the assumptions involved is not included in this report (see Ref. 10). Basically, the approach guidance system tracks the Sun, the planet and a star in order to determine the approach trajectory. Based on current estimates of the accuracy of such a system and the execution errors of two approach maneuvers, a  $\Lambda_{Y_3}$  matrix was generated following the second maneuver. Mappings similar to those of Eqs. (44) to (45) were performed, with a resulting one-sigma, rms, postencounter midcourse velocity of 50 m/sec.

A similar mapping of covariance matrices to the target planet may be determined from Eq. (41). Let

$$\left. \begin{aligned} \Lambda_2 &= (\delta Y_2^T)(\delta Y_2^T)^T \\ \Lambda_4 &= (\delta M^T)(\delta M^T)^T \end{aligned} \right\} \quad (46)$$

Then

$$\Lambda_4 = R \Lambda_2 R^T \quad (47)$$

Eq. (47) maps the distribution of impact parameter errors at intermediate planet arrival into the distribution of impact parameter errors at the target. It is instructive to evaluate Eq. (47) for a circularly distributed unit error (1 km) in the intermediate planet arrival impact parameter. Then

$$\Lambda_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and the target errors for the two sample trajectories are:

$$\begin{aligned} \Lambda_{\text{Mercury}} &= \begin{pmatrix} 0.40108 \times 10^6 & -0.35049 \times 10^6 \\ -0.35049 \times 10^6 & 0.60447 \times 10^6 \end{pmatrix} \\ \Lambda_{\text{Mars}} &= \begin{pmatrix} 8.3219 \times 10^6 & -5.1773 \times 10^6 \\ -5.1773 \times 10^6 & 31.6111 \times 10^6 \end{pmatrix} \end{aligned}$$

A diagonalization of the  $\Lambda_4$  matrix gives the principal axes of the one-sigma error ellipse in the  $T$ - $R$  plane and the angle of the major axis from the  $T$ -axis (Ref. 9,

Table 5. Principal axes of target error ellipse for 1 km circular error at arrival

	Earth-Venus-Mercury	Earth-Venus-Mars
Semimajor axis	900.6 km	5719 km
Seminor axis	441.0 km	2688 km
Angle of major axis from T-axis	143 deg	102 deg

pp. xvii-xviii). These values are given for the two sample trajectories in Table 5.

Results such as are presented in Table 5 serve as a means for comparing the sensitivity of various trajectories to a standard uncorrelated, unit error at arrival. However, a more meaningful comparison would result if the  $\Lambda_2$  matrix for each trajectory reflected the comparative accuracy obtainable at intermediate planet arrival. The target coordinate uncertainty due to orbit determination tracking errors is such a  $\Lambda_2$  matrix, since it represents the limiting accuracy that can be obtained as midcourse correction errors approach zero. Therefore, Eq. (47) has been evaluated using a  $\Lambda_2$  taken from the orbit determination accuracy group of the heliocentric conic trajectory program output. From Ref. 9, therefore, we obtain the values in Table 6.

The resulting  $\Lambda_1$  matrices are:

$$\Lambda_{1\text{Mercury}} = \begin{pmatrix} 0.07409 \times 10^{12} & 0.01575 \times 10^{12} \\ 0.01575 \times 10^{12} & 0.07978 \times 10^{12} \end{pmatrix}$$

$$\Lambda_{1\text{Mars}} = \begin{pmatrix} 5.225 \times 10^{12} & -10.924 \times 10^{12} \\ -10.924 \times 10^{12} & 25.774 \times 10^{12} \end{pmatrix}$$

It can be seen from the above results that the one-sigma rms misses at Mercury and Mars are  $0.4 \times 10^6$  km and  $5.6 \times 10^6$  km, respectively, if no approach or post-encounter maneuvers are made.

Table 6. Orbit determination target accuracy from HECON

	Earth-Venus-Mercury	Earth-Venus-Mars
$\sigma_r$	702.1 km	370.6 km
$\sigma_h$	433.4 km	995.4 km
$\rho_{rk}$	-0.7716	-0.7231

A more complete statistical description of the target errors is obtained from Eq. (38)

$$\Lambda_1 = (\delta M)(\delta M)^T = (KB')(\delta Y_2)(\delta Y_2)^T(KB')^T \tag{48}$$

$$= (KB')\Lambda_2(KB')^T$$

No numerical results are presented for Eq. (48), however.

### 5. Discussion

The results of this limited study show that the accuracy of multiplanet trajectories should not be a serious obstacle to their use in future missions. Certainly, post-encounter maneuvers will be necessary, and also probably one or two approach maneuvers dictated by an approach guidance system. An analysis, in some depth, of promising missions is warranted, and the equations developed in this report should serve as a partial basis for comparing and selecting nominal trajectories. A multiplanet version of HECON is being developed, and many of the error equations of this report will be included as subroutines.

Other considerations must be investigated, of course, and will affect the total feasibility of these types of trajectories. It was noted in the course of the analysis, for example, that neither of the sample trajectories pass through a region during the Venus encounter such that the Earth, Sun or Canopus is occulted. However, there is no freedom to choose the aiming point at the intermediate planet, and it may be found that other trajectories pass through regions such that an important reference body is occulted. The two sample trajectories also had distances of closest approach at Venus that were greater than average for these type missions (Ref. 7). Closer trajectories may have difficulty meeting constraints on probability of impacting the intermediate planet. Since a successful midcourse maneuver considerably lessens the probability of impact, it would be beneficial to improve the reliability and accuracy of the midcourse correction operation.

Other missions, such as out-of-ecliptic and solar probes, have been proposed which also use the gravitational perturbation of an intermediate planet. These missions have less stringent targeting constraints than those having a planetary target, and should be seriously considered. It may prove possible to accomplish these missions without the need for approach guidance or correction; that is, using only two or three midcourse corrections.

## References

1. Pfeiffer, C. G., "On the Second Variation of an Optimal Trajectory," SPS 37-26, Vol. IV, pp. 8-12, Jet Propulsion Laboratory, Pasadena, Calif., April 30, 1964.
2. Bliss, G. A., *Lectures in the Calculus of Variations*, University of Chicago Press, Chicago, Ill., 1946.
3. Kalman, R. E., *On the General Theory of Control Systems*, paper presented at the First International Congress on Automatic Control, Moscow, 1960.
4. Pfeiffer, C. G., "On the Second Variation of an Optimal Trajectory," Technical Report No. 32-566, Jet Propulsion Laboratory, Pasadena, Calif. (to be published).
5. Lovitt, W. V., *Linear Integral Equations*, Dover Publications, New York, N. Y., 1950.
6. Liusternik, I. A., and Sobolev, V. J., *Elements of Functional Analysis*, Ungar Publishing Company, New York, N. Y., 1961.
7. Minovitch, M. A., "The Determination and Characteristics of Ballistic Interplanetary Trajectories Under the Influence of Multiple Planetary Attractions," Technical Report No. 32-464, Jet Propulsion Laboratory, Pasadena, California, October 31, 1963.
8. Hunter, M. W., "Future Unmanned Exploration of the Solar System," *Astronautics and Aeronautics*, Vol. 2, No. 5, pp. 16-26, May 1964.
9. Clarke, V. C., Jr. et al., "Earth-Venus Trajectories, 1970," Technical Memorandum No. 33-99, Vol. 5, Parts A-B, Jet Propulsion Laboratory, Pasadena, California, October 1, 1963.
10. Gordon, H. J., "Study of an Approach Guidance System," SPS 37-19, Vol. IV, pp. 2-5, Jet Propulsion Laboratory, Pasadena, California, February 28, 1963.

## GUIDANCE AND CONTROL DIVISION

### II. Spacecraft Electrical Power

#### A. Power Sources

*P. Rouklove and D. W. Ritchie*

##### 1. Solar Energy Thermionic (SET) Electrical Power Supply Development

The development of the SET I and II electrical power supplies is continuing with emphasis on converter and generator development under JPL contracts with Electro-Optical Systems, Inc., (EOS) Pasadena, California and Thermo-Electron Engineering Corporation, (TEEco) Waltham, Massachusetts. The Missile and Space Division of General Electric Company is studying, under contract to JPL, the thermal energy storage properties of various materials and their possible application to an advanced thermionic generator.

The converters manufactured by EOS and TEEco are to be incorporated into a 4-converter generator for use with a 5-ft-D mirror (SET I) and a 16-converter generator in conjunction with a 9.5-ft-D mirror for a SET II system.

*a. Tests of TEEco Converters VIII-P-1 and VIII-P-2(a).* Two converters, manufactured by TEEco, were tested at JPL. These converters are the prototypes of a series of

converters which incorporate several improvements, such as high power density, novel and improved attachment system, better resistance to vibration and shock, and lower cesium vapor pressure. Four of these converters of the new design will be assembled in a SET generator now in the design phase. Both converters have a 2-cm<sup>2</sup> rhenium emitter and a molybdenum collector. The emitter of Converter VIII-P-1 was obtained by pyrolytic vapor deposition of rhenium on a tantalum substratum using rhenium pentachloride, while Converter VIII-P-2(a) used a rhenium emitter, pressure-bonded to a tantalum slug. The pyrolytic deposition of rhenium on tantalum presents difficulties and the resulting bonding is unreliable, pressure-bonding appearing to be a better process. The radiator in the prototype Converter VIII-P-1 was made to the same size as used in previous SET converters, and it was found that this radiator was insufficient for the high current densities obtained. To achieve the optimum collector temperature of approximately 950°K required in a rhenium-molybdenum system operating at an emitter temperature of 2000°K and at high current densities, the radiator used in Converter VIII-P-2(a) was increased by 110 cm<sup>2</sup>. The measured temperature of the collector face, operating at 40 amp<sup>2</sup>/cm<sup>2</sup> and with an emitter temperature of 1700°C, was 960°C. To reduce the large thermal

drop from the collector face to its base, the cross section of the collector used in VIII-P-2(a) was increased to a maximum compatible with the size of the converter envelope. The results of the performance tests of Converters VIII-P-1 and VIII-P-2 together with the results of converters of Series VII, built at the end of 1963, are presented in Figs. 1 (a) and (b).

**b. Tests of EOS converters.** The effort of the converters built by EOS was mostly directed toward an improvement of the resistance to vibrations and stress expected during the launch period of the thermionic generator. The converters were tested, using sinusoidal vibration.

As the main goal in the test of the prototype converters is to obtain design and analytic information, this mode of vibration was applied, as opposed to random vibration testing. The sinusoidal mode imposes greater stress on the converter as a greater energy density is applied to the structure than with a random scan. The advantage gained by observing at all times the frequency, the applied load, and the possibility of immediate observation of any resonance effect favored the vibration sinusoidal test. The converters were tested at frequencies varying between 0 and 2000 cps in two steps, 0 to 500 cps, and 500 to 2000 cps, and in the  $x$ - $y$  and  $z$ -axis at a peak load of up to 10  $g$ .

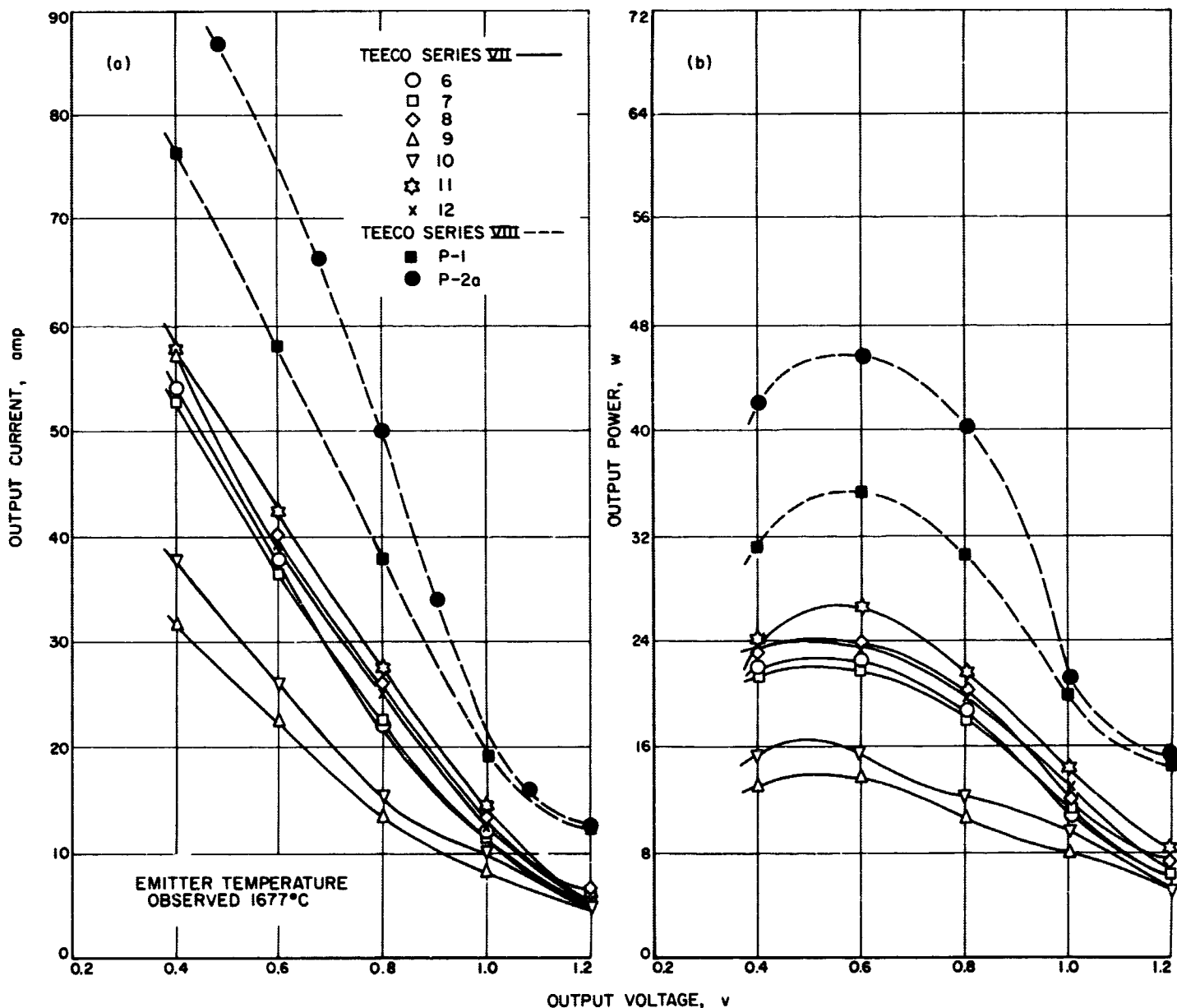


Fig. 1. Output current and power vs output voltage for SET converters

During the tests, the following observations were made:

**Prototype 1.** Resonance of the collector flange at 1050 cps resulted in a "chatter" between emitter and collector surfaces. This observation led to a redesign of the converter structure.

**Prototype 2.** Resonance observed at 475 cps (5-g peak) of the cesium reservoir tubulation resulted in displacement of approximately  $\frac{1}{8}$  in. It was felt that the reservoir heater wire may have acted as an off-center weight. Once the reservoir was tied down, the resonance disappeared.

**Prototype 3.** A resonance observed at the tip of the radiator heater wire resulted in material shaken out of the sheath. This resonance was observed at 1800 cps 10-g peak load.

Prototypes 2 and 3 were performance-tested by the manufacturer before and after the vibration test. No adverse conditions were observed. Prototype 2 was shaken twice without ill effects.

These converters are characterized by a one-piece collector radiator obtained by a drop-forge method from a solid bar of molybdenum. It was postulated that such an approach will greatly increase the resistance of the converter to environmental vibrations and eradicate a possibility of failure in the brazed section between the molybdenum collector and the copper radiator.

**c. Development of 9.5-ft-D electroformed solar concentrator.** The 9.5-ft-D mirror replica obtained from the 9.5-ft-D nickel master was tested at Table Mountain. The results of the test were reported in Ref. 1. To reduce the amount of available energy to a quantity equivalent to the one available from a 60-in.-D mirror in space, and to remove the effects of the damaged peripheral zone, the mirror was shadowed to reduce its effective diameter to 7.5 ft. This shadowing reduces the area from 70.882 ft<sup>2</sup> to 44.179 ft<sup>2</sup>. The effects of the area reduction are presented below.

Calorimeter aperture 1 in.	7.5-ft D	9.5-ft D
Power input (standardization to 90 w/ft <sup>2</sup> ), w	2460.4	3298.3
$\eta$ , %	61.88	51.70
Area, ft <sup>2</sup>	44.179	70.882
Shadow, ft <sup>2</sup>	3.129	3.598
Shadow factor	0.92917	0.94925
Reflectance EM/SF	0.666	0.5446

The results of the test of the mirror for both areas are presented graphically in Fig. 2. A 60-in.-D mirror performance is presented for comparison purposes.

## 2. Development of Photovoltaic Standards

**a. Introduction.** Measurements of photovoltaic devices, either individual or complete arrays, have been obtained by comparing the photovoltaic cell output against a total energy measuring device. Terrestrial measurements using such methods can result in errors ranging from 5 to 15%. The development of photovoltaic standards directly correlated to proven high-altitude balloon flight standard

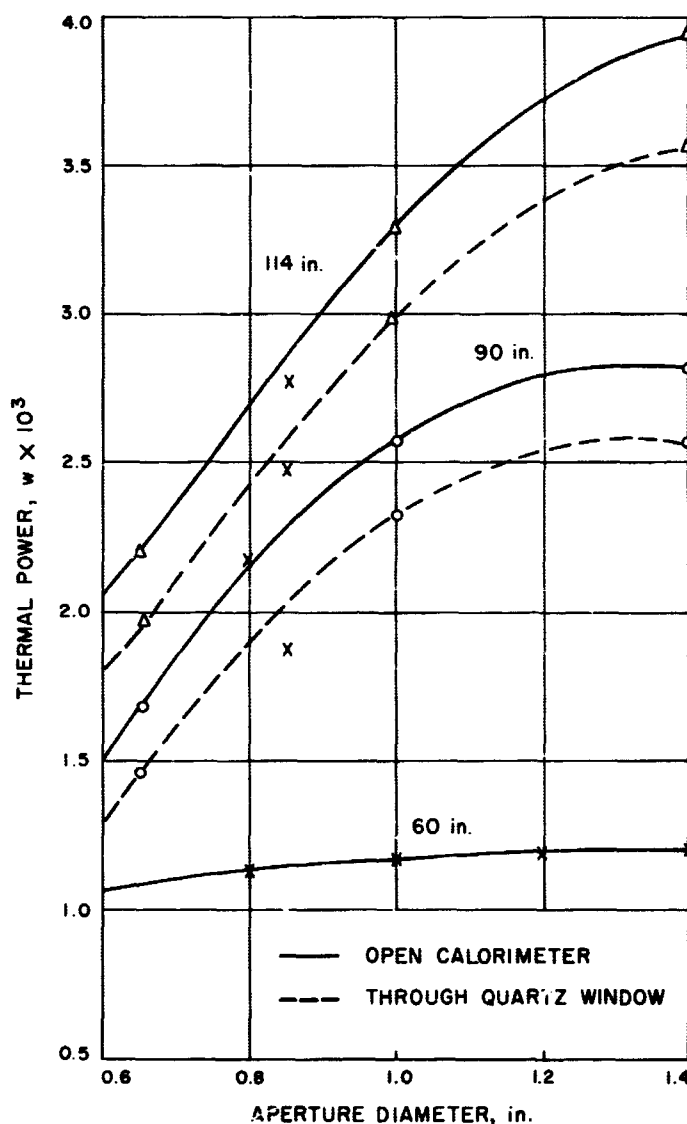


Fig. 2. Effect of mirror aperture diameter on thermal power output



cells eliminates the requirement for total energy measuring devices and provides extrapolated space short-circuit current with an accuracy greater than  $\pm 2\%$ .

The design of the standard cell is in accordance with specifications set forth by the Solar Working Group of the AIEE. The calibration methods presented herein are compared to actual high-altitude balloon flight measurements. The standardization methods incorporate both boron and phosphorus diffused silicon solar cells.

**b. Description of calibration method.** The method presented here for the calibration of the short-circuit current of standard cells is based on comparative measurements of the cell to be calibrated with solar cells whose short-circuit current in space is known. Such space standard cells have been obtained by JPL. The calibration involved simultaneous measurements of the short-circuit currents of the space standard cell with the standard cell to be calibrated under a uniform source which varies in spectral distribution and intensity. Such measurements are performed in sunlight at Table Mountain, California.

**c. High-altitude balloon flight measurements.** Since June 13, 1962, the Jet Propulsion Laboratory has successfully launched three high-altitude balloons for solar cell calibrations. The balloon flights have been correlated to one another and repeatability of measurements is better than 0.5%.

The balloon flights have provided space standards for solar panel measurements of the *Ranger* and *Mariner* programs, as well as for the development of NASA standards. A total of thirty-eight such space standard solar cells are maintained at JPL for balloon correlation measurements and standardization development.

**d. Solar cell design.** The solar cell used as standard is of a special design and is manufactured under normal production techniques. The bulk material is doped for a desired resistivity of 7 to 10  $\Omega$ -cm. The *p*-doped crystal is sliced to size and polished. The polished wafer is shallow-diffused using phosphorus gas. The resistive contacts on the cell are vacuum-deposited silver-titanium. The polished cell was not coated with the silicon monoxide layer normally used on *n-on-p* type solar cells. The silicon monoxide coating is mainly used to increase the efficiency of the solar cell, but, for this application, cell efficiency is not important.

**e. Standard cell package.** The solar cell standards housing has a brass water-cooled heat sink. The solar cell is soft-soldered to a Kovar base which attaches directly to the water-cooled base. The Kovar cell substrate will allow the standard cell to be used at various temperatures without damage to the cell from thermal expansion. The complete standard cell housing is encapsulated with an optical polished fused quartz window  $\frac{1}{16}$  in. thick. Prior to hermetically sealing, the complete unit is purged several times and then filled with dry inert gas.

The locating holes in the bottom of the housing are designed for accurate solar cell positioning when used in an artificial test source. A thermocouple is physically connected to the *n*-contact of the solar cell for accurate cell temperature measurements. The complete housing assembly is shown in Fig. 3.

**f. Use of developed standard cells.** The developed standard cells can be used for obtaining space short-circuit current values in solar arrays using terrestrial measurements. Such standards could reduce present problems with solar cell procurements and could provide correlation of photovoltaic power systems.

The short-circuit current output of solar arrays can be calculated using such standards if the spectral response of the standard cell is established to be similar to that of the test panel. If the standard cell is attached directly to the solar array being tested, and the two units "see" the same effective total energy, the use of the "collimation" system is not required.

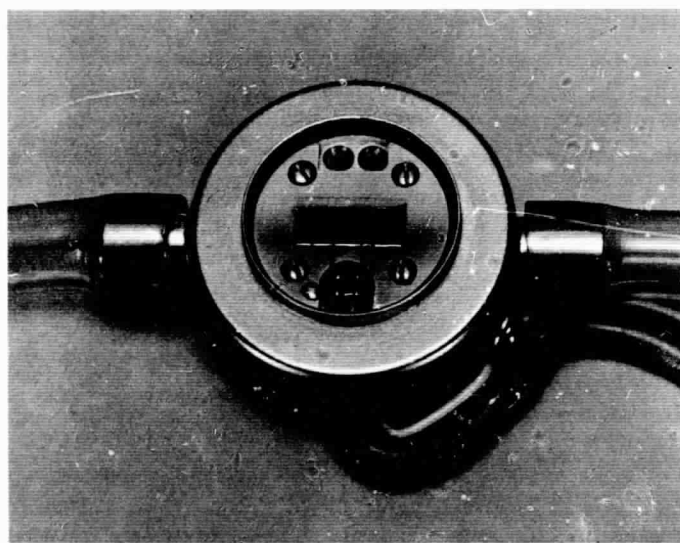


Fig. 3. Photovoltaic standard cell

## Reference

1. Rouklove, P., "Power Sources," SPS 37-26, Vol. IV, pp. 31-38, Jet Propulsion Laboratory, Pasadena, California, April 30, 1964.

### III. Guidance and Control Analysis and Integration

#### A. Satellite Orbit Determination Accuracy Using On-Board Instruments

*R. A. Virzi and R. V. Morris*

##### 1. Introduction

A future mission of interest will be to put a spacecraft in orbit around another planet such as Mars or Venus. Such an orbiting spacecraft can be used to take direct observations of the planet or to relay information from experiments on the planet surface. Many experiments that an orbiting spacecraft may perform will require accurate satellite orbit determination for the following reasons:

- (1) For orbital control to obtain optimum data gathering conditions, such as for photographing surface features where lighting and contrast conditions are important.
- (2) For the mapping of location-dependent data, such as surface features, magnetic fields, and particle densities.

In addition, the satellite orbit determination can be used to obtain the mass of the planet.

A satellite orbit is determined by seven independent parameters. One set useful for our purposes is the orbital period  $T$ , the eccentricity  $e$ , the time of periapsis passage  $T_p$ , the planet gravitational constant  $GM$ , the argument of periapsis  $\omega$ , the inclination  $i$ , and the argument of the ascending node  $\Omega$ . Some of these are illustrated in Fig. 1. These seven parameters give a complete time description of the spacecraft orbital motion. Therefore, the accuracy of this time history may be derived from the accuracy of estimates of the parameters.

A study of the accuracy to which on-board instruments can determine a satellite orbit is currently being performed. In this study, accuracy is being determined as a function of the following factors:

- (1) The spacecraft orbit.
- (2) The distribution of measurement times in an orbit.
- (3) The types of measurement taken at each measurement time.
- (4) The measurement error model.

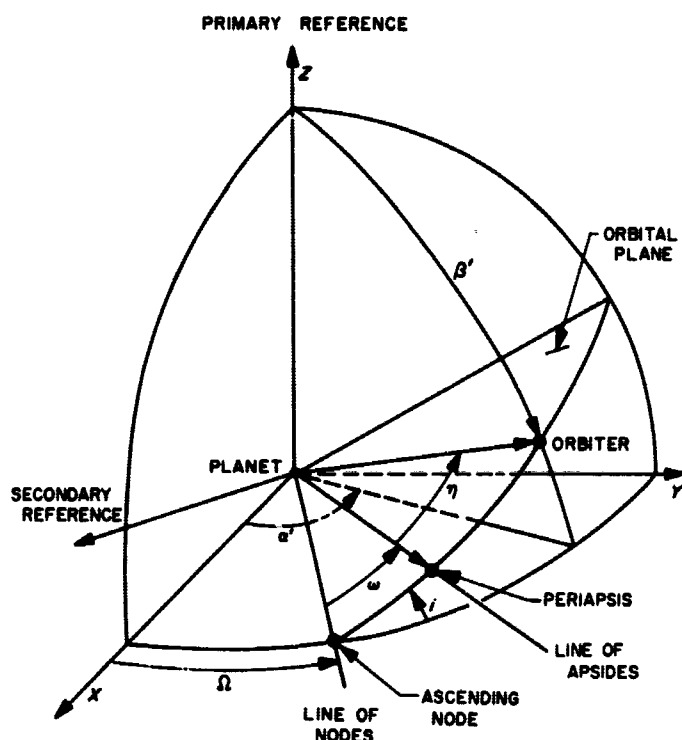


Fig. 1. Coordinate system and orbital plane

(5) The total measurement time or the number of orbits.

The results to date of this study are described below.

## 2. Measurement Types

In this section, on-board measurements useful in satellite orbit determination are resolved into categories, and the characteristics of these categories are discussed.

### a. Measurement types using on-board instruments.

There are many different types of measuring instruments which can be carried by a spacecraft. The measurements by these instruments give information which will fall into one, or a combination, of the following categories:

- (1) *Angle information*, the information on the angle subtended at the spacecraft between the central planet and a reference body, e.g., the Mars-spacecraft-Sun angle.
- (2) *Range information*, the information on the distance from the center of the central planet to the spacecraft. Measurements which contain only range information are angular diameter and altitude. Both measurements require knowledge of the planet radius.
- (3) *Rate information*, the information on the rate of change of position relative to the central planet.

Measurements which contain only rate information are orbital rate and range rate. The orbital rate measurement requires knowledge of the planet center direction. The range rate measurement may also require the planet center direction.

- (4) *Event time information*, e.g., times of occultations, and times of a given angle measurement.

In the present study, measurements containing information in the fourth category are not considered. All following statements in this report apply only to the first three categories.

*b. Orbit determination properties of measurement types.* The orbit determination properties of measurement types depend on the motion of the spacecraft. If an inverse-square central force field is assumed, then the spacecraft motion is planar and follows Kepler's laws, in which case the following statements can be made<sup>a</sup>:

Angle measurements, Category (1), are the only measurements which can give information about the orientation of the spacecraft orbit. The orientation of the orbit cannot be determined unless the spacecraft is able to make angle measurements with respect to at least two known inertial directions. This can be done by using two reference bodies or by using a single reference body having angular motion relative to the planet. Furthermore, angle, range, and rate measurements, Categories (1), (2) and (3), can be used to determine the size and shape of the orbit and the location of the spacecraft in the orbit. (In this report, size is determined by the orbital period and planet gravitation constant, shape by the eccentricity, and in-plane location for a given time by the time of periaxis passage, period, and eccentricity.) The measurements in Categories (1), (2) and (3) can always be expressed in terms of the four in-plane coordinates: true anomaly, range, range rate and orbital rate. Therefore, the in-plane orbit determination properties of measurements can be obtained from the properties of the four in-plane coordinates.

Some important orbit determination properties of true anomaly, range, range rate and orbital rate measurements are:

- (1) True anomaly—gives no information of size but always gives information of shape and in-plane location.

<sup>a</sup>Deviations due to oblateness, etc., which may be encountered at a planet such as Mars, will cause perturbations in the spacecraft motion. These perturbations can be handled in orbit determination as variations in the orbital parameters.

- (2) Range—always gives information of size and shape, and gives information of in-plane location *except* for circular orbits.
- (3) Range rate—always gives information of shape, and gives information of size and in-plane location *except* for circular orbits.
- (4) Orbital rate—gives no information of size but always gives information of shape and in-plane location.

The above orbit determination properties apply when only measurement data is used without making use of prior information (*a priori* knowledge). In practice, prior information may be available that may allow determination of size, shape or in-plane location when it is not possible with measurement data alone.

Because of the periodic nature of the spacecraft orbital motion, very accurate determination of the orbital period can be obtained if the measurement used has a periodic variation. All the above four measurements are periodic except for circular orbits when only the true anomaly, i.e., the central angle, has any variation.

### 3. Study Description

The following is a description of the work performed.

*a. The spacecraft reference coordinate system.* A spacecraft reference coordinate system is necessary for defining spacecraft based angle measurements. In this study the reference system is an inertial right-hand Cartesian coordinate system with mutually orthogonal axes labelled X-Y-Z. It is centered at the spacecraft, and its coordinate axes are determined by two reference bodies. The positive Z-axis is directed from the spacecraft toward the primary reference body. The XZ-plane contains the secondary reference body in the +X half-plane. The Y-axis completes the right-hand coordinate system X-Y-Z. This coordinate system, centered at the planet, is illustrated in Fig. 1.

*b. The spacecraft orbit.* The spacecraft is assumed to move in an elliptical orbit. The orbital plane is inertially fixed and contains the planet center of mass. The orientation of the orbital plane in the reference coordinate system is illustrated in Fig. 1. The orientation of the plane is defined by two angles. They are the argument of the ascending node  $\Omega$  and the inclination angle  $i$ . The in-plane orientation of the ellipse is given by the argument of periapsis  $\omega$ , which is the angle from the ascending node to periapsis, measured in the direction of motion

of the spacecraft. The size (i.e., semimajor axis) of the ellipse is determined by the orbital period  $T$  and the planet gravitational constant  $GM$ . Its shape is determined by the eccentricity  $e$ . The location of the spacecraft in the orbit at a given time is determined by the time of periapsis passage  $T_p$ , the period  $T$ , and the eccentricity  $e$ .

To date only two orbits (about Mars) have been studied. These orbits, identified as Orbits 1 and 2, are described in Table 1. The angles which describe the orbital orientation are referred to a spacecraft coordinate system based on the Sun and the south ecliptic pole as primary and secondary references, respectively. Although the spacecraft reference system is not inertial (i.e., it rotates at the planet's orbital rate) when the Sun is used, the accuracy results should still be valid because of the short time durations considered in the study. If the effects of the planet's motion about the Sun and its finite distance were included, it is expected that they would improve the orbit determination accuracy results for long time periods.

*c. The measurement types.* Measurements are made of both the cone and clock angles of the central planet in the spacecraft reference system at each measurement time. The cone angle  $\beta$  is the angle from the positive Z-axis to the planet direction ( $0 \leq \beta \leq 180^\circ$ ). The clock angle  $\alpha$  is the angle from the positive X-axis to the projection of the planet direction on the XY-plane. The positive sense of  $\alpha$  is a right-hand rotation about the +Z-axis. The angles  $\beta$  and  $\alpha$  are related to the angles  $\beta'$  and  $\alpha'$  (shown in Fig. 1) as follows:

$$\beta = 180^\circ - \beta' \quad (1)$$

$$\alpha = 180^\circ + \alpha' \quad (2)$$

Table 1. Parameters of two satellite orbits about Mars

Orbital parameter	Symbol	Orbit 1	Orbit 2	Units
Gravitational constant	$GM$	42977.8	42977.8	$\text{km}^3/\text{sec}^2$
Orbital period	$T$	27390.568	13094.296	sec
Eccentricity	$e$	0.42091	0.01	.....
Time of periapsis passage	$T_p$	0.0	0.0	sec
Semimajor axis	$A$	9347.5	5715.0	km
Periapsis range	$R_p$	5413.0438	5657.85	km
Apoapsis range	$R_A$	13281.956	5772.15	km
Argument of periapsis	$\omega$	-142.51455	177.08134	deg
Inclination	$i$	32.314388	32.48401	deg
Argument of the ascending node	$\Omega$	1.9102133	47.68956	deg

where  $\beta$  and  $\alpha$  are the cone and clock angles of the orbiter-to-planet direction, and  $\beta'$  and  $\alpha'$  are the cone and clock angles of the planet-to-orbiter direction.

**d. The measurement error model.** The measurement error model assumes that each measurement has an independent random error of 0.266 mrad, 1  $\sigma$ . A fixed bias in the cone angle measurement was assumed and was estimated as an auxiliary parameter of the orbit. A fixed bias in the clock angle measurement and an error in the argument of the ascending node  $\Omega$  have identical effects on the measurements. Therefore, no auxiliary parameter is necessary to account for the bias. If prior information of either is available, the bias and the error in  $\Omega$  may be separated.

**e. Measurement times.** Cone and clock angle measurements are taken every 5 min, starting at the nominal time

of a periapsis passage. The total measurement period is 50 hr.

#### 4. Results

Accuracy results have been obtained at intervals of 40 measurement times (200 min) and plotted in Figs. 2(a) and (b). They give the standard deviations, without prior information, of the argument of the ascending node  $\Omega$ , the inclination angle  $i$ , the argument of periapsis  $\omega$ , the eccentricity  $e$ , the orbital period  $T$ , and the time of periapsis passage  $T_p$ . The plotted points are connected by straight lines only for ease of identification, and do not represent the true accuracies for intermediate times.

Several conclusions that may be drawn from the results plotted in Figs. 2(a) and (b) are:

- (1) The accuracy of all the estimated orbital elements increases with measurement time. The rate of im-

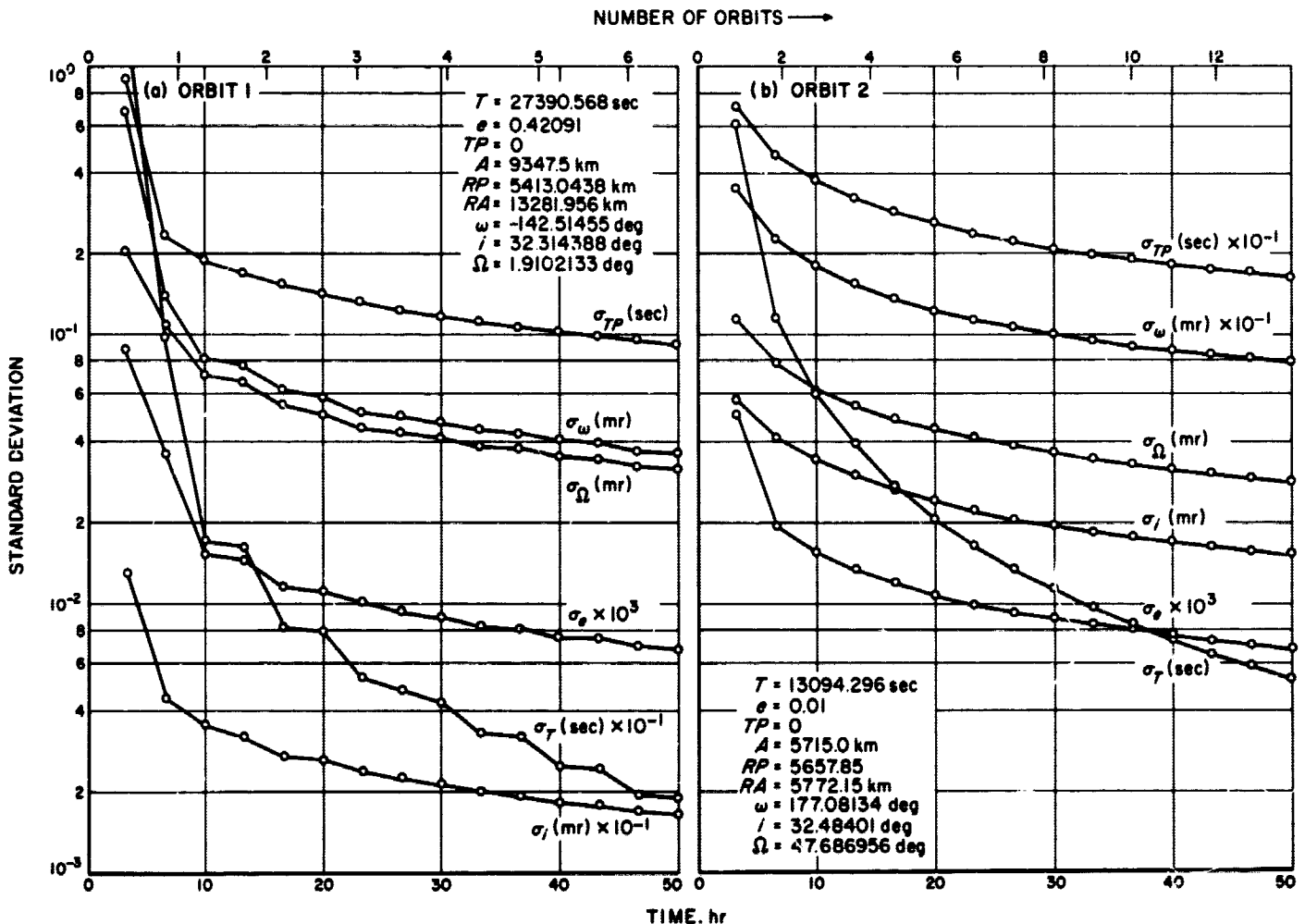


Fig. 2. Standard deviations of elliptical orbit elements vs measurement time

provement after the first orbit is about the same for all the elements except for the orbital period  $T$ , which has a greater rate.

- (2) The accuracy of determining the orientation of the orbital plane, i.e., the angles  $\Omega$  and  $i$ , appears to be relatively independent of the size and shape of the elliptical orbit, e.g., at 10 hr,  $\sigma_i = 0.0354$  and  $0.0346$  mrad, and  $\sigma_\Omega = 0.0713$  and  $0.0634$  mrad, respectively, for Orbits 1 and 2.
- (3) The accuracy of determining the orbital period appears to depend more on the number of orbits than on the total measurement time, e.g., at about 2.65 orbits, elapsed time = 20 and 10 hr, and  $\sigma_T = 0.0784$  and  $0.0608$  sec, respectively, for Orbits 1 and 2.
- (4) The accuracy of determining the eccentricity  $e$  appears to be independent of the orbit, e.g., at 10 hr,  $\sigma_e \times 10^3 = 0.0152$  and  $0.0156$  for Orbits 1 and 2, respectively.
- (5) The accuracy of determining the argument of periaapsis  $\omega$  and the time of periaapsis passage  $T_P$  appears to decrease with decreasing eccentricity  $e$ . This conclusion is not strictly supported by the data presented, because of the large changes in other parameters. Other unpublished studies do support this conclusion.

Orbital size cannot be determined from the measurement data alone because the measurements depend in time only on the true anomaly  $\eta$  (see Fig. 1). Size can be obtained from the orbital period  $T$  if the gravitational constant  $GM$  is known. The effects of the accuracy of prior knowledge of  $GM$  are discussed below.

The standard deviations of the periaapsis range  $R_P$  and the apoapsis range  $R_A$  for Orbit 1 are presented in Fig. 3. Both  $R_P$  and  $R_A$  are functions of  $T$ ,  $GM$  and  $e$ , and therefore, contain size and shape information. Errors in  $R_P$

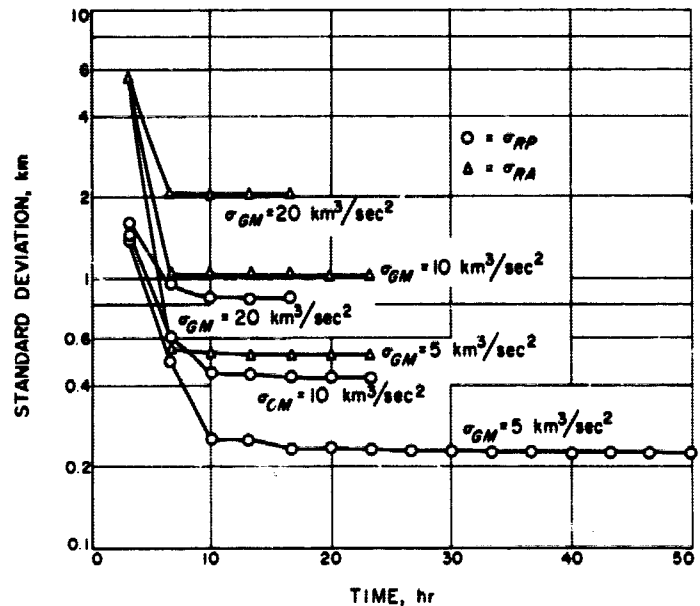


Fig. 3. Standard deviations of  $R_A$  and  $R_P$  for Orbit 1 vs measurement time for different values of  $\sigma_{GM}$

and  $R_A$  result from errors in  $T$ ,  $GM$  and  $e$ . Since  $GM$  cannot be obtained from the measurement data, the standard deviations of  $R_P$  and  $R_A$  will be dependent on, and be limited by, the prior knowledge of the standard deviation of  $GM$ . The limiting standard deviations of  $R_P$  and  $R_A$  are given by Eqs. (3) and (4).

$$\sigma_{RP} (\text{Limit}) = \frac{R_P}{3GM} \sigma_{GM} \quad (3)$$

$$\sigma_{RA} (\text{Limit}) = \frac{R_A}{3GM} \sigma_{GM} \quad (4)$$

In Fig. 3 most of the curves terminate at times less than 50 hr, because the computation used to obtain the covariance matrix (i.e., the second moment matrix of the errors) seems to become very sensitive to digital computation roundoff. The values at early termination points are found to be very close to their corresponding limiting values.

## IV. Guidance and Control Research

### A. Optical Physics Research

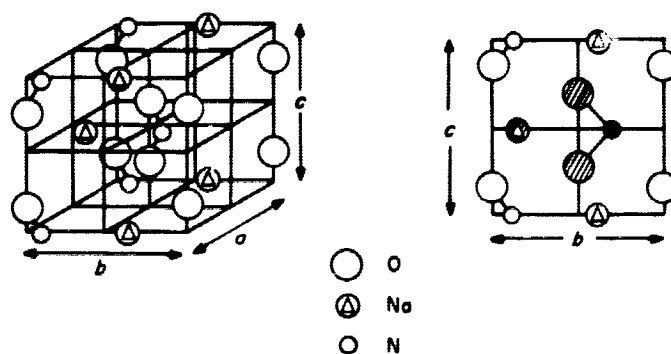
T. Tomotsu and A. R. Johnston

#### 1. Correlation Between Optical and Dielectric Properties of the $\text{NaNO}_2$ Crystal

**a. Introduction.** In 1958 it was discovered that  $\text{NaNO}_2$  was ferroelectric below the transition point of  $163^\circ\text{C}$  (Ref. 1); since that time, many detailed investigations have been carried out on its various properties, such as its dielectric (Refs. 2, 3, and 4), thermal (Refs. 5 and 6), optical (Refs. 7 and 8), and other properties (Refs. 9 and 10). Above the transition point, its crystal structure (Refs. 11 and 12) is nonpolar and belongs to the orthorhombic system, space group  $D_{2h}^{25} - I m m m$ . Below the transition temperature, the crystal structure is still orthorhombic, but is polar along the  $b$ -axis, space group  $C_{2v}^{20} - I m m$ .

The  $\text{NO}_2$  ion, a symmetric nonlinear group, lies parallel to the  $b$ - $c$  plane, as shown in Fig. 1. The bisectrix of the angle O-N-O is aligned parallel to either the  $+b$ - or  $-b$ -axis with equal probability above the Curie point, whereas it is aligned along, say, the  $+b$ -axis in a single domain below the critical point. This is a kind of order-disorder transition brought about by thermal agitation. This change of the long-range order with temperature has

recently been investigated (Ref. 4). The disorientation of the  $\text{NO}_2$  group at the transition may occur through: (1) the oscillation of the  $\text{NO}_2$  ions around the  $a$ -axis, (2) the oscillation of the  $\text{NO}_2$  ions around the  $c$ -axis, or (3) the tunneling of the N atoms between the two O atoms.



SHADED ATOMS ARE AT  $1/2 a$ ;  
 N-O DISTANCE = 1.23 Å;  
 O-N-O ANGLE =  $116^\circ$  (REFS. 11 AND 12)  
 $a = 3.55$  Å  
 $b = 5.56$  Å  
 $c = 5.38$  Å

Fig. 1. Crystal structure of  $\text{NaNO}_2$



Sato, et al. (Ref. 9) supported Process (1), based upon their observation of polarized infrared absorption. The absorption of the fundamental  $\nu_2$  of the nitrite ions in the  $\text{NaNO}_2$  crystal has a peak at about  $830\text{ cm}^{-1}$  and is effective for light polarized with its electric vector  $E$  along the bisectrix of the O-N-O angle. The absorption spectra were investigated over the temperature range between 25 and  $190^\circ\text{C}$  with polarized infrared radiation. The intensity of absorption for light polarized in the  $b$ -direction became smaller and that in the  $c$ -direction became larger with increasing temperature, whereas that along the  $a$ -direction did not show significant change. These results supported Process (1), i.e., vibration of the  $\text{NO}_2$  ion around the  $a$ -axis. No evidence could be found to prove the existence of Process (3).

**b. Proposed experiments.** The purpose of the work reported here was to confirm the above-mentioned process by means of an investigation of the ultraviolet absorption of  $\text{NaNO}_2$ . The absorption spectrum of the  $\text{NaNO}_2$  single crystal and its polarization were investigated in detail (Ref. 8).  $\text{NaNO}_2$  has an absorption band at around  $350\text{ m}\mu$ . This is due to an electron transition within the  $\text{NO}_2$  ion and is mainly polarized in the direction of the  $a$ -axis; i.e., polarized light with  $E$  along the  $a$ -axis is absorbed. Therefore, it is expected that information about the behavior of the  $\text{NO}_2$  ion might be obtained from the change of the polarized absorption; i.e., if Process (1) is the actual one, there will be no significant change in the  $a$ -polarized absorption, but if Process (2) is the actual one, the absorption in the  $b$ -direction should become larger. In the case of Process (3), presumably there will be a large change in the absorption in every direction.

For measurement of optical absorption, the apparatus shown in Fig. 2 was constructed. The high-intensity

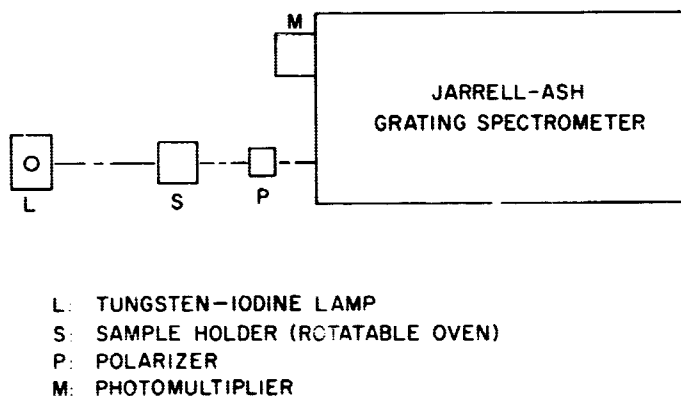


Fig. 2. Optical apparatus for measurement of absorption spectrum

tungsten-iodine lamp (General Electric Company airport lamp, Model 6.6A/T3Q/ICL) is adequate as a near-ultraviolet continuous light source (300 to  $400\text{ m}\mu$ ). The sample crystal is put inside a rotatable oven capable of maintaining constant temperatures up to  $180^\circ\text{C}$ . Under various conditions of crystal orientation and temperature, the polarized absorption spectrum is measured with a Jarrell-Ash grating spectrometer and detector system having a 1P28 photomultiplier. A Glan-Thompson prism is used as a polarizer. The measurement of polarized absorption is of special interest in a narrow temperature range around the Curie point, because anomalous dielectric properties have been reported there (Ref. 4).

Since the  $\text{NaNO}_2$  crystal is a new ferroelectric material, its electro-optical properties have not yet been investigated. Another purpose of this work is to get new data on the electro-optical properties of  $\text{NaNO}_2$  crystals using a polarimeter method (Ref. 13).

**c. Crystal growth.** The elements of the Bridgman furnace (Fig. 3) constructed for growing single crystals of  $\text{NaNO}_2$  are as follows: (1) innermost is the heater assembly which is wound of Nichrome wire, spaced to produce a specific temperature gradient; (2) enclosing the heater assembly is a glass Dewar to provide thermal insulation for the heater assembly; (3) outermost is a cylinder of Lucite tubing to protect the glass Dewar from breakage.

The furnace was made transparent in order to visually monitor the crystal-growing process. Occasionally, the crystal growth will be started and will have many inclusions which are visible. Rather than having to let the crystal grow to completion (70 to 80 hr), the inclusions can be detected early in the growing process, the crystal re-melted, and the process started anew.

A typical temperature distribution of the furnace is shown in Fig. 4. Temperature is controlled by a proportional temperature controller (Winsco proportional temperature controller, Model 4053-1) at the melting point of  $\text{NaNO}_2$  ( $277^\circ\text{C}$ ). Temperature fluctuation was less than  $0.1^\circ\text{C}$  over a period of 2 to 3 days.

The crucible is drawn from Pyrex glass tubing, a hook is formed on the upper end, and the lower end is slowly tapered to a point. Although the lower tapered end is essential to start the crystal growth from a single nucleus, the shape of the taper does not appear to be important.

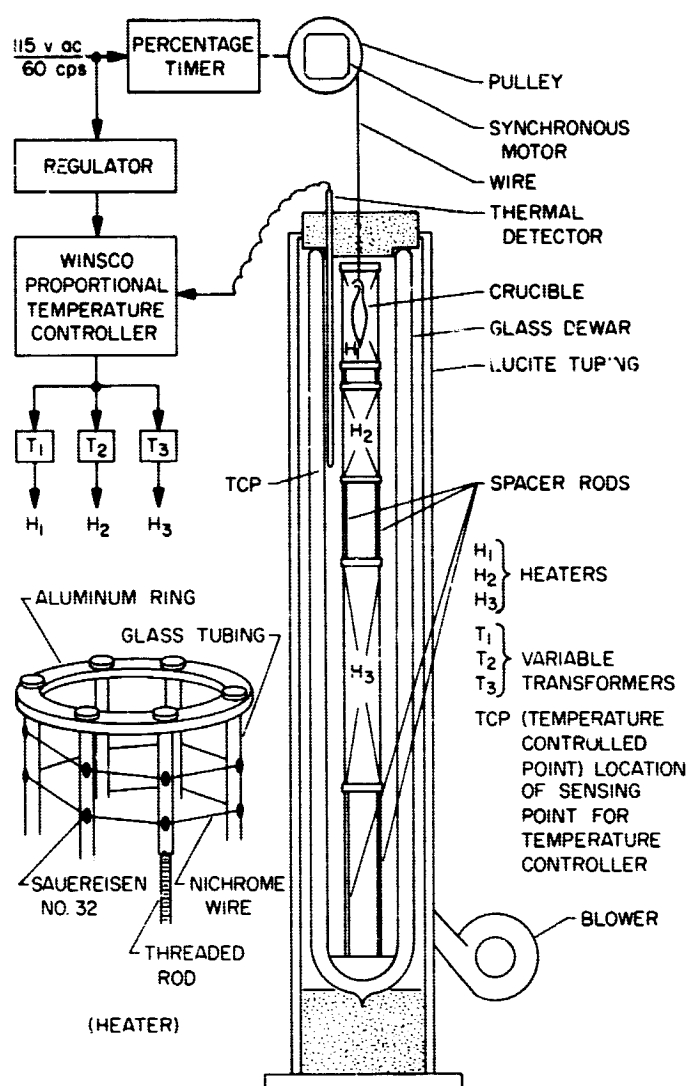


Fig. 3. Construction of Bridgman furnace

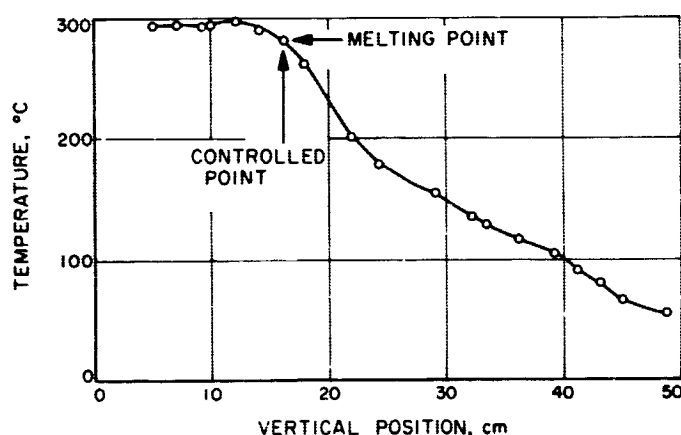
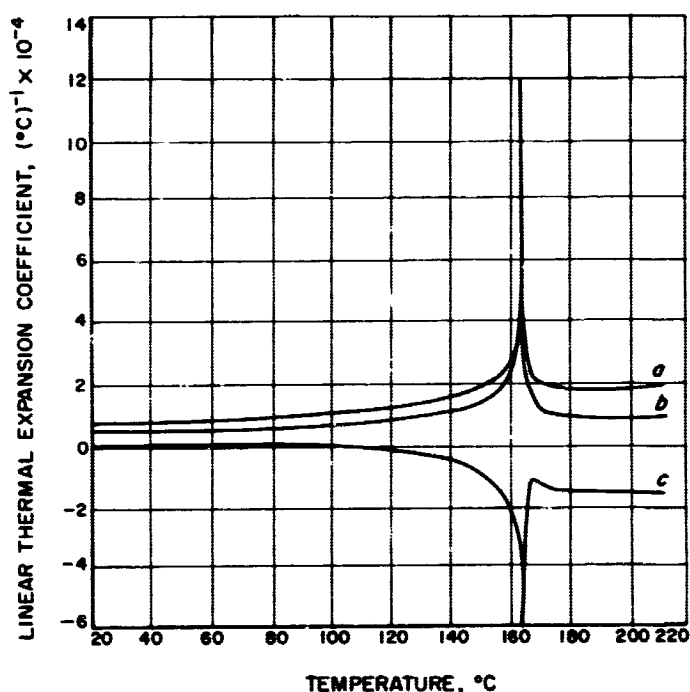


Fig. 4. Temperature distribution of the Bridgman furnace

The crucible is suspended in the furnace by a thin wire cable; the upper end of the cable is wound on a spool which is mounted on the shaft of a synchronous motor. The crucible is lowered in the furnace when the cable unwinds from the rotating spool. In order to achieve various lowering rates, the motor was pulsed on and off by means of a percentage timer, a time-switching device which has an on-time variable from 4 to 96%. It was possible to obtain lowering rates from 2 to 20 mm/hr using this system. In most cases the rate of 2 mm/hr, corresponding to  $0.5^{\circ}\text{C/hr}$ , was used during the recrystallization process. After complete crystallization, the lowering rate was increased up to a factor of 10. The  $\text{NaNO}_2$  crystals have a singular expansion coefficient (Fig. 5); i.e., the crystal expands in the direction of the  $c$ -axis when the temperature is decreased, especially near  $165^{\circ}\text{C}$ . Therefore, some cracks will appear near that temperature, but, since the fragments are still large enough for this experiment, no precautions have been taken to prevent crystal fractures.

The  $\text{NaNO}_2$  used was first purified by means of recrystallization and was outgassed at 80 to  $100^{\circ}\text{C}$  for several hours. The crucibles were not sealed off during crystal growth. In most cases the crystals grew with the  $c$ -direction roughly vertical.

Fig. 5. Linear thermal expansion coefficients in the direction of the crystal axes of  $\text{NaNO}_2$

Orientation of samples could be conveniently accomplished because the single crystal has cleavage planes along  $(101)$  and  $(10\bar{1})$ . The angle between these two planes is 67 deg, with the  $b$ -axis being the intersection of these two planes. The  $c$ -axis is the bisector of the angle between the two planes, and the  $a$ -axis is perpendicular to the  $b$ - and  $c$ -axes (Fig. 6).

Since the absorption of  $\text{NaNO}_2$  at  $360\text{ m}\mu$  is very strong, unusually thin samples ( $< 0.1\text{ mm}$  thick) are required for the absorption measurement. The crystal can be cut by a wetted thread-string saw as thin as  $1\text{ mm}$ . Then, thin plates are polished on filter paper moistened with alcohol containing a small amount of water. By this method it has been possible to make a  $0.07 \times 3 \times 3\text{-mm}$  plate of  $\text{NaNO}_2$ .

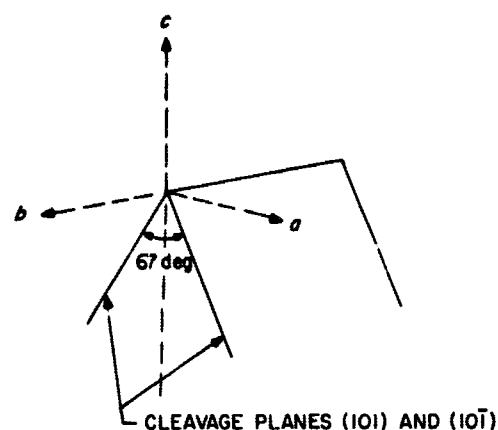


Fig. 6. Relationship between the crystal axes and cleavage planes

## References

1. Sawada, S., Nomura, S., Fujii, S., and Yoshida, I., *Physical Review Letters*, Vol. 1, p. 320, 1958.
2. Sawada, S., Nomura, S., and Asao, Y., *Journal of the Physical Society of Japan*, Vol. 16, p. 2207, 1961.
3. Nomura, S., *Journal of the Physical Society of Japan*, Vol. 16, p. 2240, 1961.
4. Yamada, Y., Shibuya, I., and Hoshino, S., *Journal of the Physical Society of Japan*, Vol. 18, p. 1594, 1963.
5. Nomura, S., *Journal of the Physical Society of Japan*, Vol. 16, p. 1352, 1961.
6. Yoshida, I., and Sawada, S., *Journal of the Physical Society of Japan*, Vol. 16, p. 2467, 1961.
7. Nomura, S., Asao, Y., and Sawada, S., *Journal of the Physical Society of Japan*, Vol. 16, p. 917, 1961.
8. Makishima, S., Tomotsu, T., et al., *Proceedings of the International Conference on Luminescence*, New York, p. 443, 1962.
9. Sato, Y., Gesi, K., and Takagi, Y., *Journal of the Physical Society of Japan*, Vol. 16, p. 2172, 1961.
10. Asao, Y., Yoshida, I., Ando, R., and Sawada, S., *Journal of the Physical Society of Japan*, Vol. 17, p. 442, 1962.
11. Carpenter, G. B., *Acta Crystallographica*, Vol. 5, p. 132, 1952.
12. Truter, M. R., *Acta Crystallographica*, Vol. 7, p. 73, 1954.
13. Williamson, S. J., Weingart, J. M., and Andrews, R. D., *Journal of the Optical Society of America*, Vol. 54, p. 337, 1964.

## ENGINEERING MECHANICS DIVISION

### V. Materials

#### A. Effect of Plastic Deformation on Preferred Orientation of Synthetic Carbons

*D. B. Fischbach*

The easy mode of plastic deformation of graphite and turbostratic (disordered stacking) carbons is generally believed to be shear on the layer planes. Shear on a single set of slip planes causes these planes to rotate toward the tensile axis. Tensile deformation, then, may be expected to produce an orientation texture in which the layer planes tend to lie parallel to the tensile axis. Observations have been made of the effect of high-temperature heat-treatment alone, and high-temperature tensile deformation on the preferred orientation of three types of carbon: (1) glassy carbon [a vitreous ungraphitized material which is nearly isotropic (Refs. 1, 2)], (2) conventional pitch-coke composite carbons (fairly well graphitized with an initial orientation texture determined by the fabrication parameters), and (3) pyrolytic carbons, commonly called pyrolytic graphites (initial ungraphitized, transversely isotropic "wrinkled sheet" structure).

The preferred orientation texture was investigated by two techniques. For the pyrolytic carbons, an X-ray diffraction method (Refs. 1, 3) was used. The normalized

orientation density distribution of the layer plane normals is determined as a function of the angle  $\omega$  measured with respect to the normal to the deposition plane. The function  $\cos^n \omega$  is fitted to the distribution and  $n$  employed as an index of preferred orientation. While preferred orientation is easily determined by this method in pyrolytic carbons, X-ray techniques are more difficult for the more nearly randomly oriented glassy and conventional carbons and were not used for them. For all three types of carbon, measurements of the diamagnetic susceptibility  $\chi$ , were employed. The susceptibility of turbostratic carbons and graphite is highly anisotropic. For single crystal graphite the susceptibility measured perpendicular to the layer planes  $\chi_c$  is about 21.5, while parallel to the layer planes  $\chi_a$  it is only 0.32 (all susceptibility values are reported in units of  $-10^{-6}$  emu/g). The single crystal anisotropy ratio  $\chi_c/\chi_a$  is 68. Therefore, magnetic susceptibility measurements in various directions in a polycrystalline sample provide a sensitive method of detecting preferred orientation.

It is well established that  $\chi_c$  is a function of the structure of the sample (crystallite diameter and layer stacking order). Although  $\chi_a$  is expected to be insensitive to structure, this has not been firmly established experimentally. Therefore, care must be used in interpreting the results. The structure dependence of the susceptibility may be monitored by observing the total (trace) susceptibility  $\chi_T$

(sum of the  $\chi$  values measured in three orthogonal directions), which is independent of anisotropy. For single crystal graphite  $\chi_T = 22.1$ . For pyrolytic carbons  $\chi_T$  of as-received materials may be as high as 34. It decreases with increasing layer ordering to a minimum value of about 20 and finally recovers to the graphite value in well-ordered material. The pyrolytic carbon used here had been heat-treated for a half hour at 2800°C prior to testing and was in the minimum susceptibility condition. For glassy and conventional carbons  $\chi_T$  is initially low and increases with increasing crystallite diameter. There is sometimes a slight maximum ( $\approx 22$ ), but no minimum and  $\chi_T$  levels out at a value less than that of graphite (usually in the range 19–21).

The preferred orientations of samples taken from the undeformed butt sections were compared with those of samples from the corresponding gage sections of tensile specimens. Since both butt and gage samples had undergone heat treatment during deformation, measurements were also made in some cases on as-received samples to determine the effect of heat treatment alone. The susceptibility was measured parallel to the tensile axis and in

two orthogonal directions at right angles to this axis. The samples were not, in general, isotropic in the plane perpendicular to the tensile axis, so maximum and minimum values in this plane were determined. Two anisotropy ratios were computed from these data:  $A_1$ , the ratio of the maximum susceptibility perpendicular to the tensile axis to the susceptibility parallel to this axis, is a measure of the tendency of the layer planes to lie parallel to the tensile axis.  $A_2$ , the ratio of the maximum susceptibility to the minimum susceptibility, both measured perpendicular to the tensile axis, measures any tendency for the layer planes to lie parallel to a particular plane (such as the deposition plane in pyrolytic carbons) containing the tensile axis. For the X-ray diffraction measurements on the pyrolytic carbons, rod-shaped samples were cut perpendicular and parallel to the tensile axis in the deposition plane. The perpendicular samples yield  $n_1$  and the parallel samples (the type used in previous investigations) yield  $n_2$  where the significance of the subscripts is similar to that for the  $A$ 's.

Representative results are given in Table 1. While heat treatment alone affects the total susceptibility, it produces

Table 1. Effect of heat-treatment and tensile deformation on the preferred orientation texture of various carbons and graphites

Type of carbon	Pretest heat-treatment, °C	Sample	Deformation		$-10^{-3} \chi_T$ , emu/g	$A_1$	$A_2$	$n_1$	$n_2$
			Temp, °C	Elongation, %					
Glassy	2000	As rec'd	—	—	9.0	1.02	—	—	—
		Butt	2700	—	12.4	1.01	—	—	—
		Gage	2700	24	14.6	1.82	—	—	—
Pitch-coke B	2200	As rec'd	—	—	19.8	1.20	1.14	—	—
		Butt	2700	—	20.3	1.18	1.14	—	—
		Gage	2700	30	20.0	1.45	1.01	—	—
		As rec'd	—	—	19.2	1.26	1.23	—	—
		Butt	2475	—	20.3	1.23	1.23	—	—
		Gage	2475	7	20.9	1.29	1.17	—	—
		As rec'd	—	—	20.3	1.23	1.17	—	—
		Butt	2475	—	20.0	1.23	1.17	—	—
		Gage	2475	$\sim 1.5$	20.2	1.23	1.16	—	—
Pitch-coke C	2500	As rec'd	—	—	19.0	0.81	1.0	—	—
		Butt	2700	—	20.4	0.82	1.03	—	—
		Gage	2700	45	20.5	1.34	1.01	—	—
		As rec'd	—	—	19.2	0.86	1.03	—	—
		Butt	2475	—	20.2	0.88	1.04	—	—
		Gage	2475	15	20.5	1.02	1.01	—	—
Pyrolytic (regenerative)	2800 ½ hr	Butt	2900	—	19.8	8.3	—	12	—
		Gage	2900	100	21.7	41.3	11.1	3400	29
		Butt	2700	—	20.2	8.2	—	—	—
		Gage	2700	17	21.4	42.4	9.6	1150	18
		Butt	2500	—	20.4	8.0	—	—	—
		Gage	2500	6	20.3	19.8	11.3	200	18.5
		Butt	2200	—	19.9	8.6	—	—	—
		Gage	2200	$\sim 2.7$	20.2	9.0	8.4	12.5	11.5
		Scrap	—	0	20.6	7.8	—	12.5	—

little change in the preferred orientation for the glassy and conventional carbons. This is also true for the pyrolytic carbon used here, which had a regenerative microstructure. However, in all cases, tensile deformation of  $\approx 5\%$  elongation produced a definite change in preferred orientation which increased with increasing deformation. The deformation-induced orientation consisted of a tendency for the layer planes to align parallel with the tensile axis, as expected. Frequently, deformation also increased  $\chi_r$ .

The glassy carbons were initially nearly isotropic, but deformation produced anisotropies as high as 1.8. The marked increase in  $\chi_r$  is probably a result of crystallite growth. The pitch-coke materials had a definite initial preferred orientation resulting from the uniaxial pressure used to form the blocks prior to baking. The layer planes tended to the oriented perpendicular to the pressing direction. Samples in Group B were cut with the tensile axis perpendicular to the pressing direction (parallel to the layer planes), so that  $A_1 > 1$ . The C samples were cut with the tensile axis parallel to the pressing direction (perpendicular to the layer planes) making  $A_1 < 1$ . Tensile deformation increased the existing preferred orientation texture in the B samples. In the C samples, deformation tended to rotate the preferred orientation axis by 90 deg, making  $A_1$  approach and finally exceed 1.0.

The behavior of the pyrolytic carbon is especially interesting. The structure of the undeformed material may be represented by a stack of wrinkled layer planes. It is transversely isotropic in the plane of the deposit, and the tensile axis lies in this plane. The magnetic and X-ray determinations of preferred orientation are in good qualitative agreement at small elongations. However, above 15% elongation the  $A$  values show little change, whereas the  $n$  values continue to increase rapidly. The reason for this is not known, though in this range  $n$  is a much more rapidly increasing function of preferred orientation than  $A$  is. Both types of measurement show that tensile deformation is much more effective in dewrinkling the planes in a direction parallel to the tensile axis than perpendicular to it. The structure of the deformed gage sections might therefore be represented by a stack of corrugated layer planes with the corrugations parallel to the tensile axis. Less extensive tests on a pyrolytic carbon with a substrate nucleated structure gave qualitatively similar results, though the  $n$  values were somewhat higher. Although this anisotropic effect of tensile deformation might have been anticipated on the basis of the present understanding of the structure and deformation processes in pyrolytic carbons, this appears to be the first experimental demonstration of it.

Because of the strong anisotropy of most of the physical properties of graphite, the degree of preferred orientation is an important factor in applications of this material. The present results demonstrate that qualitative information on preferred orientation can be obtained rather easily from magnetic susceptibility measurements. However, before this technique can be employed confidently for quantitative work, a more detailed understanding of the dependence of the principal susceptibilities on crystallite size and layer stacking order is required. The results reported here on the effect of tensile deformation on preferred orientation in glassy and conventional carbons are consistent with the hypothesis that basal plane shear is an important component of the deformation of disoriented carbons and graphite. For pyrolytic carbons the results may be explained more conveniently in terms of dewrinkling (removal of tilt boundaries), though basal plane shear is probably also involved.

## B. Growth and Perfection of Crystals

P. J. Shlichta

### 1. Comparison of the Phase Rule, the Crystallographic Contact Rule, and Euler's Boundary Rule

It has recently been discovered (Ref. 5)<sup>1</sup> that, when a crystal structure is considered as a space-group array of rigid spheres in contact, the relationship

$$F = P - C + 1$$

holds, wherein  $F$  equals the number of degrees of freedom of the structure in parameter space,  $P$  is the number of variable parameters (atomic-radius ratios, unit cell axial ratios and angles, and atomic position parameters), and  $C$  is the number of contact modes. This "Contact Rule" bears a close resemblance to Gibbs' Phase Rule

$$F = C - P + 2$$

wherein  $F$  is the number of degrees of freedom of the system in composition-temperature-pressure space,  $C$  is the number of components in the system, and  $P$  is the

<sup>1</sup>A more detailed exposition is in preparation.

number of phases present. Moreover, Euler's Boundary Rule,

$$F = E - C + 2$$

wherein  $F$ ,  $E$ , and  $C$  are respectively the number of faces, edges, and corners of any polyhedron, has a similar form. These *sum rules* are much simpler than the relationships generally encountered in mathematics and physics, and it seems appropriate to determine whether their superficial similarity is indicative of a generic relation.

The Phase Rule and the Contact Rule are, in fact, special cases of a trivial axiom of mathematics which might be termed the Simultaneous-Equation Sum Rule, namely that for  $n$  single-valued equations involving  $m$  variables, the number of independent variables is equal to  $m - n$ . The Contact Rule is merely an expression of the fact that, for each contact mode in the structure, there exists an equation stating that the sum of two atomic radii is equal to the distance between their centers. Similarly the Phase Rule rests on the assumption that, at equilibrium, the chemical potentials of all phases present are equal (Ref. 6); hence there must be  $P$  equations of the form

$$K = \mu_p(C_1, C_2, C_3 \cdots C_c, T, P)$$

The constants in each of these rules stand for implicit variables: temperature and pressure in the case of the Phase Rule and volume expansion (i.e.,  $a_0/R_n$ ) in the case of the Contact Rule.

A sum rule of the simultaneous-equation type is possible in any physical or mathematical system involving a set of simultaneous equations, each of which denotes the existence of a member of a class of enumerable entities. Most of these, however, are probably too trivial or of too little utility to merit expression.

Euler's Boundary Rule, however, despite assertions to the contrary (Ref. 7), is generically unrelated to the simultaneous-equation sum rules. Not one of the many derivations of this rule involves simultaneous equations (Ref. 8). There is no term corresponding to degrees of freedom and the constant cannot be equated to any implicit variables. Euler's Rule is in fact the special three-dimensional case of a more general  $n$ -dimensional relation (Ref. 9)

$$\sum_{d=0}^{n-1} N_d (-1)^d = 1 - (-1)^n$$

so that even the superficial resemblance to simultaneous-equation sum rules is merely coincidental.

The conceptual uniqueness of Euler's Rule is of some significance in connection with its application to other sum rules. Any simultaneous-equation sum rule can be represented diagrammatically as an  $n$ -dimensional space divided into polytopes. Therefore any such sum-rule diagram must obey some form of Euler's Rule. For example, the contact diagram of a two-parameter crystal structure may be regarded as a polyhedron in which the area outside the border of the diagram counts as a face, the segments of the border proper to each contact field count as edges, and all line intersections count as vertices. Moreover, since (with a few special exceptions) all vertices are trihedral, a special case of Euler's Rule holds

$$2F = C + 4$$

Therefore, if the number of contact modes is known, it is possible to predict the number of triple points that will be found within the diagram.

Another possible joint application of Euler's Rule and the Contact Rule involves those complex metallic crystal structures which can be interpreted as the packing of spheres into successive shells of concentric polyhedra. By a suitable choice of assumptions (e.g., that each sphere constitutes a corner (or face) of a polyhedron, that each intralayer contact constitutes an edge, that each sphere must be in contact with at least three atoms in the shell beneath it, etc.) it may be possible to use the two sum rules to determine those sets of polyhedra which correspond to optimal packing.

## C. Graphite

W. V. Kotlensky

### 1. Glassy Carbon

In the previous summary tensile results were presented for two lots of glassy carbon, one reported by the manufacturer to have been heat-treated at 2000°C (Lot 228) and the other at 3000°C (Lot 229). The strengths of these two lots over the range from room temperature to 2900°C were not markedly different. The ductility of Lot 228, on the other hand, was appreciably greater than Lot 229.

To test the effect of annealing, Lot 228 specimens were heat-treated for 1 hr at 3000°C in a graphite, resistance tube furnace, cooled to room temperature, and then pulled to fracture in the tensile testing apparatus. The results of tests made at 2500, 2700, and 2900°C are given in Table 2. The tensile data are in fair agreement with Lot 229 reported in the last issue. The lack of ductility in glassy carbon which has been heat-treated at 3000°C cannot be explained on the basis of thermal annealing alone. Unannealed Lot 228, tested at 2900°C, shows considerable ductility, and the difference observed cannot be attributed to the 100°C higher heat treatment. Other factors, such as crack formation during cyclic heating, may contribute to the lack of high ductility in glassy carbon Lot 229 and Lot 228 heat-treated at 3000°C.

Pitch-coke graphites show a decrease in bulk density and an increase in porosity with tensile deformation (Ref. 10). Recent structural studies on the gauge section of deformed specimens show that glassy carbon undergoes a similar change in bulk density and porosity. Such data for glassy carbon, as well as Knoop microhardness and unit cell height ( $C_0$ ) and width ( $a_0$ ) are given in Table 3. The density data for Lot 228 may be interpreted to indicate that pore formation accompanies deformation and that the pores which are formed are open pores. The lack of appreciable density change for Lot 229 is due to the low deformations in this material.

Hardness measurements on carbon and graphite are difficult to make because of the resiliency in these materials. One method which has been widely used involves the spraying of a thin layer of paint or plastic on the surface and relating hardness to the imprint left in the coated surface. This method has the disadvantage that the coating is often too thick, irregular, and leads to erroneous values. In the present method, a thin film of silver approximately 1000 Å thick was flash-coated in a vacuum onto the glassy carbon surface. This method was found to give better reproducibility than the painted surface method.

The Knoop microhardness data (Table 3) show that both lots of glassy carbon become softer with deformation, reaching a limiting hardness number of approximately 114. A decrease in hardness in graphite has been associated with graphitization. Glassy carbon, as indicated by the unit cell dimension, does not graphitize. Franklin (Ref. 11) has reported that carbons showing a limiting unit cell height ( $C_0$ ) of approximately 6.88 Å are nongraphitic and contain appreciable crosslinking.

**Table 2. Tensile properties of glassy carbon Lot 228, heat-treated at 3000°C for 1 hr prior to testing**

Test temperature, °C	Ultimate strength, psi	Recorded elongations, %
2500	11,400	0.5
2500	17,400	4.1
2700	20,200	4.7
2900	16,900	8.7

\*Corrected for deformation in the filleted regions. Strain rate  $2 \times 10^{-4}$  in./in./sec.

**Table 3. Structural properties of glassy carbon**

Treatment	Density, g/cm <sup>3</sup>		Knoop micro-hardness (100 g)	C <sub>0</sub> , Å	a <sub>0</sub> , Å
	Bulk	True*			
Lot 228					
As received	1.50	1.68	192	7.00	2.43
Heated to 3000°C	1.50	1.66	127	6.87	2.44
13.5% strain at 2200°C	1.41	1.71	145	6.91	2.44
23.5% strain at 2700°C	1.28	1.76	117	6.90	2.44
33.6% strain at 2700°C	1.29	1.82	114	6.84	2.43
23.6% strain at 2900°C	1.37	1.74	112	6.87	2.44
Lot 229					
As received	1.50	1.68	127	6.86	2.45
1.1% strain at 2200°C	1.54	1.65	—	6.86	2.45
2.2% strain at 2500°C	1.46	1.66	119	6.87	2.45
5.5% strain at 2700°C	1.44	1.66	116	6.86	2.44
4.8% strain at 2900°C	1.44	1.65	115	6.87	2.45
*Measured by kerosene displacement.					

\*Measured by kerosene displacement.

The change in hardness with heat-treating and deformation of glassy carbon may in some way be associated with the extent of crosslinking.

## 2. Pyrolytic Graphite

To the writer's knowledge, no report heretofore has been made of any maximum in the strength of pyrolytic graphite as a function of temperature. Tensile tests are reported to have been made at temperatures as high as 3058°C (Ref. 13) without showing any dropoff in strength with temperature. Recent improvements in the graphite tensile testing apparatus has made it possible to extend the test temperature from a previous maximum of 2760 to 2900°C. Preliminary tensile tests on pyrolytic graphite made at temperatures up to and including 2900°C indicate that the strength parallel to the substrate definitely decreases at temperatures above 2700°C.



## D. Properties of Tungsten at High Temperatures

J. L. Taylor

Study of the high temperature tensile behavior of tungsten (Refs. 13, 14) has continued.

The effect of grain size and impurity level on the tensile strength and ductility of powder-metallurgy tungsten was determined by comparing recrystallized coarse- and fine-grained undoped tungsten and recrystallized fine-grained doped tungsten. Tests were conducted in vacuum at a strain rate of 0.02/min from 2500 to 5000°F. Both fine-grained materials, about the same strength, were

stronger than the coarse-grained. Ductility for the doped fine-grained tungsten (higher impurity level) was lower than for the undoped fine-grained tungsten in which stress-induced grain growth occurred more readily. Stress-induced grain growth correlates in a general way with ductility measured by either reduction in area or elongation. High and low values of stress-induced grain growth are associated with high and low values of ductility, respectively. In the intermediate temperature region between approximately 50 and 65% of the absolute melting point where a ductility minimum occurs, grain boundary sliding and its effects, void nucleation and growth, and intercrystalline fracture appear to predominate to lower ductility in both fine-grained materials. A detailed discussion of this recently completed work appears in Ref. 15.

## References

1. Kotlensky, W. V., "Glassy Carbon," SPS 37-24, Vol. IV, p. 45, December 31, 1963; SPS 37-26, Vol. IV, p. 70, April 30, 1964, Jet Propulsion Laboratory, Pasadena, California.
2. Yamada, S., Satoh, H., and Ishii, T., "Eigen schaften und Verwendung von 'Glassy Carbon'," Tokai Electrode Manufacturing Co., Tokyo, October, 1963.
3. Fischbach, D. B., "Kinetics of Graphitization," SPS 37-24, Vol. IV, pp. 44, 45, December 31, 1963, Jet Propulsion Laboratory, Pasadena, California.
4. Guentert, O. J., and Klein, C. A., "Preferred Orientation and Anisotropy Ratio of Pyrolytic Graphite," *Applied Physics Letters*, Vol. 2, pp. 125-127, 1963; O. J. Guentert, "X-ray Study of Pyrolytic Graphite," *Journal of Chemical Physics*, Vol. 37, pp. 884-891, 1962.
5. Shlichta, P. J., "The Contact Rule in Crystallography," *Research Summary No. 36-14*, pp. 121-122, Jet Propulsion Laboratory, Pasadena, California, May 1, 1962.
6. Findlay, A., et al., *The Phase Rule and Its Applications*, 9th Ed., pp. 17-19, Dover Publications, New York, 1951.
7. Ruuel, O., "The Phase Rules and the Boundary Rule of Euler," *Zeitschrift für Elektrochemie*, Vol. 35, p. 54, 1929.
8. Sommerville, D. M. Y., *An Introduction to the Geometry of N Dimensions*, pp. 141-160, Dover Publications, New York, 1958.
9. Hilbert, D., and Cohn-Vossen, S., *Geometry and the Imagination*, p. 292, Chelsea Publishing Co., New York, 1952.
10. Martens, H. E., Jaffe, L. D., and Jepson, J. E., "High Temperature Tensile Properties of Graphite," *Proceedings of the Third Conference on Carbon*, Pergamon Press, 1957, pp. 529-542.

## References (Cont'd)

11. Franklin, R. E., "Crystallite Growth in Graphitizing and Non-Graphitizing Carbon," *Proceedings of the Royal Society*, Vol. A-209, pp. 196-218, 1951.
12. Bragg, R. H., Crooks, D. D., Fenn, R. W., and Hammond, M. L., "The Effect of Applied Stress on the Graphitization of Pyrolytic Graphite," *Carbon*, Vol. 1, pp. 171-179, 1964.
13. Taylor, J. L., "Tensile Properties of Pyrolytic Tungsten from 1370 to 2980°C in Vacuum," SPS 37-23, Vol. IV, pp. 53-57, (also appears as Technical Report 32-463), Jet Propulsion Laboratory, Pasadena, California, October 31, 1963.
14. Taylor, J. L., "Tensile Properties of Tungsten 3% Rhenium from 1400 to 2900°C in Vacuum," Technical Report 32-588, Jet Propulsion Laboratory, Pasadena, California, April 1, 1964.
15. Taylor, J. L., Boone, D. H., and Simmons, O. W., "Effect of Grain Size and Impurities on the Tensile Strength and Ductility of Tungsten from 2500 to 5000°F," Technical Report No. 32-632, Jet Propulsion Laboratory, Pasadena, California (in process).

## VI. Applied Mechanics

### A. Thermochemistry and Normal Shock Program

*T. Horton*

Current studies of the feasibility of planetary entry at high velocities have created a need for rapid solution techniques of aerothermodynamic problems involving the high-temperature properties of gas mixtures. Although most problems in the field of aerothermodynamics are characterized by large pressure, chemical composition, and temperature gradients (and thus by nonequilibrium phenomena), there are many processes of interest for which either the gradients disappear or the reaction rates are sufficiently fast to justify the assumption of chemical equilibrium. The thermodynamic properties and chemical composition of multicomponent, high-temperature, real gas mixtures which satisfy the conditions of chemical equilibrium, have been used in investigating normal shock, hypersonic flow field, communication-blackout, convective heat transfer, and radiative heat transfer problems.

Several numerical techniques have been adapted to the high-speed digital computer for the calculation of chemical equilibrium composition properties of gas mixtures. A survey of these techniques is given in Ref. 1.

From the aerothermodynamic point of view one of the most important, yet simplest, applications of the chemical equilibrium calculation is the normal shock problem. In the past, for air calculations, the procedure has been to generate thermodynamic data and use it for a separate calculation of normal shock parameters. However, because of the diversity of gas mixtures proposed for planetary atmospheres, and the need for data both in shock tube experimentation and free-flight entry calculations, a completely mechanized normal shock and chemical equilibrium scheme was required.

The JPL Normal Shock and Chemical Equilibrium Program is such a scheme which is programmed in Fortran for use on the IBM 7090 computer. The program is capable of calculating chemical equilibrium properties associated with traveling, standing, and reflected nor-

mal shocks with the initial gas mixture, temperature, and pressure easily varied. The program can also be used to generate Mollier thermodynamic data and chemical composition at selected temperatures and densities. The program can handle mixtures of gases in which both dissociation and ionization are important in the temperature range of 2000 to 20,000°K. The constituents of the gaseous mixtures are limited to the 36 species (Table 1), for which molal free energy and molal enthalpy (Ref. 2) are tabulated in the program library. (The range of temperature and number of chemical species are limits of the available library data and not the program.)

The chemical composition which satisfies the criterion of chemical equilibrium is obtained by solving the familiar mass balance and mass action equations using a Newton-

Raphson procedure similar to that of Ref. 3. This numerical procedure for obtaining chemical composition is incorporated into the equilibrium block of the program, as shown in Fig. 1, so that it can be used to supply Mollier and chemical composition data or be used in an iteration loop to obtain normal shock solutions. As programmed, this procedure requires less than 2 sec to yield solution concentrations for a mixture containing 36 constituents.

Two normal shock options are available. One option, the moving normal shock case, yields the velocities, thermodynamic properties, and equilibrium chemical composition behind the normal portion of the bow shock of a high velocity body. The other option, the combination normal shock case, supplies the velocities and thermodynamic properties associated with both the normal portion of the standing bow wave of a test model mounted in a shock tube, and the reflection of a shock from a plane surface, such as the end wall of a shock tube.

Table 1. Chemical species listed in the program library

No.	Formula	Molecular weight	Heat of formation, kcal/mole
1	CN	26.019	109.0
2	NO <sup>+</sup>	30.008	234.8
3	CO	28.011	-27.202
4	NO	30.008	21.477
5	NO <sub>2</sub>	46.008	8.766
6	N <sub>2</sub> O	44.016	20.309
7	CO <sub>2</sub>	44.011	-93.9643
8	N <sub>2</sub>	28.016	0
9	O <sub>2</sub>	32.000	0
10	C <sup>+++</sup>	12.0094	2095.69
11	C <sup>++</sup>	12.0099	991.689
12	C <sup>+</sup>	12.0105	429.537
13	C <sup>-</sup>	12.0115	141.0
14	CO <sup>+</sup>	28.011	295.977
15	C <sub>2</sub> N <sub>2</sub>	52.038	73.4
16	C <sub>2</sub>	24.022	196.3
17	C <sub>2</sub>	36.003	188.0
18	N <sup>+++</sup>	14.0064	2224.25
19	N <sup>++</sup>	14.0069	1130.22
20	N <sup>+</sup>	14.0075	447.564
21	N <sub>2</sub> <sup>+</sup>	28.016	359.306
22	O <sup>+++</sup>	15.9984	2449.47
23	O <sup>++</sup>	15.9989	1182.60
24	O <sup>+</sup>	15.9995	372.942
25	O <sup>-</sup>	16.00055	25.1935
26	O <sub>2</sub> <sup>-</sup>	32.000	-23.00
27	O <sub>2</sub> <sup>+</sup>	32.000	277.918
28	O <sub>2</sub>	48.000	34.639
29	A <sup>+++</sup>	39.9423	1943.49
30	A <sup>++</sup>	39.9428	1000.28
31	A <sup>+</sup>	39.9434	363.345
32	C	12.0110	169.99
33	N	14.008	112.507
34	O	16.000	58.9725
35	e	0.000549	0
36	A	39.944	0

The results of the standing and reflected shock calculation are useful in interpreting convective heating and thermal radiation experiments performed in a shock tube.

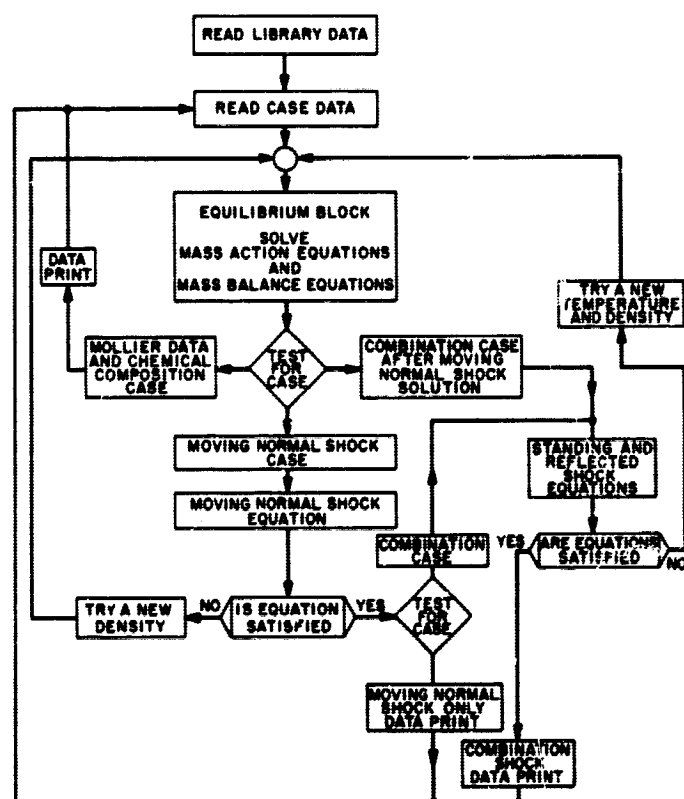


Fig. 1. Flow diagram of the thermochemistry and normal shock program

## B. Thermal Scale Modeling: Arthur D. Little Contract

J. M. F. Vickers

The third phase of the work carried out by Arthur D. Little, Inc., of Cambridge, Massachusetts, under Contract No. 950252, to demonstrate the feasibility of thermal scale modeling in steady state, has now been completed. This phase completes Contract No. 950252.

The Phase III tests were performed in three steps, using: (1) the full- and half-scale models of Phase II (Ref. 4), under modified input power conditions; (2) the Phase II models after physical modification; and (3) a new model (Ref. 5), built at half scale from Armco Iron, the prototype material, in order to check the materials preservation technique discussed in Ref. 6.

The first tests carried out in Phase III used the prototype and half-scale models of Phase II, but with a power distribution designed to produce large thermal gradients within the model. These tests have been designated the "forced gradient" tests. The same total power input was used as for the full power in Phase II, that is about 165 w into the prototype. In Phase III, however, no power was supplied to Box A (originally the box carrying approximately two-thirds of the total power input to the model), while the power was divided between the other two boxes so that the power into Box B was twice that into Box C. Since Box A was totally black, while Boxes B and C were gold-plated and only partially black, this meant that a considerable thermal gradient was developed within the model; the temperature difference between the closest thermocouples on the faces of Boxes A and B was approximately 65°C.

The second tests carried out in Phase III were made after the prototype and half-scale models of Phase II were modified. The modification consisted of milling out approximately 75% of the material forming the twelve tabs which were used to connect the three boxes of each model to the frame members. These tabs were simulating a relatively high joint conductance between boxes in Phase II, and this was reduced to about one quarter of its previous value for Phase III. Since the paint pattern on the models was designed for approximately isothermal conditions, the tests were carried out using the forced gradient power distributions, so that the effect of decreased joint conductance on modeling could be examined. The net effect on the junctions between the boxes was to decrease the conductive heat flow between

boxes by about 25%, as shown by a decrease in temperature drop across the frame, and to increase the temperature drop across the tabs themselves by about a factor of two.

The third tests carried out in Phase III were on a new model, Model 4, which was built at approximately one-half scale of the same materials as the prototype. This provided a check on the materials preservation technique proposed in Ref. 6. With this technique, the model operates at a temperature considerably above that of the prototype, the absolute temperature at any point in the model being equal to that of its homologous point in the prototype multiplied by the inverse of the scaling ratio to the one-third power. In this case, for the exact scale used (determined in Phase II by the ratios of the thermal conductivities of Armco Iron and SAE 4130), which was 1/1.78, the absolute temperature of the model was 1.212 times the absolute temperature of the prototype. The total power input to the model was 0.68 times the total full power input to the prototype, since the model laws state that the power into the model should be equal to the power into the prototype times the scaling ratio to the two-thirds power. For comparison, the model temperatures were reduced to those which the model predicted for the prototype; these predictions were then compared with the results obtained with the prototype during Phase II using the full-load power distribution.

These results from Phase III and the relevant results from Phase II (Ref. 5) are shown in Table 2.

Several important points should be noted here. First, for all three tests in Phase III it was possible to obtain satisfactory results with thermal modeling. The average error, and the errors in the maximum error regions and minimum error regions, are all higher than for Phase II, but the increase is not great considering the increased gradients experienced in this phase. Under forced gradient conditions the maximum-to-minimum temperature for a given model was 100°C, which is considerably greater than the range encountered in the bus portion of present spacecraft. Second, the test using the materials preservation technique experienced difficulties with the model running too cold, by about 4.8°C on the average. A subsequent error analysis by the contractor accounts for 3.6°C, approximately, from systematic sources, while random variations in surface emittances could account for a further 1.5°C, so that it is considered that this technique is actually better than the results in Table 2 indicate. It should, however, be pointed out that theoretically the materials preservation technique is a good one at

Table 2. Half-scale results, Phases II and III

	Full load	Twice full load	Three dimensional	Forced gradient	Reduced tab conductance	Materials preservation
Average total error, °C	0.9	2.4	1.7	3.3	3.5	4.4
Total error, %	0.31	0.67	0.55	1.02	1.09	1.39
Maximum error area (16 thermocouples)						
Average error, °C	1.8	4.7	3.1	6.7	7.3	4.7
Error, %	0.58	1.28	1.03	2.1	2.2	1.6
Minimum error area (32 thermocouples)						
Average error, °C	0.5	1.3	0.9	1.5	1.6	4.0
Error, %	0.17	0.37	0.31	0.48	0.49	1.32

half scale, but runs into considerable difficulties at smaller scales, as indicated in Ref. 6. Third, the so-called maximum error area, consisting of the sixteen thermocouples in the vicinity of the corner where the two gold-plated portions of Boxes B and C meet, gives very much worse results under forced gradient conditions, which is to be expected, since the error is due to the thermal conductivity of the plating (Ref. 5). The increased conductance heat flow in these models, where the over-all thermal conductivity is incorrectly scaled, would be expected to lead to increased errors. Fourth, in the materials preservation technique model, where the same thickness of plating was used on Armco Iron, a considerably more conductive material than SAE 4150, the so-called maximum error area is a misnomer. The difference in behavior between this area and the thermocouples in the rest of the model is not significant, since this model is far less affected by the thermal conductivity of the plating.

In summary, this contract demonstrated that it was possible to model successfully, at one-half scale, a single material prototype under vacuum and cold-wall conditions, (1) for two-dimensional flow at various power settings and power distributions; (2) for two different values of simulated joint conductance; and (3) for 3-dimensional flow, using both the temperature preservation and the materials preservation techniques. It has been possible to account for most of the discrepancies that occurred in the experiments by systematic error analyses.

Tests will also be carried out, in-house, on the models developed in Contract No. 950252, using solar simulation. A new contract, No. 950789, has now been initiated with Arthur D. Little, Inc., to build and test, at steady state and with solar simulation, a one-half scale model of the *Mariner C* Temperature Control Model.

## C. Development of Radio Frequency Transparent Energy Absorbing Structural Elements

R. K. McFarland

The first phase of a contract effort with the General Electric Company to develop fiberglass-reinforced plastic honeycomb energy-absorbing elements has been completed. The objective of this phase was to develop and test honeycomb elements which have a capability to develop high specific energy values with low electric attenuation in the S-band radio-frequency range.

Preliminary tests indicated that phenolic or polyester honeycomb possessed the ability to develop a controlled and repeatable mode of collapse when subjected to an axial loading, thus denoting a potential for use as energy absorbing elements.

A number of test specimens were fabricated using phenolic and polyester resins, with various fiberglass cloth types and layups, and in various cell sizes and resin contents. The purpose being to determine the variables which

affect the energy-absorbing characteristics of the plastic honeycomb elements. Axial static and dynamic tests were performed on the specimens, along with a determination of radio-frequency attenuation through the specimen.

Test results indicated that the phenolic honeycomb possessed the greatest energy-absorbing capability. It is of primary interest to note that the fiberglass reinforcement type or layup had little or no effect on the energy-absorbing capabilities, and that the resin type, resin content and cell size are the primary variables.

Referring to Fig. 2, the relation between specific energy and unit weight is shown for plastic honeycomb with different cell sizes, for phenolic and polyester resins. Tests are presently being conducted to determine the upper limit on these curves. Results tend to indicate that the plastic honeycomb elements, when optimized, will be able to absorb approximately 20,000 to 22,000 ft-lb of energy per pound of honeycomb.

Efforts are presently being initiated to develop techniques for forming the plastic honeycomb in singly or doubly curved arrays, for possible application to anticipated planetary capsule landing systems. In addition, tests are being performed to determine the effect on nonaxial loading on plastic honeycomb elements.

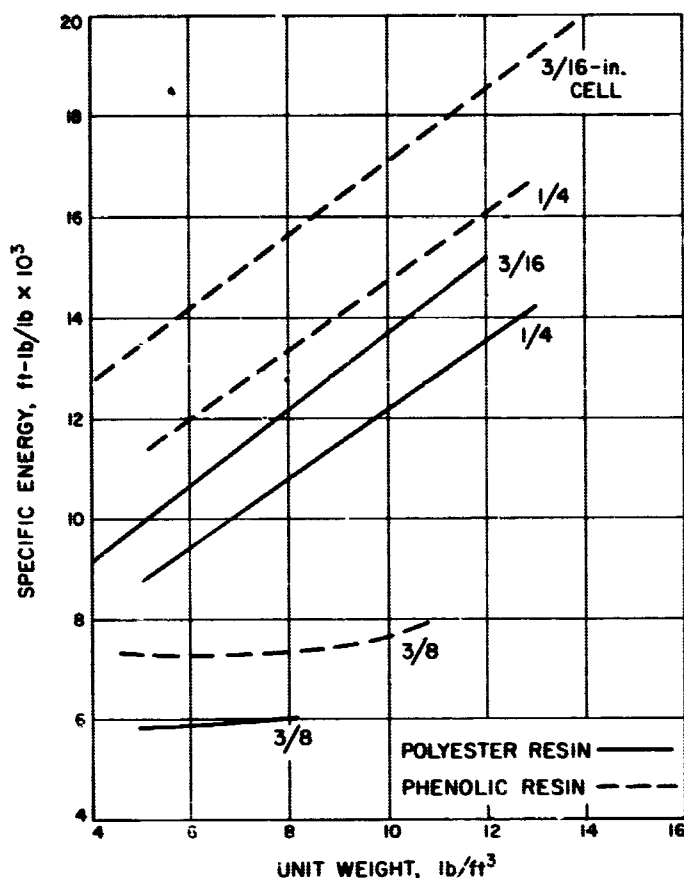


Fig. 2. Specific energy versus honeycomb density

## References

1. Brandmaier, H. E., and Harnett, J. J., "A Brief Survey of Past and Current Methods of Solution for Equilibrium Composition," *Kinetics, Equilibria, and Performance of High Temperature Systems*, Proceedings of the First Conference, Western States Section, The Combustion Institute, pp. 69-72, Butterworth's, 1960.
2. Browne, W. G., "Thermodynamic Properties of Some Atoms and Atomic Ions," Engineering Physics Tech. Memo 2; "Thermodynamic Properties of Some Diatomic and Linear Polyatomic Molecules," Engineering Physics Tech. Memo 3; "Thermodynamic Properties of Some Nonlinear Polyatomic Molecules," Engineering Physics Tech. Memo 4; "Thermodynamic Properties of Some Diatoms and Diatomic Ions at High Temperatures," Engineering Physics Tech. Memo 8; "Thermodynamic Properties of the Species, CN, C<sub>2</sub>, C<sub>3</sub>, C<sub>2</sub>N<sub>2</sub>, and C," Advanced Aerospace Physics Tech. Memo 9; "Comparison of Thermal Functions Generated for Species in High Temperature Air Systems With Literature Values," Advanced Aerospace Physics Tech. Memo 10; General Electric Co. Missile and Space Vehicle Dept., Philadelphia, Pa., 1962.

## References (cont'd)

3. Huff, V. N., Gordon, S., and Morrell, V. E., "General Method and Thermodynamic Tables for Computation of Equilibrium Composition and Temperature of Chemical Reactions," National Advisory Committee for Aeronautics Report 1037, 1951.
4. Vickers, J. M. F., "Thermal Scale Modeling: Arthur D. Little Contract," SPS 37-21, Vol. IV, pp. 48-50, Jet Propulsion Laboratory, Pasadena, California, June 30, 1963.
5. Vickers, J. M. F., "Thermal Scale Modeling: Arthur D. Little Contract," SPS 37-24, Vol. IV, pp. 57-59, Jet Propulsion Laboratory, Pasadena, California, December 31, 1963.
6. Vickers, J. M. F., "Thermal Scale Modeling: Basic Considerations," SPS 37-18, Vol. IV, pp. 80-83, Jet Propulsion Laboratory, Pasadena, California, December 31, 1962.
7. "RF Transparent, Energy Absorbing, Structural Elements," Phase I, Final Report to Jet Propulsion Laboratory, Document 69SD565, Contract 950564, General Electric, MSD.



**BLANK PAGE**

## ENGINEERING FACILITIES DIVISION

## VII. Technical Facilities

## A. An Intense Magnetic Field Facility

R. Rust

A high magnetic field facility is nearing completion at JPL. This project was initiated by the Physics Section of the Space Science Division. Engineering responsibility for the magnet design, fabrication, and subsequent laboratory completion was vested with the Technical Facility Engineering Group. This report gives descriptive information summarizing major systems design considerations and recent progress toward making the facility operational.

The first and most important item in a magnet laboratory for investigation of superconducting materials is the choice of a stable power supply. Magnets producing field strengths of 100,000 gauss cannot contain any ferritic types of materials. In order to produce continuous fields near the above strength one must use about 2 Mw of DC power (Fig. 1). Considerations of the design of continuously operating magnets having such field strength indicate that input power should be available in the form of heavy currents at low voltage, so the resistance of the magnet must be low. Preferable total magnet resistance is of the order of  $220 \mu\Omega$ , as this minimizes the temperature requirements placed on the electrical insulation spacers between the coil turns.

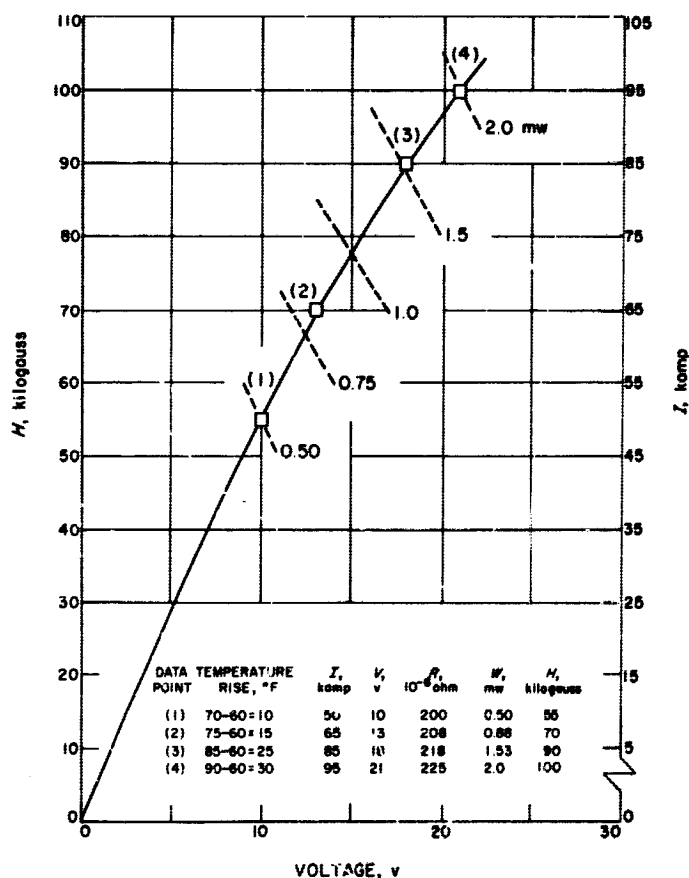


Fig. 1. Magnet performance data

For this installation a new unipolar motor generator set developing 2.4 Mw is used. The high-frequency voltage ripple stability is constant to one part in  $10^5$ . The test room containing the magnet is visible for communication from the control panel of the motor generator set, since it is not in the same room. The magnet is kept as near to the generator as possible to minimize the amount of copper required for electrical busses to carry approximately 95,000 amp at 21 v, during operation.

An existing closed loop system using clean (distilled) water is used. The complete system contains the magnet, a cooling tower, pumps, and a heat exchanger. The cost of water pumps was saved by proper choice of the laboratory site in the vicinity of the wind tunnel facilities. Water pressure gages and an orifice flowmeter are mounted near the instrument control panel where the operator has them under observation at all times.

Since the noise level of this machinery is excessive, the generator was fully enclosed in a separate room. An

appreciable magnetic field due to magnet and bus bars will penetrate nearby buildings about 30 ft in depth. This stray field will amount to several hundred gauss just outside the magnet, and one must be careful to remove any iron which might be attracted to the magnet. Attraction forces are very large when one moves inside the air core of the magnet. It is generally assumed that exterior magnetic fields have no biological effect on the occupants of the laboratory.

### 1. The Power Supply

The central item in the power supply is the unipolar motor generator set (Fig. 2). The direct current generator nominally puts out 1800 kw. It has a novel liquid metal (Nak) current collection system replacing conventional dc generator brushes. The Nak reduces internal energy dissipation so that power output can be increased from 1.8 to about 2.4 Mw. A 2500-hp induction motor operating with 140% load factor on three-phase current, at 4160 v, is capable of delivering about 3500 hp continu-

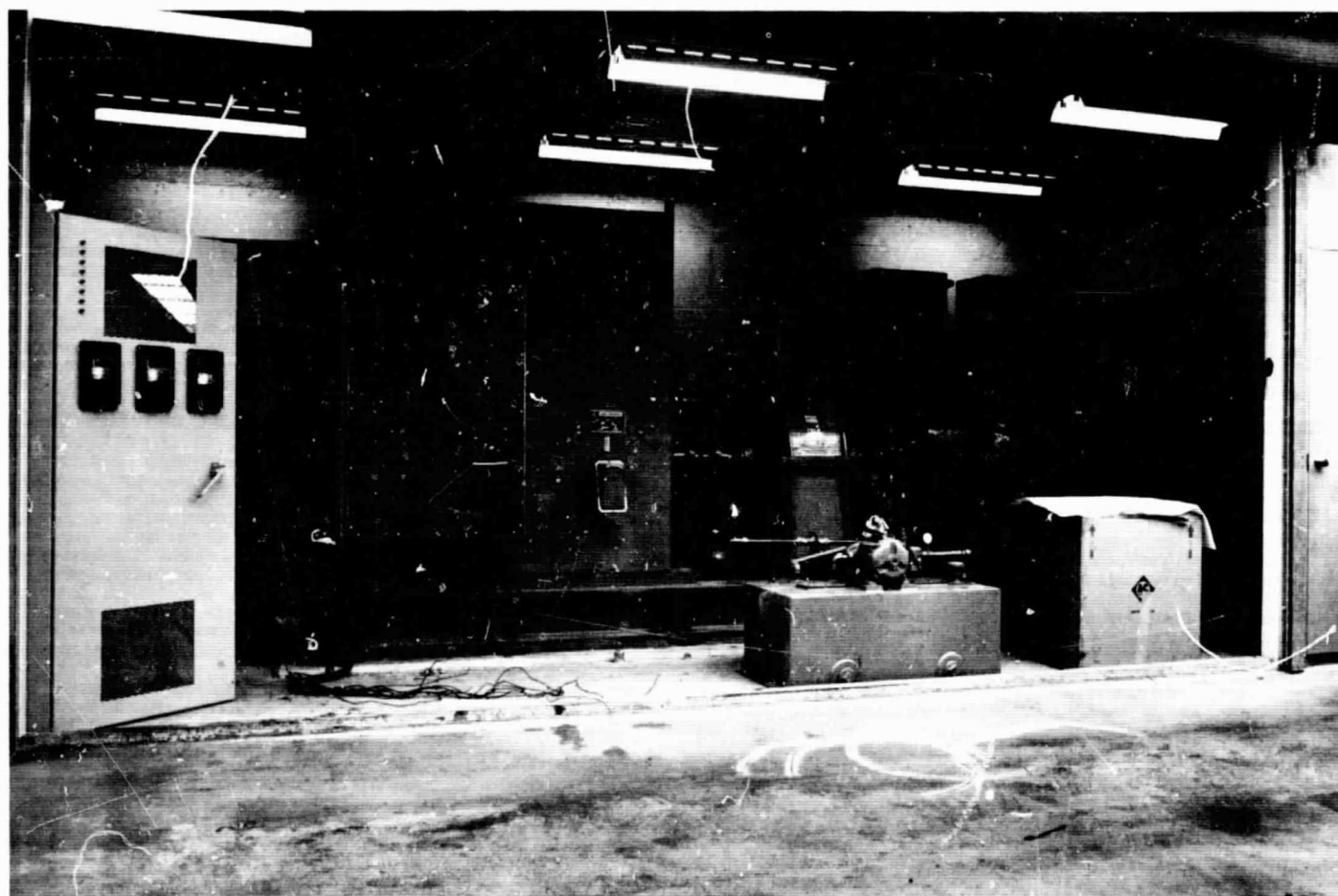


Fig. 2. Motor generator set power supply

ously. The motor turns at 1800 rpm and is physically attached to the unipolar generator via a 1:4 speed increase gear box all attached to a common bedplate. Electrical controls are interlocked with such necessary items as water flow to the magnet and magnet-bus, Nak flow and temperature in the generator, over-temperature, lube-oil flow and temperature, to bearings.

## 2. The Water Supply

Distilled water is pumped through the magnet and the heat exchanger at a rate of about 430 gal/min. The capacity of the cooling circuit is of sufficient size to keep the maximum distilled water temperature between 100 and 120°F, depending on initial water temperature. This will be sufficient to keep the maximum temperature of the copper in the magnet to within a few degrees of the boiling point of the water.

In order to keep the magnet near any particular average bulk temperature the following volume flow rates are required for water having the temperature differentials stated below.

Average magnet temperature, °C	300 gal/min, °C	600 gal/min, °C
25 (77°F)	26	15
50 (122°F)	21	12
75 (167°F)	18	10

The water contains from 500 to 1500 ppm sodium nitrite for corrosion control. The low voltage magnet system can tolerate such contaminants without very great electrical losses as follows:

Water sample	Resistivity, ohm-cm	Power loss, kw	Power loss, %
Pure	20,000	0.036	0.0024
City water	4,000	0.179	0.012
550 ppm sodium nitrite	238	3.02	0.20
1400 ppm sodium nitrite	200	3.46	0.24

The magnet is designed so that a pump must maintain a pressure of about 120 psi in order to circulate the required amount of water. The distilled water circuit

and cooling tower circuit contain pressure switches which cut the excitation for the armature circuit of the DC generator, when the water pressure falls below 100 psi.

## 3. The Magnet Solenoid

The magnet is intended to give high magnetic field strengths for continuous operation over indefinite periods. The magnet (Fig. 3) is mounted with its axis vertical. The central tube for the magnet is 2-in. ID.

Since the magnet has to dissipate 2.0 Mw, the coil is about 17-in. OD and provision must be made for internal cooling. This is done by forcing water radially through 0.025-in.-high spaces between successive turns of the coil. These passages are maintained by spacers in such numbers as to allow removal of heat generated in their immediate vicinity at the rate of about 200 w/cm<sup>2</sup> of cooling surface. Thus, the temperature at the inner portion of the magnet can be held below the boiling point of water. The flow of water is turbulent. The Reynolds number is a million, hence well above its critical value. The resistance of the magnet is about 200  $\mu\Omega$ , and the field constant is just over 1 gauss/amp.

Each of the copper turns of the magnet is 0.318-in. thick and forms a continuous solenoid of 27.2 turns. The solenoid and electrical insulation spacers are held together simply by mechanical pressure. Only the end turns of the solenoid are backfilled with copper sheet and attached to heavier copper plates which are bolted to the bus. Micarta gaskets keep the coil from touching the

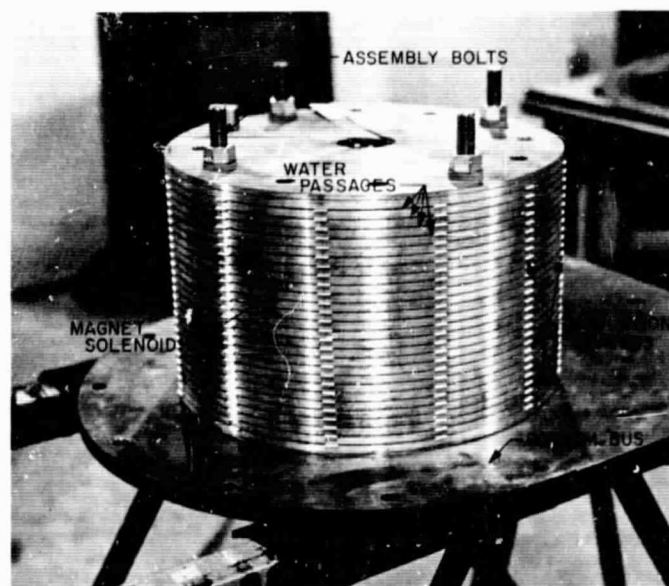


Fig. 3. Solenoid and bottom bus

inner tube, or the outer case, and allow water to circulate over the surfaces of the coil. Holes 0.75-in. in diameter penetrate the coil near its outer edge. Micarta tubes in these bolt holes prevent the turns of the coil from moving or touching each other. These tubes have holes machined in cross-section to allow water to circulate and prevent local burning. The present design for an average heating rate of only 200 w/cm<sup>2</sup> is quite safe from the conditions for forming a steam layer at the cooling interfaces. The surface heating rate at the inner radius position of the solenoid is calculated to be about 1,200 w/cm<sup>2</sup>. This is about one-fourth of the heating rate for forming steam layers. Other magnets could be designed with much less cooling surface if one were willing to go to higher operating temperatures. An upper limit heating rate of about 5,000 w/cm<sup>2</sup> might be obtainable before steam layer formation would cause the magnet to blow up.

The design parameters for the solenoid are as follows:

Current  $I = 95,000$  amp

Power  $W = 2.0 \times 10^6$  w

Inside radius  $r_1 = 1.25$  in.

Outside radius  $r_2 = 8.5$  in.

Turn thickness  $(t + t_1) = 0.343$  in.

Voltage  $V = 21$  v

Resistivity  $= 2.0 \times 10^{-6}$  ohm cm

At 60°C  $\rho = 7.85 \times 10^{-7}$  ohm in.

Coil length  $L = 10$  in.

Plate thickness  $t = 0.318$  in.

Water channel  $t_1 = 0.025$  in.

Parameter  $\alpha = r_2/r_1 = 8.5/1.25 = 6.8$  Eq. (1)

Parameter  $\beta = L/2r_1 = 10/2(1.25) = 4.0$  Eq. (2)

Turns  $N = [L - 2(t + t_1)]/(t + t_1) = 27.2$  Eq. (3)

Space factor  $\lambda = t/(t + t_1) = 0.318/0.343 = 0.925$  Eq. (4)

Current density ( $J_1$ ) at inner radius position, with ( $r_1$ ) given in centimeters.

$$J_1 = (4\pi\beta \ln \alpha)^{-1/2} (W/\rho\lambda r_1^3)^{1/2} = 18,000 \text{ amp/cm}^2 \quad \text{Eq. (5)}$$

The total effective magnet resistance  $R_1$ .

$$R_1 = 2\pi\rho [L - 2(t + t_1)]/t(t + t_1) \ln(r_2/r_1) = 219 \mu\Omega \quad \text{Eq. (6)}$$

First axial field strength equation ( $H$ ) with ( $r_1$ ) given in centimeters.

$$H = G(W\lambda/\rho r_1)^{1/2} = 101,000 \text{ gauss} \quad \text{Eq. (7)}$$

Dimensionless Fabry factor  $G$  used with Eq. (7).

$$G = \left(\frac{\pi^{1/2}}{5}\right) \cdot (\beta \ln \alpha)^{-1/2} \ln \left\{ \frac{\alpha [\beta + (1 + \beta^2)^{1/2}]}{\beta + (\alpha^2 + \beta^2)^{1/2}} \right\} = 0.196 \quad \text{Eq. (8)}$$

Second axial field strength equation ( $H$ ) with coil height ( $L$ ) in centimeters.

$$H = 4\pi N I R / 10 L = 102,000 \text{ gauss} \quad \text{Eq. (9)}$$

Dimensionless shape factor  $R$  used with Eq. 9.

$$R = 1 - (\log r_2/r_1)^{-1} \log \left\{ \frac{1 + [1 + (2r_2/L)^2]^{1/2}}{1 + [1 + (2r_1/L)^2]^{1/2}} \right\} = 0.802 \quad \text{Eq. (10)}$$

Solenoid axial compression force ( $F_1$ ) at the inner radius ( $r_1$ ).

$$F_1 = 2 \times 10^{-9} \left[ \frac{\pi N r_1 I}{L} \right]^2 \frac{\text{joules}}{\text{cm}} \left[ \frac{10^7 \text{ dyne cm}}{\text{joule}} \right] \left[ \frac{2.25 \times 10^{-6} \text{ lb}}{\text{dyne}} \right]$$

$$F_1 = 2 \times 10^{-9} \left[ \frac{\pi (27.2) (1.25) (95,000 \text{ amp})}{(10)} \right]^2$$

$$(10^7) (2.25 \times 10^{-6}) = 46,300 \text{ lb}$$

Eq. (11)

This 23-ton axial compression must be supported by the dielectric spacers placed radially throughout the coil 30 deg apart. These spacers are made of silicone-fiberglass. They are  $\frac{3}{16}$  in. wide at the inner radius, 0.50 in. wide at the outer radius, and 0.025 in. thick. No deformation of the spacers was observed.

Solenoid radial stress ( $F_2$ ) at the inner radius ( $r_1$ ).

$$F_2 = (14.7 \times 10^{-6}) (H \text{ gauss})^2 / 8\pi = 5,850 \text{ lb/in.}^2 \quad \text{Eq. (12)}$$

No permanent set was observed for the copper coil near the spacers for either of these two stresses just computed.

#### 4. Cryogenic Magnet Dewar

A commercial cryogenic Dewar was modified for use with the electromagnet. This Dewar has a nitrogen-cooled shield enclosing a helium container. The exterior of the tail section of the Dewar forms the wall of the air core magnet as shown by Fig. 4. The entire Dewar is removable by unseating an O-ring gasket water seal, on the tail section of the Dewar and removing the cover-plate bolts at the top of the magnet.

The cryogenically cooled magnet test section shown in Fig. 4 is 14 in. long and has an inside diameter of 1.5 in. reaching to the center of the magnet. The helium container has a radiation shield cooled by 4.7 l of liquid nitrogen. A vacuum pump is mounted on the roof of the laboratory and will be used to supercool the helium volume so that experiments can be carried out below 4°K. This vacuum pump will be operated continuously during such tests. The helium-filled tail section mentioned above is supplied with 2.25 l of liquid stored in the upper portion of the container. The Dewar forms an integral part of the magnet and can be used to simul-

taneously test materials under intense magnetic fields and cryogenic temperatures.

#### 5. Magnet Facility Use Program

This facility will be used first for low temperature research. Specifically, the current density versus magnetic field will be measured for superconductors in the vicinity of 100 kilogauss.

The availability of a 100 kilogauss field of fairly large test volume suggests a number of experiments in the demagnetization region for calorimetry, superconductivity, etc. Such a high field also offers an opportunity for measurements of time resolved nuclear magnetic resonance.

The facility could be used for cyclotron resonance investigations of new materials. Fundamental information about basic electronic properties can be obtained which would not be possible at lower magnetic field strengths. Experiments in solid-state physics with intermetallic compounds of the Types III-V, and II-VI hold promise in high magnetic fields.

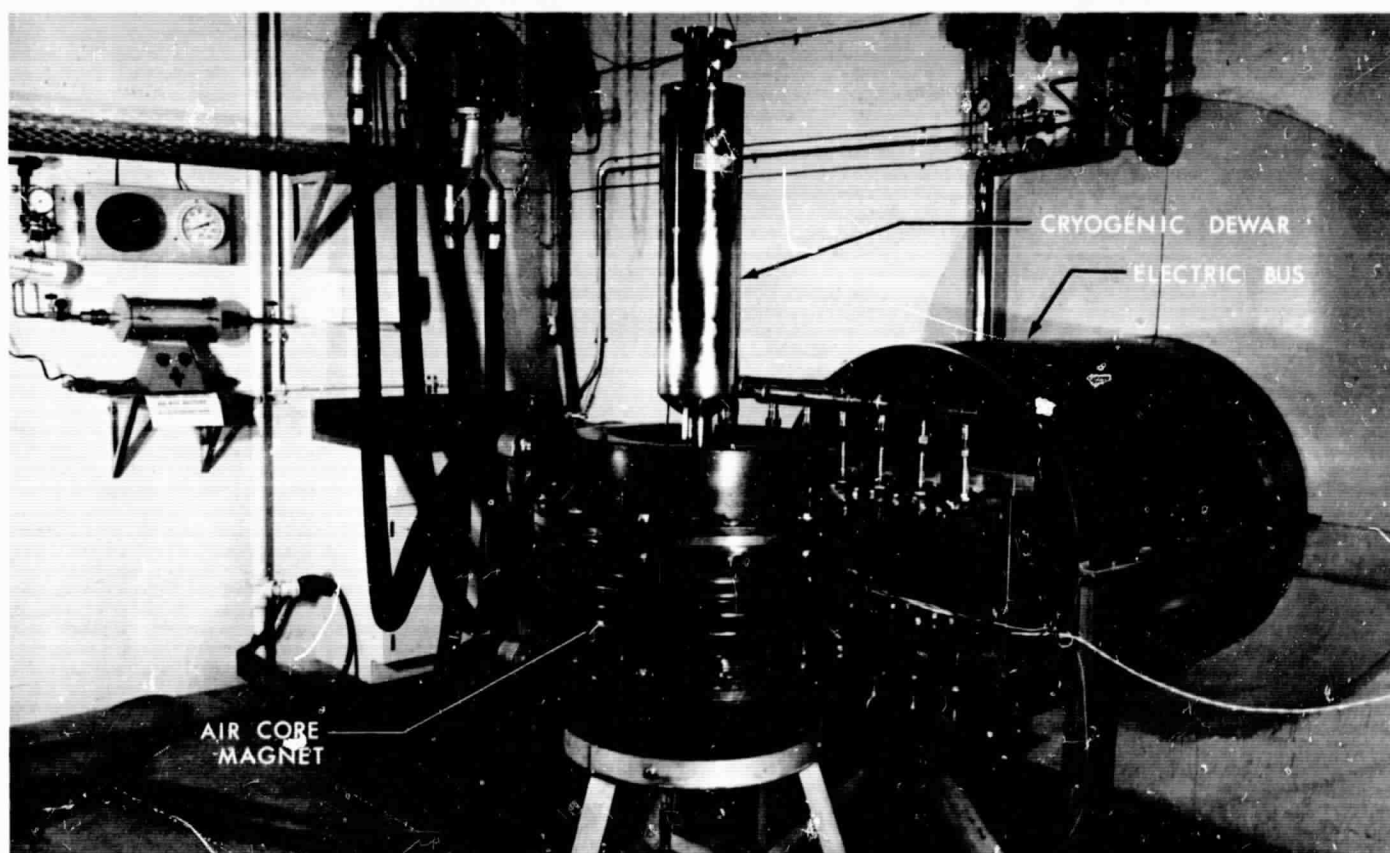


Fig. 4. Dewar, magnet and electric bus assembly

**BLANK PAGE**

## PROPULSION DIVISION

## VIII. Solid Propellant Engineering

## A. Low-Pressure Combustion Studies

L. D. Strand

## 1. Influence of Aluminum Coarseness on Low-Pressure Unstable Combustion

Two modified versions of JPL-540 propellant are being used in the transparent motor studies of low-pressure unstable combustion. In the JPL-540 Modified 1 propellant, aluminum of 31- $\mu$  average particle diameter has been substituted for the 16% of 7- $\mu$  aluminum additive in the regular propellant formulation. JPL-540 Modified 2 propellant contains aluminum of approximately 15- to 17- $\mu$  average particle size. Vacuum system test firings at an initial propellant temperature of 80°F have been conducted with these propellants in standard 3-in.-D test motors. By comparing test results with similar unmodified JPL-540 test firing data, an attempt has been made to determine any changes in low-pressure unstable combustion characteristics resulting from the change in aluminum particle size. The following is a brief summary of the findings.

Figs. 1 and 2 are  $K_n-P_c$  log-log plots for the two modified propellants, obtained from the 3-in.-D motor firing data. For comparison, the unmodified propellant  $K_n-P_c$  plot is shown as a dashed line. The curves for the modified propellants are nearly identical, and as will be seen, the unstable combustion characteristics also seem to be approximately the same.

Fig. 3 shows a log-log plot of the motor  $L^*$  values plotted versus the mean chamber pressures at the time of the combustion extinction for the three propellants. The  $L^*$  value after extinction of combustion has previously been shown to be a correlating parameter for low-pressure combustion extinction (Ref. 1). The data points for the two modified propellants appear to fall on the same correlation curve, within experimental scatter. The curve has a smaller negative slope than that for JPL-540 propellant and lies in a lower pressure region. A log-log plot of the pressure oscillation frequency prior to combustion extinction versus the mean chamber pressure at combustion extinction for the three propellants is shown in Fig. 4. Again the data points for the two modified propellants appear to fall on a straight line of smaller



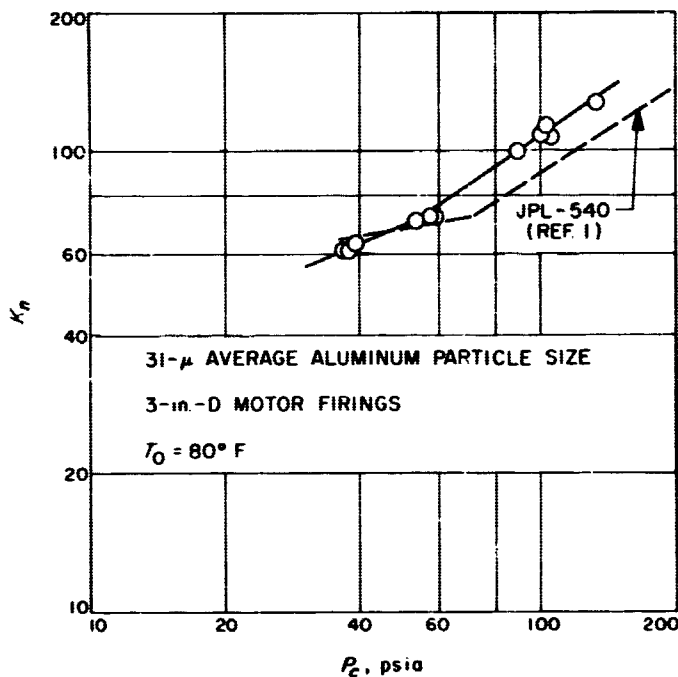


Fig. 1. Log-log plot of  $K_n$  versus  $P_c$  for JPL-540 Modified 1 propellant

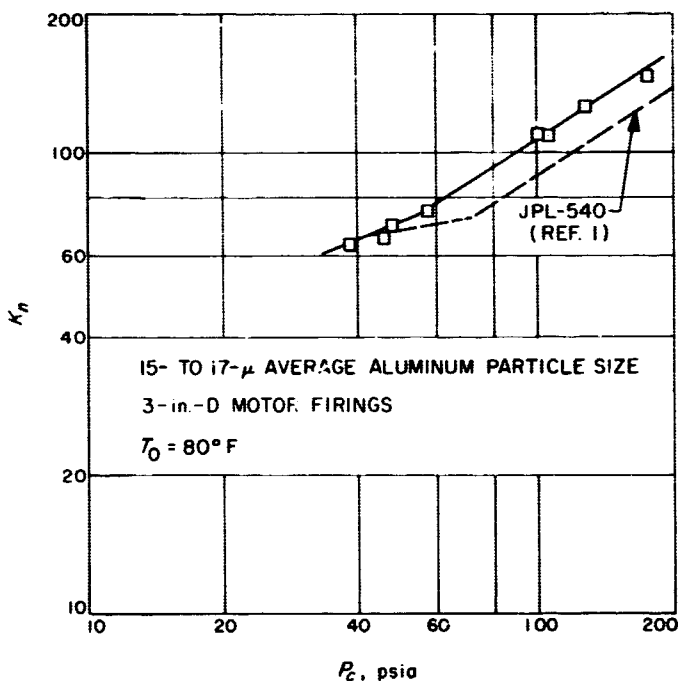


Fig. 2. Log-log plot of  $K_n$  versus  $P_c$  for JPL-540 Modified 2 propellant

slope and in a lower pressure region than the curve for JPL-540. The  $K_n$ - $P_c$  graphs predict a slight change in the slope of the modified propellant curve in Fig. 3 and possibly also in Fig. 4 (change in burning rate pressure

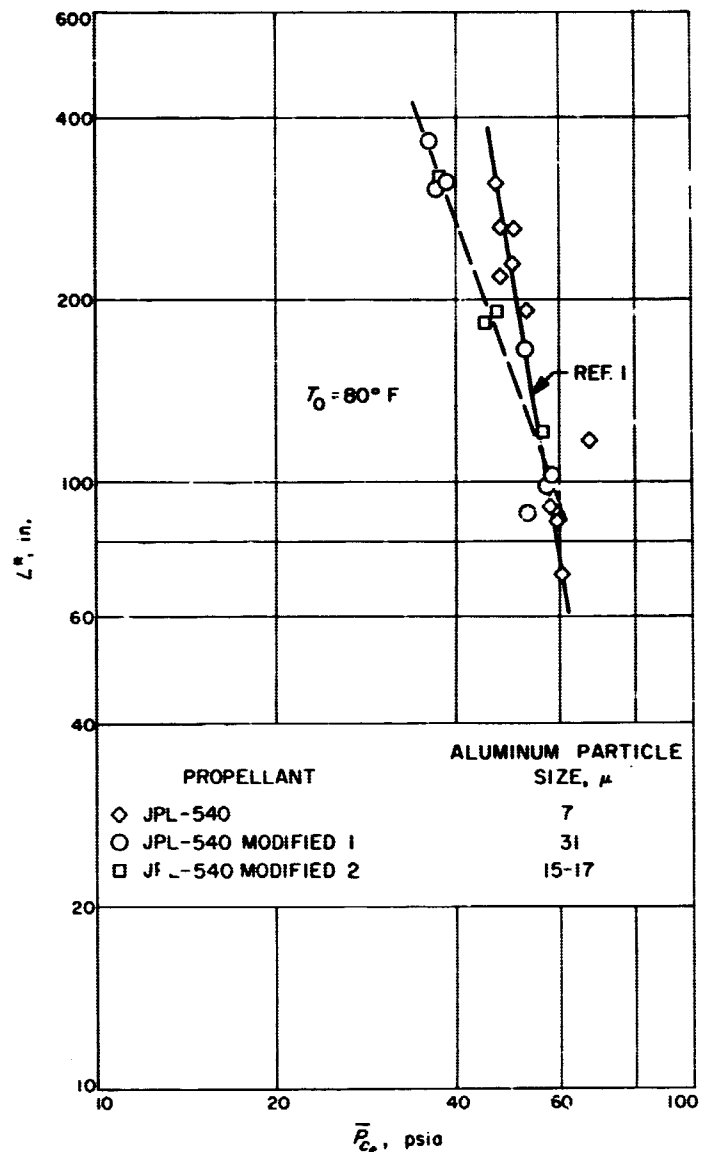


Fig. 3. Log-log plot of  $L^*$  versus mean chamber pressure at combustion extinction

exponent) at a pressure of approximately 54 psia. Not enough data points are available in the higher pressure region to substantiate any such changes in slope and therefore none are shown in the figures.

The shifts in the modified propellant curves are in the direction of similar curves for more rapid burning propellants of low aluminum composition, and this is felt to be a possible explanation of the test results. Due to the incomplete low-pressure combustion of the coarser aluminum in the modified propellants, the propellants' burning characteristics approach those of a non-aluminized propellant. As a check on this postulation, small motor firings are being planned with these three

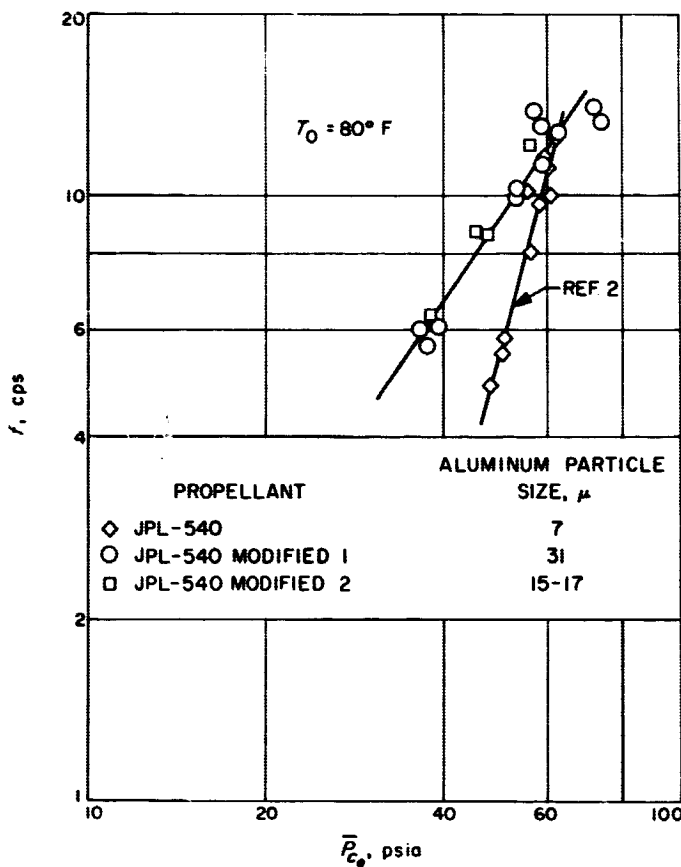


Fig. 4. Log-log plot of pressure oscillation frequency versus mean chamber pressure at combustion extinction

propellants to determine the  $c^*$  efficiencies at low operating pressures.

## 2. Pressure Oscillation Amplitudes of Low-Frequency Combustion Instability

As was reported in Ref. 2, in a majority of the experimental test firings that have and are being conducted in the investigation of low-pressure combustion the propellant charges are regressive in burning nature. These charges ignite in the stable operating pressure region, burn regressively with decreasing pressure into the unstable pressure region, and ultimately cease burning. In a majority of test firings, low-frequency, low-amplitude oscillations in pressure occur prior to extinction of combustion. The oscillation amplitudes will at first increase linearly at a slow rate with decreasing chamber pressure. With continuing pressure decrease the pressure oscillation amplitudes begin to grow at an increasing rate of increase and continue to do so until combustion extinction occurs. The base pressures and oscillation amplitudes were measured from oscillograph records at several

points over the time span of linear amplitude increase for a number of test firings. The purpose was to see if any relationship between the oscillation amplitudes and base pressures existed over the range of base pressures measured. Figs. 5, 6, and 7 are graphs of the oscillation

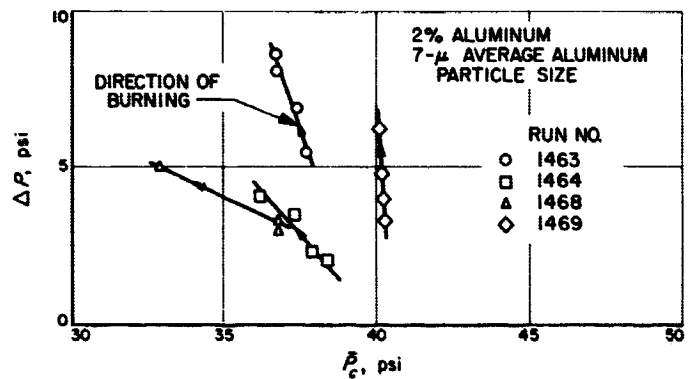


Fig. 5. Amplitude of pressure oscillations versus mean chamber pressure, JPL-534 propellant

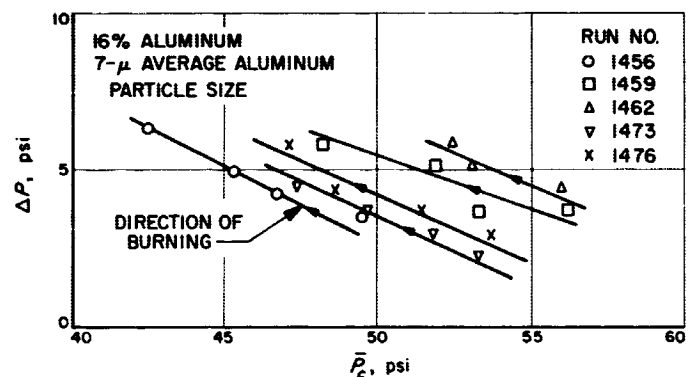


Fig. 6. Amplitude of pressure oscillations versus mean chamber pressure, JPL-540 propellant

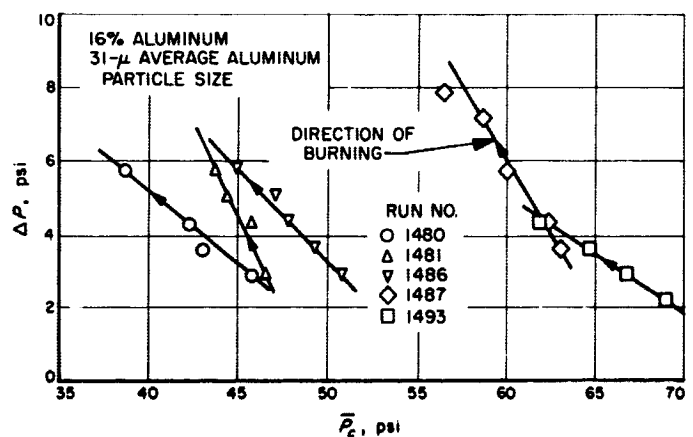


Fig. 7. Amplitude of pressure oscillations versus mean chamber pressure, JPL-540 Modified 1 propellant

amplitudes ( $\Delta P$ ) plotted versus the mean chamber pressures for several test firings of JPL-534, JPL-540, and JPL-540 Modified 1 propellants, respectively. Over the narrow span of base pressures measured, the range of oscillation amplitudes for the test runs seems to be independent of base pressure for the three propellants. The linear rate of increase in amplitude with decreasing mean pressure seems to vary randomly among the test firings. The conclusion drawn is that the amplitudes of low-frequency combustion instability do not appear to be an inherent property of the propellant and its pressure of combustion, as is the oscillation frequency.

## B. Silicone Propellant Development for Heat Sterilization

L. C. Montgomery

In consonance with the requirements for a heat sterilizable propulsion unit, JPL has undertaken the development of a propellant using high temperature silicone resins.

Resins were selected from those now commercially available. Physical tests of the resins were made and a candidate resin for a propellant system was selected. Ammonium perchlorate (AP) was selected as the oxidizer.

With extreme safety precautions small laboratory samples of the binder and oxidizer were mixed by hand. Impact tests were run on these crudely mixed samples. These tests indicated a sensitivity of the propellant to the impact of a 2-kg steel ball when dropped from a minimum height of 27.9 cm onto the 1/2-in. thick propellant sample located on a steel base. This sensitivity did not indicate any greater safety hazard than other ammonium perchlorate propellants.

It was then acceptable to mix the propellant in the small sigma mixer. From 1000-g batches made in the sigma mixer other tests of impact sensitivity, burning rate, and physical test specimens were made. Tests of these specimens indicated a lesser impact sensitivity (30.5 cm for the ball drop). The burning rate tests were run at 80°F and on propellants containing 70% and 80% oxidizer. The burning rates of the 70% mixture were 0.28 in./sec at 250 psig, 0.38 in./sec at 500 psig, and 0.59 in./sec at 1000 psig. The burning rates of the 80% mixture measured 0.46 in./sec at 250 psig, 0.61 in./

sec at 500 psig, and 0.90 in./sec at 1000 psig. Density of the propellant was not measured for the 70% oxidizer propellant, but it has averaged about 0.0597 for the 80% batches. Typical data from the small sigma mixer batches is given in Table 1. It should be noted that sterilization treatment does cause a slight increase in tensile strength with associated reduction in elongation. The density is not measurably changed by the sterilization treatment.

Before charges were made for rocket motors an investigation was made of liner materials for case bonding the silicone/AP propellant. No material was found to which the propellant would bond, with or without the recommended primers. However, it was found that RTV 11, a General Electric Corporation product, would adhere to the cured propellant and to most other surfaces and would maintain a high tensile strength and high elongation even under sterilization temperature cycling.

15- and 30-lb batches of the silicone/AP propellant have been made in the 5-gal Planetex mixer. Initially, dummy batches were made to checkout the mixing procedures. From the first live batch of propellant two cartridge charges were made by casting into 4 1/2-in.-ID teflon-coated sleeves with a tapered 2-in.-D mandrel for the center perforation. After curing these charges at 250°F for 2 hr (other samples used this same cure temperature), they were trimmed to a 5-in. length and prepared for static test. One charge was inhibited on both ends and outer surface with the silicone binder and then wrapped with glass tape around the outer cylindrical surface. The charge was then secured in the chamber by clamps and static tested. The second charge was potted in the chamber with RTV-11. Both static tests were successful and gave a higher performance than was expected. The charges gave an average  $C^*$  (calculated from head end pressures) of 4652 ft/sec and 4625 ft/sec, respectively. Data from the physical specimens from this batch indicate the propellant to have the same physical characteristics as that made in the smaller batches.

Two 30-lb batches of the silicone/AP propellant have been made and from each of these batches four propellant charges were cast. These charges are to be used to evaluate the effect of the three sterilization cycles on integrity of the charge and/or motor performance.

After a minimal evaluation program of this propellant in a rocket motor under sterilization conditions, its development will be carried along only as necessary as a backup program for sterilizable propulsion systems for Mars landing missions.

Table 1. Data from tests of the small sigma mixer batches

Batch number	Oxidizer, %	Sterilization treatment	Density, lb/in. <sup>3</sup>	Tensile strength, lb/in. <sup>2</sup>	Elongation, %
S-2176	70	None	—	111.6	68.2
				123.3	73.6
				116.8	73.9
				118.4	73.6
S-2180	80	None	0.05984 0.05968 0.05971	87.1	69.6
				87.3	70.3
				87.2	64.0
				84.7	66.5
				89.6	69.0
				87.8	68.1
S-2183	80	300 $\pm$ 8°F for 36 hr left in mold	0.05997 0.05985 0.05971	90.1	60.7
				89.4	58.8
				88.8	57.5
				89.5	57.4
				88.6	58.0
				84.1	53.1
S-2184	80	None	0.05977 0.05950 0.05980	79.6	75.5
				78.7	74.6
				81.6	79.0
				75.3	69.8
				81.3	80.7
				84.4	75.5
S-2185	80	300 $\pm$ 8°F for 36 hr; removed from mold; wrapped in aluminum foil	0.05970 0.05957 0.05980	90.5	57.7
				85.7	53.3
				91.0	61.0
				88.8	59.9
				87.2	57.0
				87.7	56.2

## References

1. Anderson, F. A., Strand, L. D., and Strehlow, R. A., *An Experimental Investigation of the Low-Pressure Combustion Limits of Some Solid Propellants*, Technical Memorandum No. 33-134, Jet Propulsion Laboratory, Pasadena, California, June 3, 1963 (Confidential).
2. Anderson, F. A., and Strand, L. D., "Low Pressure Combustion Studies," SPS 37-22, Vol. V, pp. 13-15, Jet Propulsion Laboratory, Pasadena, California, August 31, 1964 (Confidential).

## IX. Polymer Research

### A. Compatibility of Polyethylene Oxide and Poly- $\beta$ -Vinyl naphthalene

*J. Moacanin and E. F. Cuddihy*

Previously, properties of blends of poly- $\beta$ -vinyl naphthalene (PVN) and polyethylene oxide (PEO) have been discussed (Ref. 1). It was found that these two polymers exhibit limited compatibility under certain conditions. New studies on these systems have included dynamic measurements of mechanical properties and some examinations of polymer morphology by phase-contrast microscopy (Ref. 2). In addition, a more detailed study was made of a graft copolymer.

For this study, a torsional pendulum constructed according to a previously reported design (Ref. 3) was used to obtain the damping given in log decrement. Test specimens ( $2 \times \frac{1}{4}$  in.; 0.05–0.08 in. thick) were prepared by compression molding at 1000 to 4000 psi. The molding temperature was 55°C for the polyblends, and 40°C for the graft copolymers.

The results of dynamic measurements for the 50-50 blend are shown in Fig. 1. The transition at 54°C is due to crystalline melting of PEO, and the broad shoulder between 85 and 120°C corresponds to the region of slowly increasing Young's modulus (Fig. 2); whereas the peak at 138°C coincides with the dilatometric  $T_g$  for PVN. After having been heated above 120°C, the damping constant exhibited a continuous gradual increase from about 40°C until the peak at 138°C was reproduced. The loss component of the dynamic modulus which can be calculated from dynamic data may resolve the shoulder for the initial sample into a separate peak, as can be seen by considering the shape of the Young's modulus curve (Fig. 2). Additional results are needed to more accurately resolve this point. For the "heated" specimen, however, the loss component should have the same appearance as the damping curve.

The broad shoulder for the damping constant (Fig. 1) is consistent with the idea of phase changes taking place over the temperature range of the "well" (Ref. 1) until  $T_g$  of PVN (138°C) is reached, and irreversible microphase separation has taken place. The gradual rise in the damping constant with increasing temperature for the reheated

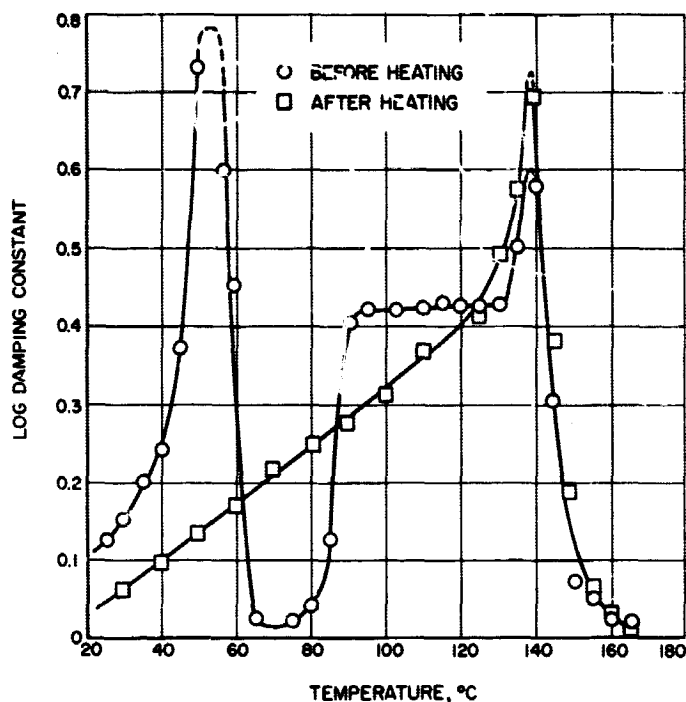


Fig. 1. Log damping constant vs temperature for a 50-50 polymer blend of PVN and PEO

polymer may be due to localized changes taking place at the microphase boundaries.

Phase-contrast micrographs are shown in Fig. 3. The granular structure and the changes with temperature in the relative content of light versus dark regions are noticeable, most of the change taking place after heating from 25 to 90°C, i.e., the temperature region for the "well." Also, it is to be noted that the specimen heated above 120°C retained its appearance when cooled to room temperature, consistent with the irreversible changes in behavior which are observed for heated specimens.

A different situation should exist for the graft copolymers in which one end of the PEO chain is attached to the PVN backbone. The modulus curve for a 20% PEO copolymer was depressed relative to pure PVN. The graft copolymer having the composition of the "complex" (46% PEO, Fig. 4) was transparent, rubbery, and had a glass transition temperature below -20°C. Furthermore, it retained its rubbery properties after being heated beyond 120°C. An additional interesting property is the existence of a plateau extending from about 60 to 180°C characteristic of a crosslinked or very high molecular weight polymer. Yet this material is readily soluble in most aromatic and ether-type solvents. This behavior

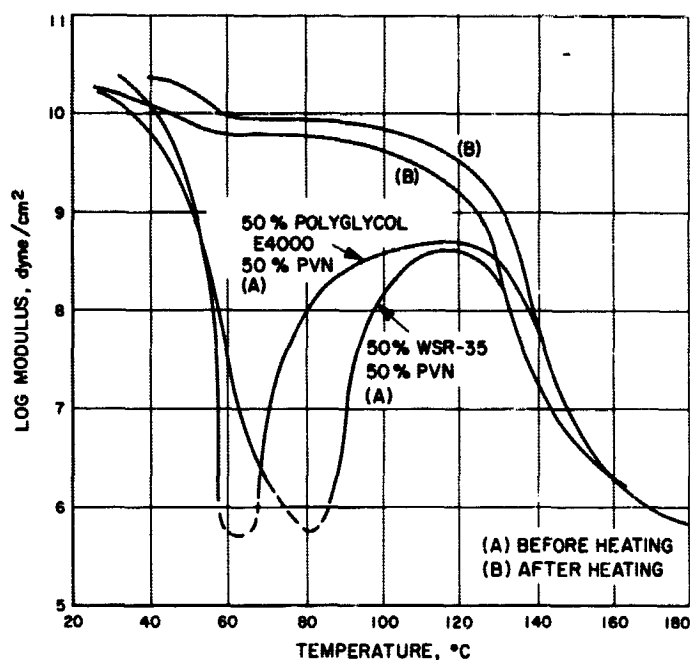


Fig. 2. Log modulus vs temperature for two 50-50 polymer blends of PVN with PEO of different molecular weights

may originate from entanglement of the PEO side chains of adjacent molecules.

Another graft copolymer with the same composition, but presumably somewhat shorter PEO side chains (probable molecular weight of side chains is 5 to 6,000), had the same transparent appearance and low  $T_g$ . The modulus, as well as dilatometric data, exhibited a hysteresis loop between -20 and 60°C. These results as well as X-ray data indicate presence of some crystallinity in contrast to the first completely amorphous copolymer. For given composition the relation between the size of the side chains and their ability to crystallize is not clear yet.

Both Young's modulus and damping constant curves for this copolymer are shown in Fig. 5. During the first heating cycle (dashed lines), the sharp drop in modulus between 40 and 54°C coincided with the peak for the damping constant, the transition being due to crystalline melting. For several subsequent heating cycles, the upper modulus curve was followed on heating and the lower on cooling. The peaks for the damping constant for both heating and cooling are displaced to a lower temperature; the heating curve exhibiting a shoulder. These curves doubtless are also due to crystalline melting, their shapes depending on distribution of crystallite size. In either case the absence of a transition corresponding to  $T_g$  of

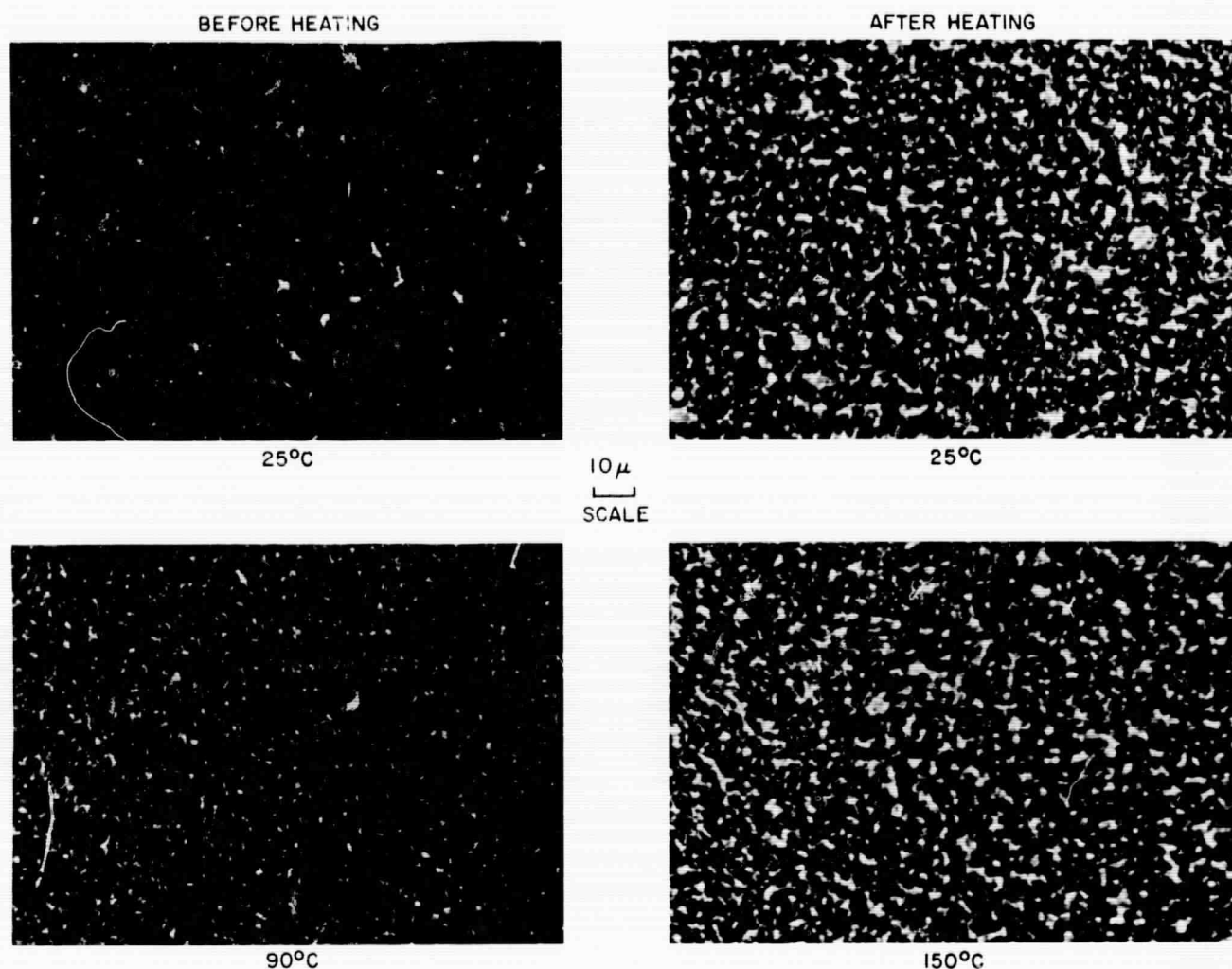


Fig. 3. Phase-contrast photomicrographs of a 50-50 polymer blend of PVN and PEO

PVN ( $138^{\circ}\text{C}$ ) is to be noted. The dilatometer data (Fig. 6) parallel the dynamic measurements; following a slightly different path on initial heating (dashed line), a reproducible hysteresis loop in the melting range was followed on subsequent heating cycles. The specific volume at  $25^{\circ}\text{C}$  was  $0.864$  to  $0.869\text{ cm}^3/\text{g}$ . The presence of crystallinity at  $43$ – $54^{\circ}\text{C}$  was corroborated by X-ray measurements. Again, the transition at  $138^{\circ}\text{C}$  is absent.

The position of the damping maxima shows that the melting temperature of the PEO crystallites was depressed. It is interesting to note that the "heating" curves were followed only if the specimen was first cooled below  $-20^{\circ}\text{C}$ ; otherwise the cooling curve was followed for both descending and ascending temperatures, irrespective of rate (the slowest rate was  $\frac{1}{4}^{\circ}\text{C}/\text{min}$ ). Doubtless this behavior is due to a large increase in the rate of nucleation below a critical temperature which seems necessary to complete the crystallization process. Curiously enough, this temperature of  $-20^{\circ}\text{C}$  coincides with the secondary peaks in the damping constant, suggesting a change in the nucleation mechanism associated with a polymer transition. This transition may be due to freezing-in of the side chains, since it is known that  $T_g$  for PEO of molecular weight  $1,000$  to  $10,000$  is approximately  $-20^{\circ}\text{C}$  (Ref. 4).

The observation that  $T_g$  for PVN has disappeared, as can be seen by examining both the damping constant and the dilatometric curve, gives unequivocal evidence that the identity of PVN has been lost and that the whole copolymer behaves as an entity. The specific volume values for temperatures in the amorphous region (above  $60^{\circ}\text{C}$ ) show that additivity of volumes for the two components is not followed, but that a contraction of about 2% has taken place. Using molecular models it is apparent that one can readily entwine PEO and PVN chains. Optimum spacing is obtained with three ethylene

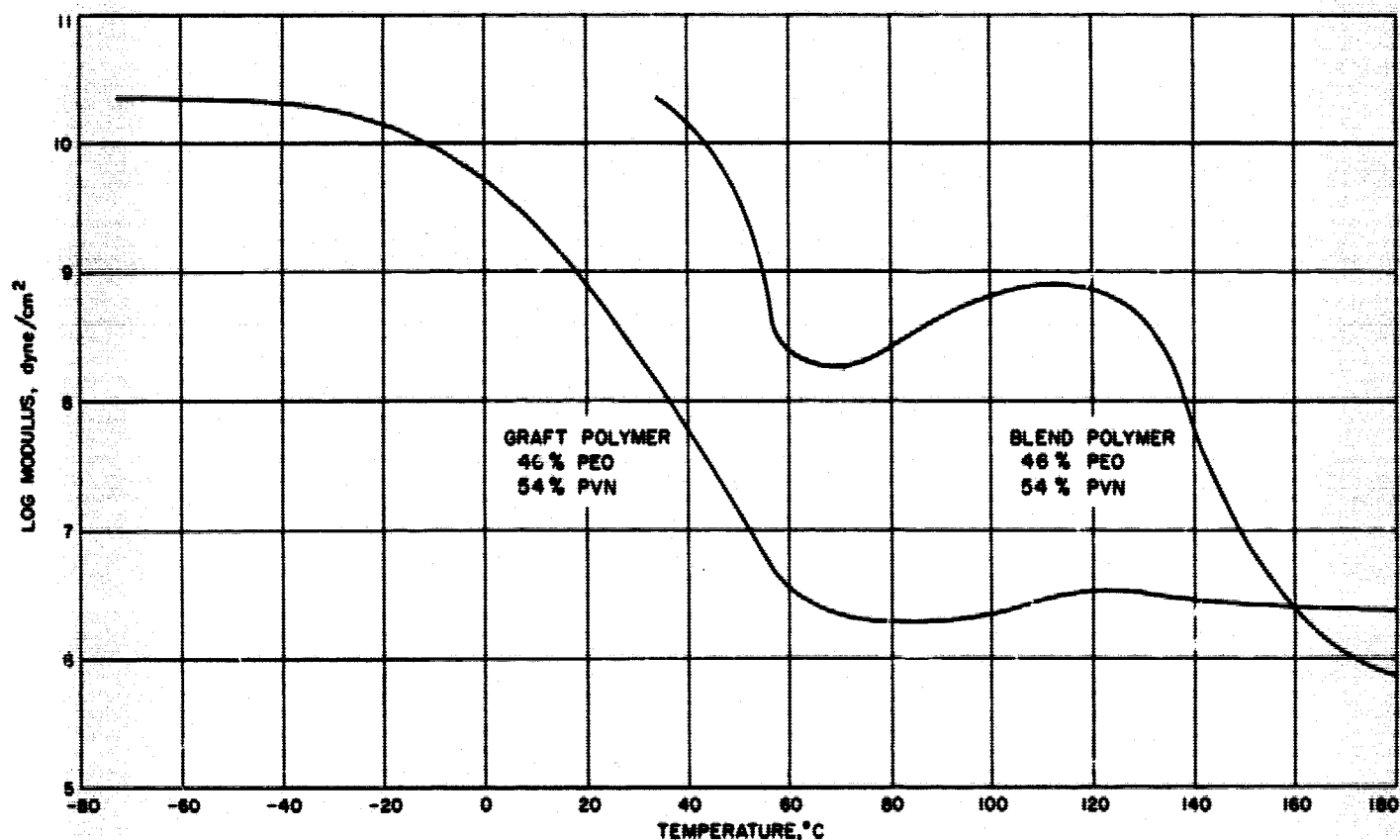


Fig. 4. Log modulus vs temperature for a blend and graft polymer

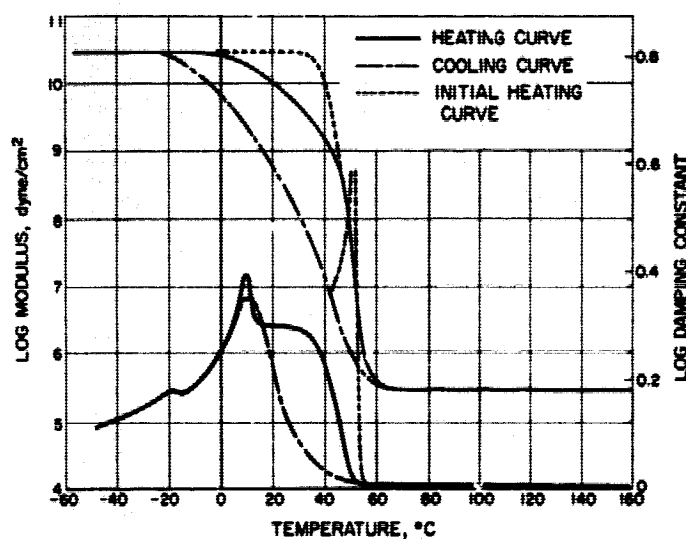


Fig. 5. Log modulus and log damping constant vs temperature for a graft copolymer of PVN and PEO

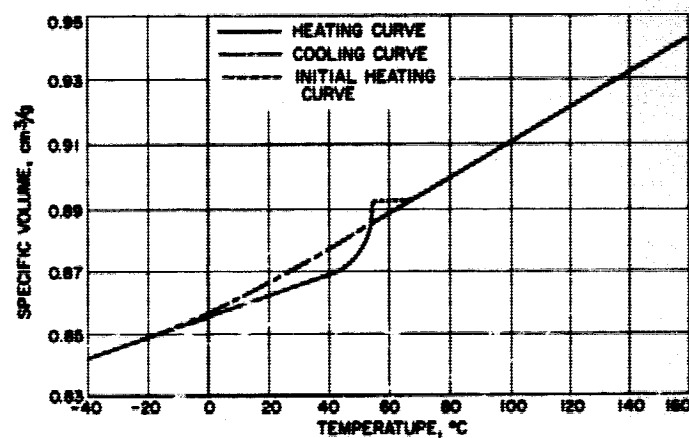


Fig. 6. Dilatometric curve of a graft copolymer of PVN and PEO

favorable but probably weak energetic interactions is evident.

oxide to one naphthalene moieties, namely the "complex" composition. Thus, here as in the case of polyblends, the importance of conformational behavior coupled with

Phase-contrast micrographs show that initially the graft copolymer exhibited layers (Fig. 7) which were perpendicular to the direction at which pressure was



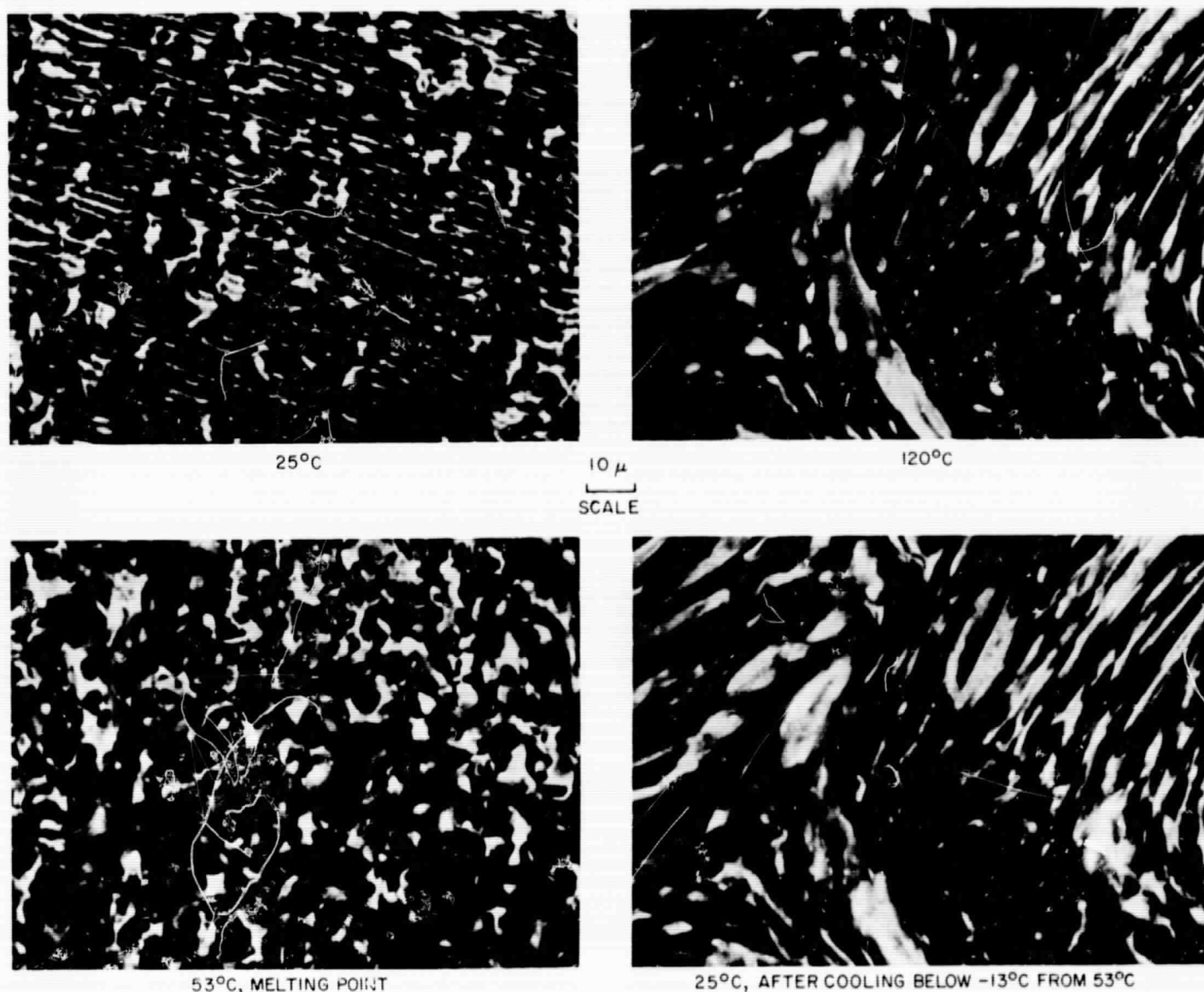


Fig. 7. Phase-contrast photomicrographs of a graft copolymer of 54 wt % PVN and 46 wt % PEO

applied to form the specimen. The layered structure disappeared above the melting temperature of PEO. After cooling to  $-20^{\circ}\text{C}$ , the structure shown in Fig. 7 (lower right) was formed. For the copolymer the granular structure characteristic of polyblends is absent. Moreover, the rather uniform appearance achieved once the graft had been heated remained unchanged upon further temperature cycling. This offers additional evidence that at higher temperatures microphase separation cannot take place.

Evidently, for copolymers the existence of the homogeneous "complex" phase is an equilibrium property.

Any effects resulting from the history of specimen preparation (micrographs and dashed lines in Fig. 5) are eliminated by annealing. In contrast, for polyblends the compatible phase is a metastable state which can exist under restricted conditions.

In summary, our studies have shown that graft copolymers of PVN with PEO form homogeneous, rubbery materials. The properties of copolymers having low PEO content remain to be investigated in order to assess the potential of using PEO as an internal plasticizer. Also, the effect of the size of side chains will have to be considered.

## B. Kinetics of the Anionic Polymerization of Acenaphthylene

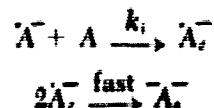
J. Moacanin and A. Rembaum

We have observed that chain transfer occurs during the anionic polymerization of acenaphthylene in a manner similar to that reported recently for 9-vinylanthracene (Ref. 5). The polymerization rate of acenaphthylene in tetrahydrofuran was sufficiently slow to permit conventional dilatometric procedures to be used to study kinetics (Ref. 6). Furthermore, since chain transfer may apply to the anionic polymerization of other systems (Refs. 7, 8, 9), we have derived theoretical kinetic expressions based on the reaction mechanism given below. These expressions could be readily adapted to other similar mechanisms.

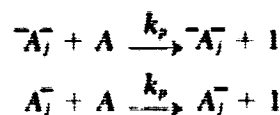
From the present study three important facts emerged: (1) the over-all rate was found to follow pseudo first-order kinetics; (2) the activation energies, as compared to other anionic systems, were high, and the propagation rates were exceptionally low; (3) the highest molecular weight which could be obtained was about 8000; electron spin resonance (ESR) measurements showed that the acenaphthylene radical anion initiator was the sole radical species which could be observed, and that its

concentration remained nearly constant during the polymerization. In order to account for these observations, we postulated a slow rate determining initiation step, followed by fast dianion formation, and a chain transfer to monomer mechanism which competes with the propagation step, according to the following scheme:

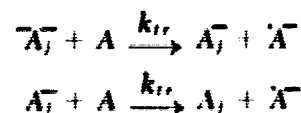
Initiation



Propagation



Chain transfer



where  $\dot{A}^-$  is the monomer radical anion,  $A$  the monomer, and  $A_j^-$ ,  $\bar{A}_j^-$ , and  $\bar{A}_4^-$  are the "dead," monoanion, and dianion  $j$ -mer, respectively. The alkali metal counterions are not shown in the scheme.

The results of this study are summarized in Table 1, and in Table 2 a few results on the free radical poly-

Table 1. Anionic polymerization of acenaphthylene

Initiator	Temperature, °C	Initial initiator concentration, mole/liter <sup>a</sup>	Polymer, % yield	Rate constant <i>K</i> , liter/mole-sec × 10	Molecular weight <sup>b</sup>	
					Measured	Calculated
Lithium	49.8	0.096	98	2.5	2,540	4,000
		0.120	35	4.1		
		0.150	96	5.4	1,300	2,600
	74.8	0.073	93	11	2,620	5,200
		0.135	90	22	1,280	2,800
		0.036	95	24	4,740	10,800
	95	0.071	95	36	2,780	5,400
		0.096	84	54	1,600	4,000
Sodium	49.9	0.054	93	0.14	1,900	7,200
		0.034	91	6.6	2,260	11,400
		0.043	89	7.8	2,020	9,000
	95	0.056	94	6.5	2,280	6,800
		0.157	88	23.0	1,340	2,400
		0.031	97	16	3,870	12,400
	115	0.047	95	29	2,960	8,200
		0.061	88	36	2,460	6,400

<sup>a</sup>The initial acenaphthylene concentration was 1.27 moles/liter.

<sup>b</sup>Measured by means of Mechrolab (Mountainview, California) vapor pressure osmometer; calculated =  $2(152)(1.27)/I_p$ .

Table 2. Free radical polymerization of acenaphthylene

Benzoyl peroxide	Temperature, °C	Solvent	Yield, %	Time, hr	Molecular wt <sup>a</sup>
✓	100	—	67.5	24	285,000
✓	125	—	74.3	24	335,000
✓	150	—	86.0	24	180,000
✓	100	Tetrahydrofuran	46.0	24	160,000
—	100	—	42	96	~2,000,000
—	125	—	65	96	820,000
—	150	—	78	96	160,000
—	200	—	82	96	140,000

<sup>a</sup>Estimated from intrinsic viscosity determinations in benzene, using the relationship  $[\eta] = 5.3 \times 10^{-5} \times M^{0.68}$  (Ref. 14).

merization are given for comparison. The kinetics data were obtained from dilatometric measurements, which were carried out using the usual vacuum techniques (Ref. 10). The concentration of "living ends" was determined by both methyl iodide and alkali metal titrations. The linearity of the monomer concentration versus time semilog plots is illustrated in Fig. 8.

The observed pseudo first-order kinetics and the upper limit for the polymer molecular weight are both consistent with the postulate of chain transfer to monomer. The exact mechanism by which it takes place cannot be ascertained on the basis of existing experimental results, but it seems quite certain that electron transfer to monomer, probably coupled with hydrogen abstraction from solvent, must occur. Also, the detection of the acenaphthylene radical anion ESR spectrum when butyl lithium is used to initiate the polymerization is consistent with

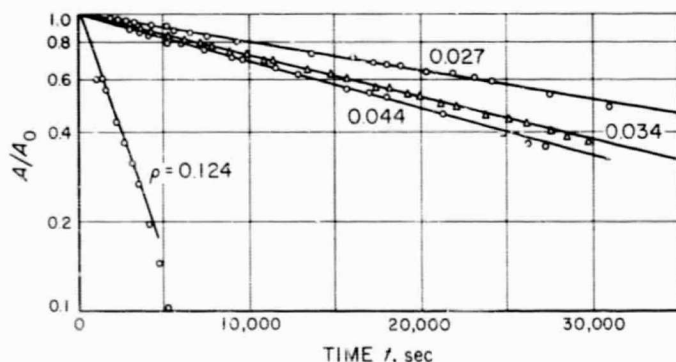


Fig. 8. Sodium-initiated polymerization of acenaphthylene at 95°C ( $\rho$  is the initial initiator-to-monomer ratio)

the postulated mechanism. Disproportionation of two anionic chain ends is unlikely, because such mechanism would lead to a continuous depletion of growing centers, and this is contrary to the observed kinetics.

The solution of differential equations<sup>1</sup> corresponding to the above reaction scheme leads to the following expression for the monomer concentration as a function of time:

$$A/A_0 = 1 - \rho \frac{\kappa_i (1 + 2\kappa_{tr})}{(\kappa_i + \kappa_{tr})} \left\{ \tau - \frac{(1 + \kappa_{tr} - \kappa_i)}{(1 + 2\kappa_{tr})(\kappa_i + \kappa_{tr})} \times [1 - e^{-(\kappa_i + \kappa_{tr})\tau}] \right\} \quad (1)$$

where  $A_0$  is the initial monomer concentration; and  $\kappa_i = k_i/k_p$ ,  $\kappa_{tr} = k_{tr}/k_p$ , and  $\rho$  is the initial initiator-to-monomer ratio. The auxiliary variable  $\tau$  was introduced to linearize the differential equation (Ref. 11) and is defined by  $d\tau = k_p A d\tau$ . The condition for first order kinetics in the monomer consumption is

$$(d/d\tau)(A/A_0) = -\rho(1 + 2\kappa_{tr})/(1 + \kappa_{tr}/\kappa_i) \quad (2)$$

which can be rewritten as a function of time as

$$dA/dt = -k_{eff}A \quad (3)$$

where

$$k_{eff} = k_p(1 + 2\kappa_{tr})/(1 + \kappa_{tr}/\kappa_i)$$

In practice this condition will be satisfied if at low conversion the exponential term in Eq. (1) becomes negligible.

Again neglecting the exponential term, the reciprocal of the number-average degree of polymerization  $P_N$  for complete conversion is given approximately by

$$(1/P_N - 4) = \kappa_{tr}/2 + \rho/[2(1 + \kappa_{tr}/\kappa_i)] \quad (4)$$

The constant 4 enters these expressions because, according to the mechanism, polymerization starts with the tetramer dianion. A plot of the left-hand side versus  $\rho/2$  should be linear, the intercept giving  $\kappa_{tr}/2$ , and the slope  $1/(1 + \kappa_{tr}/\kappa_i)$ . On the basis of the present experimental data,  $\kappa_{tr}$  is approximately  $1/10$  and the slope is definitely positive, indicating that  $\kappa_{tr}/\kappa_i < 1$ ; otherwise the slope would tend to vanish. A direct measure of this ratio can be obtained from analyses of the initiator radical anion concentration, since  $A_0^-/A_\infty^- = 1 + \kappa_{tr}/\kappa_i$ . Attempts to quantitatively interpret the ESR spectra as well as chemical determinations of the initiator content are underway.

<sup>1</sup>Moacanin, J., and Rembaum, A., unpublished results.

Accepting tentatively that  $\kappa_{tr} < \kappa_i$  and  $\kappa_{tr} \approx 0.1$ , we conclude that  $k_p$  is approximately equal to  $k_{tr} \approx 10^{-3}$  (liter/mole-sec). This is an extremely low value in comparison to the magnitude of the polymerization rate constant reported for other systems. Even if we allowed for our data  $\kappa_{tr}/\kappa_i = 10$ , the value for  $k_p$  would still be several orders of magnitude smaller than that for styrene ( $k_p \approx 600$ ) (Ref. 12).

A change in the counterion from sodium to lithium (Table 1) resulted in an acceleration of rate. There appears to be some increase in molecular weight at 115°C, but the interpretation of results at elevated temperatures is complicated because of the probable occurrence of free radical reactions. A control run showed, however, that in absence of initiator no polymerization takes place up to 95°C. Examination of Table 2 shows that very high molecular weights can be obtained by free radical methods. With increasing polymerization temperature, however, molecular weight decreases rapidly.

A study of the temperature effect indicates an over-all activation energy of 11–22 kcal/mole, presumably representing mainly the propagation step. This is a very high value in comparison to  $\sim 1$  kcal/mole reported for styrene (Ref. 12). The estimation of the relative contributions of initiation and chain transfer must await further results. The activation energy is of the same magnitude as that observed by Imoto and Soematsu (Ref. 13) for the acenaphthylene polymerization by a Ziegler catalyst ( $\sim 15$  kcal/mole). Intrinsic viscosity values indicate low molecular weights for their polymers; their osmotic pressure data, which suggest high molecular weights, appear to be in error (Ref. 14). It is of interest to note that the activation energy for the anionic polymerization of 9-vinylanthracene is also high (8–15 kcal/mole) (Ref. 5).

Surprisingly, powder X-ray patterns for the polyacenaphthylenes prepared by either anionic, cationic, or free radical methods exhibited crystallinity, although the polymerization conditions used normally do not lead to stereospecificity. This result indicates that the placement of the monomer is controlled by the geometry of the monomer unit (Ref. 15). Examination of molecular models suggests that acenaphthylene should add in the trans position relative to the penultimate monomer unit, the cis addition being sterically impossible. Furthermore, for the trans configuration, isotactic placement appears more likely than syndiotactic placement. Trans-isotactic chains should favor helical conformations of reasonable flexibility. These speculations are consistent with our studies on the solution properties of polyacenaphthylene

(Ref. 14). For low molecular weight polymers there is evidence of the existence of the trans-syndiotactic configuration.<sup>2</sup>

<sup>2</sup>Story, H., private communication.

## C. Kinetics and Mechanism of the Cyanoethylation of 1- and 2-Propanols

S. H. Kalfayan

Some rate studies on the cyanoethylation of 1- and 2-propanols in p-dioxane (Dx) were reported previously (Refs. 16 and 17). It was noted then that precipitation of the catalyst, sodium propoxide or sodium isopropoxide, started to take place when the reaction was about half complete. Up to the point of catalyst precipitation, the kinetic data fitted first order rate expressions. It is desirable, however, to follow the reaction well over 50% conversion. Cyanoethylations were, therefore, carried out in absence of Dx. When the initial molar ratio of alcohol to acrylonitrile (AN) was 1:1, the catalyst still precipitated before equilibrium was attained. This, however, occurred when the reactions were 70–75% complete. At higher dilutions (2:1, 3:1) the catalyst remained in solution during the entire course of the reaction. Kinetic data obtained show a first order rate dependence on the acrylonitrile concentration and a first order dependence on the catalyst concentration. The rates are independent of the alcohol concentration. Energies and entropies of activation are reported. A mechanism for the cyanoethylation of alcohols is proposed.

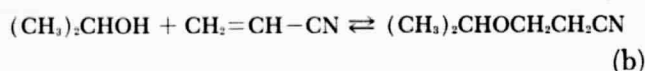
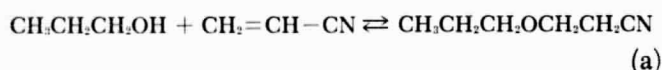
### 1. Experiments

The gas chromatographic equipment used during the present analyses was described previously (Ref. 16). Columns were made of 12-ft  $\times$  1/4-in. aluminum tubing packed with 20% nonpolar Ucon fluid on Chromosorb "W." Column, injection port, and detector temperatures were set at 100, 120, and 265°C, respectively. The bridge current was 150 ma and the helium flow-rate was 86 ml/min. Peak areas were measured by a voltage-to-frequency type integrator (Infotronics Corporation, Houston).

For kinetic studies, calculated amounts of alcohol and AN were placed in 50-ml glass-stoppered volumetric flasks and thermostated for a minimum of 3 hr before addition of the catalyst. Hallikainen low-temperature baths, capable of controlling the temperature within  $\pm 0.05^\circ\text{C}$ , were used. Samples were chemically quenched prior to analysis by gas chromatography. The quenching procedure consisted of drawing samples from the reaction flask into a 1-ml graduated syringe which contained a measured amount of dry HCl in Dx. Neutralization of catalyst was thus rapid and temperature changes during sampling were minimized. 1- $\mu\text{l}$  samples were introduced into the gas chromatographic apparatus.

## 2. Results and Discussion

a. *Apparent rate constants.* Cyanoethylation of 1- and 2-propanols are represented as follows:



The rates of both Reactions (a) and (b) were determined at several initial reactant ratios, catalyst concentrations, and at two different temperatures. The solvent Dx was omitted. The alkoxide catalysts, sodium propoxide and sodium isopropoxide, are practically insoluble in Dx and their solubility in the corresponding alcohol is limited. When the initial molar ratio [alcohol]/[AN] was 1:1, the catalyst still precipitated before equilibrium was reached. This occurred when the reaction was well over 50% complete [75–80% for Reaction (a) and 65–70% for Reaction (b)]. Up to these extents of reactions the standard deviations for the apparent rate constants were less than 4%. When the initial ratios were 2:1 or 3:1, no catalyst precipitated for the entire duration of the reaction and apparent rate constants could be calculated with standard deviations of less than 4%, up to 90% reaction.

The rate constants were calculated from the following integrated first order equations:

$$k = \frac{2.303}{t} \log \frac{A_0}{A_t} \quad (1)$$

$$k = k_1 + k_{-1} = \frac{2.303}{t} \log \frac{A_0 - A_e}{A_t - A_e} \quad (2)$$

where  $k$  = apparent first order rate constant,  $k_1$  and  $k_{-1}$  are the first order rate constants for the forward and reverse reactions.  $A_0$ ,  $A_t$ , and  $A_e$  are the normalized

(against Dx) peak areas of AN at  $t = 0$ ,  $t$ , and  $t$  at equilibrium.

Fitting the kinetic data into Eq. (1) gave consistent apparent first order constants with less than 4% deviation to about 75% completion (excluding the 1:1 molar ratio). Beyond this range there was a downward drift. Equilibrium was attained at about 95 and 94% conversions for Reactions (a) and (b), respectively.

When the same data was fitted into Eq. (2), which is the integrated expression for a first order forward step and a first order reverse step, apparent rate constants could be calculated with standard deviations less than 4%, to about 90% conversion.

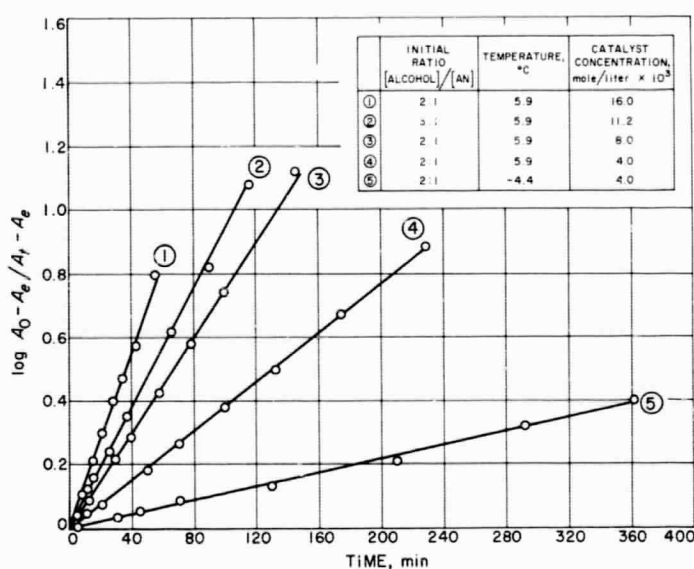


Fig. 9. Cyanoethylation of 1-propanol

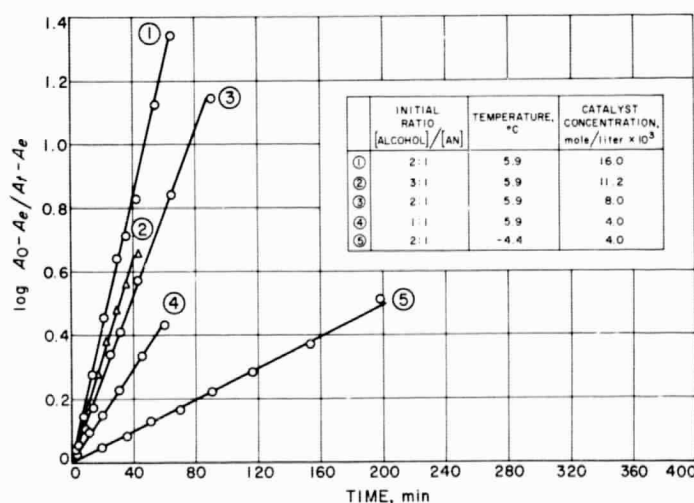


Fig. 10. Cyanoethylation of 2-propanol



The apparent rate constants given in Table 3 were calculated according to Eq. (2). Some of the first order plots are shown in Figs. 9 and 10.

It is seen from Table 3 that the rates of both Reactions (a) and (b) are catalyst dependent, first order in AN, and independent of the alcohol concentration.

Both Reactions (a) and (b) were slower when carried out in Dx (Ref. 17). Reaction (b) was slower than (a) in presence of Dx. The reverse is observed in absence of Dx.

Plots of  $\log k$  at 5.9°C vs  $\log C$  (catalyst concentration) for initial mole ratios of 1:1, 2:1, and 3:1 are shown in Fig. 11. These are linear with slopes of 1.0 and 0.80 for Reactions (a) and (b), respectively. This suggests that the rates are first order with respect to catalyst concentration.

**b. Energy and entropy of activation.** Activation energies  $\Delta E^*$  were calculated from Arrhenius plots. Values obtained were 18.0 and 15.7 kcal/mole for Reactions (a) and (b), respectively. Consequently, the enthalpy  $\Delta H^*$  and entropy  $\Delta S^*$  at 5.9°C were 17.4 kcal/mole and  $-5.5$  cal/mole-deg for Reaction (a) and 15.1 kcal/mole and  $-12.4$  cal/mole-deg for Reaction (b). These values were obtained from the following equations (Ref. 18):

$$\Delta H^* = \Delta E^* - RT$$

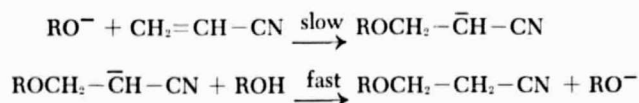
$$k = \frac{RT}{Nh} e^{\frac{\Delta H^*}{RT}} e^{\frac{\Delta S^*}{R}}$$

where  $R$  is the gas constant,  $h$  is the Planck constant, and  $N$  is Avogadro's number.

Entropy values could have been calculated using the activation energies  $\Delta E^*$  since the numerical value of the  $RT$  term in the enthalpy equation is about the same (or less) than the uncertainty in the  $\Delta E^*$  values.

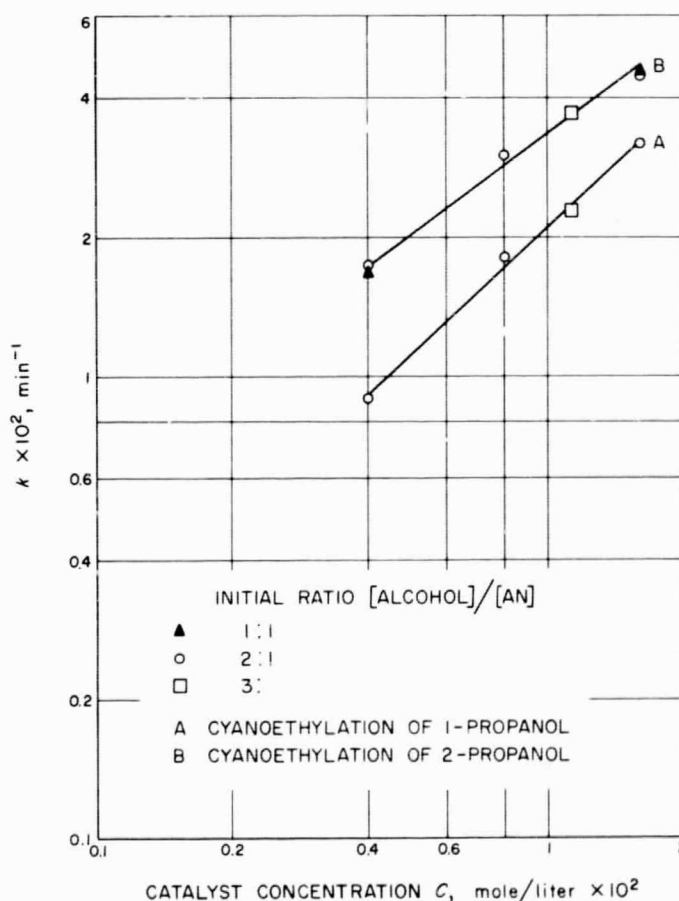
### 3. Mechanism

Since the cyanoethylations of both 1- and 2-propanols are independent of alcohol concentration and are first order in AN and first order in catalyst concentration, a likely mechanism for these reactions is as follows:



**Table 3. First order rate constants for the cyanoethylation of 1- and 2-propanols**

Temperature, °C	Catalyst concentration, mole/liter $\times 10^3$	Initial ratio [alcohol] / [AN]	$k \times 10^2$ , min <sup>-1</sup>	$\sigma$
<b>1-propanol</b>				
-4.4 +5.9	4.0	2:1	0.250 $\pm$	0.010
	16.0	2:1	0.866 $\pm$	0.02
	4.0	2:1	0.886 $\pm$	0.01
	8.0	2:1	1.79 $\pm$	0.07
	11.2	3:1	2.27 $\pm$	0.09
	16.0	2:1	3.15 $\pm$	0.10
<b>2-propanol</b>				
-4.4 +5.9	4.0	2:1	0.572 $\pm$	0.02
	4.0	1:1	1.74 $\pm$	0.05
	4.0	2:1	1.73 $\pm$	0.04
	8.0	2:1	2.98 $\pm$	0.10
	11.2	3:1	3.69 $\pm$	0.10
	16.0	2:1	4.74 $\pm$	0.15
	16.0	1:1	4.77 $\pm$	0.15



**Fig. 11. Alcohol and catalyst dependence**

Thus, the rate-controlling step is the addition of the alkoxide anion to the  $\alpha - \beta$  double bond of acrylonitrile.

Ogata et al. (Ref. 19) found the cyanoethylation of methanol at very low concentrations of AN, with sodium methoxide catalyst, to be a pseudo-first-order reaction. Their proposed mechanism is similar to the one above.

The extent of conversion is not changed in the absence of solvent, about 4.5% of the alcohol remaining unreacted. Therefore, hydrolysis of the cyano-ether gives the corresponding carboxylated product contaminated with the alcohol. Preliminary efforts have indicated that the alcohol and undesired hydrolytic side-products can be removed by isolating the carboxylated product as an insoluble salt.

## D. Polymer Chemistry: C<sup>14</sup> Distribution in Labeled 2,4-Toluene Diisocyanate (TDI)

E. F. Kopka and J. D. Ingham

### 1. Introduction

Well-characterized polyurethanes prepared from TDI labeled with C<sup>14</sup> in the isocyanate groups are extremely useful for the study of polymer degradation mechanisms (Refs. 20 and 21). The unique feature of tracer applications is that, because of the high sensitivity of radioactivity measurements, very small extents of bond scission can be detected.

However, for an unsymmetrical diisocyanate such as TDI, the distribution of C<sup>14</sup> in the isocyanate groups depends on the kinetics of isocyanate formation and on the method of preparation. Ideally, the relative C<sup>14</sup> activity in the 2- and 4-positions should be equal, or at least known, to reasonably interpret results of degradative studies. One potential method for determining the C<sup>14</sup> distribution is the subject of this report.

### 2. Reaction of TDI with 2,6-Dimethylaniline; Subsequent Hydrolysis and Methanolysis

The C<sup>14</sup> distribution in isocyanate-labeled TDI can be established by following a sequence of reactions in which

the carbon atom of one isocyanate group is removed exclusively followed by specific radioactivity determinations of TDI and its derivative. Such a reaction sequence is shown in Fig. 12. The work presented here is a study of these reactions and an attempt to establish the yields and purity of the products using unlabeled TDI. Subsequently, labeled TDI will be employed; from its specific activity and the specific activities of I, and either II, III, or IV, the C<sup>14</sup> distribution can be calculated. Obviously, yields should be nearly 100% for valid determination of the distribution.

Previous work (Ref. 22) has shown that equimolar amounts of aniline and TDI react to give ~70% yield of 4-methyl-3-isocyanatocarbanilide (V), which is the same as I but without the methyl groups in the 2' and 6' positions. For the present work 2,6-dimethylaniline was used to increase the probability of exclusive reaction at the isocyanate group in the 4-position. The yield of a product believed to be I was ~87%. The melting point of I was 207–209°C; however, II, III, and IV have melting ranges >250°C, which is above their decomposition temperatures, so that sharp melting points for them were not obtained. A sample of the tentative product [I]<sup>1</sup> has been submitted for elemental analysis. Since infrared and thin-layer chromatography indicated the presence of some undesired side-products, possibly resulting from reaction of the free isocyanate, and since past experience has shown that isocyanates are difficult to count in the scintillation spectrometer, a sample of [I] was treated with methanol to give the methanolysis adduct [VI] in 94% yield. Samples of [I] were also hydrolyzed to give [II] and [IV], as shown in Fig. 12, in ~98% and ~47% yield, respectively. Possibly the low yield of [IV] can be increased by scaling up the reaction; only ~0.1 g of product has been prepared thus far.

### 3. Thin-Layer Chromatography of Carbanilide Products

Since the melting points of derivatives such as II and IV were near or above their decomposition temperature, they could not be used as evidence that substantially pure products were prepared. Therefore, thin-layer chromatograms were obtained; a single spot under resolving conditions should indicate a single major compound in any reaction product. By trial and error it was found that a binary solvent mixture, consisting of 80:20 by volume of benzene:pyridine, was a good developing

<sup>1</sup>Elemental analyses of the products are not yet available. To indicate tentative identification, brackets are used: [I] for I, etc.

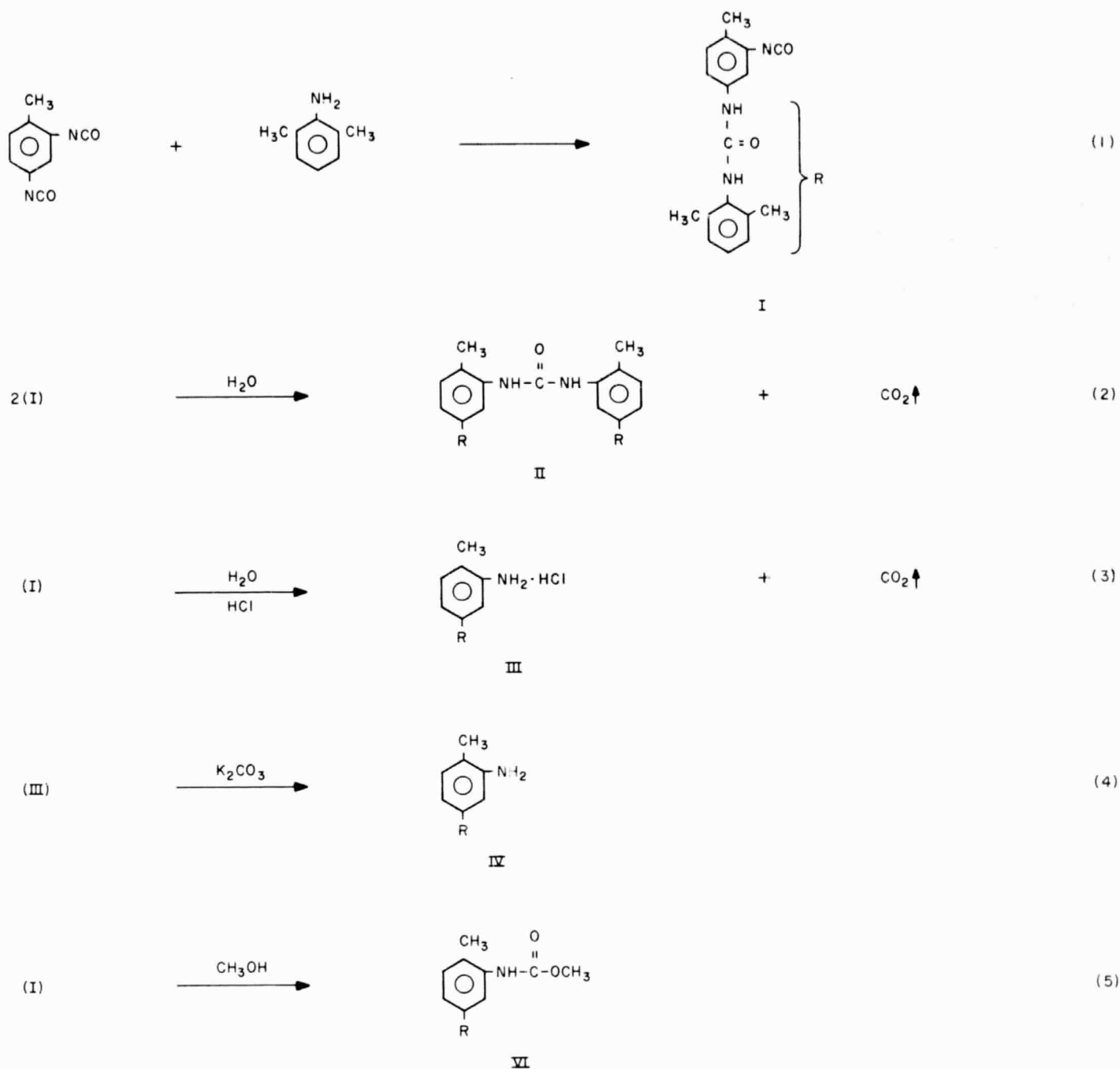


Fig. 12. Reactions of 2,4-toluene diisocyanate with 2,6-dimethylaniline and subsequent hydrolysis and reaction with methanol

system. The spots could be observed by spraying with ninhydrin solution and heating the plates at  $100^\circ\text{C}$ . A better method was found to be simple heating at  $100^\circ\text{C}$  and observation under ultraviolet light. A typical plate is shown in Fig. 13. The products used and corresponding  $R_f$  values are given in Table 4. A comparison of the first two chromatograms shows that [I] is not stable in chloroform so that its initial purity could not be

established directly. Apparently [I] is hydrolytically or otherwise transformed from a product that is mainly retained at the origin to one with an  $R_f$  value of 0.6 in this solvent system. The third chromatogram of Fig. 13 was a minor product removed from the filtrate after isolation of [I]. It appears to be a mixture of the compositions of the first two chromatograms. The two remaining chromatograms (Fig. 13) were products obtained by placing



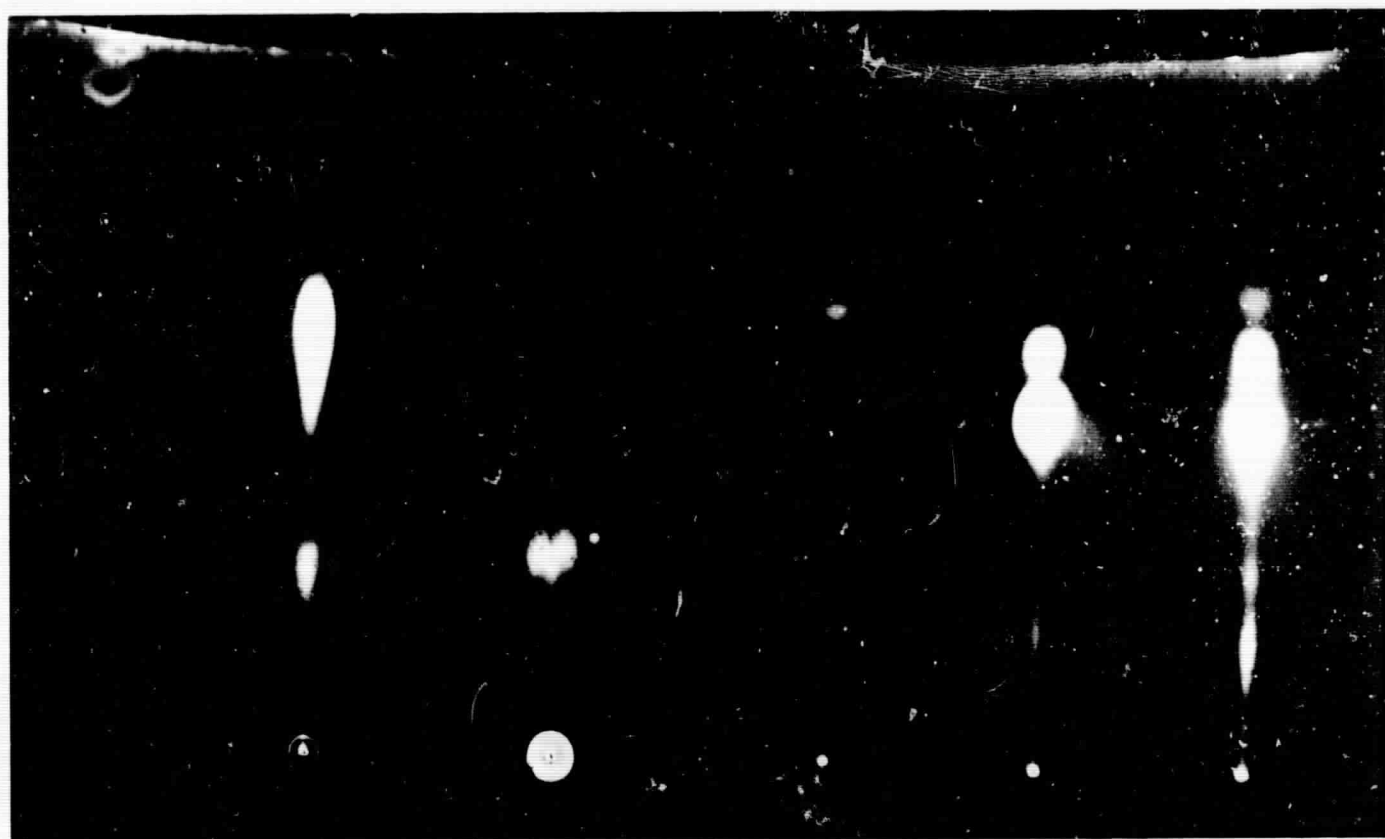


Fig. 13. Thin-layer chromatoplate of carbanilide products on silica gel

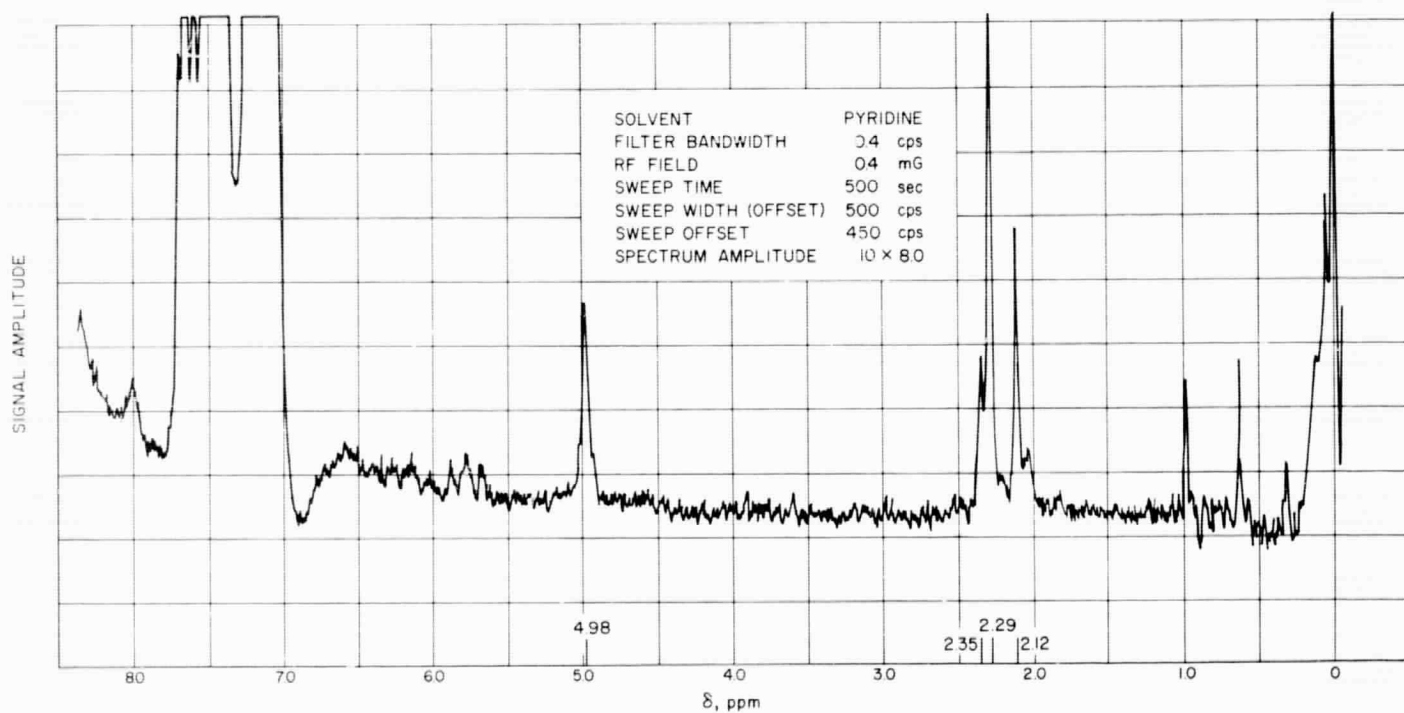


Fig. 14. Proton NMR spectrum of 4,2',6'-trimethyl-3-aminocarbanilide (3% in pyridine)

**Table 4.  $R_f$  values from thin-layer chromatographs of carbanilide products**

Product (chloroform solution)	$R_f$ values <sup>a</sup>
[I] TDI + dimethylaniline	Origin: very strong 0.29: weak 0.43: weakest 0.63: weaker
[I] TDI + dimethylaniline (solution aged 12 days)	Origin: weak 0.27: weak 0.61: strong
TDI + dimethylaniline (reaction time 3 days; solution aged 12 days)	0.29: weaker 0.63: weak
[I] soxhlet; refluxing methanol 6 days; recovered from thimble	0.50: strong 0.58: strong
[I] soxhlet; refluxing methanol 6 days; recovered from flask	0.49: strong 0.57: strong 0.66: weak
<sup>a</sup> Developed with benzene: pyridine (80:20 by volume; solvent front 100 mm from start) (data from Fig. 13).	

a sample of [I] in a soxhlet extractor and extracting with refluxing methanol for 6 days. This was the first attempt to obtain the methanolysis adduct VI. Apparently methanol did add to convert the isocyanate to methyl urethane but, because of the relatively harsh conditions, methanolic cleavage of the urea bond apparently occurred to give the two intense spots observed.

The methanolysis adduct [VI] (mentioned in Part 2, above) was obtained by refluxing [I] in methanol for 30 min. The yield was 94% and only one strong spot ( $R_f = 0.61$ ) was obtained by chromatography with a faint spot occurring at  $R_f = 0.69$ ; this is very good evidence that [I] is indeed the desired 4,2',6'-trimethyl-3-isocyanatocarbanilide and [VI] is the corresponding methyl urethane. Thin-layer chromatography of the hydrolysis product [IV] also showed an intense spot ( $R_f = 0.26$ ) and a faint spot ( $R_f = 0.35$ ), indicating that [IV] is primarily one product, presumably 4,2',6'-trimethyl-3-aminocarbanilide.

#### 4. Proton NMR of the Product IV: 4,2',6'-Trimethyl-3-Aminocarbanilide

To prove that [IV] was 4,2',6'-trimethyl-3-aminocarbanilide, its proton nuclear magnetic resonance (NMR)

**Table 5. Integrals of NMR signals from the proton resonance of amino and methyl protons in 4,2',6'-trimethyl-3-aminocarbanilide**

Integration number	Amino <sup>a</sup> protons	2',6'-dimethyl <sup>b</sup> protons	4,2',6'-trimethyl <sup>c</sup> protons
1	1	3.12	4.05
2	1	3.24	4.11
3	1	2.98	3.83
Average	1	3.11	4.00
Calculated	1	3	4.5
<sup>a</sup> At $\delta = 4.98$ ppm, area ratio normalized so that amino area is 1. <sup>b</sup> At $\delta = 2.29$ ppm. <sup>c</sup> At $\delta = 2.29$ ppm + signal at $\delta = 2.12$ ppm.			

spectrum was obtained with the Varian A-60 spectrometer. Initially, spectra of deuterated chloroform solutions were obtained; however, the solubility was limited to ~0.5% and the amino proton signal was not observed, possibly because of deuterium exchange from a trace amount of deuterium oxide or deuterated methanol that may have been present in the solvent. Although the solubility was limited to ~3% in pyridine and the aromatic protons were obscured by the solvent, pyridine solutions gave satisfactory spectra (Fig. 14). Electronic integration of the signal areas gave the results shown in Table 5. The ratio of amino protons to 2',6'-dimethyl protons is 1:3 in IV. The average ratio found experimentally was 1:3.11, which agrees quite well. However, the ratio of amino protons to the total number of methyl protons in IV is 1:4.5, whereas a ratio of 1:4.0 was found because of the relatively small integral from the resonance signal assigned to the 4-methyl protons. It is tentatively believed that the 4-methyl chemical shift assignment is correct and that [IV] is the desired 4,2',6'-trimethyl-3-aminocarbanilide. A possible explanation for the low apparent 4-methyl proton integral is that a systematic integration error resulted from the low concentration and resulting high noise level.

More complete characterization of the urethane derivatives should be obtained from their elemental analyses. Experimental procedures have been sufficiently established for the reaction sequence shown in Fig. 12 and will be applied to determinations of the  $C^{14}$  distribution in labeled TDI.

## References

1. Cuddihy, E., Moacanin, J., and Rembaum, A., "Some Unusual Properties of Blends and Grafts of Poly- $\beta$ -Vinylnaphthalene and Polyethylene Oxide," SPS 37-22, Vol. IV, pp. 100-104, Jet Propulsion Laboratory, Pasadena, California, August 31, 1963.
2. Cuddihy, E., and Moacanin, J., "Phase-Contrast Microscopy Studies on Heterogeneous Polymer Systems," SPS 37-25, Vol. IV, pp. 67-70, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.
3. Koppelman, V. J., *Kolloid-Zeitschrift*, Vol. 144, p. 12, 1955.
4. Read, B. E., *Polymers*, Vol. 3, p. 529, 1962.
5. Eisenberg, A., and Rembaum, A., *Journal of Polymer Science*, Part B, Vol. 2, p. 157, 1964.
6. Moacanin, J., Rembaum, A., and Adler, R., "Kinetics of Anionic Polymerization of Acenaphthylene," SPS 37-23, Vol. IV, pp. 94-98, Jet Propulsion Laboratory, Pasadena, California, October 31, 1963.
7. Litt, M., and Szwarc, M., *Journal of Polymer Science*, Vol. 42, p. 159, 1960.
8. Zilkha, A., Barzakaz, S., and Ottolenzhi, A., *Journal of Polymer Science*, Part A, Vol. 1, p. 1813, 1963.
9. Koral, J. N., *Makromolekulare Chemie*, Vol. 62, p. 148, 1963.
10. Tobolsky, A., Rembaum, A., and Eisenberg, A., *Journal of Polymer Science*, Vol. 14, p. 1735, 1963.
11. Ginell, R., and Simha, R., *Journal of the American Chemical Society*, Vol. 65, p. 706, 1943.
12. Szwarc, M., *Advances in Chemical Series*, Vol. 34, p. 96, 1962.
13. Imoto, M., and Soematsu, I., *Bulletin of the Chemical Society of Japan*, p. 34, 1961.
14. Moacanin, J., Rembaum, A., and Laudenslager, R. K., *American Chemical Society, Polymer Preprints*, Vol. 4, p. 179, 1963.
15. Dall'Asta, G., and Mazzanti, G., *Makromolekulare Chemie*, Vol. 61, p. 178, 1963.
16. Kalfayan, S. H., and Havlik, A. J., "Polyoxyalkylenes Terminated with Different Functional Groups: Rate and Equilibrium Studies of Cyanoethylation," SPS 37-22, Vol. IV, pp. 91-94, Jet Propulsion Laboratory, Pasadena, California, August 31, 1963.
17. Havlik, A. J., and Kalfayan, S. H., "Cyanoethylation-Propanols and Polyoxypropylene Glycol," SPS 37-23, Vol. IV, pp. 84-86, Jet Propulsion Laboratory, Pasadena, California, October 31, 1963.
18. Frost, A. A., and Pearson, R. G., *Kinetics and Mechanism*, John Wiley and Sons, N. Y., p. 96, 1956.
19. Ogata, Y., Okano, M., Furuya, Y., and Tabuski, I., J., *American Chemical Society*, Vol. 73, p. 5426, 1956.

## References (Cont'd)

20. Rapp, N. S., and Ingham, J. D., "Polymer Degradation Mechanisms: C<sup>14</sup>-Labeled Polyoxypropylene Glycol-Toluene Diisocyanate Polymers," SPS 37-23, Vol. IV, pp. 104-108, Jet Propulsion Laboratory, Pasadena, California, October 31, 1963.
21. Ingham, J. D., and Rapp, N. S., "Polymer Degradation Mechanisms: C<sup>14</sup>-Labeled Polyoxypropylene Glycol-Toluene Diisocyanate," SPS 37-25, Vol. IV, pp. 70-75, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.
22. Simons, D. M., and Arnold, R. G., "Relative Reactivity of the Isocyanate Groups in Toluene-2,4-diisocyanate," *Journal of the American Chemical Society*, Vol. 78, pp. 1658-1659, 1955.

## X. Liquid Propulsion

### A. Advanced Liquid Propulsion System

*R. N. Porter, H. B. Stanford, L. K. Jones, L. H. Peebles,  
W. F. MacGlashan, Jr., and W. H. Tyler*

#### 1. Introduction, *R. N. Porter*

The Advanced Liquid Propulsion System (ALPS) program is investigating selected problems generated by spacecraft operational requirements for propulsion systems capable of high inherent reliability, long-term storage in space, multiple start in free fall (zero gravity), and engine throttling. The solutions proposed to satisfy these requirements have been incorporated in the ALPS.

Periodic reports in SPS, Vol. IV, follow the progress of work on the various ALPS projects. The following paragraphs present accomplishment of significance which has occurred during the last reporting period. This includes the simultaneous expulsion of both propellants hydrazine ( $N_2H_4$ ) and nitrogen tetroxide ( $N_2O_4$ ) from a single tank using metal diaphragms as the expulsion device, two successful firings of a 2000-lb, free-standing pyrolytic graphite uncooled rocket chamber, and a graphic demonstration of hypergolic propellant streams blowing apart upon contact within a plexiglass rocket chamber.

#### 2. Diaphragm-Type Expulsion Device Development, *H. B. Stanford*

Existing slosh and expulsion test equipment in use at JPL was described in Ref. 1. This equipment was designed for use with test tanks 18 in. in diameter or smaller; and for reasons of compatibility and limitations on propellant handling, it was intended to be used only with water as the expelled fluid. Obviously, in a program to develop expulsion devices for use with rocket propellants, it eventually becomes necessary to conduct both slosh and expulsion tests using the actual propellants in quantities which the end objective may require. For this reason, a contract was let to Lockheed Aircraft Services of Ontario, California, to design and fabricate a device capable of supporting and sloshing (oscillating) tank shapes up to 64-in.-D spheres weighing as much as 6,000 lb, combined weight of tank and propellant. The support structure is a cage of heavy construction in the general shape of a tetrahedron with a circular mounting ring at the center. It pivots at the lower end on a mono-ball joint and is laterally driven at the upper end by an eccentric arm from a controlled power source. The equipment is capable of oscillation frequencies of 0 to 16 cycles/sec and linear amplitudes up to  $\pm 4$  in. at the support ring, but is not intended to withstand acceleration forces in excess of 2 g. Weight measurements for the determination of expulsion efficiency are obtained by

means of a load cell mounted beneath the monoball joint. A spider-type adaptor ring is required to install tanks appreciably smaller than the 64-in.-ID mounting ring. Such was used to mount the 18-in.-ID stainless steel tank used in the initial expulsion tests.

Three slosh and expulsion tests, using the described equipment and live propellant, have been conducted. The expulsion devices used were uniform thickness 0.015-in. aluminum diaphragm, 18 in. in diameter (Ref. 1). In all of these tests sloshing was conducted at a linear amplitude of  $\pm 1\frac{1}{2}$  in. and a frequency of 128 cycle/min or approximately 1 g. These conditions had been determined in previous tests with the small slosh rig to be the most severe (Ref. 2).

The first and second tests were conducted with each propellant individually so that techniques for filling, purging, and general propellant handling could be worked out separately before attempting to handle both propellants at the same time. These first diaphragms were not of the nesting variety so they were backed up by a divider plate which is a physical barrier but is not gas tight. This divider allows the use of one diaphragm at a time without the danger of fill pressure extending the diaphragm the wrong way. Filling was accomplished through the outlet ports by venting the top of the propellant compartments. In the case of  $N_2H_4$ , filling was done the day before the run so that an inspection for leakage could be made after a considerable storage time. Nitrogen tetroxide was remotely filled immediately before running. In all cases expulsion pressure was provided by a single source of  $N_2$  at 250 psi. Weight readings were taken from the load cell before and after expulsion to determine expulsion efficiency.

Some difficulty was encountered during the first two runs because of a faulty installation of the monoball mounting bolt and because of excessive friction at the slosh rig guide rails. Both of these defects affected weight readings and consequent determinations of expulsion efficiency. In both cases, however, expulsion was conducted in approximately 90 sec without incident under sloshing conditions as previously explained.

The third run, which involved both propellants simultaneously, was handled in the manner described above with  $N_2H_4$  being filled the day before and  $N_2O_4$  being filled, remotely, immediately before expulsion. In this run a mutual start switch failed to open the oxidizer valve. 20 sec later the oxidizer flow was started by manual over-

ride, allowing approximately 70 sec of simultaneous expulsion time. No other difficulties were encountered.

Examinations of the diaphragm after expulsion revealed the characteristic patterns caused by traveling creases (Ref. 2) but no damage that could be attributed specifically to the expulsion of live propellants.

A second dual propellant expulsion test was made at a later date under similar conditions. No difficulties were encountered and no leakage through the diaphragms occurred.

Dual propellant expulsion tests will be made with nesting diaphragms without a supporting divider as soon as necessary hardware has been obtained and fill techniques have been worked out.

### 3. Expulsion Control Device Development, L. K. Jones

Positive expulsion devices capable of meeting the 1- to 3-yr space environment storage specifications of the ALPS must be completely compatible with and impermeable to both hydrazine ( $N_2H_4$ ) and nitrogen tetroxide ( $N_2O_4$ ). To date only certain metals, metal foils, and composite materials containing metal foils tested in a static condition have met these fundamental requirements. The basic reason for an expulsion device is of course to separate the pressurizing gas from the fluid to be expelled. To do this adequately, the material from which the expulsion device is made must be capable of either conforming to or controlling a movable liquid surface and of eventually assuming the contour of the containing vessel in order that high expulsion efficiency may be obtained. Experience has shown that failures, such as tears and pinholes, in metal expulsion device materials occur within the first few expulsion cycles when the device is allowed to collapse or expand in a random fashion. Studies are being made of methods to control the folding modes of various expulsion devices; it is hoped the control methods will eliminate the random bends and creases that result in material failure. The initial phases of these studies on the bladder configurations are being conducted, using both compatible and non-compatible polymeric materials. Studies on the diaphragm-type expulsion device will generally use metals.

Fig. 1 shows one method of controlled expulsion collapse of a hemispherical bladder. This method makes use of the fact that the hemispherical bladder is a portion of a surface of revolution. The flat surface of the bladder folds along the diameter perpendicular to the direction



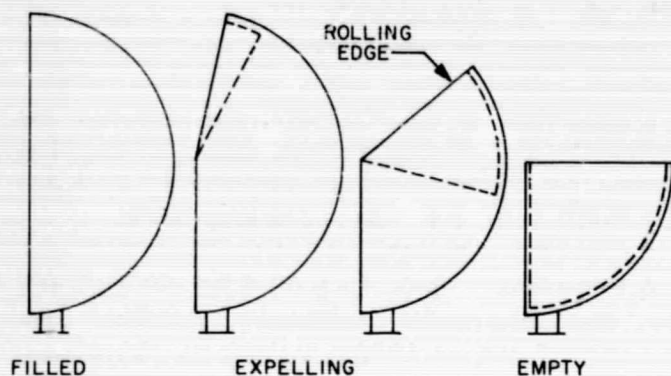


Fig. 1. Hemispherical bladder during controlled expulsion

of propellant discharge. As the top half of the flat, circular surface rotates about its fold axis, it "sweeps" the inside of the hemispherical bladder wall. Ideally, the rotating flat surface pulls the hemispherical surface behind it in a continuous, rolling motion. A smooth, rolling motion is not nearly as destructive to bladder walls as random creasing, especially if the rolling radius is relatively large. The final configuration shows the upper half of the hemisphere inverted within the lower half, a position capable of high expulsion efficiency.

An 18-in. hemispherical Teflon bladder was stiffened to induce the expulsion collapse control mentioned above. The lower half of the hemispherical surface was bonded to an aluminum shell with epoxy. The flat side of the bladder was stiffened with two semicircular plates, leaving an unstiffened "hinge" along the fold axis. Stiffening these surfaces prevents them from becoming distorted and interfering with the collapsing walls. The upper half of the hemispherical surface was left unstiffened to allow it to roll through 180 deg during expulsion.

The stiffened bladder was filled with water and expulsion tested, using the equipment described in Ref. 1. During expulsion, the upper semicircle rotated as desired. However, the upper half of the hemisphere collapsed and wrinkled ahead of the rotating semicircle, instead of being pulled behind it. At the end of expulsion, the top half of the bladder was correctly inverted within the bottom half. A second expulsion test, with induced sloshing, revealed the strengthening and damping qualities of the stiff, rotating semicircle. Bladder motion was greatly reduced from that of unstiffened bladder expulsion tests.

Methods of temporarily attaching the upper half of the hemisphere to the inside of the propellant tank wall are being studied. Rather than collapsing ahead of the rotat-

ing semicircle, the unstiffened wall would adhere to the propellant tank until it would be pulled off by the downward motion of the semicircle. For recycling capability, the surface should re-adhere during refilling.

Hemispherical bladders with other types of expulsion control and stiffening are being fabricated for testing.

#### 4. Heat Exchanger Development, I. H. Peebles

Two heat exchangers are used in the ALPS to cool the gases from the monopropellant gas generator to a temperature compatible with the propellant tank environment.

The fuel heat exchanger, which is fabricated from Type 347 corrosion resistant steel, to withstand the initial 1500°F gas inlet temperature, has co-current fluid flows with nucleate boiling over a large fraction of the transfer area at nominal flow conditions. The oxidizer heat exchanger, which is downstream of the fuel heat exchanger with respect to gas flow, is fabricated of 6061 aluminum and is designed to operate in the range of 250°F. It is of multipass, radial crossflow configuration with nucleate boiling over a small portion of the heat-transfer area. This flow sequence was selected to capitalize upon the high allowable heat flux in fuel ( $N_2H_4$ ) boiling, and the lower boiling point of the oxidizer ( $N_2O_4$ ). Typical ALPS heat exchanger assemblies are shown in Fig. 2, in which the directions of gas flow are indicated by arrows.

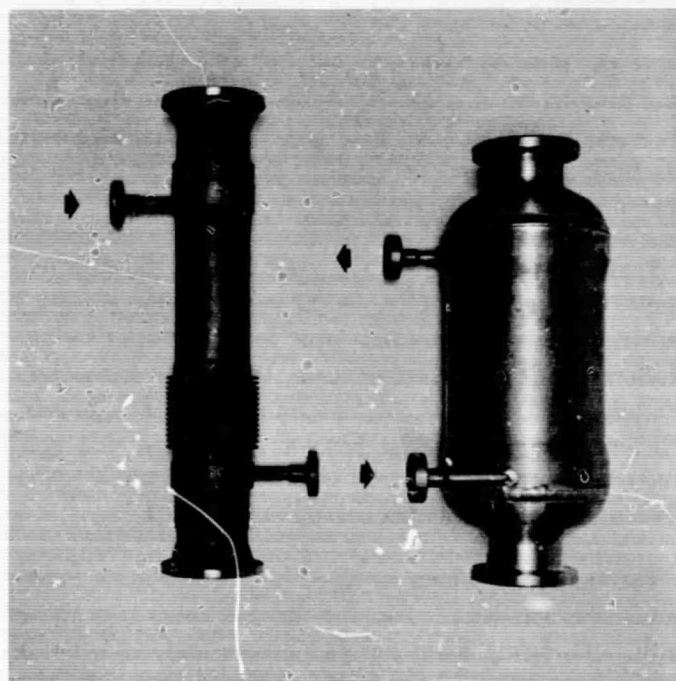


Fig. 2. ALPS fuel and oxidizer heat exchanger assemblies

Two heat exchangers of each type were assembled for test purposes, with enough spare parts available to build additional assemblies as the program advances. Proof and leak testing revealed minor leaks in some weld joints which were later repaired by rewelding. The propellant passages were flow-tested with water to determine pressure drop as a function of flow rate, and the gas passages were flow-tested with gaseous nitrogen. Hot gas testing is planned for the future.

### 5. Propellant Fill Valve

**Development,** W. F. MacGlashan, Jr.

Two goals of the ALPS component development program have been the accomplishment of weight saving through unique design concepts, and reliability through simplicity. A propellant fill valve under development for some time has achieved some success in these respects.

These valves, which weigh approximately 1 oz, consist of a floating ceramic ball that seats on a sharp edge orifice. Each valve is in two parts, a flight portion and a ground portion which may be connected or disconnected by means of a threaded coupling. They have been made in 1/2- and 3/4-in. sizes with flow rates of 1.0 and 0.2 lb/sec, respectively. Working pressure for the 1/2-in. valve is 300 psi, and for the 3/4-in. valve the pressure is 1500 psi.

To date four of these propellant fill valves (which may also be used for gas fill) have successfully passed proof, pressure drop, handling shock, and low temperature and humidity tests. Two leaked after the high temperature and humidity test and one after a salt fog test. The seats of the valves that leaked were coined wider than normal due to pressure from the ball. It is assumed that reduction in unit pressure on the seat was responsible for the leakage.

From these failures it can be deduced that seat pressure is a critical factor with this component. Work is continuing to determine what this pressure should be and how to improve the general reliability of the valve.

### 6. Combustion Device Development, W. H. Tyler

Pyrolytic graphite (PG) is being investigated as a promising material for use as the ALPS system radiation-cooled thrust chamber material.

Successful demonstration tests with a 100-lb thrust injector and free-standing PG chamber assembly were reported in Refs. 3 and 4. The results of these tests indicated a problem of nonsymmetrical local erosion of the chamber, apparently due to the arrangement of the 10 pairs of orifices in the test injector assembly. This orifice arrangement places the orifices in rows across the circular face of the injector rather than in a circular symmetric pattern. To reduce the nonsymmetrical heat and mass distribution, a new injector assembly was designed and fabricated with 10 doublet elements and long orifices ( $L/D = 100$ ). Six elements were circularly spaced around four elements in two rows in the center area of the injector.

Short duration testing of this new water-cooled injector assembly with nitrogen tetroxide hydrazine propellant combination has yielded a characteristic velocity of 5500 ft/sec at a mixture ratio ( $w_o/w_f$ ) of 1.17 to 1.21 and a chamber pressure range of 145 to 151 psia. Preparations are being made to test this injector with an ablative thrust chamber. The resulting erosion pattern will be compared with the results of a similar test made with the original 100-lb injector before firings in a PG chamber are attempted.

Two short duration tests have been conducted with the ALPS Mod III pyrolytic graphite chamber assembly at the 2000-lb thrust level. These tests were run with a 21-element splash plate injector assembly using a propellant combination of nitrogen tetroxide and 50-50 fuel blend of unsymmetrical dimethylhydrazine and hydrazine. The duration of the tests was limited by the fact that the heavy-weight mild steel injector face and splash plate were uncooled. Test duration was sufficient, however, to obtain steady-state combustion.

Both tests were run at 148-psia chamber pressure, at a mixture ratio of 1.62, and produced  $c^*$  figures of approximately 5500 ft/sec. The duration of the first test was 2 1/2 sec and that of the second test was 4 1/2 sec. After this total firing time of 7 sec, there is no obvious erosion of the chamber wall nor damage of any kind to the chamber assembly.

Tests with longer duration firing times are anticipated and will be conducted as soon as an adequately cooled splash plate and injector are available.



## References

1. Porter, R. N., et al., "Advanced Liquid Propulsion Systems," SPS 37-21, Vol. IV, pp. 109-113, Jet Propulsion Laboratory, Pasadena, California, June 30, 1963.
2. Porter, R. N., et al., "Advanced Liquid Propulsion Systems," SPS 37-22, Vol. IV, pp. 126-134, Jet Propulsion Laboratory, Pasadena, California, August 31, 1963.
3. Rowley, R. W., "Investigation of Combustion-Ablation Phenomena," RS 36-13, pp. 128-133, Jet Propulsion Laboratory, Pasadena, California, March 1, 1962.
4. Rowley, R. W., *An Experimental Investigation of Uncooled Thrust Chamber Materials for Use in Storable Liquid Propellant Rocket Engines*, Technical Report No. 32-561, Jet Propulsion Laboratory, Pasadena, California, February 15, 1964.

## XI. Advanced Propulsion Engineering

### A. Liquid MHD Power Conversion

*D. Elliott, D. Cerini, D. O'Connor, and E. Weinberg*

The long lifetimes required of electric-propulsion powerplants make nonrotating cycles attractive. A nonrotating powerplant under investigation at JPL is the liquid magnetohydrodynamic (MHD) system shown schematically in Fig. 1. In this cycle a fluid, such as cesium, circulates in the vapor loop and causes a liquid metal, such as lithium, to circulate through an MHD generator in the liquid loop. The cesium leaves the radiator as condensate, flows through an electromagnetic (EM) pump to the mixer, vaporizes on contact with the lithium, atomizes and accelerates the lithium in the nozzle, separates from the lithium in the separator, and returns to the radiator. The lithium leaves the separator at high velocity (typically 500 ft/sec), decelerates through the production of electric power in the MHD generator, and leaves the generator with sufficient velocity (typically 300 ft/sec) to return through a diffuser to the reactor where the lithium is reheated.

Results of nozzle, separator, and diffuser tests with nitrogen and water, and of generator tests with NaK, were reported in Refs. 1-5. The initial phase of feasibility evaluation has now been completed with the achievement of closed-loop operation and constant-pressure power generation.

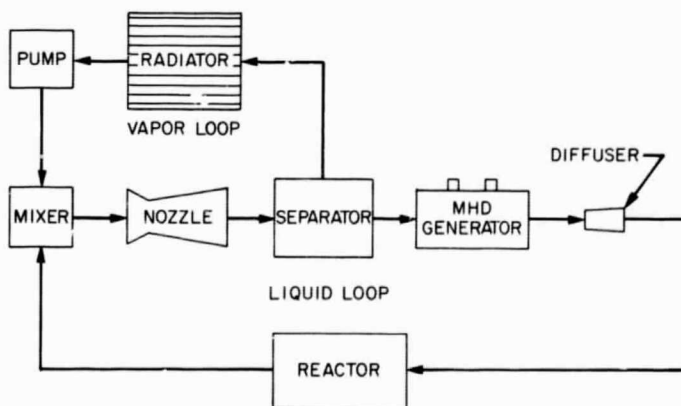


Fig. 1. Liquid MHD power conversion cycle

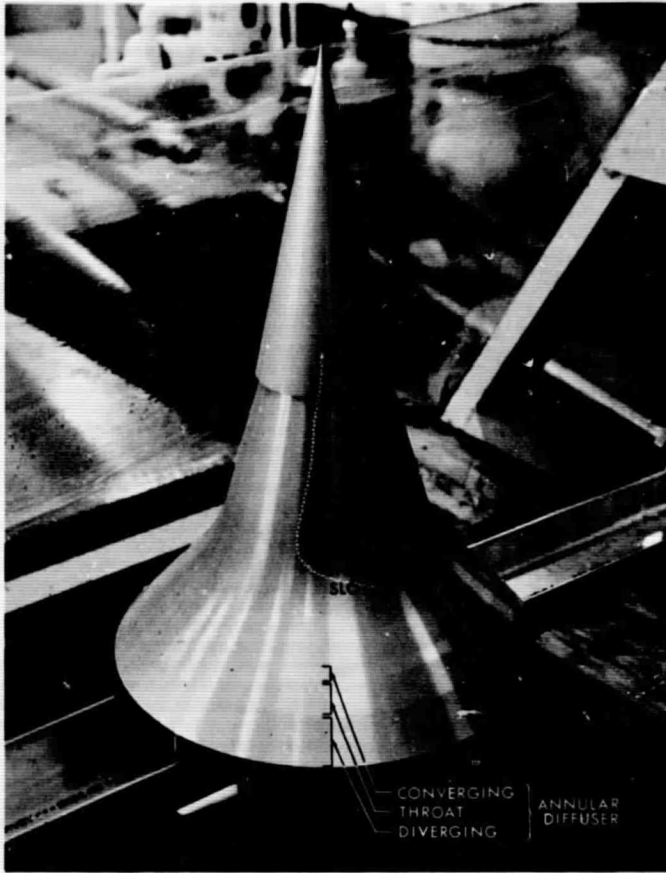


Fig. 2. Modified separator cone

Closed-loop operation was obtained with the addition of a secondary flow return slot and annular diffuser to the straight conical separator described in Ref. 4 (Fig. 2). Fig. 2 shows the separator cone with these two modifications. The secondary injection slot reinjects the flow from the secondary capture slot, and the converging-diverging annular diffuser provides the pressure for returning the primary liquid flow to the injector of the two-phase nozzle.

### 1. Secondary Injection Theory

The reinjected secondary flow enters at less than 20 ft/sec. This reduces the velocity of the main flow, but by less than the percentage of injected flow, since skin friction is also reduced. The actual reduction to be expected can be calculated by applying the momentum equation to the situation illustrated in Fig. 3. A two-phase jet of velocity  $V_n$  and liquid flow rate  $\dot{m}_n$  impinges at angle  $\theta$  on a truncated cone of projected area  $A_n$ . Liquid of flow rate  $\dot{m}_1$  and velocity  $V_1$  is already flowing along the cone. The velocity  $V_2$  of the combined flow  $\dot{m}_1 + \dot{m}_n$  leaving the cone is to be determined.

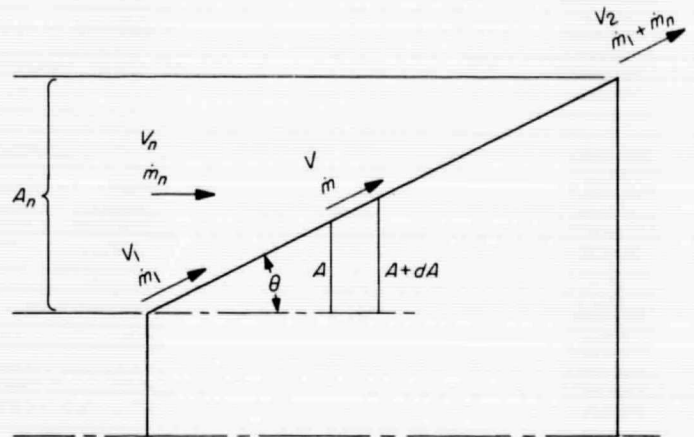


Fig. 3. Nomenclature for conical separator analysis

At a particular projected area  $A$  along the cone the velocity is  $V$  and the flow rate is  $\dot{m}$ . The surface area for a further increment  $dA$  of projected area is  $dA/\sin \theta$ . Hence the retarding force on the liquid due to skin friction in this increment is

$$dF = \frac{C_f \rho V^2}{2 \sin \theta} dA \quad (1)$$

where  $C_f$  is the skin friction coefficient and  $\rho$  is the liquid density.

The momentum of the flow entering increment  $dA$  is

$$M_{in} = \dot{m}V + \frac{\dot{m}_n V_n \cos \theta}{A_n} dA \quad (2)$$

The momentum of the flow leaving increment  $dA$  is

$$M_{out} = \left( \dot{m} + \frac{\dot{m}_n}{A_n} dA \right) (V + dV) \quad (3)$$

With no pressure variation along the cone,

$$M_{in} - M_{out} = dF \quad (4)$$

Substituting Eqs. (1), (2), and (3) into Eq. (4), the differential equation for the velocity variation along the cone is

$$\frac{d\left(\frac{V}{V_n}\right)}{\sin \theta \left(\frac{V}{V_n}\right)^2 + \frac{V}{V_n} - \cos \theta} = - \frac{d\left(\frac{\dot{m}}{\dot{m}_n}\right)}{\frac{\dot{m}}{\dot{m}_n}} \quad (5)$$

where

$$Q = \frac{C_f \rho V_n A_n}{2 \dot{m}_n} \quad (6)$$

Integration of Eq. (5) yields the following expression for the separator exit velocity  $V_2$ :

$$V_2 = \frac{\lambda(1-\alpha)V_0 + (V_0 + \alpha\lambda)V_1}{\alpha V_0 + \lambda + (1-\alpha)V_1} \quad (7)$$

where

$$\alpha = \left( \frac{\dot{m}_1}{\dot{m}_1 + \dot{m}_n} \right)^{(1+4Q/\tan\theta)^{1/2}} \quad (8)$$

$$\lambda = \frac{V_n \sin\theta}{2Q} \left[ \left( 1 + \frac{4Q}{\tan\theta} \right)^{1/2} + 1 \right] \quad (9)$$

$$V_0 = \frac{V_n \sin\theta}{2Q} \left[ \left( 1 + \frac{4Q}{\tan\theta} \right)^{1/2} - 1 \right] \quad (10)$$

It can be seen from Eq. (7) that the separator exit velocity with no initial flow ( $\alpha = 0$ ) is simply  $V_0$ .

## 2. Secondary Injection Tests

The separator velocity reduction with secondary injection was determined experimentally from the small changes in thrust of the liquid leaving the capture slot as secondary flow was injected. Fig. 4 compares these reductions with the values calculated from Eq. (7) for nozzle liquid/gas mixture ratios of 20, 40, and 62. Both the theoretical and experimental results, which agree within the accuracy of the data, show that the velocity reduction is 50–70% of the secondary injection fraction.

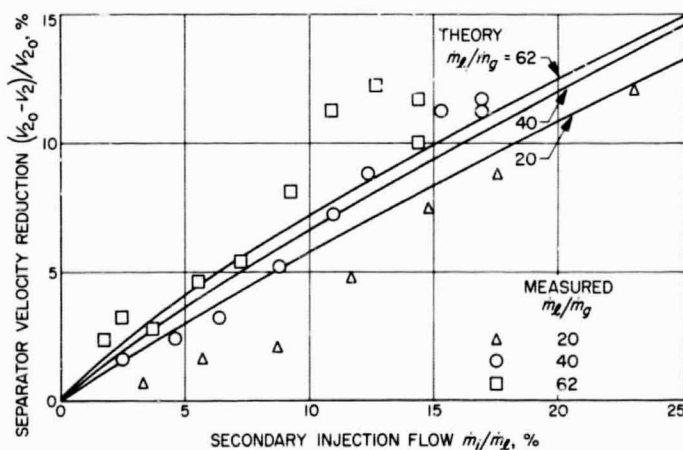


Fig. 4. Comparison of experimental and theoretical velocity reduction due to secondary injection

Injection of an equal volume of nitrogen with the secondary flow had no significant effect on the separator velocity.

## 3. Annular Diffuser Performance

The annular diffuser was added by machining a converging section, throat section, and diverging section into the existing capture slot (Ref. 4, Fig. 2). One wall of the diffuser is visible in Fig. 2. With the capture slot inlet set at a 0.1-in. gap the diffuser had a contraction area ratio of 0.75 and a throat length/gap ratio of 15, values found in previous tests of rectangular diffusers (Ref. 4) to give the best efficiencies at the desired inlet gas/liquid volume ratios.

Fig. 5 presents the measured efficiency of the annular diffuser (ratio of maximum discharge pressure to inlet isentropic stagnation pressure) as a function of inlet gas/liquid volume ratio. The data was obtained at inlet velocities between 220 and 300 ft/sec, liquid flow rates between 100 and 180 lb/sec, and inlet slot widths of 0.1 (contraction ratio = 0.75) and 0.121 in. (contraction ratio = 0.82). As shown in Fig. 5, the efficiency depended mainly on inlet volume ratio, varying from 75% at 0.9 to 57% at 3.4.

For comparison, Fig. 5 also shows the efficiencies of the best rectangular diffusers tested previously (Ref. 4) and the efficiencies reported in Ref. 6 for volume ratios near zero. It is seen that the best two-phase diffuser efficiencies obtained to date can be represented by a straight line from  $\eta = 0.86$  at a volume ratio of zero to  $\eta = 0.50$  at a volume ratio of 4.0. These efficiencies are satisfactory for early conversion system operation, but an increase of 1 or 2 percentage points in cycle efficiency

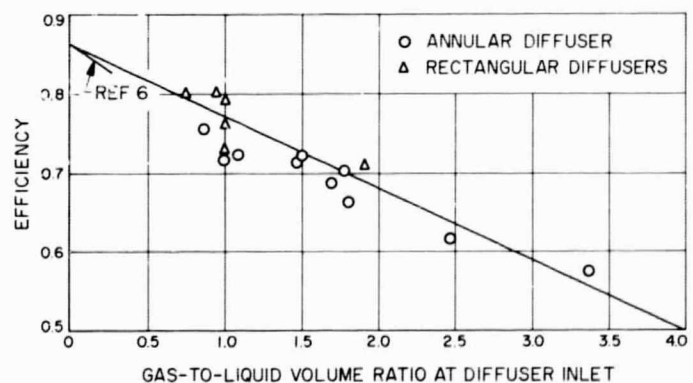


Fig. 5. Two-phase diffuser efficiencies

would be possible if diffuser efficiency could be maintained at 80–85% out to a volume ratio of 2.0. Diffuser studies aimed at such an improvement are continuing.

#### 4. Closed-Loop Operation

The discharge pressures with the annular diffuser exceeded the liquid injection pressure of the two-phase nozzle at all mixture ratios from 20 to 62. The margin was greatest with a 0.1-in. capture slot and a mixture ratio of 20; at this condition the diffuser discharge pressure was 288 psia and the nozzle injection pressure was 168 psia. With a 0.121-in. slot the discharge pressure was 250 psia, and when secondary flow return lines were added for returning the 2.3% secondary slot flow, the discharge pressure dropped to 241 psia. This provided a 73-psi margin for returning the main flow to the injector. (In a conversion system the generators upstream of the diffuser would reduce this margin to about 10 psi.)

The system was then connected for closed-loop operation, as shown in Figs. 6 and 7. Six 1.5-in. return lines were connected from the diffuser outlet to the injector inlet through check valves to prevent back-flow during starting. The main pump provided starting flow through three 1.5-in. lines with check valves to prevent back flow after starting. A throttling valve was provided for dumping the diffuser discharge flow prior to switching to closed-loop operation. Makeup water flow (about 1.0 lb/sec) to compensate for the water lost with the nitrogen exhaust was provided from an auxiliary pump through a pressure regulator.

To start the system, the makeup pressure regulator was first set to the desired starting pressure (about 85 psia), and the auxiliary pump was started, providing about 5 lb/sec of water flow through the open regulator. The nitrogen was then turned on full flow (typically 4.3 lb/sec corresponding to an operating mixture ratio of 30), and the starting water flow rate was set at 60 lb/sec, the lowest value permitting closed-loop operation. The throttle valve was then closed gradually until the diffuser discharge pressure exceeded the injector inlet pressure, causing the return line check valves to open. Further closing of the throttle valve, accompanied by reduction of the main pump flow, caused an increasing proportion of the injector flow to be supplied from the return lines. Finally, the throttle valve was fully closed, the main pump was turned off, the start line check valves closed, and closed-loop operation was established at partial flow. To reach the desired operating point (typically a water

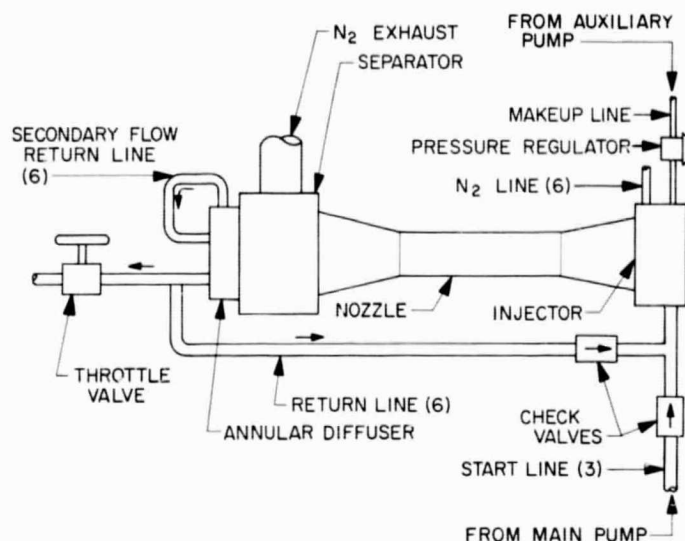


Fig. 6. Piping arrangement for closed-loop tests

flow rate of 130 lb/sec and mixture ratio of 30) the pressure regulator was reset to 150 psia. This caused the makeup flow to exceed the liquid loss until enough water inventory had been added to establish the new operating point. The water inventory at the operating point was about 120 lb.

Operation under closed-loop conditions was found to be smooth and stable, and any operating point within the closed-loop range could readily be reached by varying the nitrogen flow rate and the makeup feed pressure.

The system is now being modified for closed-loop tests with NaK.

#### 5. Constant-Pressure Generator Test

The cycle of Fig. 1 requires approximately equal inlet and outlet pressures from the MHD generator. Power is derived from deceleration of the liquid. To confirm the feasibility of this mode of operation, a NaK generator was tested which was identical to the generator described in Ref. 5 except that the flow channel was divergent, with an area ratio of 2.0, instead of straight. In preliminary tests an output power of 3.0 kw (10,000 amp at 0.3 v) was obtained at zero pressure differential with 207 ft/sec NaK inlet velocity. The efficiency (ratio of electric output power to fluid kinetic power change) was 46%, 4 percentage points below the theoretical value.

A failure of the insulation coating occurred when a higher power level was attempted. The generator is now



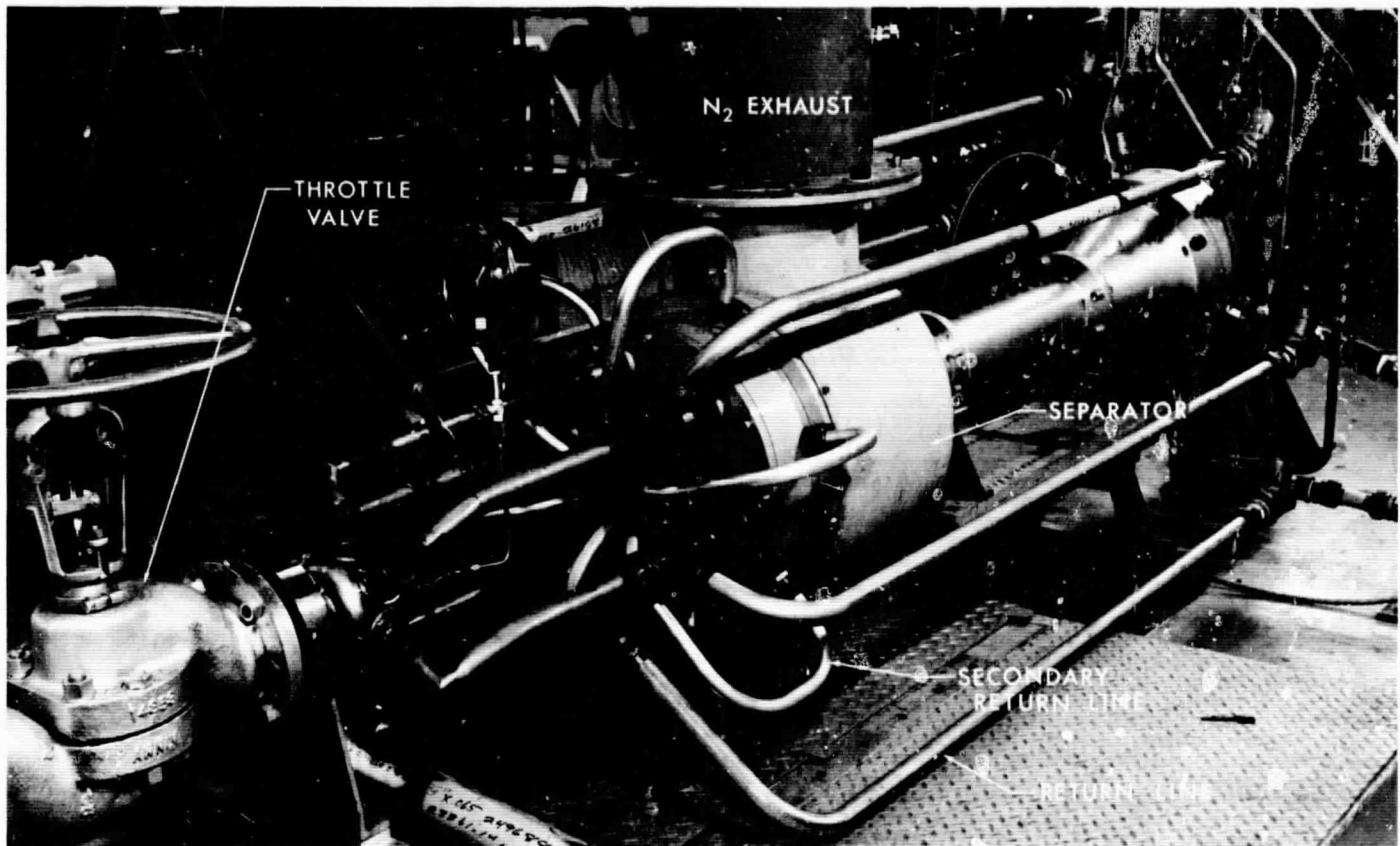


Fig. 7. Closed-loop system

being modified to minimize heating of the insulation and to eliminate a pressure peak near the inlet which may have reduced the efficiency.

## 6. Electrical Conductivity of Two-Phase Mixtures

Because the separator does not remove all of the vapor, the MHD generators in a conversion system will employ a two-phase mixture. This is desirable as well as necessary because the conductivity of the two-phase mixture can be varied through void fraction changes to reduce shunt end losses.

The conductivity of two-phase NaK-nitrogen mixtures has been measured at the velocities and void fractions of interest, using the apparatus shown in Fig. 8. NaK was injected into a 1.58- $\times$ 0.56-in. duct through a bundle of 150 hypodermic tubes at a velocity of 260 ft/sec. Nitrogen was injected between each tube to provide a two-phase mixture of 0.69 void fraction. The electrical conductivity of the mixture was determined by passing a 40-amp current in the 0.56-in. direction between 1.0-in.-D copper electrodes flush with the insulated duct

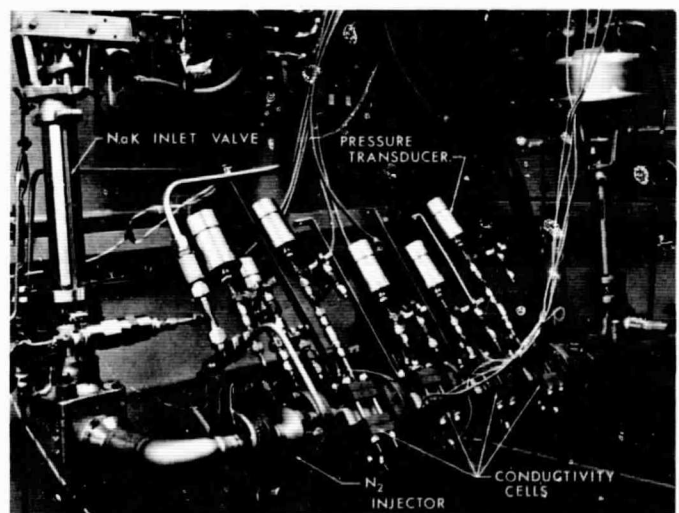


Fig. 8. Two-phase conductivity apparatus

walls. The voltage difference between the electrode faces, corrected for induced emf's, gave the mixture conductivity. After leaving the first conductivity cell, the mixture entered three narrower cells wherein the mixture was successively compressed to void fractions of 0.61, 0.50, and 0.29.

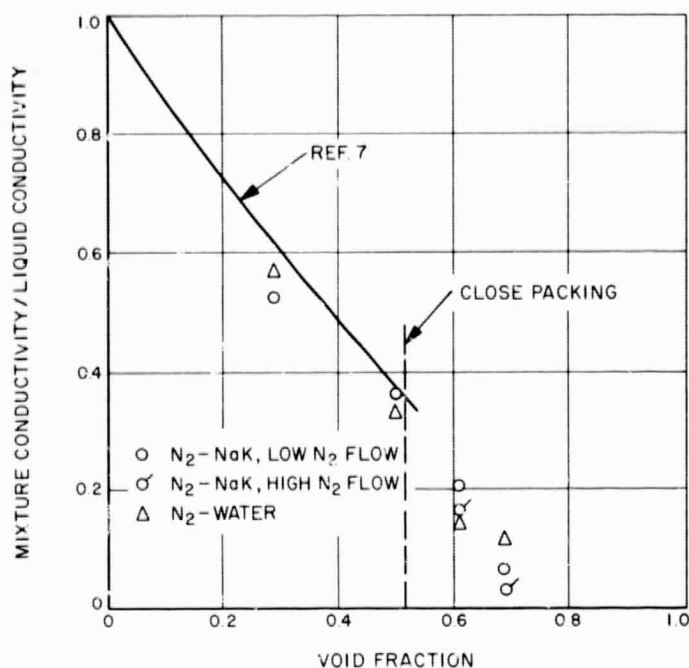


Fig. 9. Conductivity of two-phase mixtures

The measured ratio of mixture conductivity to liquid conductivity is shown as a function of void fraction in Fig. 9 and compared with the theoretical values of Ref. 7 derived for a cubical array of spherical gas bubbles up to closest packing. Also shown are nitrogen-water conductivity values obtained with the same apparatus using 1000-cps ac voltage. With both water and NaK the conductivity appears to follow the theoretical curve to closest packing, beyond which the conductivity drops to less than 10% of the liquid conductivity at the highest void fraction employed.

An additional effect was observed at the two higher void fractions where increasing the nitrogen flow rate (raising the conductivity-cell pressures but not the void fraction) lowered the conductivity. Values at two different nitrogen flows are indicated in Fig. 9.

### 7. Solubility of Cesium in Lithium

The solubility of cesium in lithium up to 2100°F has been determined under JPL contract by MSA Research Corporation (Ref. 8). The method of measurement was the same as employed for the lower temperature measurements reported in Ref. 9 except that Cb-1% Zr capsules were used. Fig. 10 shows the combined result of both studies. It is seen that the results are favorable for the operation of a Cs-Li conversion system in that the solubility is only about 7.5% at 2000°F.

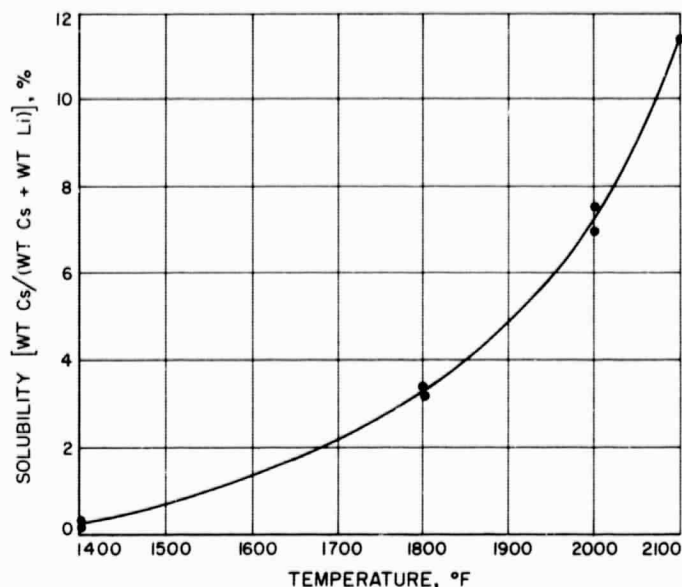


Fig. 10. Solubility of cesium in lithium (MSA)

MSA also investigated the solubility of potassium in lithium. Their measurements indicated that potassium is about 10% soluble in lithium at 1500°F and about 15% soluble at 1800°F, values too large to make this combination of interest.

### 8. Estimated Cycle Efficiency and Weight

The new Cs-Li solubility values permit a refined estimate of the efficiency of a 300-kw Cs-Li powerplant with 2000°F nozzle inlet temperature. The estimate, shown in Fig. 11, is based on the same assumptions discussed in Ref. 2, p. 115. Two of these assumptions are 85% diffuser efficiency and 70% generator efficiency, representing improvements over the values obtained to date. Fig. 11 shows a peak efficiency of 5.7% at a radiator temperature of 1250°F.

The variation of isothermal radiator area with radiator exit temperature, for an emissivity of 0.9, is shown in Fig. 12. The isothermal area is the area obtained when the radiating temperature is everywhere equal to the local fluid temperature; a minimum-weight finned radiator is 25-50% larger. The areas in Fig. 12 include the effect of lithium condensation and cesium vapor cooling in parts of the radiator hotter than the exit. It is seen that the radiator area reaches a minimum of 3.0 ft<sup>2</sup>/kw(e) at about 1400°F.

Weight estimates were made for a powerplant delivering 300 kw at 5000 v dc using a dc-to-dc converter with 20-v input from the MHD generator. The latter voltage

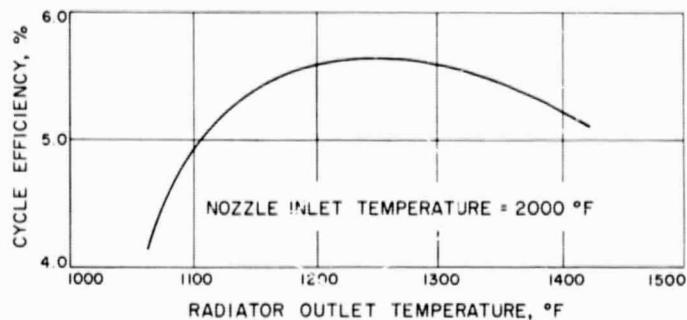


Fig. 11. Estimated efficiency of cesium-lithium cycle

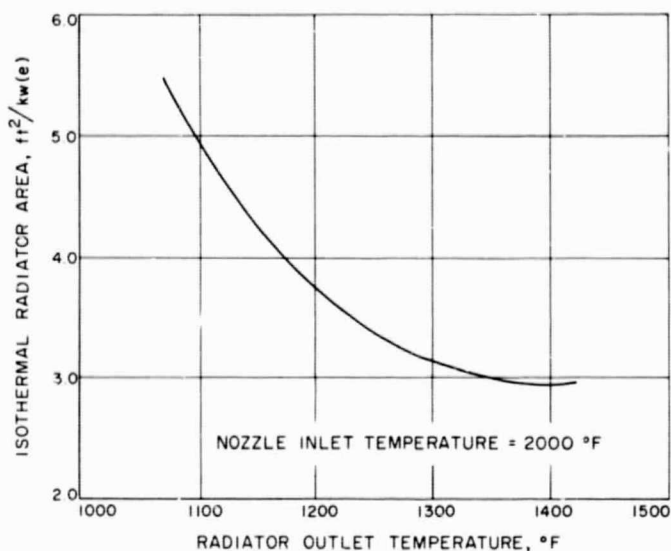


Fig. 12. Estimated radiator area of cesium-lithium cycle

is about the highest that a series arrangement of dc MHD generators can be expected to achieve with 70% efficiency. Induction ac generators are promising with liquid metals but are not considered in the present estimate.

The studies reported in Ref. 10 indicated that a 20- to 5000-v converter would have an efficiency of 86.5% and a specific weight of 8.3 lb/kw. Hence the input power required for this application is 347 kw and the converter weight is 2500 lb. This and the other weights are summarized in Table 1.

A weight analysis of the conversion system (nozzle, separator, diffusers, cesium EM pump, piping, and generators, including magnets and bus bars) yielded a weight range of 2000-3500 lb, using Cb-1% Zr with 1000-psi maximum stress, uncooled copper bus bars and magnet windings, and iron magnet poles. This weight cannot be better defined until further generator work

Table 1. Estimated weight of a 300-kw(e) liquid MHD powerplant (unshielded)

Component	Weight, lb	Specific weight, lb/kw(e)
Dc-to-dc converter	2500	8.3
Conversion system	2000/3500	6.7/11.7
Reactor	1500/3500	5.0/11.7
Radiator	2300/3900	7.7/13.0
Total	8000/13,400	28/45

has been done, and both values are given in Table 1 representing upper and lower estimates.

With 5.7% cycle efficiency, the required reactor power is 6100 kw. It is estimated that a lithium-cooled reactor of this size specifically designed for the high-burnup requirements of electric propulsion, such as the one described in Refs. 11 and 12, would weigh 1500 lb. It is estimated that a more conventional design would weigh about 3500 lb.

The isothermal radiator area from Fig. 12 is 3.3 ft²/kw(e), or 1150 ft² in the present case. Two estimates of radiator weight were used, a value based on the advanced spacecraft design described in Ref. 13 and a higher value based on the more conservative design of Ref. 14. Ref. 13 employed a minimum weight radiator for which the actual area was 1.5 times the isothermal value; the corresponding area in the present case is 1720 ft². The radiator weight in Ref. 13 was 1.35 lb/ft², giving a weight of 2300 lb here.

Ref. 14 employed a radiator of 1.26 times the isothermal area, corresponding to 1450 ft² for the MHD system. The Ref. 14 weight was 2.7 lb/ft² (after subtracting NaK inventory, since direct condensing is assumed here), yielding 3900 lb in the present case.

Adding these weights, the total unshielded powerplant weight would lie between 8300 lb [28 lb/kw(e)] and 13,400 lb [45 lb/kw(e)]. The shield weight depends strongly on the spacecraft configuration and would be 2000-4000 lb for integrated neutron and gamma doses at the payload of  $10^{13}$  n/cm² and  $10^9$  ergs/g (c), respectively.

## 9. Insulator Compatibility Tests

After a 500-hr exposure to lithium at 2000°F, four Cb-1% Zr capsules containing  $Y_2O_3$ ,  $Sm_2O_3$ , CaO, and



ThO<sub>2</sub>, respectively, were opened. The Y<sub>2</sub>O<sub>3</sub> sample showed a minor amount of dissolution. Microscopic examination of a sample cross-section indicated no change in its crystalline structure when compared to an unexposed sample, but showed a small number of voids at its outer edges.

A cross-section of Sm<sub>2</sub>O<sub>3</sub> revealed outer and inner layers of material. Microscopic examination showed large voids and changes in the crystalline structure of the outer layer. The inner layer, however, seemed unchanged when compared to a cross-section of unexposed material. The inside of this test capsule was blackened. From these results, it was concluded that mass transfer had occurred, although this has yet to be confirmed by chemical analysis.

Results of a microscopic examination of the ThO<sub>2</sub> sample are not available as yet. Visual examination, however, indicated no dissolution since all the corners and edges of the sample remained sharp. The inside of the test capsule appeared clean, indicating that no mass transfer had occurred, although this also remains to be confirmed by chemical analysis.

Of all the samples tested, only CaO disintegrated. It is impossible to determine, at this point, whether this was caused by the lithium or by water vapor which acci-

dentally entered the welding chamber and came into contact with all the samples for an extended period prior to the closure welding of the test capsules. All the samples were darkened considerably during exposure to lithium. These tests will be repeated with heated and evacuated samples to determine whether this discoloration is caused by the reaction of the lithium with the air and water on the surface of the sample or by lithium attack.

### 10. Separator Erosion Test

In order to simulate the erosion caused by a stream of high velocity lithium droplets at 2000°F impinging on a Cb-1% Zr separator, a small separator cone, made of an aluminum alloy having mechanical properties at room temperature that are similar to those of Cb-1% Zr at 2000°F (e.g., yield strength, ultimate strength, and hardness), was placed in a jet of 450-ft/sec water droplets. The cone initially had a 0.375-in.-D flat tip. This shape was used since previous tests had shown that a cone tip having a sharp point was flattened rapidly by erosion until a "critical diameter" was reached beyond which erosion proceeded at a much slower rate. After a run of 430 hr, the mass loss with the blunted tip was 0.085 g. If this rate were to continue for 10,000 hr, the total loss would be about 2 g, an acceptable amount.

## References

1. Elliott, D., and Cerini, D., "Liquid MHD Power Conversion," SPS 37-17, Vol. IV, pp. 199-203, Jet Propulsion Laboratory, Pasadena, California, October 30, 1962.
2. Elliott, D., Cerini, D., McNary, C., Stephenson, R., and Weinberg, E., "Liquid MHD Power Conversion," SPS 37-21, Vol. IV, pp. 114-121, Jet Propulsion Laboratory, Pasadena, California, June 30, 1963.
3. Elliott, D., Cerini, D., Otte, H., and Weinberg, E., "Liquid MHD Power Conversion," SPS 37-23, Vol. IV, pp. 132-135, Jet Propulsion Laboratory, Pasadena, California, October 31, 1963.
4. Elliott, D., Cerini, D., Hays, L., and Weinberg, E., "Liquid MHD Power Conversion," SPS 37-25, Vol. IV, pp. 106-111, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.
5. Elliott, D., Cerini, D., and O'Connor, D., "Liquid MHD Power Conversion," SPS 37-26, Vol. IV, pp. 124-131, Jet Propulsion Laboratory, Pasadena, California, April 30, 1964.

## References (Cont'd)

6. Schneiter, G. R., *Diffuser Studies with Single and Two-Phase Flows*, Report No. TM-62-1, Jet Propulsion Center, School of Mechanical Engineering, Purdue University, Lafayette, Indiana, April 1962.
7. Meredith, R. E., and Tobias, C. W., "Resistance to Potential Flow Through a Cubical Array of Spheres," *Journal of Applied Physics*, Vol. 31, No. 7, July 1960, pp. 1270-1273.
8. Tepper, F., Udavcak, R., and Zelenak, J., *Determination of the Solubility of Potassium and Cesium in Lithium*, Report on JPL Purchase Order No. BE-32-1544, MSA Research Corporation, Callery, Pennsylvania, April 16, 1964.
9. Cuerou, T. R., and Tepper, F., *Solubility of Cesium in Lithium at Elevated Temperatures*, ARS Space Power Systems Conference, Santa Monica, California, ARS Preprint 2537-62, September 25-28, 1962.
10. Baker, A. R., et al., *Space Electric Power Systems Study, D.C. to D.C. Converters for Nuclear-Thermionic Energy Sources*, Report No. WAED 63.11E, Westinghouse Electric Corporation, Lima, Ohio, December 3, 1963.
11. Davis, J. P., *A Nuclear Reactor Concept for Electric Propulsion Application*, Technical Report No. 32-385, Jet Propulsion Laboratory, Pasadena, California, March 4, 1963.
12. Gronroos, H. G., *Criticality Calculations for a Fast Liquid-Metal-Cooled Reactor-Phase I*, Technical Report No. 32-512, Jet Propulsion Laboratory, Pasadena, California, November 15, 1963.
13. Beale, R. J., Speiser, E. W., and Womack, J. R., "Electric Space Cruiser for High-Energy Missions," *Journal of Spacecraft and Rockets*, Vol. 1, No. 1, January-February 1964.
14. Larson, J. W., *Research on Spacecraft and Powerplant Integration Problems*, Second Quarterly Report, 26 July to 26 October, 1963, Document No. 63SD886, General Electric Company, Philadelphia, Pennsylvania.

**BLANK PAGE**

## SPACE SCIENCES DIVISION

### XII. Space Instruments Development

#### A. A Gas Chromatograph for the Analysis of the Martian Atmosphere

C. Josias, L. Bowman, and H. Mertz

In anticipation of future missions providing capsule entries into the atmosphere of Mars, a two-phase program for the development of atmospheric analysis instrumentation using gas chromatography has been undertaken. The initial phase is the development and construction of a scientific feasibility model applicable to a Mars mission. This model is currently under construction, and after its completion and evaluation, its performance will be reported in one of the forthcoming issues of the SPS. The second phase requires the formulation and construction of an engineering model of a flight-type instrument.

This report describes electronics for the signal processing and sequencing sections of an instrument proposed for use in a Mars' atmosphere capsule mission. The instrument proposed<sup>1</sup> was described as a device to be used in a *Mariner* 1966 mission, should this mission utilize a capsule. The scientific feasibility model presently under development will justify and prove premises stated in the submitted report and will be followed by development of an engineering model of an instrument whose electronics will be patterned after those described here.

The basis upon which the referenced instrument was proposed was a set of informal and highly speculative

<sup>1</sup>Lipsky, S. R. (Yale University Medical School), Bentley, K., Bowman, L., Josias, C., Mertz, H., Wilhite, F., "Mariner B 1966 Mission: Analysis of the Atmosphere of Mars by Gas Chromatography Instrumentation," proposal submitted to Space Sciences Division by S. R. Lipsky, November 15, 1963.

guesses on mission constraints. Probably the most pertinent, and at the same time most restrictive, assumptions were that (1) analysis time for the case of fastest descent would be less than 60 sec, and (2) total capsule data returned to Earth for the same descent condition would be 150 bits, 50 of which were to be allocated to this experiment. The following aspects of a Mars atmospheric chromatograph are presented in this report with the associated requirements imposed by the two basic assumptions:

- (1) Instrument design parameters
- (2) Signal processing electronics
- (3) Programming
- (4) Data considerations

Since the low data volume problems encountered by the chromatograph would similarly be encountered by many spectrographic instruments with the same data restrictions, special attention is called to the section of this report on data, where a method of wide-range data compression for spectrographic outputs is proposed.

## 1. Electronics for Programming, Signal Processing and Data Handling

In remote applications of gas chromatography such as the one described here, in which severe data restrictions are imposed, one cannot rely upon customary visual analytical techniques for identification of constituents and measurement of their values. If data capacity is insufficient to reconstruct the chromatogram, the output peaks must be described by such minimal information as retention time and peak height or peak area. The following sections will discuss some factors and parameters which affect the design of the electronic portions of a chromatographic instrument and will also indicate the capabilities of available elements.

### a. Design parameters for signal processing

*Dynamic range.* This section is concerned with range requirements for those electronics performing quantitative measurements.

- (1) Detector range. The cross-section detector has approximately a five-decade current range such that

$$10^{-3.5} < \frac{I_s}{I_{BL}} < 10^{+1.5}$$

where  $I_s$  = signal current

where  $I_{BL}$  = baseline current

and upper and lower limits are determined by 100% concentration and noise conditions, respectively. The minimum electrometer range requirement for linear reproduction of all detector peaks is therefore five decades.

- (2) Effects of sample pressure. If the chromatograph maintains its strictest resolution requirements over a specified range of sample pressures, the chromatogram will then be most accurately described by integral measurements. Consider, then, the quantitative dynamic requirements produced by a pressure range of 30:1 (a 10:1 spread in estimates of surface pressure extended by a factor of three to account for initial samples being taken at scale height). If this were the case and the instrument were designed to accommodate the lowest expected sampling pressure (i.e., the minimum surface value divided by 3), the usable integrating range should be larger than the linear detector range by a factor roughly equal to the ratio of acceptable pressures. An electronic integrator would consequently require a minimum dynamic range of 6.5 decades.

*Noise and standing current.* The noise content of the detector or detectors supplying current to an electrometer determines the threshold sensitivity required, if it is desired to measure signals just discernible in the noise. If sensitivity were equated with noise, the electrometer instability as described by its summing-point voltage drifts, input leakage currents, amplifier noise, and drifts in baseline compensation should be no greater than  $10^{-13}$  amp, the noise estimate for a single representative cross-section detector at a bandwidth of 0.1 cps.

The standing current (baseline) and its variations are principal parameters upon which the design of corrective stabilization electronics is based. Consider the example of two dual cross-section detectors each having baselines of  $10^{-9}$  amp and thermally induced drifts of  $0.02\%/^{\circ}\text{C}$ . If temperature were the only consideration in baseline drift, the aggregate variation over  $100^{\circ}\text{C}$  for two identical detectors would be 4%, or  $4 \times 10^{-11}$  amp. Allocating half of the electrometer system instability to the baseline compensation, it is seen that a corrective factor of 800 must be applied during the analysis cycle.

*Form factor.* The form factors, or variations in geometry of peak waveforms, provide important requirements for the electrometer and the signal-processing electronics.

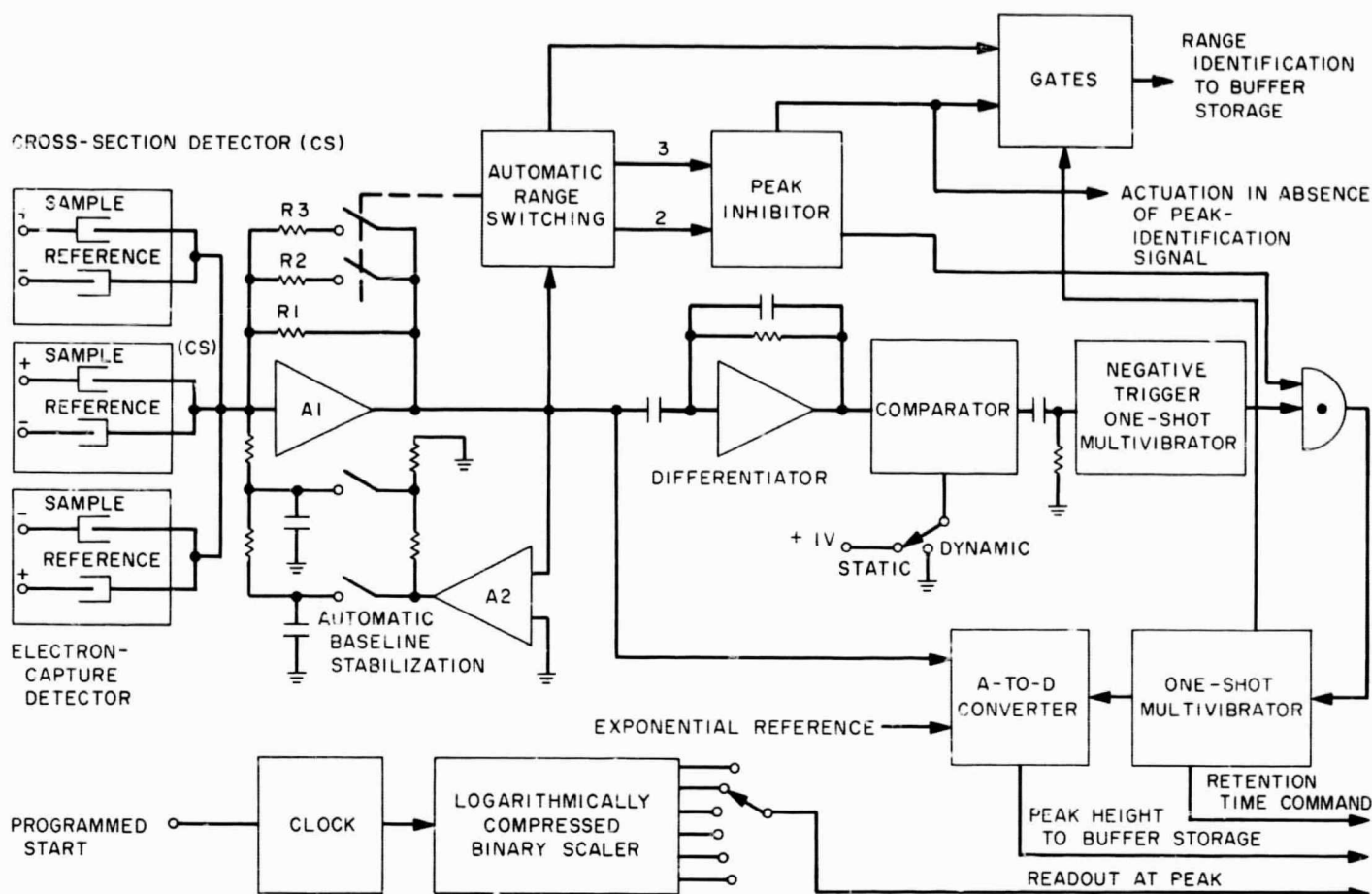


Fig. 1. Peak-height measurement system

The speed of response requirement, for instance, will be determined by the narrowest predicted peak. The widest peaks determine differentiating gain and eventually the threshold currents at which automatic retention time and peak-height readings are made. The maximum and minimum values of peak width at half-height that will be used for the prototype design will be 1 and 12 sec, respectively.

#### b. Description of signal processing electronics

**Peak height.** To measure peak height, as shown in Fig. 1, the chromatogram is first developed in the baseline-stabilized, automatic range-switching electrometer. The signal is then differentiated and transferred to a discriminator where zero crossings are extracted (Ref. 1). The discriminator uses a controlled hysteresis characteristic in such a way as to exclude differentiated noise. Firing is set at moderately regulated nonzero levels and reset occurs precisely at zero. The timing pulse thus obtained is used to simultaneously trigger the time-base measurement and initiate the analog-to-digital (A-to-D) conversion. Circuitry will be provided to exclude

false peaks produced by scale changes, and electrometer scale readout will occur even if the peak occurs immediately after scale change and the peak is still not read out. Elements shown in this block diagram that are usually associated with data systems may actually appear at either side of the instrument-data-system interface. Of prime importance though, is the desire to quantize and read out values to storage buffers on instrument command.

**Peak integration.** Fig. 2 illustrates generally how the electrometer amplifier would be used in an integrating application. Were there no scale changing, the mechanism as shown would be adequate with few embellishments. The diagram omits many of the automatic pulse-timing electronics shown in Fig. 1, which are also implicit in the integration method. Also required are logic elements to prevent readouts during scale-switching transients.

#### c. System elements

**Electrometer for linear reproduction of the chromatogram.** The electrometer system proposed for this instru-



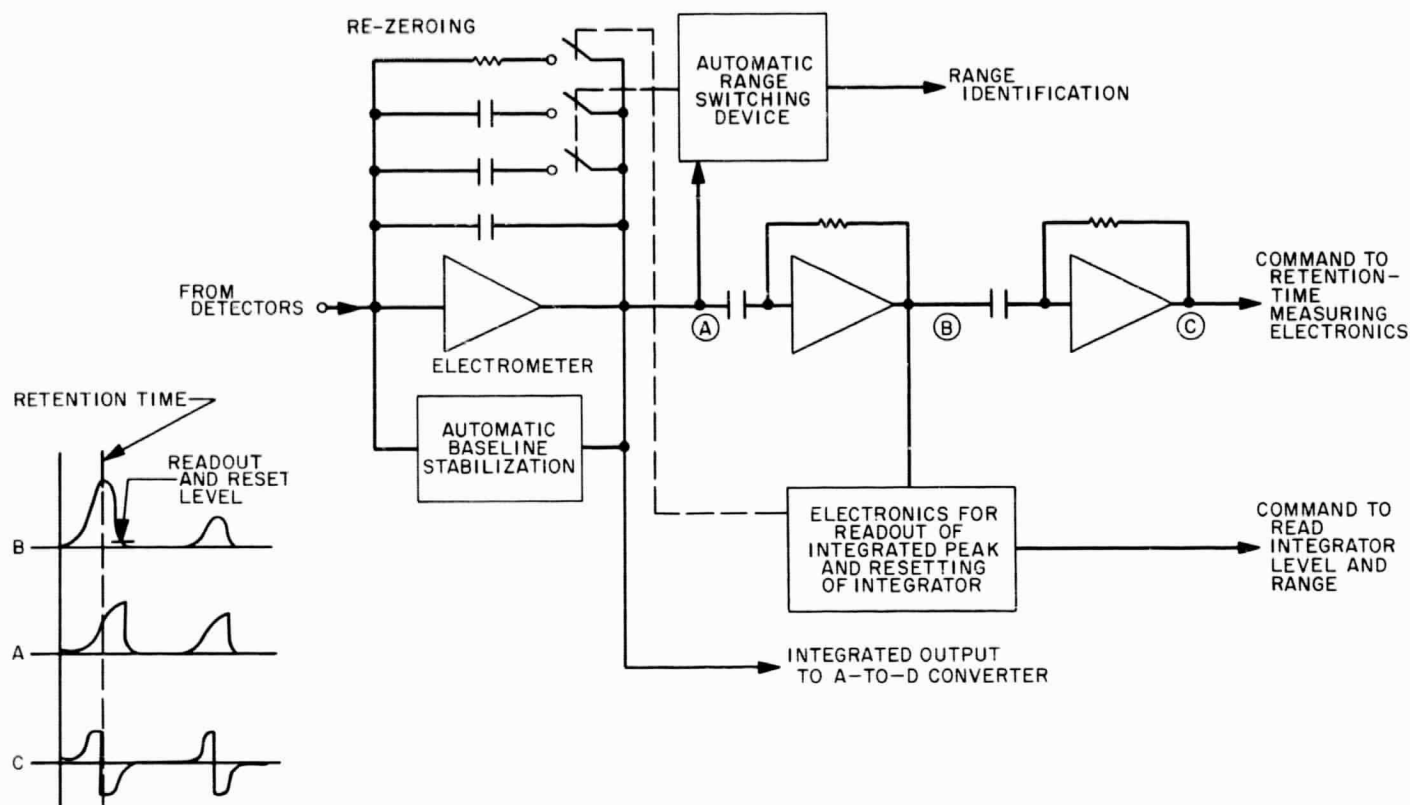


Fig. 2. Partial block diagram of peak-integrating electronics

ment is of the dynamic capacitor variety. It utilizes a carrier-type DC amplifier and is similar to units flown on *Rangers 1* and *2* and *Mariner 2* (Ref. 2). Some of its characteristics are as follows:

Input leakage current, amp	$10^{-15}$
Voltage stability, $\mu\text{V}/^\circ\text{C}$	70
Rise time (10 to 90%), msec	100
Rate limit, v/sec	100
Absolute DC accuracy (with perfect feedback elements), %	0.1

Automatic range switching of three resistors (Fig. 1) is accomplished by a scale factor device (Ref. 3) whose switching thresholds are as follows:

Resistor	Electrometer output voltage, v	Corresponding input current, amp	Dropout voltage, v	Dropout current, amp
$10^{12}$	$0.01^a$ to 10	$10^{-13a}$ to $10^{-10}$	$b$	$b$
$10^9$	0.1 to 10	$10^{-10}$ to $10^{-8}$	0.05	$5 \times 10^{-11}$
$10^7$	0.1 to 10	$10^{-8}$ to $10^{-6}$	0.05	$5 \times 10^{-9}$

<sup>a</sup> Electrometer threshold

<sup>b</sup> Not applicable

**Baseline stabilizer.** Two methods have been developed for compensation of baseline drift. The first technique maintains constant linear servo control of the electrometer against changes in detector baseline (Ref. 1). This is especially effective with short-term drifts such as might occur with limited temperature control during the transient conditions of atmospheric entry. The second technique performs linear feedback control up to the start of analysis, at which time the baseline compensation is supplied by a pre-analysis stored voltage (Refs. 4, 5). This kind of operation assumes negligible drift in detector baseline during the analysis period.

Both systems have the capacity for correction of drifts up to 10% of the nominal baseline current. The linear feedback method as presently constructed can compensate for maximum drifts of  $10^{-11}$  amp/min. Its undershoot error due to rate-limited integration of chromatographic peaks is also  $10^{-11}$  amp/min. The gated-capacitor storage system has a storage decay corresponding to  $2 \times 10^{-12}$  amp/hr. During linear pre-analysis control, it reduces baseline drift by a factor of 1000.

**Integrating electrometer.** The principal factor determining the feasibility of a Miller-type integrator in this

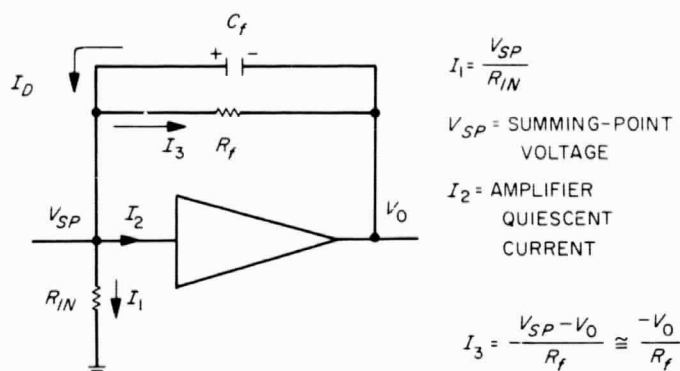


Fig. 3. Leakage diagram of Miller integrator

application is the ability of the circuit to retain accurately the integrated information. Consider the integrator in Fig. 3. The total discharge current is given by

$$I_D = I_1 + I_2 + I_3.$$

The discharge due to capacitor leakage resistance ( $I_3$ ) can be approximated for the worst case by integrating an equivalent rectangular chromatogram having a width of 20 sec. The decay in this idealized case may be expressed as follows:

$$\text{Decay} = \frac{t}{2\tau_f}$$

where  $\tau_f = C_f R_f$ .

A specification that decay not exceed 1% of the ideal integrated value at the end of 20 sec requires a minimum capacitor time constant ( $\tau_f$ ) of 1000 sec, a modest requirement for high-quality Teflon dielectric capacitors.

The following table illustrates the effects of a constant aggregate leakage current ( $I_1 + I_2$ ) of  $10^{-14}$  amp on a 20-sec integration:

$C_f, f$	$\Delta V_0, v$
$10^{-6}$	$2 \times 10^{-7}$
$10^{-8}$	$2 \times 10^{-5}$
$10^{-10}$	$2 \times 10^{-3}$

**Sequencer and programmer.** A sequencer will be needed to activate the chromatograph, initiate the calibration functions, and coordinate these functions, where necessary, with the capsule data conditioning system. A programmer will be used to initiate the repetitive functions required in the chromatographic analysis sequence.

The sequencer and programmer events will occur approximately as follows:

#### Sequencer

- (1) Sequencer reset command
- (2) Read electrometer output
- (3) Initiate programmer sequence
- (4) Inhibit sample valve
- (5) Activate peak simulator
- (6) Open gas vent
- (7) Activate carrier flow
- (8) Activate oven
- (9) Activate calibration sample after equilibrium is attained
- (10) Initiate programmer sequence
- (11) Inhibit sample valve
- (12) Initiate programmer sequence for first sample run
- (13) Repeat initiate programmer sequence

#### Programmer Sequence

- (1) Read out column-detector temperature
- (2) Zero electrometer
- (3) Open baseline compensation loop
- (4) Activate sample valve

A command from the capsule data system will be required to start the sequencer when capsule power is turned on (Fig. 4). Another command will tell the sequencer when to begin the sampling sequence. The clock that operates the logarithmically compressed binary scaler will also be used as the time base for the sequencer and programmer.

#### d. Data

**Format.** The allocation of data within this experiment is shown in a first-guess form in Table 1. It has been arbitrarily assumed that the compositional experiment will be assigned 50 of the 150 data bits available after acquisition of phase lock in the minimum time-to-impact case. The distribution of data bits includes temperature, identification, parity, electrometer scale, peak amplitude (height or area) and retention time.



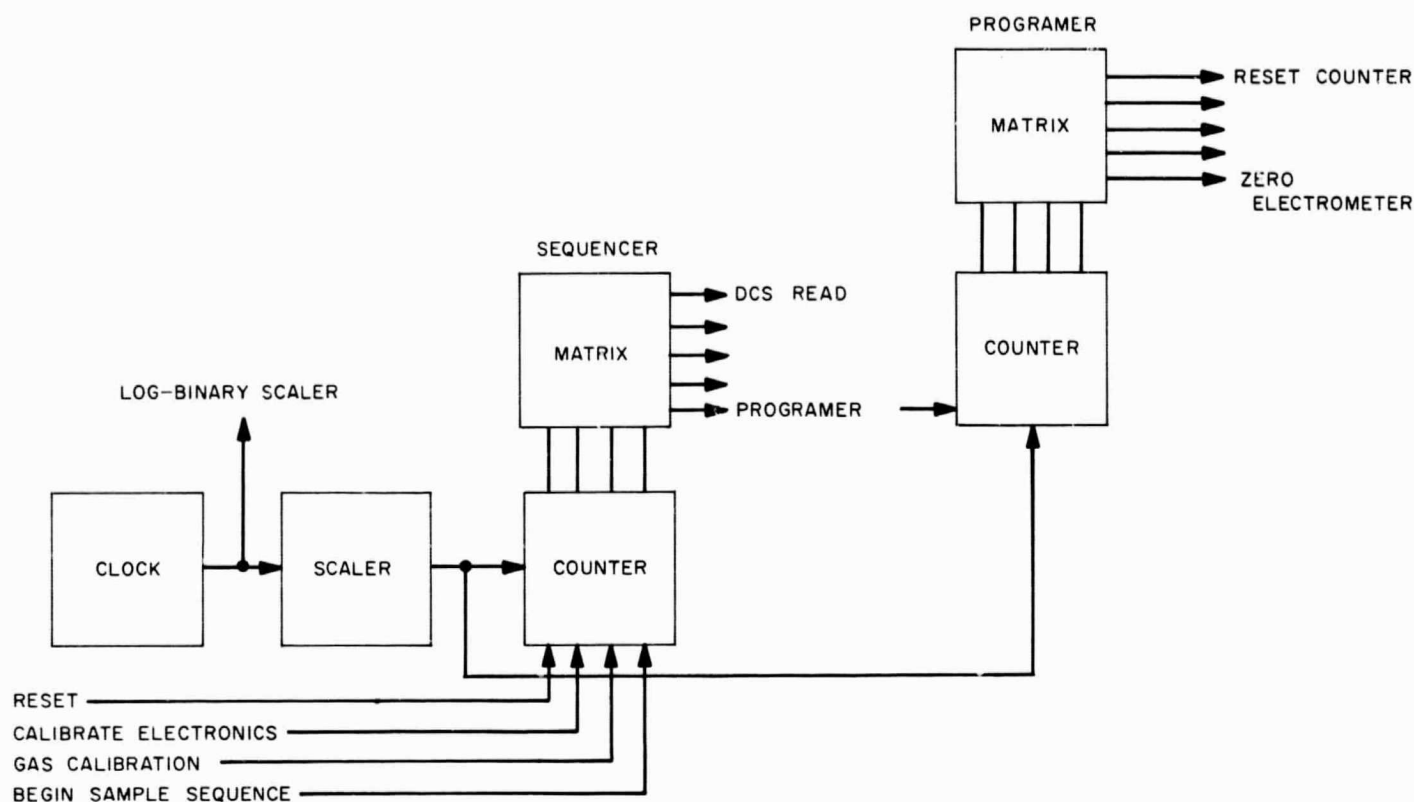


Fig. 4. Gas chromatograph sequencer and programmer

Table 1. Possible allocation of bits for first analysis (worst case)

No. of peaks	Column-detector temperature, bits	Identification and parity, bits	Electrometer range, bits	Amplitude, bits	Time, bits	Total
4	5	5	2/pk (8)	4/pk (16)	4/pk (16)	50
3	5	5	2/pk (6)	6/pk (18)	5/pk (15)	49

**Data quantization accuracy.** Since it is imperative that maximum efficiency be made of data capacity, nonlinear quantization is proposed for both the A-to-D converter and the time-base system. An exponential program similar to the one recommended in Table 2 was indeed used in the electrostatic deflection system in the *Mariner 2* solar plasma instrument (Ref. 2). Accuracies of both nonlinear and linear quantization systems are shown in Tables 2 and 3, respectively, and the conventional linear technique does not compare favorably.

Table 4 shows accuracy figures for an exponentially quantized time base. The time base could be provided by a D-to-A converter or by a precision ramp generator. The time-base voltage is then digitized into the required number of bits (4 or 5 as suggested in Table 1) when

Table 2. Three-decade (0.01 to 10 v) exponential analog quantization accuracy

Bits	Quantizing levels	Quantizing levels/decade	3-decade quantizing uncertainty, %
4	15	5	$\pm 23$
5	31	10	$\pm 11.5$
6	63	21	$\pm 5.7$
7	127	42	$\pm 3$
8	255	85	$\pm 1.5$

the A-to-D converter is commanded by the peak-sensing electronics.

**Data priority.** It is proposed that all peaks be read into a buffer storage in their sequence of acquisition by the

**Table 3. Two-decade (0.1 to 10 v) linear quantization accuracy**

Bits	Quantization interval, v	10 v, %	1 v, %	0.1 v, %
4	0.67	$\pm 3.3$	$\pm 33$	—
5	0.32	$\pm 1.6$	$\pm 16$	—
6	0.16	$\pm 0.8$	$\pm 8$	—
7	0.08	$\pm 0.4$	$\pm 4$	$\pm 40$
8	0.04	$\pm 0.2$	$\pm 2$	$\pm 20$

**Table 4. Two-decade (0.5 to 50 sec) exponential quantization of time base**

Bits	Quantization levels	Quantization levels/decade	2-decade quantization uncertainty, %
4	15	8	$\pm 14.5$
5	31	16	$\pm 7.4$
6	63	32	$\pm 3.5$
7	127	64	$\pm 1.8$
8	255	128	$\pm 0.9$

chromatograph and on command from the instrument at the time of peak acquisition. Since the first analysis should be completed at least 150 sec before earliest impact, it is suggested that

- (1) The buffer storage have the capacity to store at least six peaks,

- (2) Data not be commutated to the telemetry link until the analysis is complete,
- (3) Three or four peaks be read out after the first analysis,
- (4) Peaks be read out in amplitude priority basis, and
- (5) Additional peaks be transmitted on subsequent analysis if time permits.

#### *e. In-flight electronic calibration*

A certain amount of useful information about the instrument can be acquired during interplanetary cruise. Electronic calibration may be performed by generating triangular current waveforms against which amplitude and time measurements can be made. The health of detector ionization sources can also be determined by measuring detector baselines. This measurement can be made in either the presence or absence of carrier gas.

## **2. Summary**

The capabilities of the electronic circuits described here have been demonstrated in the form of refined, experimental breadboards. These circuits, which are key elements in the instrument, have been successfully adapted to the gas chromatograph, and are capable of handling the fast, wide-range analyses that would be available from the column-detector system described in the submitted proposal.

## **References**

1. Bowman, L., Josias, C., Marshall, J. H., "Gas Chromatograph Instrument Development," SPS 37-20, Vol. IV, pp. 169-176, Jet Propulsion Laboratory, Pasadena, Calif., April 30, 1963.
2. Josias, C., Lawrence, J., "An Instrument for the Measurement of Interplanetary Solar Plasma," Technical Report 32-492, Jet Propulsion Laboratory, Pasadena, Calif., May 1, 1964.
3. Marshall, J. H., "Automatic Scale Factor Device for Use With Spacecraft Electrometers," SPS 37-23, Vol. IV, pp. 245-254, Jet Propulsion Laboratory, Pasadena, Calif., October 31, 1963.
4. Marshall, J. H., "Gas Chromatograph Detector Quiescent Current Drift Compensation," SPS 37-22, Vol. IV, pp. 213-217, Jet Propulsion Laboratory, Pasadena, Calif., August 31, 1963.
5. Marshall, J. H., "A Capacitor Storage Scheme for Gas Chromatograph Detector Quiescent Current Compensation," SPS 37-24, Vol. IV, pp. 213-220, Jet Propulsion Laboratory, Pasadena, Calif., December 31, 1963.

## XIII. Space Instruments Systems

### A. Carrier Collection and Spectral Response of Radiation Detectors

R. Y. Wong

During early phases of the *Mariner* Mars 1964 program it became evident that a more complete understanding of the parameters affecting the performance and characteristics of radiation detectors would be of assistance in defining the various parameters of the detectors required for performing the specified tasks. This summary describes a portion of the work performed in this area.

When radiant energy is incident at the surface of a p-n junction semiconductor detector, photons of the proper energy levels penetrate the device to generate electron-hole pairs. Diffusion of the electron-hole pairs due to density gradients and the collection of minority carriers near the junction give rise to a current flowing across the junction. Not all of the minority carriers generated are being collected by the junction; only carriers created within the diffusion length are capable of contributing to the output current. Other carriers diffuse away from the junction and recombine at the surface or

deep inside the detector. The fraction of carriers being collected determines the collection efficiency of the detector; i.e., the ratio of collected carriers to the total number of carriers generated.

#### 1. Collection Efficiencies

To derive the expressions for collection efficiencies, a model of the p-n junction detector is shown in Fig. 1, with the junction depth as  $a$  and the depth of the bulk layer as  $b-a$ . The appropriate equation for the minority electronic carrier  $n$  in the  $p$  layer has the form of:

$$\frac{\partial n}{\partial t} = -\frac{n}{\tau_n} + f + D_n \operatorname{div} \operatorname{grad} n + \mu_n E \cdot \operatorname{grad} n \quad (1)$$

where the first term of the right-hand side denotes the recombination rate of the minority carrier in excess of the equilibrium density and the second term represents the volume-density excitation rate of carriers produced by the radiation. The third term is the diffusion rate into or out of the layer and the fourth term is the drift rate of the carrier due to the built-in electric field.

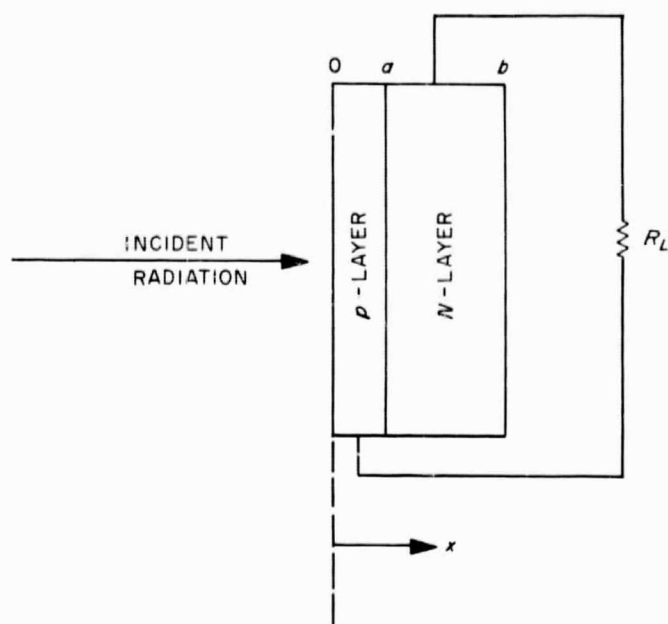


Fig. 1. Analytic model of the detector

If the detector is examined under those conditions for which the volume generation of minority carriers can be considered to be uniform, a one-dimensional analysis can be used. This one-dimensional model is valid if the detector length and width are large compared to the diffusion length for the minority carriers. If the detector is operating at a sufficiently low input energy level, the electric field generated is small and the drift rate of minority carriers due to this field is negligible when compared with that due to diffusion. It is further assumed that the space charge neutrality law applies and the detector is operating in the steady-state. With the above conditions taken into consideration

$$\frac{\partial n}{\partial t} = 0, \text{div grad } n = \frac{d^2 n}{dx^2}, \quad E = 0,$$

and therefore Eq. (1) becomes

$$D_n \frac{d^2 n}{dx^2} + f - \frac{n}{\tau_n} = 0 \quad (2)$$

where

$D_n$  = diffusion coefficient of the free electrons

$\tau_n$  = lifetime of the electrons

$$f = \alpha(\lambda) N_0 \exp[-\alpha(\lambda)x]$$

$\alpha(\lambda)$  is the absorption coefficient as a function of the wavelength of the incident energy, and  $N_0$  is the current

density of the incident photons. The general solution for Eq. (2) has the following form:

$$n = A \exp\left(\frac{x}{L_n}\right) + B \exp\left(-\frac{x}{L_n}\right) + \frac{\alpha(\lambda) N_0}{D_n \left[ \left(\frac{1}{L_n}\right)^2 - \alpha(\lambda)^2 \right]} \exp[-\alpha(\lambda)x] \quad (3)$$

where

$$L_n = (D_n \tau_n)^{1/2}$$

$L_n$  = diffusion length of free electrons

Similar equations can also be obtained for the hole,  $p$ , in the  $n$ -layer with  $D_p$  in place of  $D_n$  and  $L_p$  in place of  $L_n$ . The coefficients (i.e.,  $A$  and  $B$  in Eq. (3)) can be determined from the boundary conditions. At  $x = 0$ , surface recombination is taking place with  $S$  = surface recombination velocity. At  $x = a$ , the junction is a sink for the minority carriers

$$D_n \left( \frac{dn}{dx} \right) \Big|_{x=0} = SN(0) \quad (4)$$

$$n(a) = 0 \quad (5)$$

The diffusion current flow across the junction is

$$-q D_n \left( \frac{dn}{dx} \right) \Big|_{x=a} = I_n(a) \quad (6)$$

$$+q D_p \left( \frac{dp}{dx} \right) \Big|_{x=a} = I_p(a) \quad (7)$$

$$n(b) = N_0 \quad (8)$$

Applying the boundary conditions, the electronic current can be obtained by solving Eq. (3):

$$I_n(a) = \frac{q N_0}{\alpha(\lambda) \left[ 1 - \left( \frac{1}{L_n \alpha(\lambda)} \right)^2 \right]} \times \left\{ \frac{\beta_n}{L_n} \exp\left(\frac{a}{L_n}\right) - \frac{\gamma_n}{L_n} \exp\left(-\frac{a}{L_n}\right) - \alpha(\lambda) \exp[-\alpha(\lambda)a] \right\} \quad (9)$$

where

$$\beta_n = \frac{[h + \alpha(\lambda)] \exp\left(-\frac{a}{L_n}\right) - \left(h + \frac{1}{L_n}\right) \exp[-\alpha(\lambda)a]}{\left[h + \frac{1}{L_n}\right] \exp\left(\frac{a}{L_n}\right) - \left(h - \frac{1}{L_n}\right) \exp\left(-\frac{a}{L_n}\right)} \quad (10)$$

$$\gamma_n = \frac{\left[h - \frac{1}{L_n}\right] \exp[-\alpha(\lambda)a] - [h + \alpha(\lambda)] \exp\left(-\frac{a}{L_n}\right)}{\left[h + \frac{1}{L_n}\right] \exp\left(\frac{a}{L_n}\right) - \left[h - \frac{1}{L_n}\right] \exp\left(-\frac{a}{L_n}\right)} \quad (11)$$

$$h = \frac{S}{D_n} \quad (12)$$

Similarly, the hole current is

$$I_p(a) = \frac{-qN_0}{\alpha(\lambda) \left[1 - \left(\frac{1}{L_p \alpha(\lambda)}\right)^2\right]} \times \left\{ \frac{\beta_p}{L_p} \exp\left(\frac{a}{L_p}\right) - \frac{\gamma_p}{L_p} \exp\left(-\frac{a}{L_p}\right) - \alpha(\lambda) \exp[-\alpha(\lambda)a] \right\} \quad (13)$$

where

$$\beta_p = \frac{\alpha(\lambda) \exp\left[-\frac{a}{L_p} + \alpha(\lambda)b\right] - \frac{1}{L_p} \exp\left[-\frac{b}{L_p} + \alpha(\lambda)a\right]}{\frac{2}{L_p} \cosh\left[\frac{b}{L_p} - \frac{a}{L_p}\right]} \quad (14)$$

$$\gamma_p = \frac{-\alpha(\lambda) \exp\left[\frac{a}{L_p} - \frac{\alpha(\lambda)}{b}\right] - \frac{1}{L_p} \exp\left[\frac{b}{L_p} - \alpha(\lambda)a\right]}{\frac{2}{L_p} \cosh\left[\frac{b}{L_p} - \frac{a}{L_p}\right]} \quad (15)$$

To derive the expressions for the collection efficiencies of the  $p$ -layer ( $\eta_n$ ) and  $n$ -layer ( $\eta_p$ ), the maximum current which can be obtained from the photons is  $I_{max} = qN_0$ .

Therefore,

$$\eta_n = \frac{i_n(a)}{I_{max}} \quad (16)$$

$$\eta_p = \frac{I_p(a)}{I_{max}} \quad (17)$$

It is seen from (9), (13), (16), and (17) that the collection efficiencies of the  $p$ -layer and  $n$ -layer are a function of the detector's electronic and geometric parameters as well as the absorption characteristics of the device. Therefore, the appropriate efficiencies as a function of incident energy wavelength may be obtained by the proper selection of the electronic and geometric parameters.

## 2. Numerical Analysis of Collection Efficiency

To show the effects of the various parameters on detector performance, Eqs. (16) and (17) can be evaluated by the substitution of numerical values for these parameters. Silicon  $p$ - $n$  junction detectors have been selected for evaluation in this report, but similar evaluations can also be made for other  $p$ - $n$  junction devices.

The diffusion length  $L$ , is the average length of the path the minority carriers travel before recombination takes place. The diffusion length is a function of the diffusion coefficient  $D$ , and the lifetime  $\tau$ , of the carriers

$$L_p = (D_p \tau_p)^{1/2} \quad (18)$$

$$L_n = (D_n \tau_n)^{1/2} \quad (19)$$

The diffusion coefficient is related to the mobility  $\mu$  of the carriers by the Einstein relation

$$\mu_p = \frac{e}{KT} D_p \quad (20)$$

$$\mu_n = \frac{e}{KT} D_n \quad (21)$$

where  $K$  is the Boltzmann constant,  $T$  is temperature in  $^{\circ}\text{K}$  and  $e$  is the electronic charge.  $KT/e$  has a numerical value of  $26 \times 10^{-3}$  v at room temperature. Experiments performed by D. C. Cronemyer, E. M. Conwell, M. B. Prince, and G. Backenstoss (Refs. 1, 2, 3, and 4) indicated that the approximate numerical value of the mobility  $\mu_n$  is  $1200 \text{ cm}^2/\text{v-sec}$  for the  $p$ -type, and the mobility  $\mu_p$  is  $500 \text{ cm}^2/\text{v-sec}$  for the  $n$ -type. The body lifetime is on the order of  $10^{-6}$  sec for the electrons and  $10^{-9}$  sec for the holes. Therefore, the diffusion length, when calculated by Eqs. (18) and (19), should have the limiting range of values  $10^{-2}$  to  $10^{-4}$  cm for the purpose of numerical evaluation. The selection of these values for this analysis is

considered appropriate since more recent measurements performed by V. S. Vairilvo, L. S. Smirnav and V. M. Patskevitch (Ref. 5) indicated that the diffusion length is on the order of  $5 \times 10^{-3}$  cm for silicon.

The surface recombination velocity  $S$ , depends on the dimensions and the surface properties of the detector since the free carriers must diffuse to the surface before recombination. As indicated by E. M. Conwell (Ref. 2),  $S$  has been found ranging from  $10^2$  cm/sec to  $10^4$  cm/sec. The high value is obtained on sandblasted surfaces. If the surface is polished smooth and etched, the value of  $S$  may decrease to  $5 \times 10^2$  cm/sec. Chemical treatment of this polished and etched surface may reduce  $S$  further to  $1 \times 10^2$  cm/sec. By choosing  $\mu_n$  of the  $p$ -layer to be  $1200$  cm<sup>2</sup>/v-sec and  $S$  to be  $1 \times 10^4$  cm/sec,  $h$  is on the order of  $300$  cm<sup>-1</sup> where  $h = (e/KTS)/\mu_n$ .

Theories have been developed (Ref. 6) which relate the optical absorption coefficient to the parameters of the band structure. The numerical values of absorption coefficient as a function of wavelength have been calculated theoretically in Ref. 7 and measured experimentally (Ref. 8) for silicon. Fig. 2 shows the numerical value of the absorption coefficient  $\alpha(\lambda)$ , as a function of wavelength.

To obtain the numerical values of  $\eta_n$  and  $\eta_p$ , Eqs. (16) and (17) were solved for a number of cases using the IBM 7094 Computer. Fig. 3 shows the resulting  $p$ -layer collection efficiency  $\eta_n$  as a function of wavelength with a detector thickness of  $5 \times 10^{-2}$  cm and various junction depths as the parameter. It is noted that  $\eta_n$  starts to rise at longer wavelengths for a deep junction than for a shallow junction. Therefore, a detector with a deep junction utilizes the infrared radiation more efficiently. However, the over-all efficiency decreases as the junction depth increases. For a deep junction, a considerable percentage of electron-hole pairs are generated near the surface and away from the junction. These minority carriers are not capable of contributing to the junction current because of the recombination process taking place at the surface and inside the  $p$ -layer.

Fig. 4 (dotted curves) shows the  $n$ -layer collection efficiency  $\eta_p$  as a function of the wavelength with a detector thickness of  $5 \times 10^{-2}$  cm and the various junction depths as the parameter. The dotted curves were derived by assuming the diffusion length of the holes to be  $10^{-2}$  cm. It is seen that quite a large number of carriers are created in the  $n$ -layer. These carriers are created by radiation penetrating through the  $p$ -layer. Therefore, a

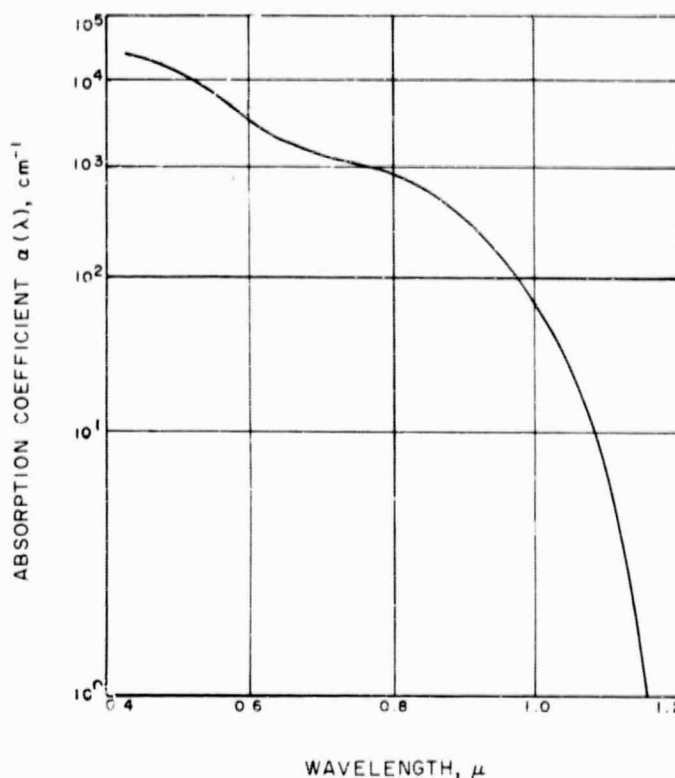


Fig. 2. Absorption coefficient of silicon as a function of wavelength

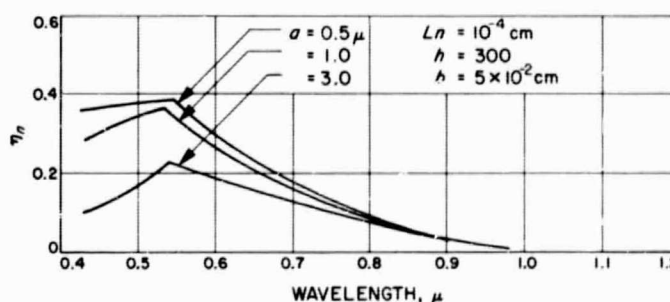


Fig. 3. Collection efficiency for  $p$ -layer,  $\eta_n$  as a function of wavelength

high percentage of the total junction current is being collected in the  $n$ -layer. To see the effects of the diffusion length and lifetime of the carrier on the  $n$ -layer collection efficiency, the curves in solid lines were generated on Fig. 4. The values of these curves were calculated by assuming the diffusion length of the holes to be  $10^{-4}$  cm instead of  $10^{-2}$  cm. It is seen that the collection efficiency is considerably reduced due to the decrease in diffusion length. To obtain high collection efficiency, therefore, the diffusion length of the  $n$ -layer must have a reasonably high value (on the order of  $10^{-2}$  cm).



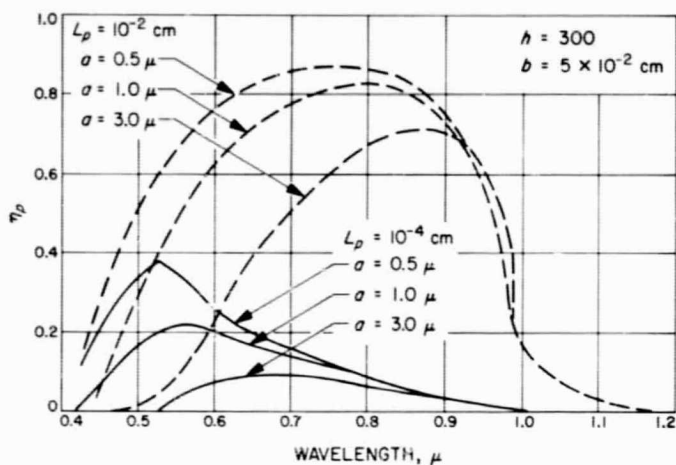


Fig. 4. Collection efficiency for  $n$ -layer  $\eta_p$  as a function of wavelength

### 3. Detector Collection Efficiency and Spectral Response

The analysis in the preceding section indicated that both the  $p$ -layer and  $n$ -layer are capable of contributing carriers to create junction current. Therefore, the detector collection efficiency  $\eta_c$  is the sum of the  $p$ -layer and  $n$ -layer efficiencies. Fig. 5 shows the detector collection efficiency as a function of wavelength with various junction depths as the parameter. Again, it is obvious from these curves that a detector with a deep junction is capable of using energy in the infrared region more efficiently while a detector with a shallower junction has a higher over-all collection efficiency. Since these curves are plotted as a function of incident energy wavelength, the shape of each curve represents the spectral response of a detector having the specified electronic and physical

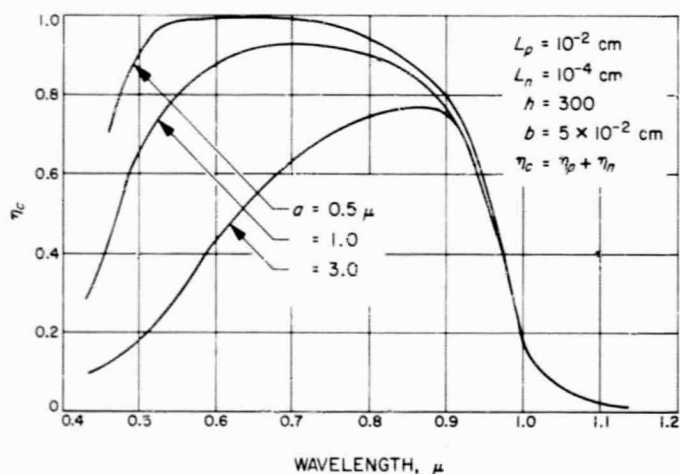


Fig. 5. Detector collection efficiency as a function of wavelength

parameters. For an energy source of known spectral distribution, it is therefore possible to select the various parameters of the detector to match its spectral response to that of the source. Such a detector is capable of detecting the incident energy most effectively.

### 4. Conclusion and Discussion

In this summary the carrier collection and spectral response were analyzed and numerical values of the detector collection efficiency were evaluated using the various parameters of silicon  $p$ - $n$  junction detectors. This method can also be used for the analysis of junction detectors of other types.

As seen from Figs. 3 and 4, the  $p$ -layer is most sensitive to radiation in the violet and low visible spectral range. Photons in the longer wavelengths penetrate and transmit through this layer due to insufficient photon-electron interaction. These photons, after transmitting through the  $p$ -layer, penetrate and create electron-hole pairs in the  $n$ -layer. Incident photons having an energy spectrum exceeding 1.0 or 1.1  $\mu$  will transmit through the detector without interacting with the structure at all.

The  $p$ -layer should be made as thin as possible for high collection efficiency in the short and medium wavelength regions. However, the series resistance of the  $p$ -layer increases as the junction depth decreases, resulting in a reduction in the efficiency improvement. A deep junction is required for high efficiency in the long wavelength region. With deep junctions, the achievement of large diffusion lengths becomes important for high over-all efficiency.

In this summary, only the efficiency due to minority carrier collection by the junction was considered. To obtain the over-all detector radiation collection efficiency, a number of factors need to be considered. When the radiation is incident upon the detector, some is reflected without entering the detector. When the photons absorbed have more energy than is necessary for the generation of electron-hole pairs, the excess energy is wasted and contributes to lattice vibration and is dissipated as heat. The energy needed for the photon-electron interaction is a function of the energy gap of the material. For every value of the energy gap, an absorption band edge is obtained beyond which the photon energy is not sufficient to create electron-hole pairs and the detector will be transparent to these photons. The over-all detector radiation collection efficiency is given by:

$$\eta(\lambda) = [1 - R(\theta)] [1 - \exp(-\alpha(\lambda)b)] p(\lambda) \eta_c(\lambda)$$

where  $R(\theta)$  is the reflection coefficient as a function of radiation incident angle and  $(1 - \exp[-\alpha(\lambda)b])$  is the factor accounting for incomplete absorption.  $p(\lambda)$  is the relative spectrum distribution of the radiation source and  $\eta_c(\lambda)$  is the detector collection efficiency which has been discussed in this summary.

## B. A New Head for Recording Twenty-Four Tracks on Quarter-Inch Magnetic Tape

W. G. Clement

### 1. Current Tape Utilization Practice

Two factors determine the tape utilization efficiency of a magnetic tape recorder as applied to digital data storage. These are the number of data bits per lineal tape inch, and the number of data tracks per inch of tape width. At the present time it is a routine accomplishment to pack  $10^3$  bits per lineal inch, provided reasonable care and thought is given to matching the head and tape characteristics. Densities up to 1500 bits/in. have been reliably achieved through the utilization of special pulse and data reconstruction schemes (Ref. 9). With regard to transverse track density, it is also routine at present to record 32 tracks per inch of tape width. Track density depends primarily upon the ability of the head manufacturer to fit together the individual transducers as closely as possible in the head stacks while preserving sufficient electrical and magnetic isolation to ensure low cross-talk. Thirty-two tracks per inch reflect the present capability of manufacturers to fit 16 channels/in. in one head stack without compromising head design integrity. The alignment of two such head stacks so that the channels interlace provides 32 equally spaced tracks on 1-in. tape.

### 2. Goals and Implications of the Study

A study is being conducted to determine the feasibility of recording bits at 10,000 per lineal tape inch and tracks at 100 per inch of tape width. If this goal, which is equivalent to  $10^6$  bits/in., can be achieved, the implications are far reaching. For example, the length of 0.50-in. tape required to store 20 TV pictures, as planned for the *Mariner C* program would be only 10 in., as

opposed to the 300 ft of 0.25-in. tape required by the *Mariner C* recorder. It would then become likely that the traditional methods of tape handling would be inappropriate. Moreover, the present design concepts of tape recorders for flight data storage might require drastic modification in order to utilize this new dimension of the storage medium with its attendant benefits in size, weight, and power. A  $10^7$  bit recorder for a space capsule might well assume the size of a cigarette pack, the minimum size being dictated principally by the drive motor.

### 3. Technical Problems Associated with High Track Density

Some work on lineal bit density has been reported previously (Ref. 10). Further effort toward higher lineal densities involving record head design is currently in progress and will be reported at a later date. Since this report is concerned with the achievement so far realized in increased track density, subsequent discussion will be limited to this aspect of the problem. The two most important effects to be considered in increasing the track density of a magnetic recording system are signal level and cross-talk. Signal level decreases in direct proportion as the track width is decreased, and cross-talk increases as the spacing between tracks is decreased. These two major effects combined with numerous lesser effects establish the signal-to-noise ratio (S/N) which in the final analysis determines the information storage capacity of the system. In applying information theory to magnetic recording systems, Eldridge has shown (Ref. 11) that the information capacity per unit area of tape is greatest when the tracks are made as narrow as possible. Eldridge and Baaba have shown (Ref. 12) that the S/N varies as the square root of track width and that, theoretically, with presently available tape and head-core materials, the minimum attainable track width is 0.001 in., at which value the signal-to-tape-noise ratio would be 30 db. This track width represents at least one order of magnitude beyond present practice. The practical significance of the conclusions reached by Eldridge and Baaba diminishes somewhat for a digital system in which other noise sources, such as cross-talk, tend to overshadow tape noise.

From a cross-talk standpoint, track spacing may be reduced until the output from any reproduce head includes the maximum permissible fraction of the signal in an adjacent track. One portion of the cross-talk signal increases with recorded wavelength and relates directly to the distance between tracks on the tape. Another portion of the cross-talk signal increases with frequency



and relates to the transformer action between core windings within a head stack. This latter source of cross-talk can be reduced by employing a multiplicity of head stacks and staggering or "interlacing" cores and tracks.

In brief, factors mitigating against higher track densities are:

- (1) A reduction of signal level.
- (2) An increase in cross-talk.
- (3) The difficulty of construction as thinner and thinner head cores are placed closer and closer together.
- (4) Increased drop-out susceptibility from tape defects and loose dirt due to the smaller area of tape being scanned at one time.
- (5) The need for more precise tape guidance for proper track registration.

#### 4. A New HTD (High Track Density) Head, Description and Test Results

Two head stacks, one record and one playback, have been manufactured to JPL specifications by the Applied Magnetics Corporation. These head stacks (Fig. 6) embody the traditional ring-core construction but reflect advanced techniques in coil winding, machine-slotting, fixturing and assembly. Twelve channels are packed into each stack assembly. Two such stacks, interlaced, provide

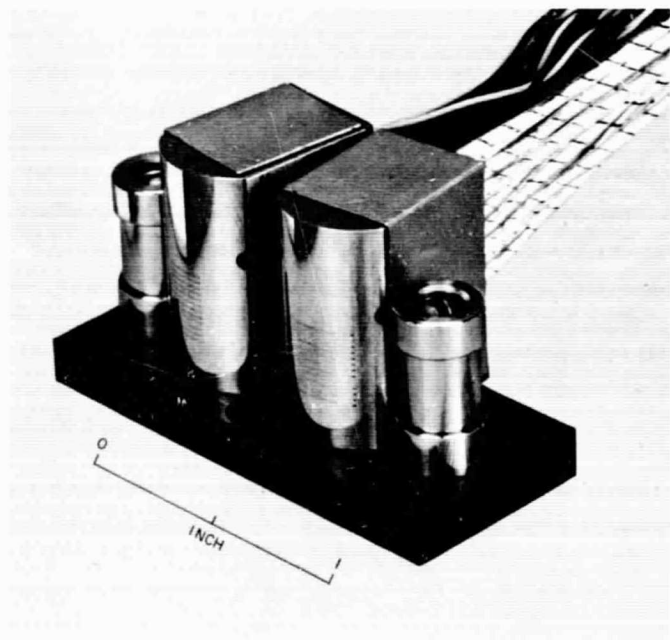


Fig. 6. High track density head assembly

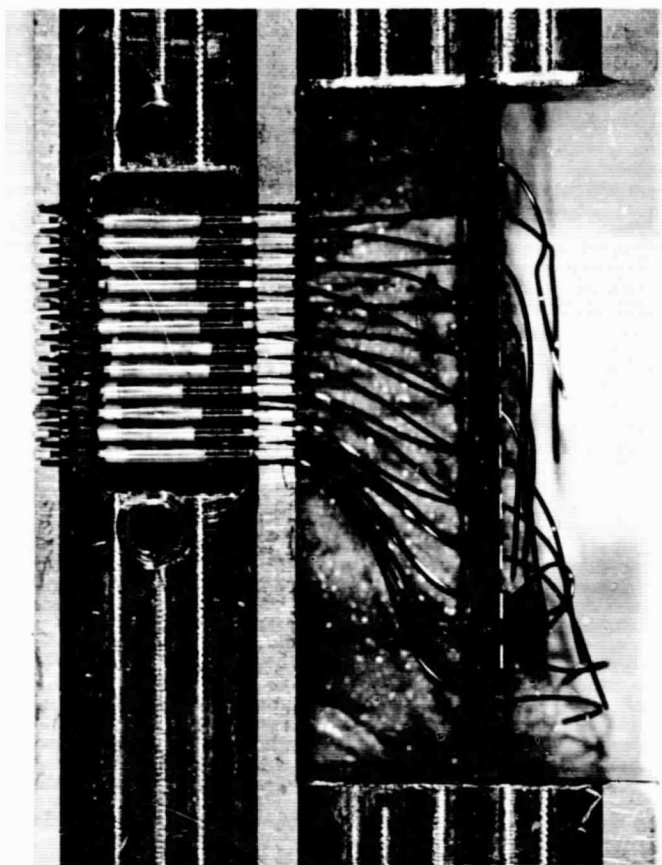


Fig. 7. High track density head half-block assembly

24 tracks on 0.25-in. tape, e.g., 96 tracks/in. Each track is 0.006 in. wide, while the pitch, or distance from track to track, is 0.020 in. The gaps are 0.25 mil (record) and 0.055 mil (playback). Fig. 7 is a view of the "half-block" assembly showing the 12 half-cores in place. Some idea of the size can be gained from the fact that the wire wound on the half-cores is No. 50 (0.001-in. D) and the pigtails connecting the half-cores to the terminal board are No. 35 (0.006-in. D) wire.

Inductance, AC resistance, and output data were measured channel-by-channel in order to determine the uniformity factor. This factor usually reflects the degree of manufacturing control over the fabrication and assembly operations. Table 1 is a compilation of these measurements. Similar measurements were made on a standard four channel head (0.025-in. track width and 0.060-in. track pitch) made by the same manufacturer, and Table 2 is a similar compilation of these measurements. Table 3 lists the variations in these factors expressed as percentages of the average for both the HTD head and the standard head. As expected, the variations are significantly greater for the HTD head, although part of this increase may be attributed to the greater quantity of channels measured.

Table 1. Inductance, resistance and output data for HTD head

Channel	100 kc <sup>a</sup>						500 kc		●Output <sup>b</sup> Mv, P-P
	Record — Leg 1		Record — Leg 2		Play — Total		Record — Leg 1		
	L	R	L	R	L	R	L	R	
1	121.05	36.0	129.05	37.0	1175	803	65	140	0.56
2	138.75	42.8	148.50	44.4	1214	843	73	160	0.60
3	145.44	45.3	150.15	45.6	1194	794	77	170	0.60
4	149.07	45.5	147.82	45.0	1032	512	79	170	—
5	148.70	44.4	144.94	43.9	1204	824	80	170	0.60
6	145.30	44.5	136.72	43.3	1016	737	77	167	0.52
7	148.64	44.6	142.52	44.0	1267	847	79	174	0.62
8	145.73	43.7	148.16	44.6	1280	841	77	170	0.60
9	147.95	45.2	145.55	44.8	1316	860	79	170	0.60
10	143.60	44.3	147.64	43.6	1257	826	75	170	0.56
11	146.64	44.4	—	—	1234	814	79	167	0.52
12	123.05	35.5	124.21	35.5	1236	848	67	138	0.46

<sup>a</sup> L in microhenries and R in ohms measured with Boonton Electronics Model 63C Bridge<sup>b</sup> Output at 7.5 ips, Memorex 62L tape, 2-kc square wave recorded at 4-amp turns peak

Table 2. Inductance, resistance, and output data for standard 4-channel head

Channel	100 kc <sup>a</sup>		500 kc <sup>a</sup>		Output <sup>b</sup> Mv, P-P
	L	R	L	R	
1	47	11.7	29	49	5.0
2	44	11.1	27	46	6.2
3	44	10.3	27	45	6.0
4	44	10.2	27	45	6.2

<sup>a</sup> L in microhenries and R in ohms, measured with Boonton Electronics model 63C Bridge<sup>b</sup> Output at 7.5 ips, Memorex 62L tape, 2-kc square wave recorded at 4-amp turns peak

Table 3. Comparison of parameter variations between HTD and standard head

Measurement	Variation, %	
	Standard head	High track density head
100-kc inductance	6.7	20.0
100-kc resistance	14.0	23.2
500-kc inductance	7.3	20.0
500-kc resistance	8.7	22.0
Output signal	20.5	31.0

It is interesting to note that there is no significantly greater variation in 500-kc parameters than in 100-kc parameters for either the HTD head or the standard head. It may be noted in Table 1, record channels 1 and 12 of the HTD head are significantly lower in inductance and resistance than the other channels. This is attributable in each case to the absence of one adjacent

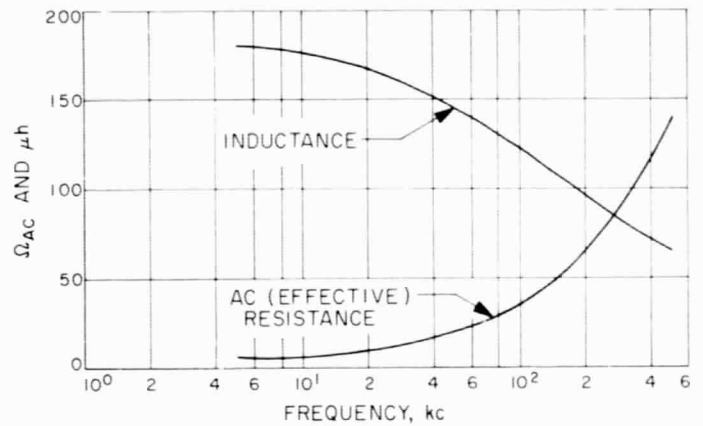


Fig. 8. Inductance and resistance versus frequency for HTD head

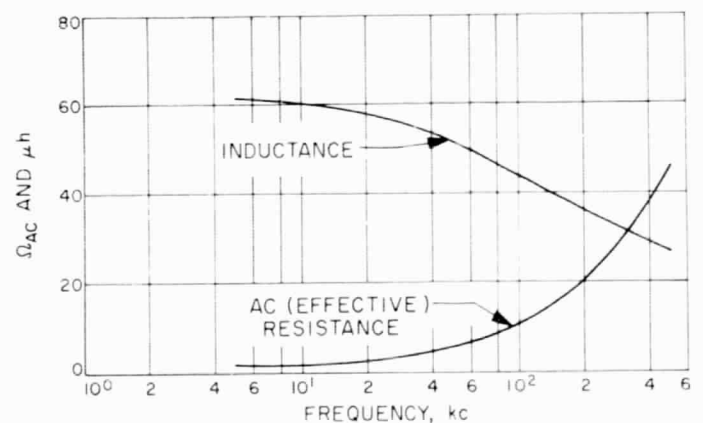


Fig. 9. Inductance and resistance versus frequency for standard head

channel which would by its presence lower the magnetic reluctance, thereby increasing the magnetic permeability of these two channels.

Figs. 8 and 9 are plots of inductance and AC resistance versus frequency for the HTD head and the standard head, respectively. It is instructive to compare the two sets of characteristics in terms of slope, e.g., rate of change with respect to frequency. The change in AC resistance between 10 and 500 kc is about the same percent for both heads; however, the inductance of the high-track density head at 500 kc has decreased to 31% of that at 10 kc, whereas the inductance of the standard head has only decreased to 44% of that at 10 kc. The lesser degradation of inductance (with increased frequency) of the standard head may be attributable to its core being made up of 3-mil laminations, whereas the core of the HTD head consists of 3- and 4-mil laminations. The generation of eddy currents with attendant reduction of effective permeability may therefore be greater in the HTD head. In any case it is not apparent that this effect is related to the HTD density aspects of the new head.

Fig. 10 indicates S/N attainable with the HTD head for given values of record current.<sup>1</sup> Noise here is, by definition, limited to that contributed by cross-talk and incomplete erasure of previously recorded pulses. The latter may be a significant source of noise in nonreturn to zero (NRZ) systems where no separate erase function is pro-

<sup>1</sup>In determining the operating record current on a characteristic curve such as shown in Fig. 10, an important consideration not indicated is that recorded pulse spreading and consequent loss of resolution accompanies an increase in record current (Ref. 13).

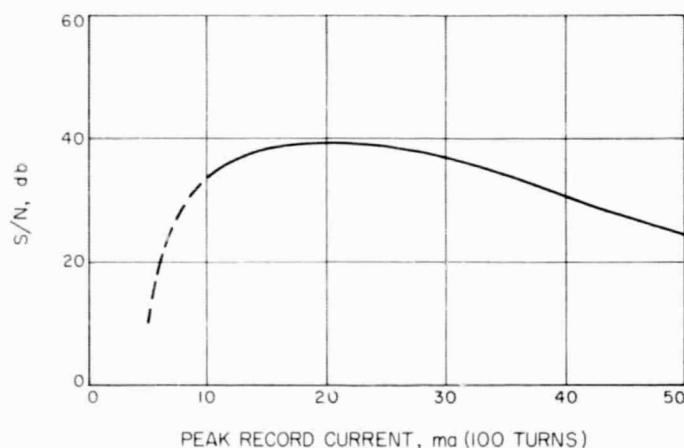


Fig. 10. HTD head S/N versus record current  
Memorex 62L tape at 15 ips

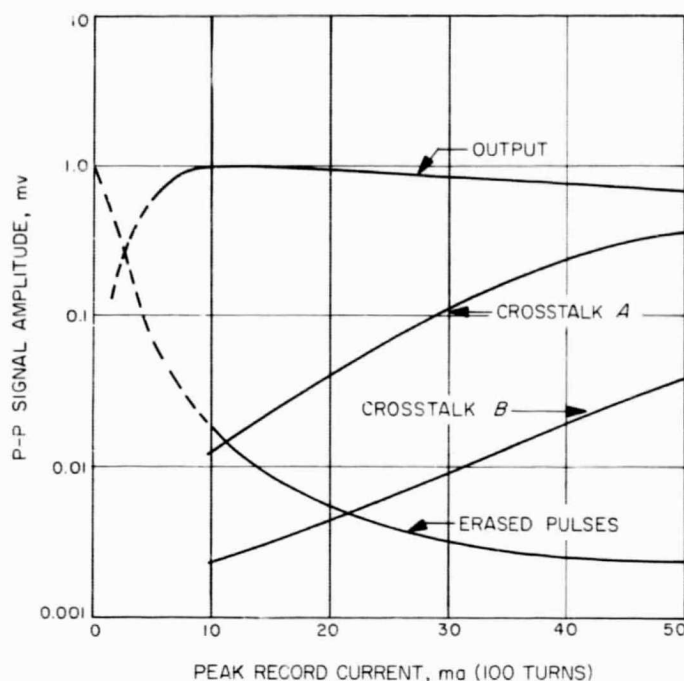
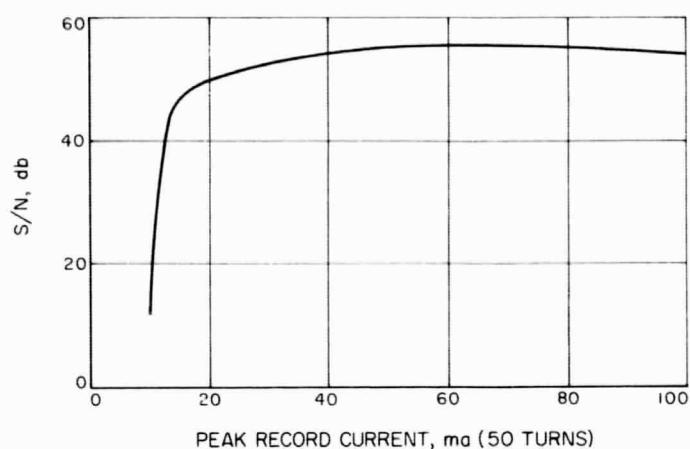
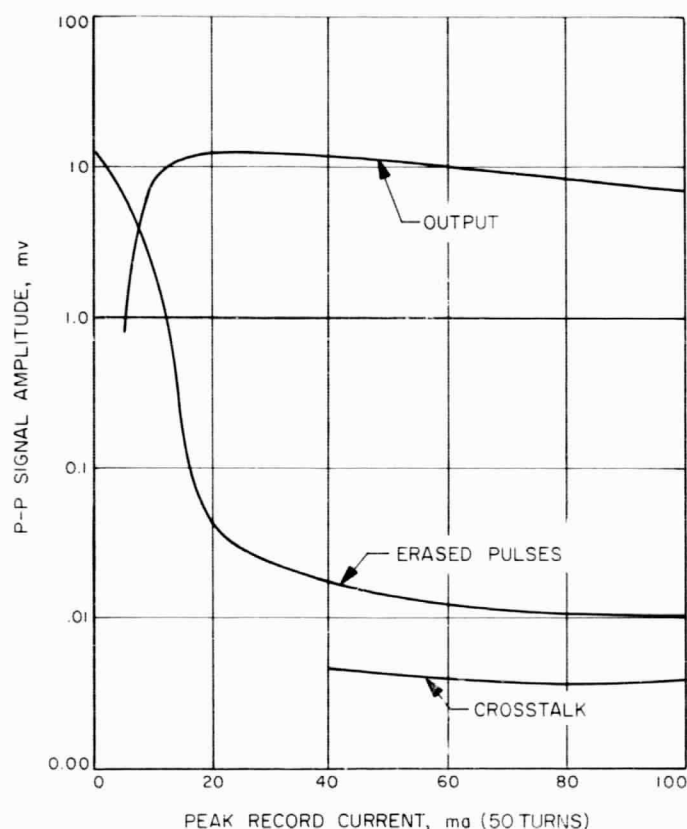


Fig. 11. HTD head characteristics

vided. Fig. 11 indicates these two noise sources as well as signal output separately, with record current as the independent variable. Figs. 12 and 13 indicate, for comparison, the same characteristics of the standard heads. Output curve shapes for the two heads are about the same, indicating equal recording efficiency. Cross-talk Curve A in Fig. 11 was made with the original version of the HTD head (as shown in Fig. 6) with no record-stack interchannel-shields. Subsequently, 4-mil slots were milled into the face of the head stack between the record channels and "tip" shields inserted. Cross-talk Curve B resulted, and the signal-to-noise ratios in Fig. 10 reflect this improved version. Cross-talk measurements were made by recording a 10-kc square wave in one channel and reading the amount of 10 kc in the adjacent channel. In view of the relatively short recorded wavelengths involved it seems reasonable to assume that most of the cross-talk in the HTD head resulted from cross-coupling in the head stacks. To the extent that this cross-coupling resulted from simple transformer action, the level would be expected to increase with frequency up to a rate corresponding to the limiting bit-packing density. This was the case only to a very minor extent, e.g., the increase of cross-talk with frequency was barely discernible up to about 10 kc, above which point it diminished rapidly. From this and other experimental evidence, it appears that a significant amount of interchannel coupling in the HTD head occurs by a process other than simple transformer action.



**Fig. 12. Standard head S/N versus record current**  
Memorex 62L tape at 15 ips



**Fig. 13. Standard head characteristics**

A factor of possible significance in comparing the erased signal curves for the HTD head and the standard head is that the HTD record tracks are 8 mils wide, thus providing a 1-mil margin each side of the 6-mil playback tracks which would minimize the effects of any track misalignment due to tape "skew." The standard record and

playback heads are of equal width and thus provide no such margin.

Another point of interest with regard to S/N comparisons is suggested by an examination of Tables 1 and 2. It may be noted that the standard head output is about 12 times that of the HTD head. The obvious factors of track width (4X) and turns (1.8X) do not completely account for this difference. The remaining factor of 1.7X may be attributable to the shunting effect of the inter-channel shields which for the HTD head are separated from the main core structure by only 0.005 in. as contrasted with a separation of 0.020 in. for the standard playback head.

## 5. Conclusions

With regard to three of the five previously enumerated factors mitigating against higher track density, the following statements can be made:

- (1) 6-mil heads spaced 20 mils apart can be physically constructed at least in a quantity of 12 in one stack, thus providing, in an interlaced set, 24 tracks on 0.25-in. tape.
- (2) Signal levels of 1.0 mv peak-to-peak at 15 ips are attainable.
- (3) Cross-talk at a reasonable record level can be held to 34 db at 15 ips when considering the interaction of two adjacent channels.

## C. A Multi-Instrument Data Conditioning, Storage and Retrieval System for Planetary Spacecraft

*D. W. Slaughter*

A system is described which is capable of storing and playing back, from a single tape transport, the data generated during planetary encounter. Preliminary to a discussion of the system developed, the characteristics of the telecommunication channel and tape storage systems are assessed as to their effect on science data recovery. Their characteristics are compared and contrasted with the data output from scientific instruments.

It is shown that the manner in which the tape transport is utilized constitutes the heart of the functional design of a planetary data system when a large amount of storage is required.

### 1. Telecommunications System Interface

The science data system may have interfaces with a telecommunication system which has the following characteristics:

- (1) Provides a single binary channel.
- (2) Transmits at a synchronous rate under the control of a stable clock (oscillator).
- (3) Rate changes may be required as a function of distance, trajectory, power availability, or other factors.
- (4) Channel capacity may be time-shared with spacecraft performance information, and for two-way doppler and ranging information.

While item (1) may create difficulties in sequencing the science data, a single binary channel is in the best interests of science data retrieval if it provides the greatest total bit rate, and the best over-all reliability.

The channel capacity projected for the 1966 Mars opportunity is  $8\frac{1}{2}$  bits/sec. Other missions may provide lower rates, while rates of 1000 bits/sec or more may be possible for some missions in the 1970s.

### 2. Instrument Data Characteristics

The on-board data acquisition system must accept data simultaneously from a variety of instruments. A payload group may possess the following characteristics:

- (1) A wide range of data rates.
- (2) A binary data which is necessarily asynchronous with a spacecraft clock.
- (3) Intermittent data (in bursts with extensive quiescent intervals).
- (4) Require the introduction of data identification.
- (5) Produce data at a rate which exceeds the telecommunications channel capacity for various intervals of time.

Item (5), the production of data at rates which temporarily exceeds the capacity of the telecommunications channel, dictates the use of data storage devices. For the

storage of millions of bits (bulk storage), the magnetic tape transport is currently in use. Other mechanisms, such as electrostatic and thermoplastic recording are currently under investigation. Below one million bits, interest in coincident current magnetic matrix storage increases. However, this report describes the development of a system which uses a tape transport for bulk storage in combination with buffer (magnetic matrix) storage, as required by the recording characteristics.

### 3. Tape Storage System Characteristics

The recording data format characteristics are a function of the playback mode. On *Mariner C*, the bit sync is locked to the telemetry clock ( $8\frac{1}{2}$  bits/sec) during playback. The motor speed is servo-controlled in order to maintain lock. The input characteristics of this machine are thus the same as that of the telemetry channel, i.e., synchronous, except that the data rate is higher by the ratio of the record to playback tape speed. If there are any gaps in the data, communication time is used up; that is, data zeros are transmitted. The machine may be started and stopped during recording, but no provision exists for removing the zeros which are recorded during tape acceleration and deceleration.

The difficulty of utilizing a single tape transport with synchronous input rates becomes readily apparent if a group of instruments having the following characteristics (refer to Sect. 2) is assumed:

- (1) Widely variant data rates.
- (2) Asynchronous sources.
- (3) Intermittent data.
- (4) Data identification required.

The following techniques may be utilized in matching the data sources to a recorder of the type previously discussed: Intermittent data may be smoothed into a continuous flow by using coincident-current magnetic matrix "buffer" storage. The read-out rate from the buffer may be made constant, while the amount of data stored increases during data bursts, and decreases during instrument quiescent intervals. The limitation on this technique is the quantity of bits in each burst, which determines the size of the buffer. An attempt may also be made to interleave or "fit" two intermittent data sources so as to improve the smoothness of data flow; however, opportunities for using this technique are very limited. Asynchronous sources may also be converted into synchronous sources by passing the data through a buffer. Utilizing



this technique, the synchronous clock frequency is set higher than the highest average rate of the asynchronous source. Should the average rate of the asynchronous clock drift lower, then zeros are added to the data flow, usually at intervals determined by the buffer data capacity. Having widely variant rates from several instruments poses a difficult problem. If each instrument is provided with a buffer whose size (capacity) is proportional to its data rate, then these buffers may be unloaded sequentially. However, it is possible for the rate from a given instrument to change as a function of the mission (trajectory) profile. In addition to requiring a tape transport speed change, the format for sequential unloading of the buffers must change.

A single instrument which produces a high percentage of all data may profitably use its own tape transport.

These machines should be simple in design and capable of starting and stopping rapidly in order to accommodate bursts of data.

#### 4. A Multi-Instrument Data Acquisition, Storage and Retrieval System

A science data system is being developed which utilizes a tape storage scheme capable of accepting data at various rates from different instruments, and recording simultaneously on independent tracks. Data gaps of arbitrary length are permitted on the tape. Continuous utilization of the telemetry channel on playback is provided for by buffering the playback. In this system, the data is organized into blocks by accumulating it in buffers. When the buffer is full, the entire block is read into the

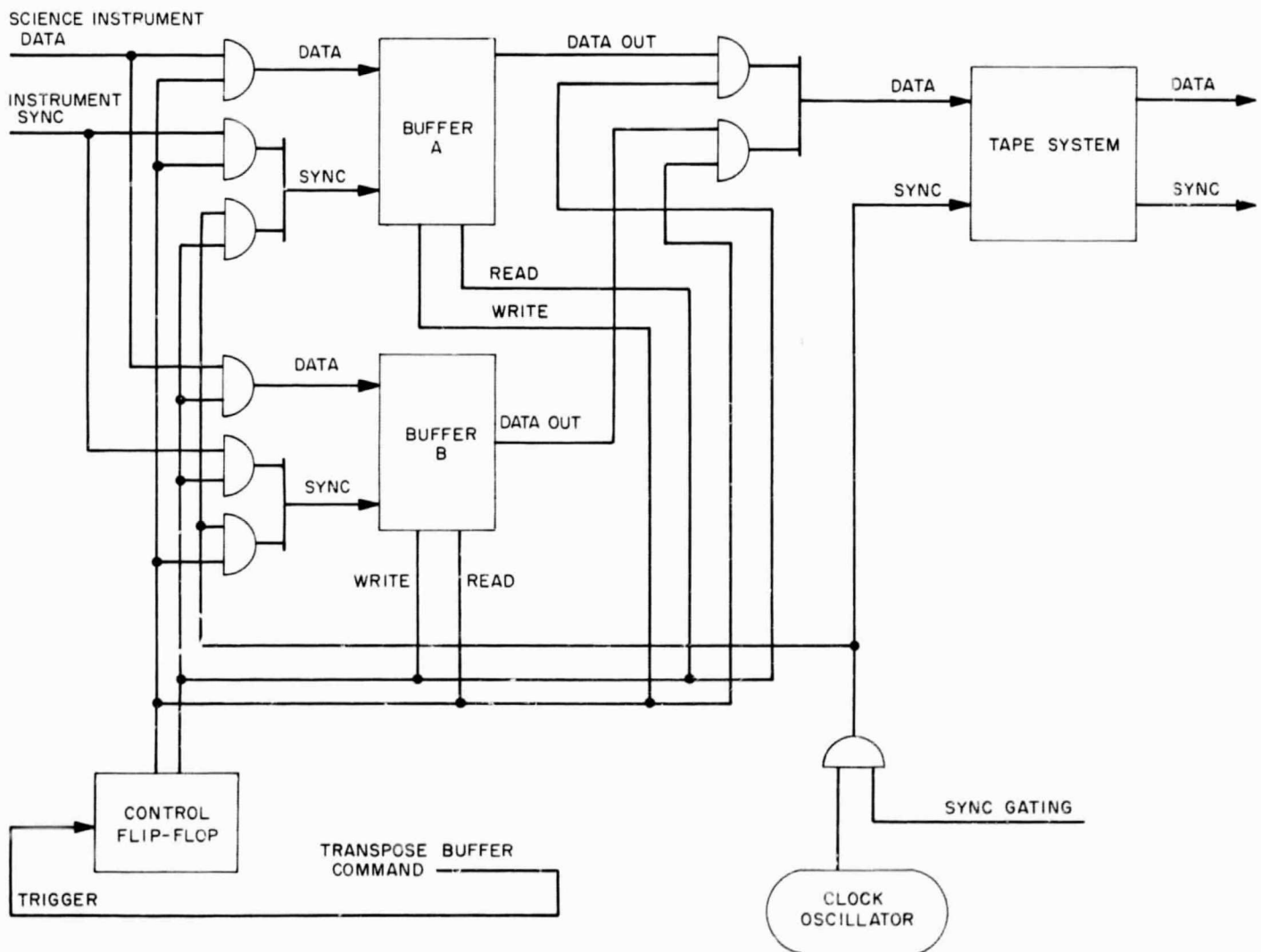


Fig. 14. Buffer interlace

recorder at a constant rate. In order to accept data continuously, if required, the buffer system is provided with interlaced read-in and read-out. For the purpose of maximum economy in the use of developmental funds, two *Mariner C* buffers (Ref. 14) have been employed, with one buffer available for loading data while the other unloads onto the tape machine. Fig. 14 is a logic diagram of this system. When the first buffer is full, then their situations are reversed. These *Mariner C* buffers are of the serial loading and unloading, first-in, first-out type; and store 1320 bits each. A pair of buffers weigh 1.25 lb and have a volume of 36 in.<sup>3</sup>; their power consumption is about 0.50 w per pair at low data rates, increasing linearly to 1.7 w at 100 kc. As stated previously, one entire buffer load is recorded continuously at a fixed clock rate, independent of the instantaneous or average rate at which the other buffer acquires data from the scientific instrument. It is intended that data gaps be inserted onto the tape between the times when one buffer has finished its read-out, and the other is fully loaded. The length of these gaps will be an inverse function of the data output rate from the instrument, averaged over one buffer load. If there is but a single recorded track, then the recorder may be stopped between blocks; however, this system has been engineered to handle several instruments (or

instrument groups) simultaneously and independently on several tracks. The number of tracks which may be recorded simultaneously and independently is limited only by the usual tape and head characteristics. It is assumed that the recorder may be running continuously. However, provisions have been made for running the tape recorder intermittently for single buffer dumps.

During playback, the tracks are played back one at a time in sequence. The recorder stops on data gaps; when it is given a start command, a block of data is read into a buffer. When the next data gap is reached, the recorder is stopped by means of a gap detector. This system is shown in Fig. 15. The gap detector is a device which looks for recorded data by sensing the data sync which marks each *one* or *zero*. When none is found for a fixed interval of time (usually several sync pulses, so as to allow for accidental data dropout), a gap indication is given. When the buffer issues an indication that it is empty, the recorder is restarted. Because the machine will pass over the remainder of the gap before encountering data, the gap detector must be inhibited until the beginning of the data block is sensed, after which the gap detector may be "armed." In practice, two buffers have been used so that a full one is always available and

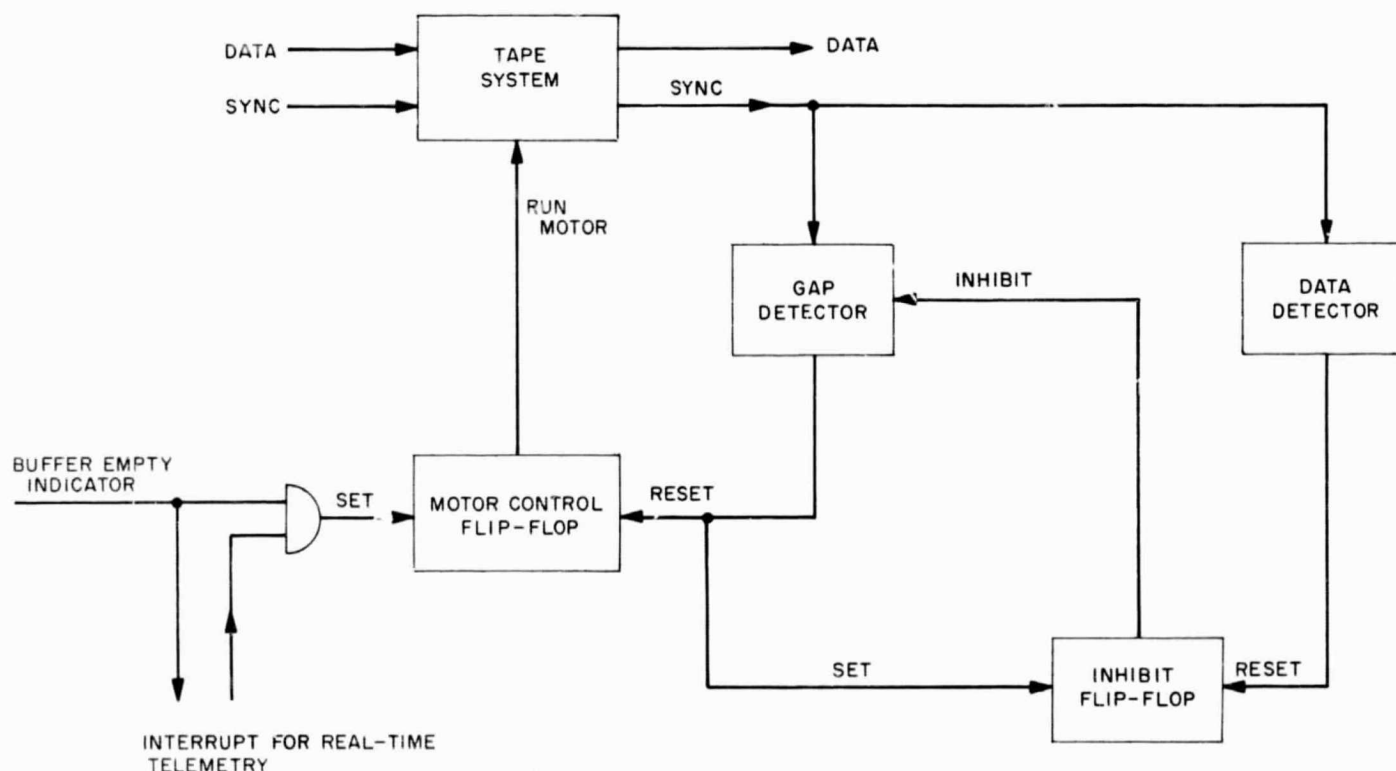


Fig. 15. Playback control circuits

ready for telemetry, and the same buffers in the same logical arrangement are used both for writing data onto the tape and reading it from the tape. The switchover is accomplished by logic gating, as shown in Fig. 16. Logic gates for Buffer B have been omitted in the diagram. The total cost is one diode per *and* gate; the logic level applied to these diodes enables the proper gates for the record or playback function.

The length of the data gap required between data blocks is a function of the motor stop-start characteristics. Data identification should normally be inserted at the beginning of each data block. A code generated by implementing a suitable binary sequence (pseudonoise) is presently being used for identification, usually followed by a reading of a binary clock. Because of the high read-

in rate capability of the buffer (100 kc in the present model), identification may be inserted between consecutive words of lower rate sources, without any loss of data. If desired, the instruments control system may command a partially loaded buffer to be dumped into the recorder at the end of some basic instrument data frame, scan, or other operating interval. The remainder of the recorded data block is filled out with *zeros*. Techniques for handling variable word length are under consideration.

Fig. 17 is a functional diagram of a data-conditioning system which may be inserted between the instrument and the buffer. The pulse width converter is an approach to analog to digital conversion used on *Mariner C* (Ref. 15). It permits pulse signals to be coupled from the instrument electronics into the data system, avoiding analog ground-loop problems.

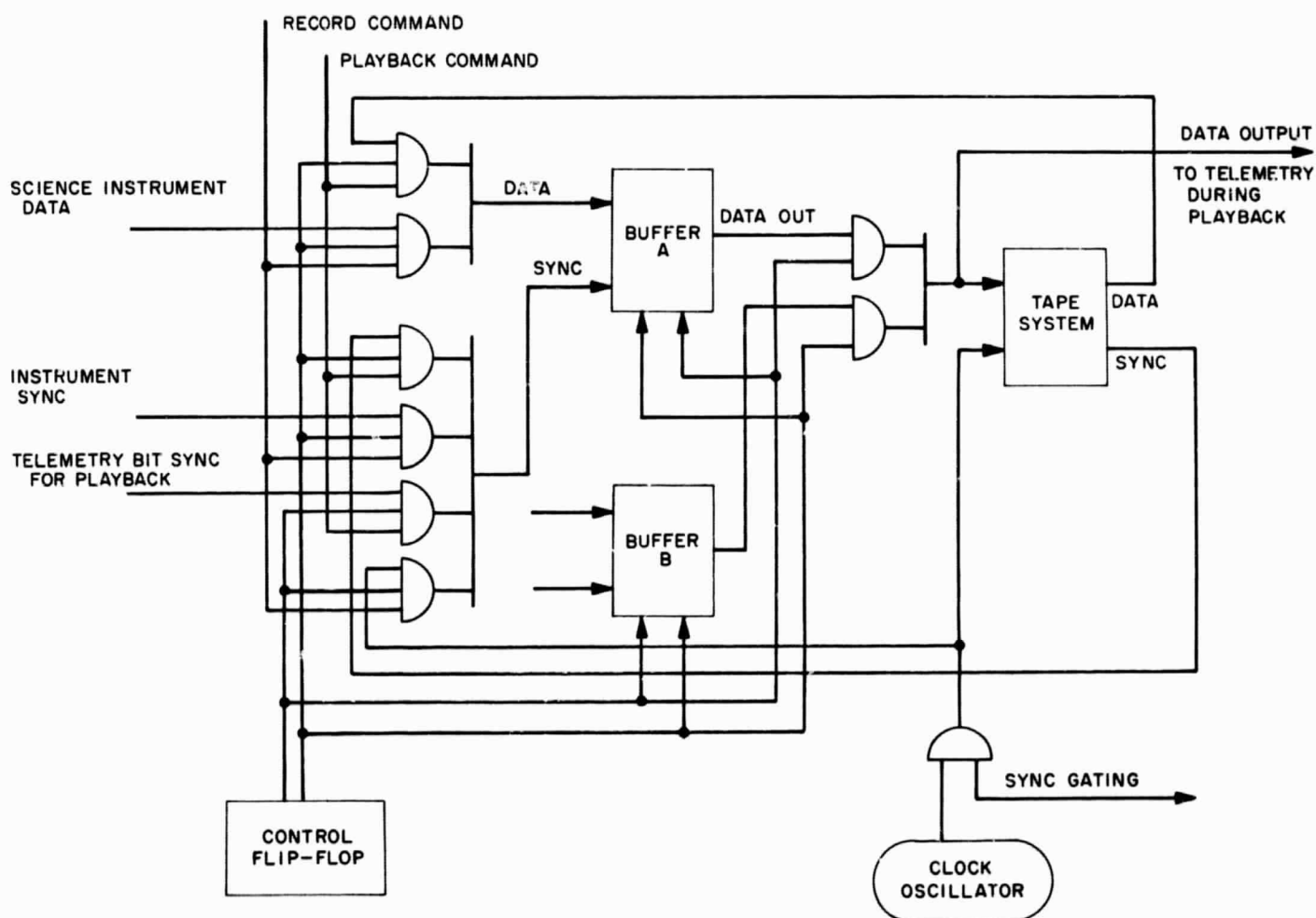


Fig. 16. Dual use of buffers for record and playback



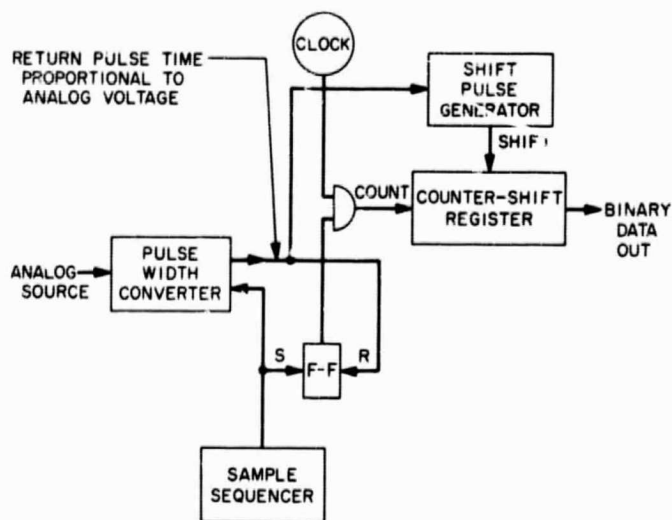


Fig. 17. Data conditioning system

## 5. System Characteristics

Sufficient developmental work has been completed to demonstrate the feasibility of a system having the following characteristics:

Data may be accepted from several different data sources for independent recording on separate data tracks. The record tape velocity is determined by the instrument which produces the highest data rate (averaged over one buffer load). More than one parallel track may be used if this rate is twice or more higher than the next highest rate instrument.

Lower data rates from other instruments are accommodated by allowing proportionately larger gaps between data blocks. Quiescent intervals in the instrument operation are accompanied by data gaps on the tape. All gaps are removed by buffered playback.

Data identification is inserted whenever required, and nonsynchronous operation is accommodated by accumulating the data into blocks by means of buffers and by allowing the instrument to command the transposition of the buffer from the read-in to the read-out (recording) mode.

The playback of recorded data may be interrupted for the playback of real-time spacecraft performance data or science data at any time simply by interrupting the telemetry bit sync to the system. However, one would

prefer to wait until the end of the current data block so that all data will be identified in whole blocks.

The data tracks (representing the various instruments) may be played back in any order. A control system which could repeat the playback of some of the tracks on a priority basis would not be difficult to mechanize.

The playback of data will be automatically accomplished at any telemetry bit sync rate supplied, up to the basic playback speed of the recorder (about 2 kc in the system constructed). This characteristic is particularly useful in spacecraft system testing. A two-week playback at 8½ bits/sec may be compressed to less than 2 hr.

## 6. Construction Progress

A breadboard system has been constructed. The tape transport used is a reel-to-reel recorder employing the isoelastic drive principle. This transport has been previously reported in Ref. 16. Fig. 7 in Ref. 16 (p. 225) is a photograph of the transport used. It weighs 2 lb, 10 oz (without electronics), contains 500 ft of 0.25-in. tape, and has passed *Mariner C*-level type-approval tests. Automatic rewind has been incorporated in the control system. Tape speed may be varied by a factor of 8:1 by changing the motor drive frequency. Speeds used in tests are 15 and 15/8 ips for record, and 15/8 ips for playback. At this playback speed, the data gap required for playback start-stop is less than 0.1 in. Recording at 1000 bits/in., a data block of 1300 bits is 1.3 in. Record and playback electronics provide for recording in NRZ format. Only one track has been mechanized, although the head is a 4-track unit. A gap detector and motor control circuits are included. The system has been tested with the dual use of a single pair of buffers as shown in Fig. 16.

The data source is a simple pattern generator. Provision was made for reducing the block length well below the 1320-bit maximum for testing purposes. In this manner, the transport may be caused to go through a more frequent succession of stop-starts; a procedure useful for accelerated life testing.

The electrical transposition of the buffers is commanded by counting the number of data words or bits with a binary counter. The buffers are currently being modified so as to bring out an indication of the buffer-full condition. It is intended that this condition be able to command the buffer transposition.

## 7. Future Work

Development is recommended in the following areas in approximately the order indicated:

- (1) A track switching system. Common playback data circuits will be used to the extent possible.
- (2) A playback sync circuit which will provide sync over a small interval of tape data dropout.
- (3) A data identification system which can read pseudo-noise and/or spacecraft time into any data channel upon demand.
- (4) The packaging of a flight-worthy prototype system.
- (5) De-skewing circuits for the simultaneous playback of several tracks, for use in instances where parallel recording is necessary in order to meet a high record data rate. Otherwise, the playback of one track at a time requires the use of a ground system which can reorder the data bits. This situation is not necessarily serious, but warrants investigation.
- (6) Provision for variable block length.
- (7) The ability to search for data blocks by means of addressing.

## References

1. Cronemyer, D. C., "Hall and Drift Mobility in High-Resistivity Single-Crystal Silicon," *Physical Review*, Vol. 103, No. 2, p. 522, 1957.
2. Conwell, E. M., "Properties of Silicon and Germanium," *Proceedings of the IRE*, Vol. 40, Part II, p. 1327, November 1952.
3. Prince, M. B., "Drift Mobilities in Semiconductors, II, Silicon" *Physical Review*, Vol. 93, No. 6, p. 1204, March 1954.
4. Backenstoss, G., "Conductivity Mobilities of Electrons and Holes in Heavily Doped Silicon," *Physical Review*, Vol. 108, No. 6, p. 1416, December 1957.
5. Vairilvo, V. S., Smirnav, L. S., and Patskevitch, V. M., "Soviet Physics, Solid State," Vol. 1, 1960, pp. 1344.
6. Hall, L. H., Bardeen, J., and Blott, F. J., "Infrared Absorption Spectrum of Germanium," *Physical Review*, Vol. 95, No. 559, p. 559, 1954.
7. Fan, H. Y., Shepherd, M. L., and Spitzer, W., "Infrared Absorption and Energy-Band Structure of Germanium and Silicon," in *Proceedings of the Photoconductivity Conference*, p. 184, John Wiley and Sons, Inc., 1956.
8. Braunstein, R., Moore, A. R., and Herman, F., "Intrinsic Optical Absorption in Germanium-Silicon Alloys," *Physical Review*, Vol. 109, No. 3, p. 695, 1958.
9. Gabor, A., "High Density Recording on Magnetic Tape," *Electronics*, October 16, 1959, pp. 72-75.
10. Clement, W. G., "Magnetic Tape Study," SPS 37-22, Vol. IV, p. 227, Jet Propulsion Laboratory, Pasadena, Calif.
11. Eldridge, D. F., "A Special Application of Information Theory to Recording Systems," *IEEE Transactions on Audio*, Vol. AU-11, No. 1, pp. 3-6.

## References (Cont'd)

12. Eldridge, D. F., and Baaba, A., "The Effects of Track Width in Magnetic Recording," *IRE Transactions on Audio*, Vol. AU-9, No. 1, pp. 10-15.
13. Miyata, J. J., and Hartell, R. R., "The Recording and Reproduction of Signals on Magnetic Medium Using Saturation Type Recording," *IRE Transactions on Electronic Computers*, June 1959, pp. 159-169.
14. "Mariner C Scientific Instrumentation," SPS 37-24, Vol. VI, p. 83, Jet Propulsion Laboratory, Pasadena, California, December 31, 1963.
15. "Mariner C Scientific Instrumentation," SPS 37-24, Vol. VI, p. 84, Jet Propulsion Laboratory, Pasadena, California, December 31, 1963.
16. Storer, W., "A Family of Tape Recorder-Reproducer Mechanisms Employing a Novel Drive Technique," SPS 37-25, Vol. IV, p. 217, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.

## XIV. Chemistry

### A. Effects of Ultrahigh Vacuum on Three Types of Microorganisms

P. J. Geiger, F. A. Morelli, and H. P. Conrow

Interest in the effects of ultrahigh vacuum on microorganisms has stemmed directly from implications to space exploration in recent years. The vacuum of interplanetary space is one of the important parameters to be considered in designing spacecraft and materials for spacecraft. In addition, it is vital to know whether extreme vacuums will be important microbiologically where travel of microorganisms from Earth to another planet might be detrimental to studies of indigenous life. Experience with culture collections seems to show that at least a nominal vacuum ( $150 \mu$ ) is necessary for the best preservation (Ref. 1). Exposure to an ultrahigh vacuum of at least  $10^{-10}$  torr, however, might produce different results.

Prince and Bakanauskas (Ref. 2) and Prince (Ref. 3) described experiments that indicated freeze-dried spores of *Aspergillus niger*, *A. flavus*, *Bacillus subtilis* var. *niger* and *B. mycoides* retained some degree of viability for

31.5 days at ambient temperature in vacuums ranging from  $10^{-5}$  to  $5 \times 10^{-7}$  torr.

Portner, et al. (Ref. 4) exposed *B. subtilis* var. *niger*, *A. fumigatus*, and *Mycobacterium smegmatis* to ultrahigh vacuum for 5 days at ambient temperature. At 5 days the system indicated  $3.6 \times 10^{-10}$  torr. Recovery of *B. subtilis* var. *niger* spores and *A. fumigatus* spores was not significantly different from that of control samples assayed at the start of the experiment after drying over calcium sulfate. *M. smegmatis* recovery, however, was only 33% of that of the control sample. It was concluded that these microorganisms would survive the vacuum of outer space.

Brueschke, Suess, and Willard (Ref. 5) carried out a qualitative study of the survival of spores of *B. subtilis*, *A. niger*, *A. terreus*, and *Penicillium citrinum* at very low pressures. Water suspensions of the spores of these organisms were mixed and dried directly in tubes attached to a manifold sealed to a high-vacuum system. After 10 days the pressure had reached  $8 \times 10^{-8}$  torr, and a tube was removed at that time. *P. citrinum* was not recovered from this tube, but all the others survived. A second tube was sealed off after 30 days of pumping, at which time the

system had dropped to  $1.2 \times 10^{-8}$  torr. No microorganisms were recovered at this time or later. It was surmised that the organisms which were used would be killed at  $1.2 \times 10^{-8}$  torr within 30 days.

Morelli, Fehlner, and Stenbridge (Refs. 6 and 7), on the other hand, showed that about one-half of the *B. subtilis* var. *niger* spores survived 35 days of exposure at  $10^{-8}$  torr at room temperature. Control preparations in air, nitrogen, and a vacuum of  $3 \times 10^{-1}$  torr gave comparable results; spores in argon and carbon dioxide atmospheres yielded somewhat lower counts.

Davis, Silverman, and Keller (Ref. 8) in a series of studies have sought to determine the combined effects of ultrahigh vacuum and temperature on the viability of some spores and soil organisms. Considerably fewer spores of *B. stearothermophilus*, *B. megaterium*, and *Clostridium sporogenes* were recovered than were spores of *B. subtilis* var. *niger* and *A. niger* after 4 to 5 days at 53 and 60°C and  $10^{-8}$  to  $5 \times 10^{-9}$  torr. There were no significant differences in the recoveries of these five organisms after exposure to 25 and -190°C in vacuum. *B. subtilis* var. *niger* spores apparently were killed after retention at 90°C and  $10^{-7}$  to  $6 \times 10^{-9}$  torr for several days, but  $10^4$  out of an initial  $10^6$  spores survived 90°C and atmospheric pressure for the same period. Molds and actinomycetes in soil were particularly resistant up to 69°C in vacuum, and actinomycetes were recovered even after exposure to 120°C.

The work reported here, which considers the effects of ultrahigh vacuum at 25°C on three widely differing types of microorganisms, was carried out in an attempt to expose these microorganisms to a better vacuum for a longer period than had thus far been achieved. The results are therefore made more meaningful in relation to space exploration.

## 1. Experimental

**a. Microorganisms.** *B. subtilis* var. *niger* spores were prepared according to the method of Davis, Silverman, and Keller (Ref. 8). A mature culture of *Mycobacterium phlei* was harvested from glycerol agar (ATCC), and cell clumps were broken up by homogenization in a solution (0.4 g/liter) of Tween 80 (Difco Laboratories, Inc., Detroit, Michigan) for 5 min in a high-speed homogenizer (Virtis, Gardiner, New York). Subsequent microscopic examination revealed that the cells were well separated and essentially free from clumping. *Aspergillus niger* spores were harvested from Trypticase Soy Agar (BBL),

washed with water, and placed in water suspension just before use.

Water suspensions of all three organisms were placed in aluminum combustion boats (Coleman Instruments, Inc., Maywood, Illinois) and were dried under vacuum over calcium sulfate the day before use. The numbers of viable organisms were counted by the standard plate count method both at the time of drying and subsequently in order to assay recovery during the actual experiment.

The three organisms used were selected because they: (1) are highly resistant to drying, (2) represent three greatly different genera, and (3) are widely distributed in nature.

**b. Ultrahigh vacuum system.** To achieve the best ultrahigh vacuum obtainable from a glass (Corning Glass Works 7740) system, baking at high temperature is necessary. The system we employed in this work is illustrated in Fig. 1. It was designed so that baking could be accomplished with heating tapes during pumping and with the samples of microorganisms in place. All of the system beyond the trap in the diffusion pump was baked for 1 wk at about 300°C.

The total volume of the system was small, and the removable sample tubes permitted us to protect the microorganisms by cooling with damp wicks during baking. Also, duplicate samples could be taken off for assay of viability at discrete intervals.

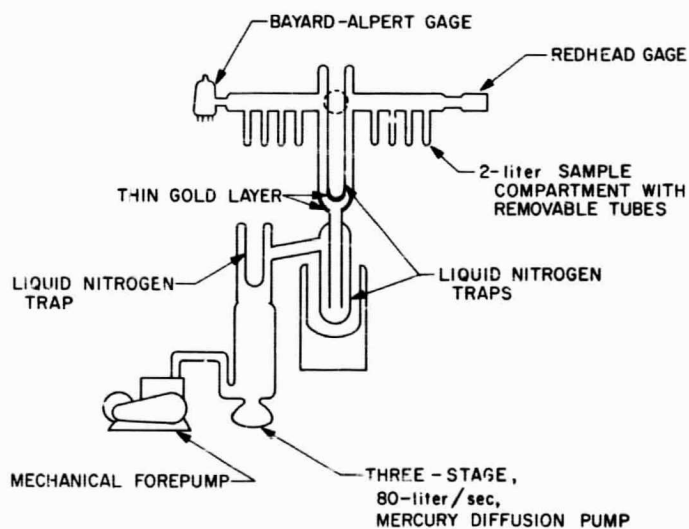


Fig. 1. System for exposing microorganisms to ultrahigh vacuum

In order to assure the best possible exposure to ultrahigh vacuum, the aluminum boats with their samples were moved into the horizontal arms of the chamber after the system had been baked. This was accomplished by attaching bits of iron wire to each boat and then moving the boats magnetically. The boats were then moved back into individual vertical tubes when the tubes were to be fused off for viability assays.

Two ultrahigh vacuum gages, a Bayard-Alpert (Vacuum Electronics Corporation, Long Island, New York) and a Redhead (National Research Corporation, Cambridge, Massachusetts), were used for obtaining an accurate measure of the vacuum as well as for safety. Advantage was taken of the high pumping speed inherent in the Redhead gage, as well as its requirement for reading pressures accurately in the range of  $10^{-10}$  torr and below.

A thin layer of gold was strategically placed in the system to keep any mercury vapor from reaching the samples during baking and throughout the remainder of the experiment.

**c. Control samples.** These samples were prepared as described, but were exposed to a number of different conditions for the duration of the experiment: air at 1 atm, nitrogen at 1 atm, air at 50% relative humidity, and a vacuum of  $5 \times 10^{-3}$  torr. All control samples except those at 50% relative humidity were kept dry over calcium sulfate.

## 2. Results

The performance of the system and the duration of maintenance of ultrahigh vacuum at about  $25^\circ\text{C}$  can be seen in Fig. 2. After baking and on approximately the

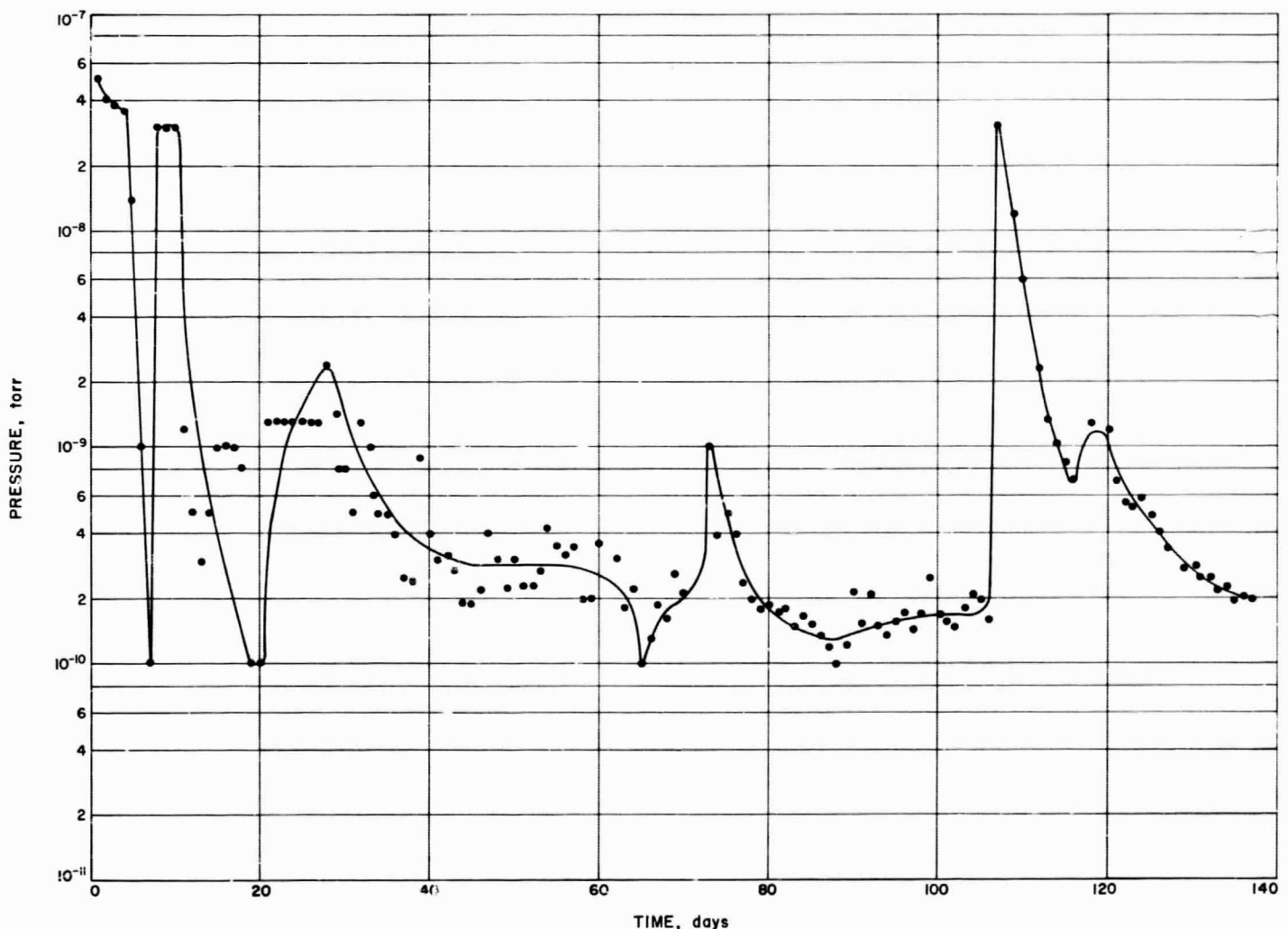


Fig. 2. Attainment and maintenance of ultrahigh vacuum at  $25^\circ\text{C}$



tenth day, the vacuum reached well into the  $10^{-9}$ -torr range. Each point on the curve is an average of two readings taken twice a day for the 137-day period of operation. The occasional rises into the  $10^{-8}$ -, and on two occasions into the  $10^{-7}$ -, torr range represent malfunctions of the system which, except for the last one, were easily remedied. The curve illustrates the successful average operation of the system in the range of  $10^{-10}$  to  $4 \times 10^{-10}$  torr for a longer period than has heretofore been achieved in such experiments.

Fig. 3 shows the apparatus as it appeared in actual operation toward the end of the experiment. Only the top of the diffusion pump can be seen behind the Dewar at

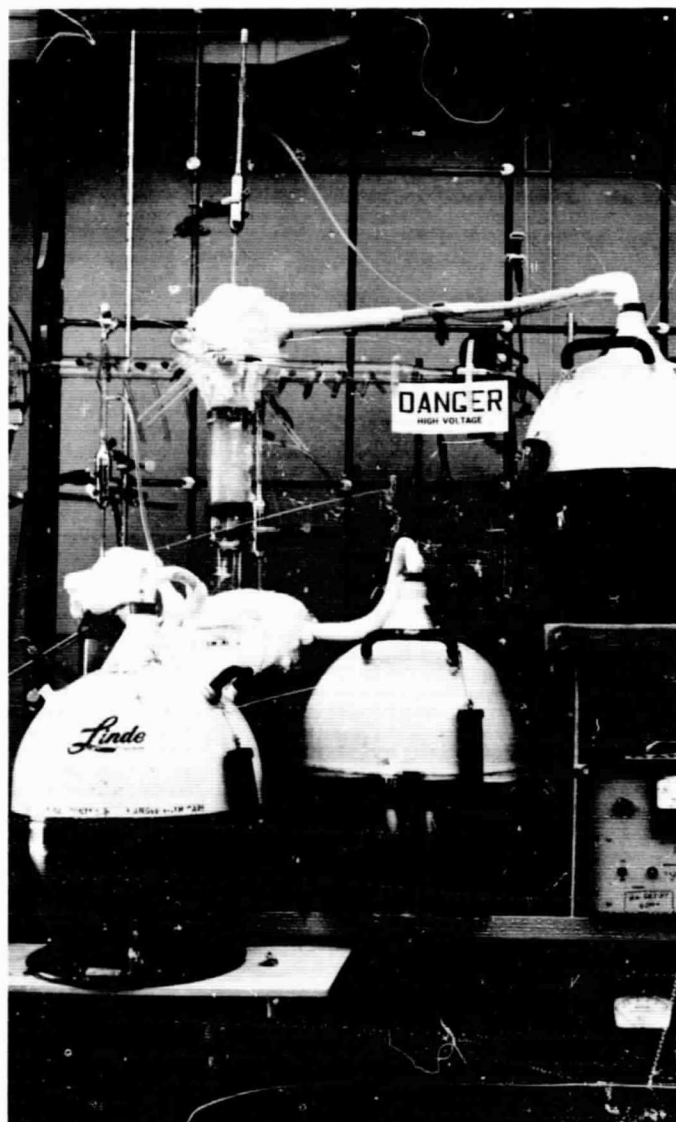


Fig. 3. Ultrahigh vacuum system during the experiment, with numerous tubes removed for viability tests

the left, but the ultrahigh vacuum chamber and removable sample tubes are clearly shown.

In testing the viability of the microorganisms, sample-containing boats were removed in duplicate. Each of these was then plated in duplicate, and, after a sufficient growth period to ensure the appearance of colonies of any viable cells, colony counts were made by the standard plate count method.

Fig. 4 illustrates the fate of cells exposed to dry air. Of an initial  $1.3 \times 10^6$  spores of *B. subtilis* var. *niger*, only two colonies were recovered after the first 39 days, at which time the first tests of viability were performed. At each time of sampling thereafter, only an occasional colony or two of *B. subtilis* appeared. *M. phlei* fared even worse; no colonies from an initial  $10^5$  cells were recovered at 39 days or subsequently. On the other hand,

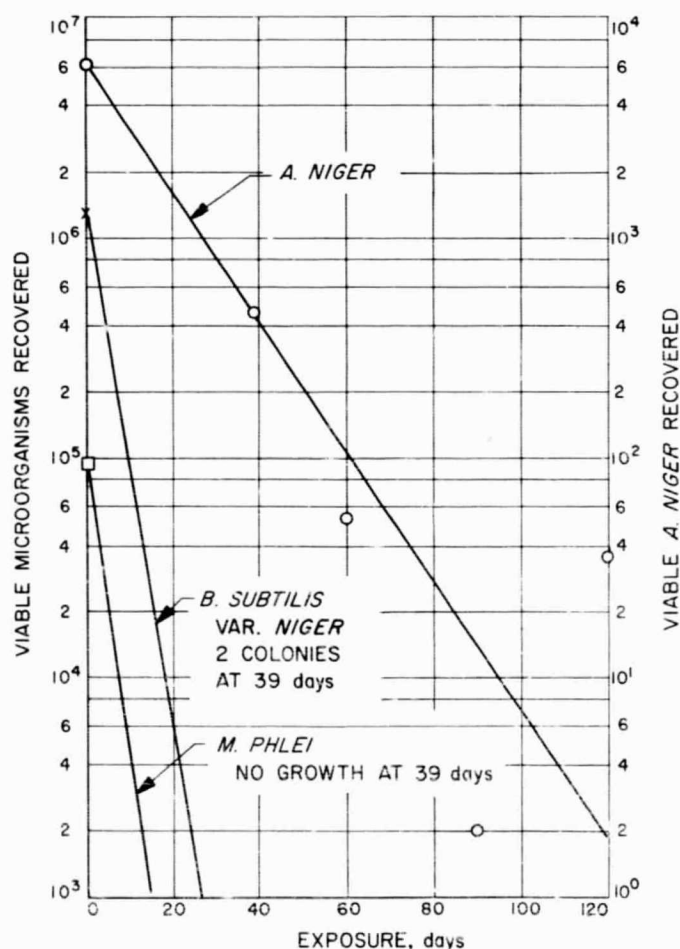


Fig. 4. Effect of dry air at 1 atm on three types of microorganisms

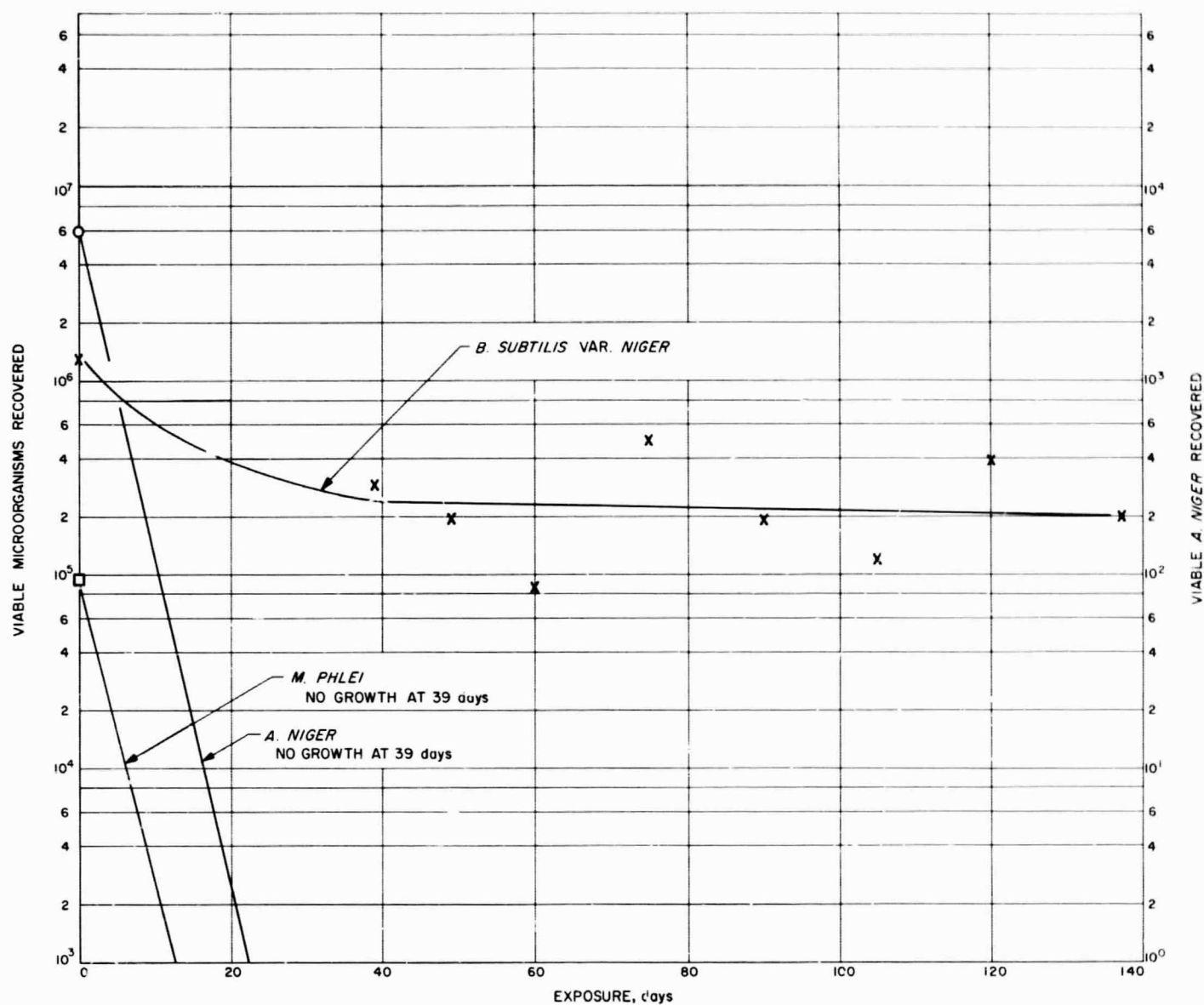


Fig. 5. Effect of air at 50% relative humidity on three types of microorganisms

*A. niger* retained viability throughout the experiment. Of an initial  $6 \times 10^3$  spores (right ordinate in Figs. 4 through 8), fewer were recovered with each plating, and the general trend was toward death after a few months of exposure.

In air at 50% relative humidity, *B. subtilis* spores retained viability after an initial drop in numbers, probably due to drying from the washed suspension. Fig. 5 illustrates this ability to survive many months and perhaps indefinitely under these conditions. The curves for *M. phlei* and *A. niger*, however, indicate an apparent rapid loss of viability somewhere between 0 and 39 days.

From this control, neither of the organisms was subsequently recovered during the experiment.

The effect of dry nitrogen on the three organisms is clearly shown in Fig. 6. Again, after an initial decrease, the recoverable numbers of *B. subtilis* remained almost constant, and viability apparently was preserved for long periods. Viability of *A. niger* was also improved; only a slow decrease in numbers was evident over the 4-mo period of observation. *M. phlei* again suffered a complete loss of viability; one colony appeared in the first sample, but none were recovered in any of the other samples assayed thereafter.



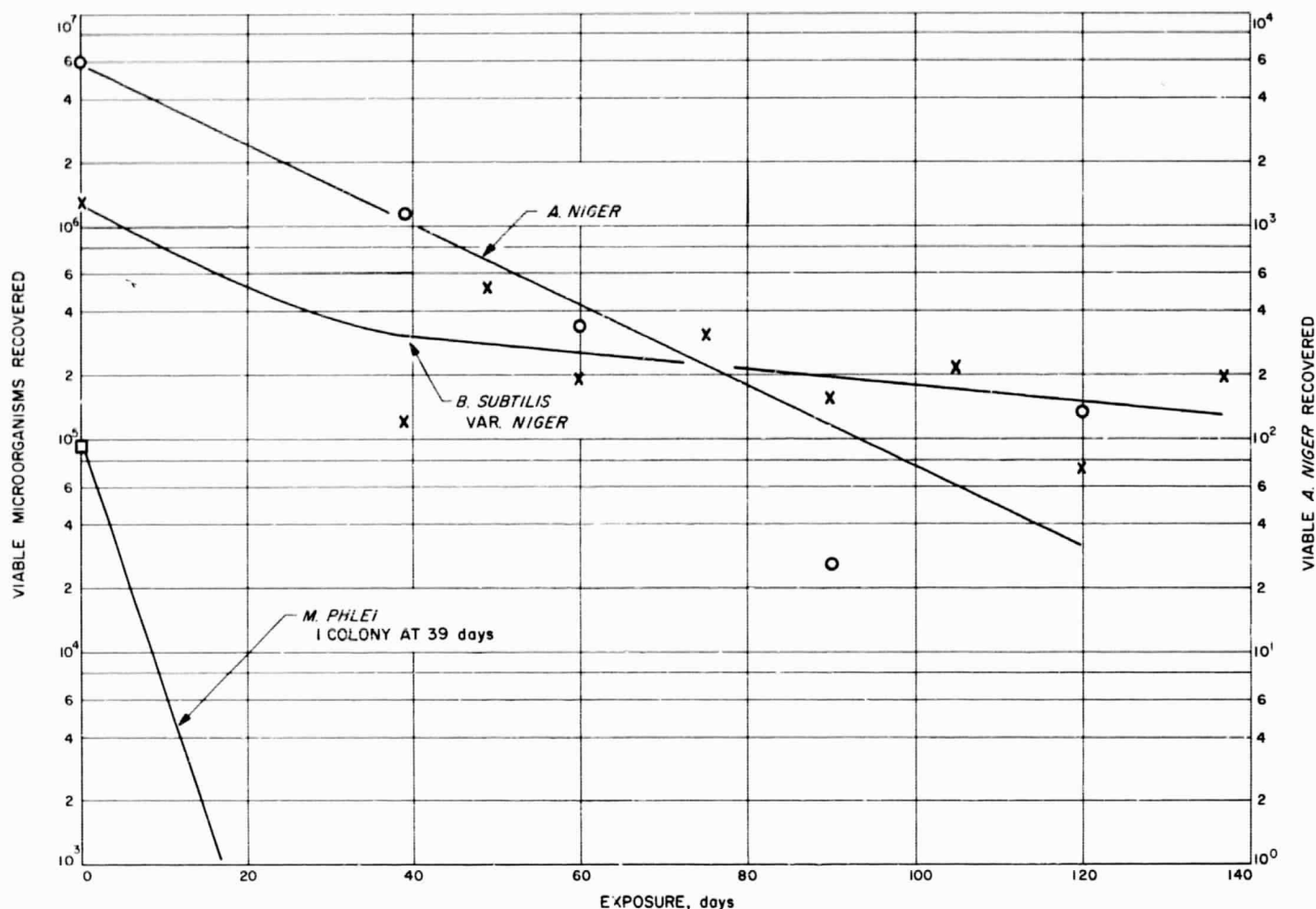


Fig. 6. Effect of dry nitrogen at 1 atm on three types of microorganisms

In a moderate vacuum of  $5 \times 10^{-3}$  torr, all organisms were preserved (Fig. 7). *B. subtilis* spores were evidently quite stable, while *A. niger* spores showed a slow trend downward just as they did when exposed to dry nitrogen. A startling difference is apparent in the behavior of *M. phlei*. Although there is some scatter in these data when compared to those from conditions previously described, the trend toward increased viability is readily apparent. These organisms will be preserved perhaps almost indefinitely by a moderate vacuum.

In ultrahigh vacuum (Fig. 8), the viability of all three organisms is stabilized. Clearly the curves could be extended and would indicate viability for many months. Although there appears to be more scatter in the data for *B. subtilis* than was previously observed, nevertheless the resultant trend seems obvious: ultrahigh vacuum does

not kill the microorganisms selected for the experiment, but, on the contrary, tends to preserve them, presumably for an indefinite period.

### 3. Discussion

In general, the present work is in accord with the results reported by most other authors working in this field. Microorganisms apparently retain considerable viability when exposed to ultrahigh vacuum. Such results have usually referred only to experiments of relatively short duration. In addition, the reported vacuums have been 1 to 2 orders of magnitude poorer than that achieved here. Further, we wished to observe in our experiment not just the survival of microorganisms at the end of a specified period, but the trends in survival over a long period of exposure to ultrahigh vacuum. The implica-

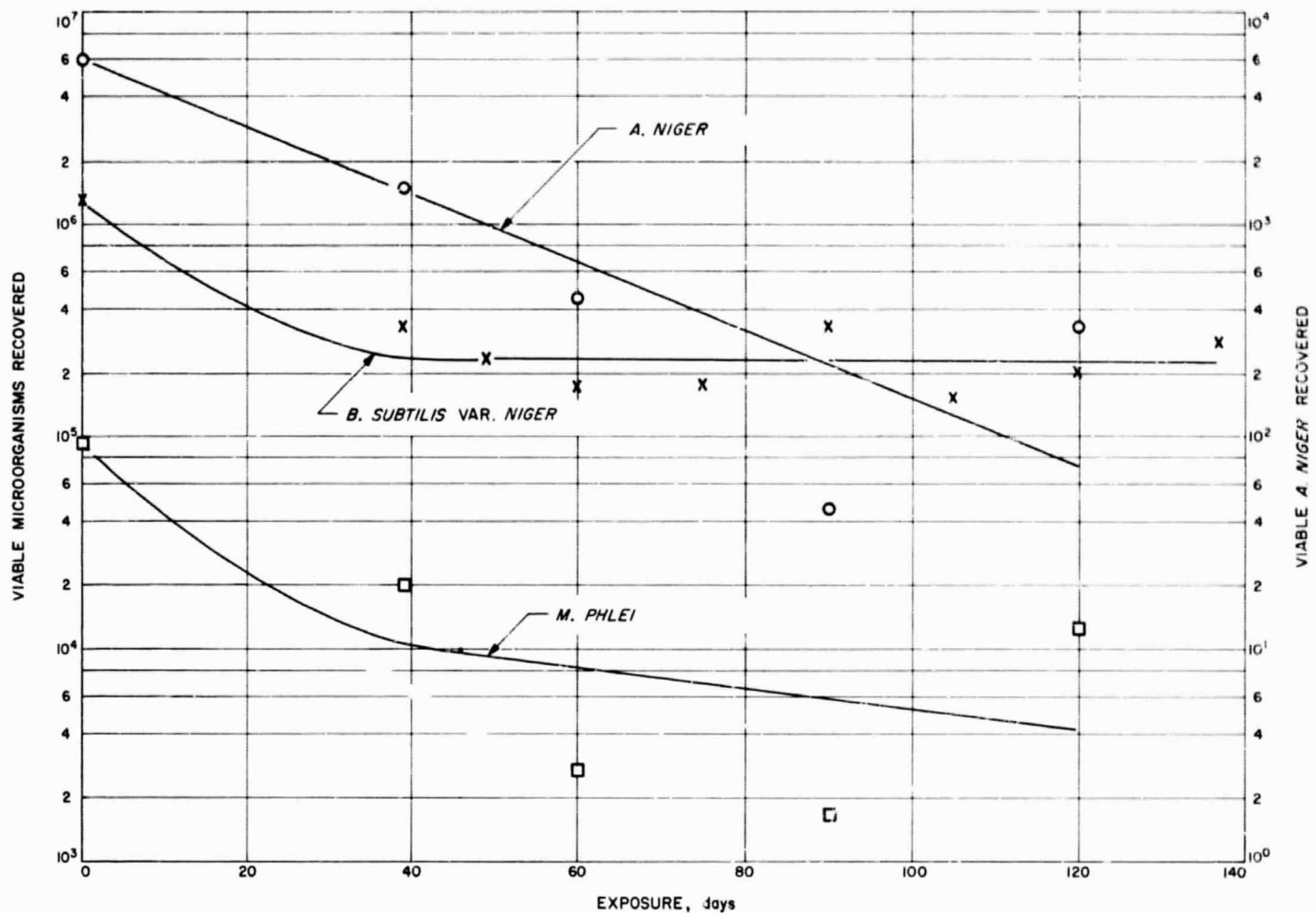


Fig. 7. Effect of vacuum ( $5 \times 10^{-3}$  torr) on three types of microorganisms

tions of these results to space exploration seem clear regarding the attendant problems of contamination and sterilization.

The results reported here are also in general agreement with experiences in freeze-drying for the maintenance of culture collections. For the most part, a wide range of microorganisms shows good survivability when freeze-dried and stored under vacuum (Ref. 1). Preservation is frequently best when some colloidal organic substances (serum albumin) or even certain organic molecules (glutamate or sucrose) are present. Viability usually decreases faster if samples are stored in air rather than in vacuum, and it decreases at some intermediate rate in nitrogen. In addition, viability is often best preserved at low temperatures, i.e., 0 to 4°C. Further, it is frequently observed that an initial drop occurs in viable numbers of organisms undergoing freeze-drying; then an apparent stabilization

takes place so that further losses may be small. We have also observed this process in the present work.

It is important to note that an initial drop in numbers of microorganisms may be less significant than the stability of the remaining viable numbers. A survivability of 0.1%, for instance, may look small; however, consider that in an initial small sample, which can contain  $10^{11}$  organisms,  $10^8$  would survive.

As a result of the present work, it would seem that the proper sterilization of spacecraft should be re-emphasized if we are to avoid contamination of the planets by terrestrial organisms. On the other hand, if certain organisms can exist in ultrahigh vacuum and, as shown in Ref. 8, are uninjured at  $-190^\circ\text{C}$ , it is not too difficult to imagine that cross contamination of the planets of the solar system might already have occurred.

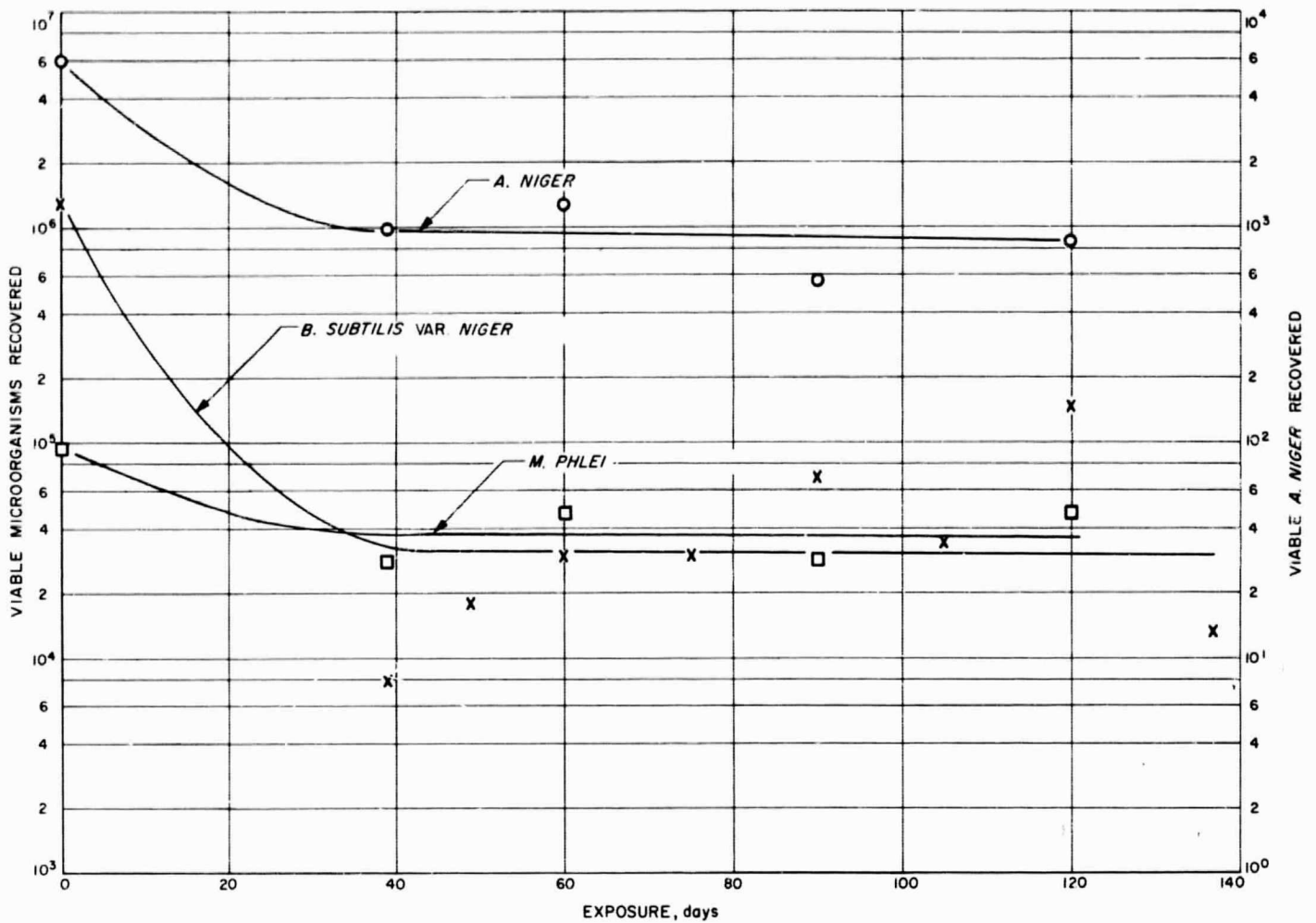


Fig. 8. Effect of ultrahigh vacuum ( $\sim 10^{-10}$  torr) on three types of microorganisms

## B. Fluorometric Measurements of Growth: III. The Interference of Soil with the Fluorescence of Proteins in a Neutral Aqueous Solution

J. H. Rho

The characteristic fluorescence of proteins is sufficiently sensitive and specific that it has been possible to utilize it in developing a quantitative assay technique

(Ref. 9). In previous work, both qualitative and quantitative assays of several different proteins and microorganisms were made, and the increase in fluorescence of consecutive samples of actively growing *Escherichia coli B* (*E. coli B*) cultures was used as a measure of growth for this organism (Ref. 10).

A method involving the use of native protein fluorescence in intact microorganisms is very attractive because of its possible direct application to a quantitative assay of microorganisms in their natural environment. For a direct detection or a quantitative analysis of microorganisms in any natural soil, the effect of soil particle suspensions on protein fluorescence is of utmost interest. The presence in a soil suspension of materials which absorb or scatter a significant proportion of the excitation or fluorescence radiation will diminish the observed fluorescence. Any other solutes present in the soil suspension may interact with the fluorescent molecule in such a

manner as to lower the efficiency and/or lifetime of the fluorescence process, thus limiting the sensitivity or reproducibility of an assay procedure.

In the present study, the effect of soil suspensions on the fluorescence of protein in neutral aqueous solutions has been studied by adding various concentrations of soils to two different concentrations of dilute protein solutions, as well as a population of *E. coli B* culture. Increase in the fluorescence of proteins of consecutive samples taken from the soil-suspension medium at hourly intervals during the growth period was used as a measure of growth for this organism. It was then possible to measure the growth of the organisms in the soil medium specifically by measuring the increase of protoplasmic proteins.

## 1. Experimental

The source of the microorganisms, *E. coli B*, and the growth medium used for the organism were described previously (Ref. 10). The instrumental specifications of the spectrophotofluorometer were also described in Ref. 10. Bovine serum albumin was obtained from the Sigma Chemical Company, St. Louis, Missouri. Other chemicals were reagent grade. The soil samples used were sandy soils collected from the surface ( $\frac{1}{8}$  in.) of the Colorado Desert, Riverside County, Thermal, California.

## 2. Results

**a. The interference of soil with the fluorescence of bovine serum albumin solutions.** The fluorescence-emission spectrum of  $10^{-4}\%$  bovine serum albumin solution in 0.05M phosphate buffer was determined by excitation with radiation of 280-m $\mu$  wavelength. A typical fluorescence spectrum of this solution is shown in Fig. 9(a), which displays the fluorescence spectrum characteristic of the amino acid tryptophan, with a maximum around 350 m $\mu$ . The addition of 1 mg of ignited soil (800°C for 24 hr) to 1 ml of this protein solution caused considerable distortion of the spectrum, as shown in Fig. 9(b). As the concentration of soil was increased, the distortion effect became more severe (Fig. 9c and 9d). The original fluorescence spectrum of protein was recovered after centrifugation at about 500 g for 10 min (Fig. 9e).

On the other hand, the addition of an equivalent amount of ignited soil to a slightly higher concentration of protein produced a less severe distortion effect on the original protein fluorescence spectrum, as shown in Fig. 10 (a through d). The addition of 1 mg of soil to a 1-ml volume

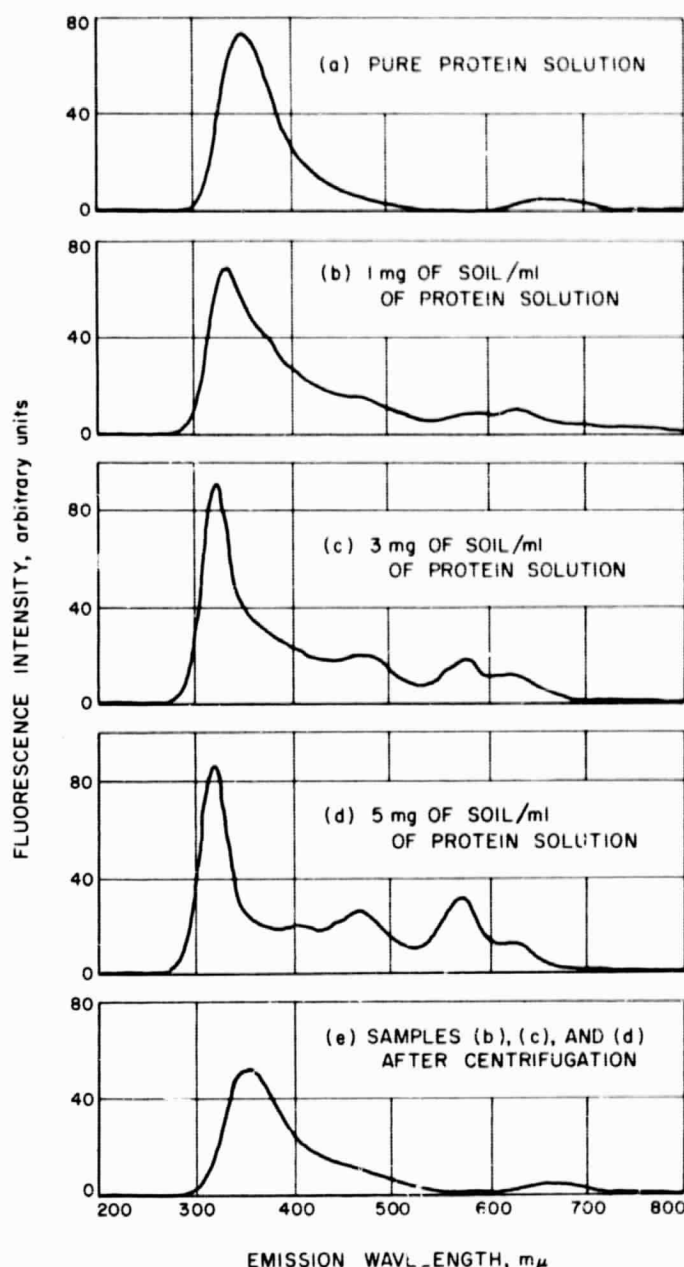


Fig. 9. Effect of ignited soil on the fluorescence of  $10^{-4}\%$  bovine serum albumin

of  $10^{-3}\%$  bovine serum albumin solution did not appreciably change the original fluorescence spectrum with a maximum around 350 m $\mu$ , although there is a shoulder around 320 m $\mu$  (an indication of light scattering from colloidal particles in the soil). As more soil was added to this solution, the Tyndall effect became more prominent and the spectrum approached that of ignited soil alone (Fig. 10e). The fluorescence spectrum of protein, however, can clearly be observed with the maximum at 350 m $\mu$ , even with the addition of 5 mg of soil/ml of

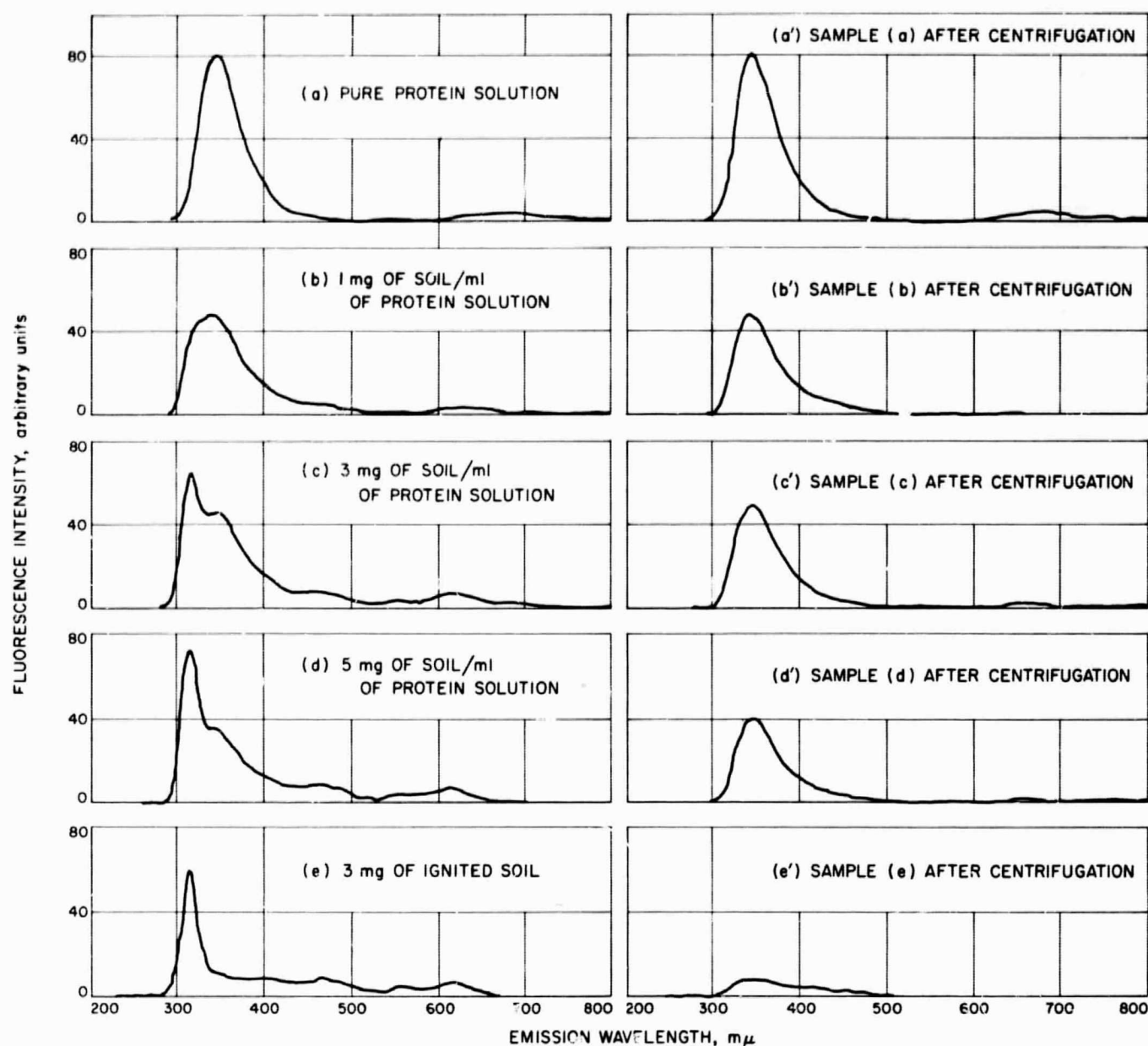


Fig. 10. Effect of ignited soil on the fluorescence of  $10^{-3}\%$  bovine serum albumin

protein solution. When these samples were centrifuged at 500 g for 10 min (Fig. 10a' through 10d'), the original fluorescence spectrum of protein was restored, although the fluorescence intensities were slightly reduced. It was also noted that not many light-scattering particles were left in the supernatant solution from ignited soil after centrifugation at 500 g for 10 min (Fig. 10e').

**b. The effect of soil on the fluorescence of *E. coli* B.** *E. coli* B cells were harvested from the exponential growth phase, washed with 0.05M phosphate buffer, pH 7.0, and then made up to a population of  $10^8$  cells/ml of the

same buffer for use in the following experiments. The fluorescence-emission spectrum of this population was determined by excitation with radiation of 280-mμ wavelength; a typical curve is shown in Fig. 11(a), which resembles the fluorescence spectrum of tryptophan-containing protein and agrees with the author's previous observations (Ref. 10). To the same *E. coli* B suspension, varying amounts of ignited soils were added and the tubes were shaken thoroughly. The fluorescence spectra of these samples are shown in Fig. 11(b through f). The addition of soil up to 1.5 mg/ml did not cause any significant distortion of the original fluorescence spectrum of

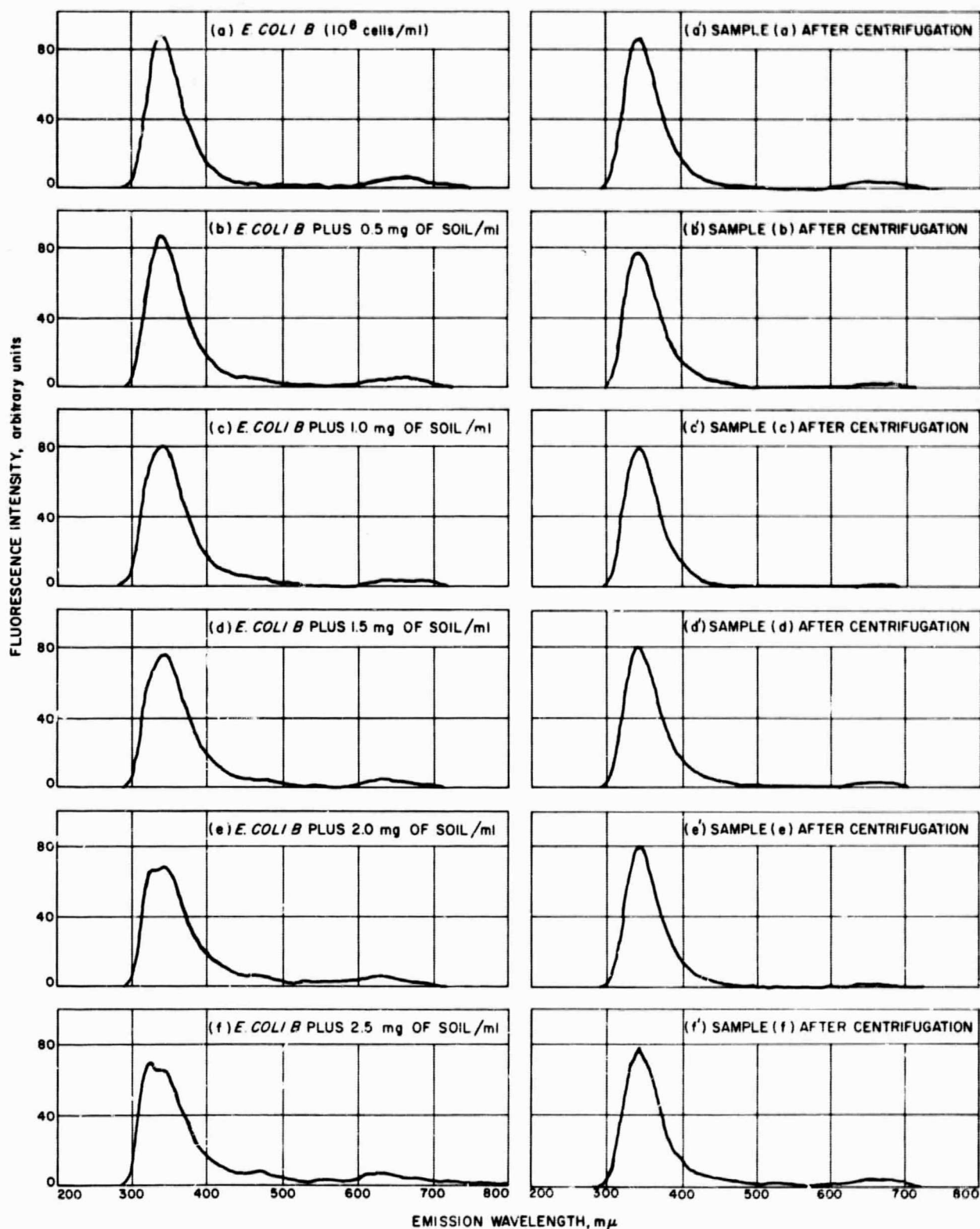


Fig. 11. Effect of ignited soil on the fluorescence of *E. coli B*



*E. coli B*, although the fluorescence intensities of these samples were reduced progressively as the concentration of added soil was increased. At a concentration of soil above 2 mg/ml, however, a definite soil-scattering shoulder began to appear around 320 m $\mu$ . Upon centrifugation at 100 g for 5 min (Fig. 11a' through 11f'), the original fluorescence spectra reappeared, as in experiments with 10<sup>-3</sup>% bovine serum albumin samples.

**c. Protein assay in natural soil samples.** Since the fluorometric assay of protein in a soil-suspension medium appeared feasible, a direct determination of protein and/or microorganisms in natural soil was attempted. First, 50 mg of sandy soil collected from the surface ( $\frac{1}{8}$  in.) of the Colorado Desert was suspended in 1 ml of glass-distilled water, and its fluorescence-emission spectrum was determined by excitation as before. The same sample was then subjected to a slight centrifugation at 50 g for 5 min. The supernatant was removed, and its fluorescence-emission spectrum was determined in the same manner. The supernatant of the latter sample was again used for a successive treatment, and the fluorescence spectrum was similarly determined. Spectra of the samples are shown in Fig. 12. As can be seen, the original sample shows very little, if any, distinct protein fluorescence, whereas soil scattering, which coincides with that of pure ignited soil, is quite prominent. When this sample was subjected to further centrifugation, however, a peak around 350 m $\mu$  was retained, whereas the light-scattering peak, which was originally prominent, diminished considerably in intensity as soil particles gradually precipitated from the samples.

**d. Determination of the growth of *E. coli B* in a soil-suspension medium.** *E. coli B* cells were grown in 100 ml of inorganic medium enriched with glucose as the sole organic carbon source. To each 250-ml flask containing this medium, various amounts of both ignited and unignited soils were added together with various sizes of inocula of *E. coli B*. The cultures were shaken in a water bath at 37°C. Aliquots of 2 ml of bacterial suspension were removed from the culture flasks at hourly intervals, and the fluorescence intensities of proteins in the medium were observed. With an inoculum of 10<sup>7</sup> cells/ml of medium to which 5 mg/ml of either ignited or unignited soil had been added, no sign of growth was detected up to 3 to 4 hr. Beyond this period, however, a definite shoulder appeared on the scattering peak at about 350 m $\mu$  and increased progressively as the period of growth was prolonged. With a larger inoculum of 10<sup>8</sup> cells/ml of medium to which 2 mg/ml of either ignited or unignited soil had been added, the fluorescence shoul-

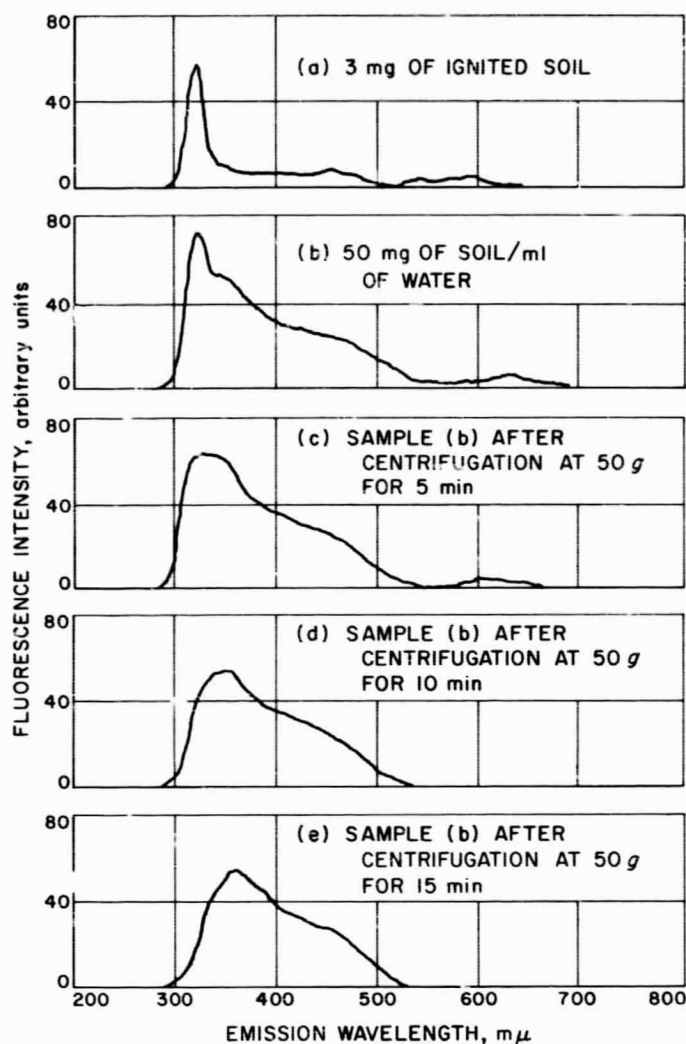


Fig. 12. Fluorescence spectra of proteins in sandy desert soil

der around 350 m $\mu$  was observed during the first 2-hr period of growth. A typical fluorescence spectrum of these cultures is shown in Fig. 13. In Fig. 13(a), the initial culture shows no fluorescence shoulder around 350 m $\mu$ , whereas the light-scattering peak caused by the soil particles is prominent. After 2 hr of incubation (Fig. 13b), however, a definite fluorescence shoulder appeared at 350 m $\mu$  which can plainly be differentiated from the control sample. In a 3½-hr period of growth, this fluorescence peak increased to about 150 times its initial height, and the light-scattering peak was completely masked (Fig. 13c). The appearance and behavior of this fluorescence maximum closely resemble those of the peaks due to protein and *E. coli B* bacteria in a soil-suspension medium. Again, for a given concentration of soil particles, a more prominent fluorescence peak is observed as the protein concentration is increased.

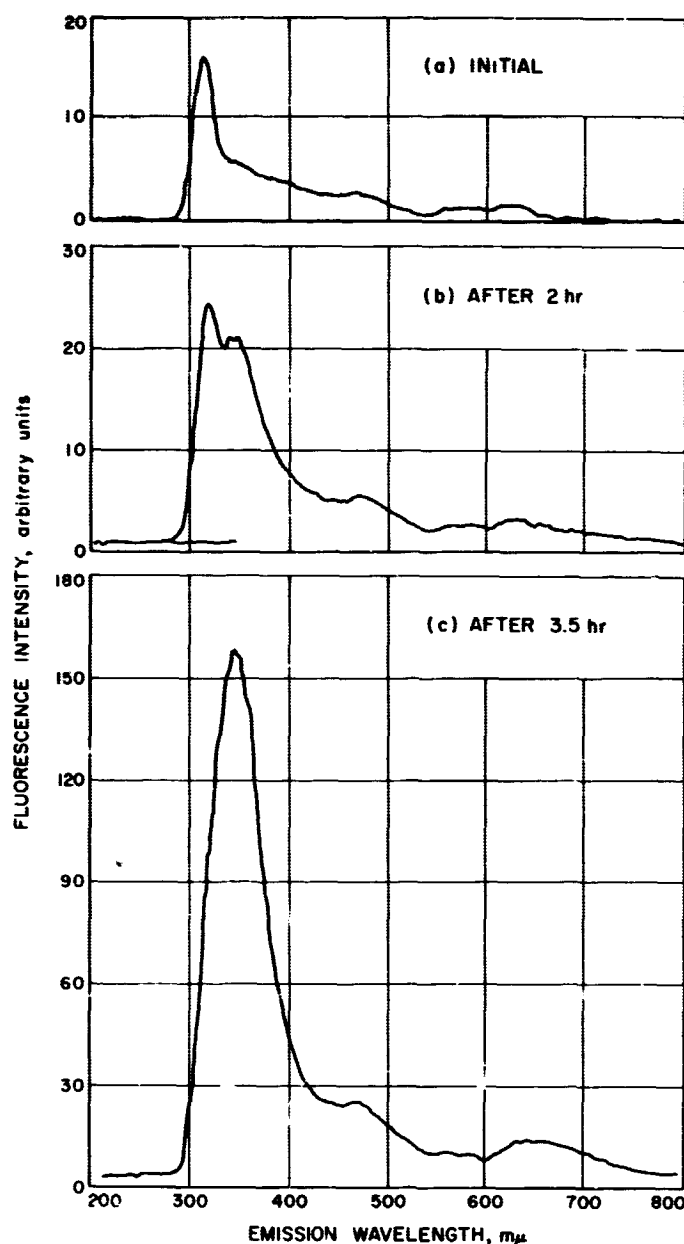


Fig. 13. Growth of *E. coli B* in a soil-suspension medium

When the fluorescence maxima of these cultures were plotted against their growth periods, characteristic growth curves were obtained. In Fig. 14, two such experiments are presented: (1) from a culture with an inoculum of  $10^7$  cells/ml of medium in which 2 mg/ml of ignited soil had been added; and (2) the other with an inoculum of the same size, but with the addition of 2 mg/ml of unignited soil. In the ignited soil-suspension medium, the fluorescence intensity of the *E. coli B* culture increased slowly for the first 2 hr and then increased exponentially during the following 2 hr. The rate of increase of the fluorescence intensity slowed down beyond a growth

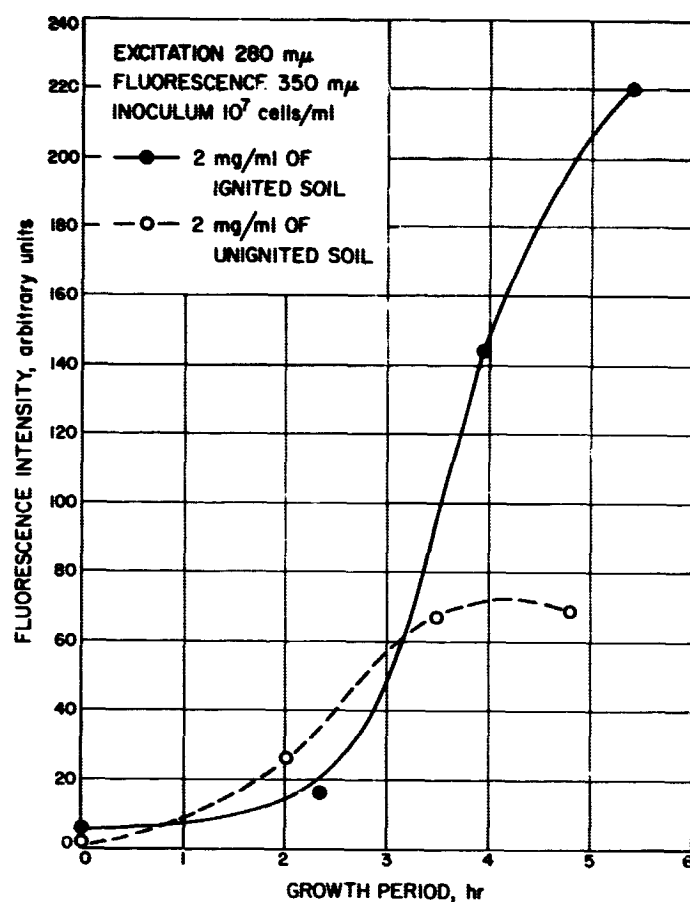


Fig. 14. Characteristic growth curves of *E. coli B* in a soil-suspension medium

period of 5 hr. In the second culture containing the unignited soil suspension, the maximum fluorescence intensity was reached much earlier and its height was much lower. The fluorescence intensity started to increase sooner with the unignited soil but stopped at 3½ hr, probably because of strong light scattering and absorption of both excitation and emission radiation by soil colloidal particles as well as bacteria. Thus, the sensitivity of fluorescence seems to be inhibited by the presence of more intact soil colloidal particles in the second culture, as compared to those in the first culture.

### 3. Discussion

The quenching of fluorescence may be due to many causes, both chemical and physical. Besides Rayleigh scattering of the solvent, the Tyndall scattering from colloidal particles and the scattering from surfaces of the container are most frequently the limiting factors in the sensitivity or reproducibility of this assay procedure. Scattered light not only distorts fluorescence spectra, but also decreases the measured fluorescence, since the scat-



tering surfaces deflect the fluorescent light from the detector as well as deflecting the exciting radiation.

When 1 mg of soil/ml of solution was added to  $10^{-4}\%$  bovine serum albumin solution, the fluorescence spectrum of protein was seriously distorted, and more or less complete distortion was observed with the addition of 3 mg of soil to the solution. Since a good recovery of the original spectrum was made by precipitation of the soil colloidal particles, most of the distortion and the diminution of the fluorescence spectrum may well be accounted for by light scattering from these colloidal particles. When the fluorophore itself is present in high concentration, as in the  $10^{-3}\%$  solution of protein but only 1 to 3 mg of soil suspension, the colloidal materials should, of course, be diluted so as to yield little distortion due to absorption or scattering. This proved to be the case in these experiments. As shown in Fig. 10(a through d), the same amount of soil seemed to cause less distortion and diminution of the original protein fluorescence spectrum in  $10^{-3}\%$  protein solution than in  $10^{-4}\%$  protein solution. A good recovery of the original spectrum after centrifugation again indicates that the interference of soil with protein fluorescence is probably due to light scattering from the soil particles themselves. Apparently there was very little chemical quenching, which is due to an interaction of the fluorescence molecules with solvent or with other solutes in such a soil suspension.

In most natural soil samples, the protein or microorganisms are present in trace quantities, whereas many other absorbing or scattering species are present in much larger amounts. Removal or dilution of such interfering substances is necessary in order to fluorometrically assay the compound of interest. In natural sandy desert soil, which was employed in this study, a light centrifugation was necessary for detection of the fluorescence from soil proteins. It was also observed that the proper speed of centrifugation of lesser amounts of the same soil made the protein fluorescence spectrum appear relatively clear. Thus, the dilution of materials which absorb or scatter the excitation or fluorescent radiation is desirable in order to obtain a maximum signal for a given fluorophore concentration.

When the fluorophore itself is of high concentration in a given solution, of course, the fluorescence spectrum of the fluorophore should be readily apparent despite the presence of interfering substances. Growth experiments were used to illustrate this here. Thus, the protein fluorescence spectrum appeared rather prominent after about 2 hr of incubation, whereas the initial culture failed to show any indication of such a spectrum.

Halide ions, heavy metals, and inorganic as well as organic nitro compounds quench fluorescence (Ref. 11). The quenching is not random, however, for each instance is indicative of a fairly specific chemical interaction. Because of the good recoveries of the fluorescence spectrum of protein in aqueous soil suspensions in most samples, it is doubtful that any significant degree of chemical interaction occurs between the proteins and other solutes present in the soil suspensions which might lower the efficiency of the fluorescence process.

#### 4. Conclusion

The interference of sandy desert soil with the fluorescence of proteins has been studied with dilute solutions of bovine serum albumin as well as with a population of *E. coli B* bacteria. It was found that scattered light from colloidal particles in the soil not only distorts the fluorescence spectra of proteins, but also decreases the measured fluorescence considerably. It was also learned that, when the protein itself is present in high concentration or the light-scattering colloidal particles are diluted, the characteristic fluorescence spectra of proteins are made clearly evident. For these reasons, adequate detection of protein fluorescence in natural desert soil first required a slight treatment by centrifugation. The quantitative determination of growth of *E. coli B* in a soil-suspension medium was made by measuring the increase of protein fluorescence intensity of the culture.

We conclude that fluorometric measurement of the increase of protein in a soil-suspension medium affords a good means for measuring specific growth of protoplasmic proteins of the microorganisms.

### C. Many-Electron-Theory Ab-Initio Calculation for the Be Atom: I. Numerical Results and Discussion<sup>1</sup>

M. Geller, H. S. Taylor, and H. B. Levine

In two papers, Sinanoğlu (Ref. 12) has developed a theory for calculating the correlation energy of atoms and molecules. In this theory, the correlation energy is

<sup>1</sup>This paper has been submitted for publication in the *Journal of Chemical Physics*.

written as a sum of electron-pair correlation energies plus a remainder. The assumption is made that the remainder is small and can be neglected in calculating the wave function. This assumption is based primarily on data obtained from the LCAO-SCF-CI calculations by Watson (Ref. 13) on Be and by Ebbing (Ref. 14) on LiH.

In this paper, we report the results of a calculation for the Be atom using Sinanoğlu's formulation (Ref. 12). A detailed discussion of the algebra needed to obtain these data will be presented in a later paper; however, it seems worthwhile to make a few statements here about the mathematical consistency of the formulation, since this question has recently been raised (Ref. 15). In Eqs. (11) through (18) of Ref. 12(a), Sinanoğlu defines a pair of operators  $e_i$  and  $m_{ij}$  which play important roles in his theory. The definitions he gives there are, in fact, inconsistent. Either one of the definitions is correct, but the two taken together are not correct. However, in his subsequent derivation of the correlation energy, he does not use these inconsistent definitions, but does in fact use different and correct definitions.

In our calculations, we have used Eqs. (17) and (20) of Ref. 16 to define the operators  $e_i$  and  $m_{ij}$ . We have then constructed effective Schrödinger equations of the type of Eq. (27) of Ref. 12(b): one for the 1s-1s electron pair, and one for the 2s-2s electron pair. These equations were solved by variation of the parameters of trial wave functions. We have not explicitly calculated the inter-shell 1s-2s correlation energy, but instead use the value given by Kelly (Ref. 17), i.e.,  $4\epsilon'_{1s-2s} = -0.004966$  au, which is based upon perturbation theory.<sup>2</sup> (Sinanoğlu suggested using perturbation theory for intershell correlation energy.)

The variational pair functions used for the 1s-1s and 2s-2s correlation calculations were<sup>3</sup>

$$\hat{u}_{ij} = Q \psi_{ij} + C_1 B(i, j), \quad (1)$$

where the operator  $Q$  orthogonalizes the  $\psi_{ij}$  to all the one-electron-occupied Hartree-Fock orbitals (see Eq. 37 of Ref. 12b), and  $B(i, j)$  is an antisymmetrized product of Hartree-Fock orbitals. (We use the compromise SCF orbitals given in Ref. 18.) In Eq. (1), the pair correlation function is

$$\psi_{ij} = \sum_{K=2}^5 C_K \phi_K(i, j) \Theta(i, j), \quad (2)$$

<sup>2</sup>Here, and throughout this discussion, we use atomic units (au)  $\hbar=m=e=1$ ; 1 au of energy equals 27.210 ev.

<sup>3</sup>The notation is that of Refs. 12 and 16.

where

$$\Theta(1, 2) = 2^{-1/2} [\alpha(1)\beta(2) - \beta(1)\alpha(2)] \quad (3)$$

is the two-electron antisymmetric spin function.

The trial two-electron symmetric space functions are

$$\begin{aligned} \phi_2(1, 2) &= \frac{z^3}{\pi} \exp[-(r_1 + r_2)z], \\ \phi_3(1, 2) &= \frac{z^4}{\pi} \left(\frac{2}{21}\right)^{1/2} (r_1 + r_2) \exp[-(r_1 + r_2)z], \\ \phi_4(1, 2) &= \frac{z^5}{\pi} \frac{1}{3} (r_1 - r_2)^2 \exp[-(r_1 + r_2)z], \end{aligned} \quad (4)$$

and either

$$\phi_5(1, 2) = \frac{z^4}{\pi} 6^{-1/2} r_{12} \exp[-(r_1 + r_2)z] \quad (5a)$$

or

$$\phi_5(1, 2) = \frac{z^5}{\pi} 3^{-1/2} r_1 r_2 \cos \gamma_{12} \exp[-(r_1 + r_2)z], \quad (5b)$$

where  $r_1$ ,  $r_2$ , and  $r_{12}$  are the two electron-nucleus and the electron-electron separations, and  $\cos \gamma_{12} = (\mathbf{r}_1 \cdot \mathbf{r}_2)/r_1 r_2$ .

The five linear coefficients  $c_i$  and the nonlinear effective nuclear charge parameter  $z$  were adjusted to minimize the pair correlation energies (Refs. 12 and 16)  $\epsilon'_{ij}$  for both cases. Following minimization, the wave functions are renormalized by setting  $c_1 = 1$  in each case. (Such renormalization does not affect the energy.) The results are given in Table 1.

Table 1. Parameter values and correlation energies

Parameter	(1s-1s) <sup>a</sup>	(1s-1s) <sup>b</sup>	(2s-2s) <sup>a</sup>	(2s-2s) <sup>b</sup>
$c_1$	1	1	1	1
$c_2$	-0.470506	0.022534	0.121151	0.018182
$c_3$	-0.686008	-0.176592	-1.720671	0.925145
$c_4$	0.071737	0.064263	0.060442	0.125450
$c_5$	0.134633	-0.030282	1.067863	-0.320297
$\langle z \rangle^c$	4.1806	5.3947	0.7513	0.9974
$\langle \epsilon'_{ij} \rangle^d$	-0.041259	-0.035033	-0.041960	-0.044072
Kelly <sup>d,e</sup>	-0.04212	-0.04212	-0.04387	-0.04387

<sup>a</sup>Eqs. (4) and (5a);  $r_{12}$ -type correlation term.

<sup>b</sup>Eqs. (4) and (5b);  $p$ -type correlation term.

<sup>c</sup> $z$  for minimum  $\epsilon'_{ij}$ .

<sup>d</sup>Values reported here are in au.

<sup>e</sup>Ref. 17.

From Table 1 we see that the lowest energy is obtained by using the  $r_{12}$ -type function (Eq. 5a) for the  $1s-1s$  correlation, but using a  $p$ -type function (Eq. 5b) for the  $2s-2s$  correlation. This may explain the relatively poor result obtained by Szasz (Ref. 19), since he used  $r_{ij}$ -type functions for all of the correlations in his calculation of the energy of the Be atom. It further suggests that the great amount of labor involved in making calculations using many  $r_{ij}$ 's simultaneously may not be worthwhile.

In Table 2, we give the results of another calculation in which the functions  $\phi_2$ ,  $\phi_3$ , and  $\phi_4$  were omitted. In this case, we varied  $c_1$ ,  $c_2$ , and  $z$  so as to minimize the energy, and then we renormalized  $c_1$  to 1 at the end of the calculation as before. This calculation gives a lower energy than the 37-configuration wave function of Watson (Ref. 13).

In Table 3 we compare our result for the total energy with the results of several other calculations. We see that

Table 2. Parameter values and correlation energies (2-term calculation)

Parameter	(1s-1s) <sup>a</sup>	(1s-1s) <sup>b</sup>	(2s-2s) <sup>a</sup>	(2s-2s) <sup>b</sup>
$c_1$	1	1	1	1
$c_2$	0.142016	-0.030618	0.881834	-0.325190
$\langle z \rangle^c$	4.3288	5.3721	0.8148	0.9958
$\langle E'_{ij} \rangle^d$	-0.039488	-0.021889	-0.037818	-0.042823

<sup>a</sup>Eq. (5a);  $r_{12}$ -type correlation term.  
<sup>b</sup>Eq. (5b);  $p$ -type correlation term.  
<sup>c</sup> $z$  for minimum  $E'_{ij}$ .  
<sup>d</sup>Values reported here are in au.

Table 3. Pair correlation energies and total energy

Term	Energy, au
Hartree-Fock <sup>a</sup>	-14.572956
1s-1s <sup>b</sup>	-0.041259
2s-2s <sup>c</sup>	-0.044072
1s-2s <sup>d</sup>	-0.004966
Total energy	-14.663253
Szasz <sup>e</sup>	-14.6337
Watson (37-configuration) <sup>f</sup>	-14.65740
Best 2-term (Table 2)	-14.660233
Weiss (55-configuration) <sup>g</sup>	-14.66090
Kelly <sup>h</sup>	-14.66337
Experimental	-14.66745

<sup>a</sup>Using compromise SCF orbitals of Ref. 18.  
<sup>b</sup>Eqs. (4) and (5a);  $r_{12}$ -type correlation term.  
<sup>c</sup>Eqs. (4) and (5b);  $p$ -type correlation term.  
<sup>d</sup>Ref. 17.  
<sup>e</sup>Ref. 19.  
<sup>f</sup>Ref. 13.  
<sup>g</sup>Ref. 22.

our energy is lower than all of those of previous calculations, with the exception of the value given by Kelly (Ref. 17). However, our value does lie within the range of possible values that he gives; therefore, the two values may be said to be identical within his computational uncertainty. Thus it appears that, for Be, the remainder term involving simultaneous pair correlations is small, as suggested by Sinanoğlu (Ref. 12). This, of course, does not settle the question as to whether it is small in other systems.

Some comments are appropriate here. First, one of the "objectives of the theory" (Ref. 12a) is to use the  $r_{ij}$  coordinate in an  $N$ -electron problem. Szasz has shown (Ref. 20) that this can be done without the Sinanoğlu theory, although very difficult integrals appear. Unfortunately, difficult integrals (in fact, many of the same integrals as in the Szasz theory) occur in the Sinanoğlu theory: the correlation potential  $m_{ij}$  contains permutation operators which, when acting upon  $\hat{u}_{ij}(r_{ij})$ , give rise to three-electron integrals (Ref. 21). (This point will be discussed in more detail in a later paper.)

Second, much computational difficulty is introduced by the necessity of orthogonalizing each  $\psi_{ij}$  to all the occupied Hartree-Fock orbitals, i.e., by the requirement that the operator  $Q$  must appear in Eq. (1). For example, in the above Be calculation, the 4 terms of  $\psi_{1s-1s}$  become 36 terms in  $Q\psi_{1s-1s}$ . For atoms or molecules with  $N > 4$ , this extra work, plus that of the above-mentioned multiple- $r_{ij}$  integrals, might readily get out of hand. But it is just for the case  $N > 4$  that the theory would be significant.

Third, some theoretical problems arise upon extension of the theory to larger systems. The total wave function must be an eigenfunction of  $S^2$  and either  $L^2$  for atoms or some lower symmetry group for molecules. We have previously discussed the spin problem (Ref. 16). For Be, the angular momentum problem is trivial (the pair correlation functions must depend upon  $r_i$ ,  $r_j$ , and  $r_{ij}$ , but not upon other electron coordinates). For larger systems, this problem, or the corresponding symmetry problem in molecules, may not be trivial, and in any case has not yet been solved. Further consideration of this point is warranted.

Fourth, the double-excitation concept, which forms the foundation of the theory, is based upon the use of a spin-free Hamiltonian and the eigenfunctions of  $S^2$  and  $L^2$  (for atoms) or a lower symmetry group (for molecules). (This assumes that one has clarified or extended

the theory, as indicated previously.) Such a procedure is only valid for real systems where the spin interactions are weak, e.g., when  $L$ - $S$  coupling applies in atoms. In such a case, the spin interactions can be handled by perturbation theory after the analysis of the spin-free problem is completed. However, as is well-known for heavy atoms, the  $L$ - $S$  coupling scheme is unsuitable, and one must use the  $j$ - $j$  coupling scheme. In such a case, the relative importance of double excitations may be drastically different, so that the theory may fail *even as a concept*, quite aside from its computational difficulty.

Finally, the question of the magnitude and sign of the remainder term in the correlation energy must be discussed. In our calculation, we have taken the total energy to be the sum of the Hartree-Fock energy plus the pair correlation energies. Thus, the numerical results indicate that the remainder is negligible for Be. In general, Sinanoğlu (Ref. 12) suggests calculating the remainder from the pair functions obtained by first ignoring it. It would be much more satisfactory if one could make a definitive statement about the sign of the remainder

term. In particular, if it could be shown that it is negative, either in general or in specific cases, then Sinanoğlu's approximate energy expression

$$E = E_{HF} + \sum_{i,j} \epsilon'_{ij}, \quad (6)$$

would be an upper bound to the true energy in the usual sense of the Rayleigh-Ritz variation principle. If the remainder term is not negative, some care will be needed in interpreting the energy in Eq. (6). Further work in this area is obviously necessary.

In conclusion, we have shown that the Sinanoğlu theory is consistent in that it gives an excellent numerical result for the Be atom. A number of problems and open questions which call for resolution have been stated. The present situation is that there is considerable hope for the "many-electron theory," but that matters are less clear than was originally suggested. A paper outlining pertinent mathematical details will be presented in the near future.

## References

1. Heckly, R. J., in *Advances in Applied Microbiology*, W. W. Umbreit, Ed., Academic Press, New York, 1961, p. 1.
2. Prince, A. E., and Bakanauskas, S., *Technical Memorandum WCLT TM 58-146*, Appendix I, Wright Air Development Center, Air Research and Development Command, U.S. Air Force, 1958.
3. Prince, A. E., *Developments in Industrial Microbiology*, Vol. 1, p. 13, 1960.
4. Portner, D. M., Spiner, D. R., Hoffman, R. K., and Phillips, C. R., *Science*, Vol. 134, p. 2047, 1961.
5. Brueschke, E. E., Suess, R. H., and Willard, M., *Planetary Space Sciences*, Vol. 8, p. 30, 1961.
6. Morelli, F. A., Fehlner, F. P., and Stembridge, C. H., "Effects of Ultrahigh Vacuum on *Bacillus Subtilis* Variety *Niger*," RS 36-14, pp. 1-4, Jet Propulsion Laboratory, Pasadena, California, May 1, 1962.
7. Morelli, F. A., Fehlner, F. P., and Stembridge, C. H., *Nature*, Vol. 196, p. 106, 1962.
8. Davis, N. S., Silverman, G. J., and Keller, W. H., *Applied Microbiology*, Vol. 11, p. 202, 1963.

## References (Cont'd)

9. Konev, S. V., and Kozunin, I. I., *Dairy Science Abstracts*, Vol. 23, p. 103, 1961.
10. Rho, J. H., "Fluorometric Measurements of Growth: II. Fluorescence of Proteins as a Measure of Bacterial Growth," SPS 37-25, Vol. IV, pp. 243-248, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.
11. Udenfriend, S., *Fluorescence Assay in Biology and Medicine*, Academic Press, New York, 1962.
12. Sinanoğlu, O., *Journal of Chemical Physics*, Vol. 36, (a) p. 706, (b) p. 3198, 1962.
13. Watson, R. E., *Physical Review*, Vol. 119, p. 170, 1960.
14. Ebbing, D. D., *Journal of Chemical Physics*, Vol. 36, p. 1361, 1962.
15. Nesbet, R. K., *Electron Correlation in Atoms and Molecules*, Research Paper RJ-284, International Business Machines Corporation, New York, New York.
16. Levine, H. B., Geller, M., and Taylor, H. S., *Journal of Chemical Physics*, Vol. 40, p. 595, 1964.
17. Kelly, H. P., *Physical Review*, Vol. 131, p. 684, 1963.
18. Roothaan, C. C. J., Sachs, L. M., and Weiss, A. W., *Review of Modern Physics*, Vol. 32, p. 186, 1960.
19. Szasz, L., *Physics Letters*, Vol. 3, p. 263, 1963.
20. Szasz, L., *Journal of Chemical Physics*, Vol. 35, p. 1072, 1961.
21. Krauss, M., and Weiss, A. W., *Journal of Chemical Physics*, Vol. 40, p. 80, 1964.
22. Weiss, A. W., *Physical Review*, Vol. 122, p. 1826, 1961.

## XV. Fluid Physics

### A. The Structure of a Supersonic, Low-Density, Free Jet

H. I. Ashkenas

The underexpanded supersonic jet has been a subject of interest to low-density wind tunnel operators since Sherman (Ref. 1) suggested the possibility of using such a jet in place of a conventional nozzle in order to avoid the low-density nozzle boundary-layer problem. Stimulated by Sherman's suggestion, an attempt has been made to evaluate the flow parameters in a supersonic, low-density, free jet by means of extensive pitot-tube surveys.

#### 1. Apparatus and Methods

The experiments were performed in Leg I of the JPL Low-Density Gas Dynamics Facility (a conventional wind tunnel operated by three oil-diffusion-ejector pumps having a total capacity of approximately  $10^4$  liters/sec at  $10^{-3}$  torr). The tunnel has been described previously (Ref. 2). The jet which was studied issued from an orifice formed by a  $\frac{3}{4}$ -in. D square-edged hole machined in a 0.010-in.-thick stainless-steel orifice plate.

Pitot pressure measurements were made using externally chamfered pitot tubes in conjunction with an unbonded-strain-gage-diaphragm pressure transducer. The geometry of the tubes was similar, namely,

- (1) 10-deg external chamfer.
- (2) Ratio of OD to ID = 1.25.
- (3) Ratio of length to outside radius = 100.

The primary problem to be faced in this experiment was that of deducing pitot-tube corrections which would be applicable to the low-density, high  $M$  (Mach number) flow that exists in the jet. The solution adopted was as follows:

The pitot tube corrections noted in Ref. 3 (which were obtained in a conventional low-density nozzle flow at  $M = 4$ ) were used as a starting point for a bootstrap procedure using three geometrically similar pitot tubes; the small region in the underexpanded free jet between  $3.5 \leq M \leq 4.5$  was systematically covered with these three pitot tubes at stagnation pressures ranging from 0.1 to 7.0 torr. From these measurements, in addition to the pitot-tube correction curve, the data shown in Fig. 1 were obtained. Fig. 1 compares the measured axial  $M$

with the method-of-characteristics calculation of Owen and Thornhill (Ref. 4). It is noted that, for the low-density flows under consideration here, a small but definite reduction in effective orifice diameter is required in order that the measured  $M$  distribution coincide with that of Owen and Thornhill.

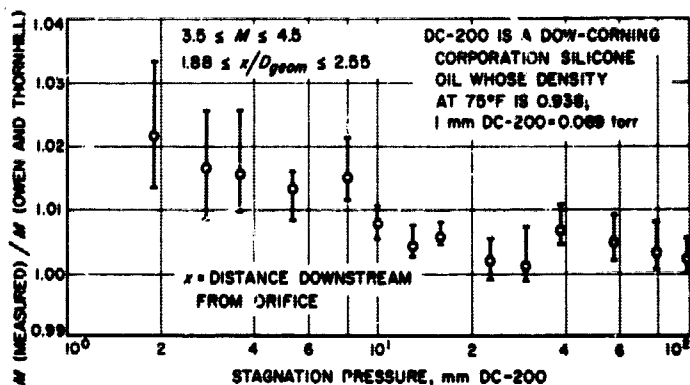


Fig. 1. Relationship between measured  $M$  and Owen and Thornhill solution

Thus, using the Owen and Thornhill distribution, it is possible to calibrate pitot tubes over the entire range of  $M$ 's available in the jet. In the present case, due to the limited pressure ratio across the orifice ( $p_1/p_2 \approx 120$ , where  $p_1$  and  $p_2$  are the stagnation chamber and test chamber pressures, respectively), the maximum attainable  $M$  before the first Mach disk is felt is of the order of 7. The pitot-tube correction curves obtained in this manner are shown in Fig. 2. Note that these curves were obtained for integer  $M$ 's; linear interpolation between curves was used in reducing data for non-integer  $M$ 's.

## 2. Results and Discussion

The pitot-tube corrections having been obtained, it was then possible to survey the jet in a radial direction and obtain off-axis  $M$  data. It will be noted in Fig. 2 that the pitot-tube corrections to be applied can be as large as 300%; i.e., the pitot pressure sensed by the tube is 3 times greater than the ideal value. Such corrections are applied with some apprehension on the part of the

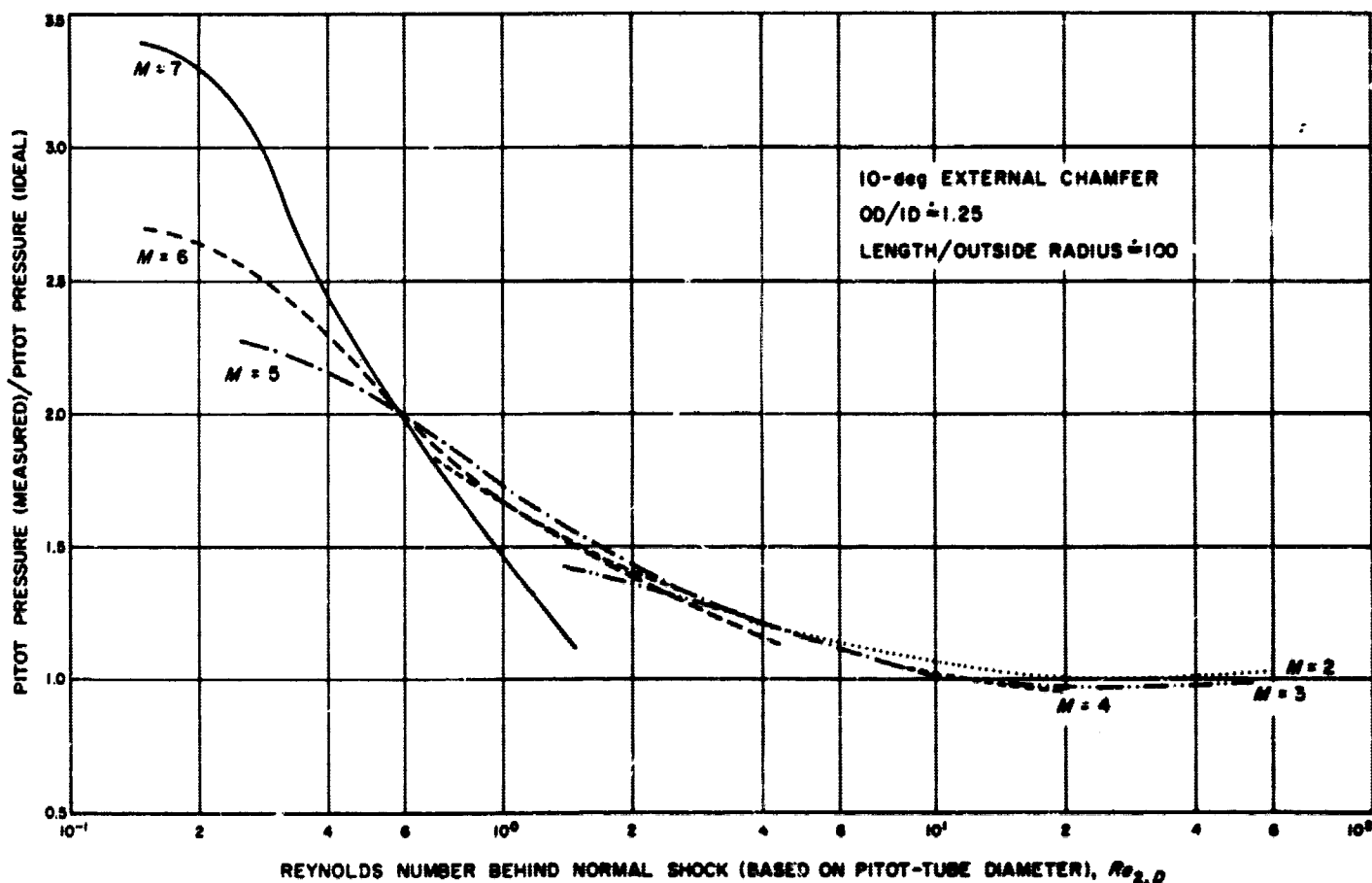


Fig. 2. Pitot-tube corrections



author. In order to allay some of the fears attendant to such a procedure, Fig. 3 is presented. Here, the axial gradient computed by Owen and Thornhill is again compared with the measured values, the measured data being based on the interpolation of the correction curves shown in Fig. 2. To be sure, Fig. 3 merely indicates that internal consistency has been maintained, but, in the face of 200 to 300% corrections, this in itself is worthy of note.

A sampling of the data is shown in Fig. 4, where radial  $M$  distributions for the jet in question, at a stagnation

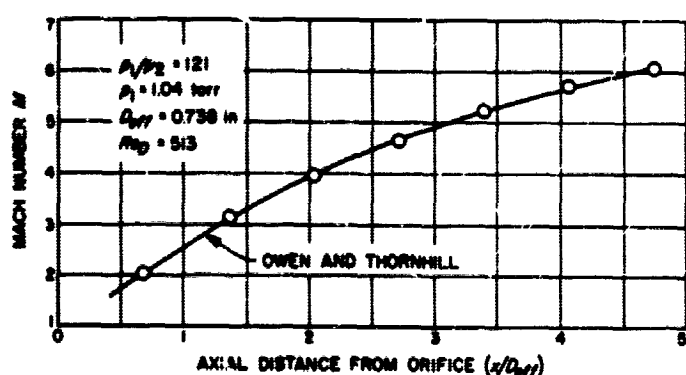


Fig. 3. Centerline data compared with Owen and Thornhill solution

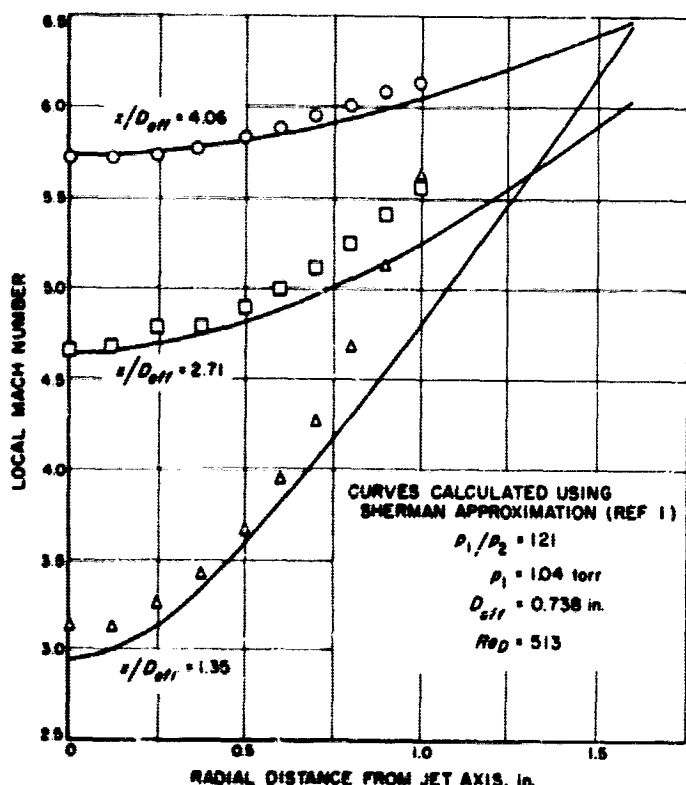


Fig. 4. Radial  $M$  distribution

pressure of 1.04 torr, are drawn. Also in Fig. 4, for comparison purposes,  $M$  distributions are presented which were computed from Sherman's (Ref. 5) analytic approximation to the characteristics solution for the free jet.

Considering the approximations made in the analysis, together with the magnitude of the corrections applied to the experimental data, the agreement shown in Fig. 4 is remarkable. It is noted that agreement is best: (1) on the axis of the jet, where both the analysis and experiment reflect the full characteristics solution; and (2) at the downstream stations, where the analytical approximations are likely to be most valid and where the pitot-tube data do not suffer from the effects of large cross-stream  $M$  gradients.

The departures of the measurements from the calculated curves apparently do not result from any shortcoming of the self-similarity approximation; the characteristics solution shows the same general trend. Rather, it is felt that, at the low densities under consideration, the influence of the viscous regions must tend to extend into the flow field to a greater degree than that anticipated by the analytical solution.

## B. The Flow Between a Rotating Disk and a Coaxial, Stationary Disk

T. Maxworthy

It has been observed that, when a tank of water is rotated, the contents eventually come to a state of solid-body rotation. When a stationary disk is then placed into such flow, one might initially believe that the boundary layer formed on the plate can be determined from the solutions of Mack (Refs. 6 and 7) for a boundary layer with solid-body outer flow. However, such layers have one unique feature which is missing in conventional boundary-layer solutions. The boundary layer, instead of acting as a sink for the outer flow, actually injects some of its fluid into the outer flow. This fluid, with its lower angular momentum, must alter the nature of the outer flow in a very drastic way. Previous theoretical work by Batchelor (Ref. 8) and Stewartson (Ref. 9) on flow between infinite disks brought to light a difficulty in deciding the exact nature of this outer flow. Lance and Rogers

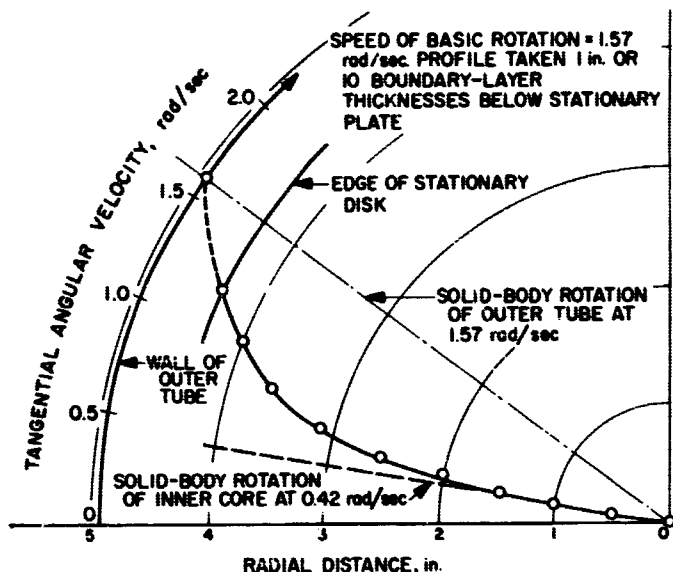


Fig. 5. Polar diagram of tangential angular velocity profile vs radial distance

(Ref. 10) with a very complete numerical solution have found that the external flow should rotate at about one-third of the angular velocity of the rotating wall.

The present experimental study is the first of several to investigate the nature of such interactions; initially we wish to observe the modification to the outer flow.

A long cylinder filled with water is rotated at constant speed, with a stationary disk placed in the flow at the open end of the tube. A thin platinum wire attached to a micrometer traversing mechanism is stretched across a diameter of the stationary plate. Short duration, 50-v pulses are supplied to the wire by a high voltage, Schmitt trigger circuit. During each pulse, hydrogen bubbles are formed on the wire by electrolysis, and these are swept off by the flow to mark a line of fluid particles at a given time. By photographing this line of particles a short time later, direct knowledge of the tangential velocity profile can be obtained.

Fig. 5 shows a typical tangential velocity profile when the wire is approximately 10 boundary-layer thicknesses from the stationary wall. It is one of the unique characteristics of rotating flows that motions perpendicular to the axis of rotation tend to be suppressed, so that only flow along vortex lines can take place easily. Such a restriction means that flows tend to depend only on the radial coordinate, and not on the coordinate in the direction of the axis of rotation. This characteristic can be seen in the present flows, for the flow profiles are found

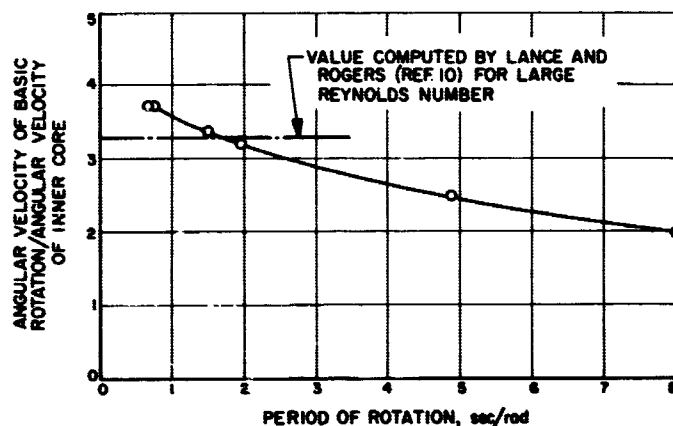


Fig. 6. Angular velocity of basic rotation/angular velocity of inner core vs period of rotation

to be independent of the axial coordinate until the end-wall boundary layers are reached.

The dependence of the angular velocity of the central solid-body rotation on the speed of rotation of the tank is shown in Fig. 6. Note that these values are very close to the value obtained by Lance and Rogers for high-Reynolds-number flows. Any difference is due mainly to the presence of the cylindrical tube wall, which also diffuses vorticity into the central region and causes the outer flow to be something other than a pure solid-body rotation.

## C. The Stability of the Compressible Laminar Boundary Layer According to a Direct Numerical Solution, Part V

L. M. Mack

In previous issues of SPS, Vol. IV, neutral stability curves and a few contours of constant  $c_i$  ( $c_i$  = the imaginary part of the complex disturbance velocity, with  $c_i = 0$  being the contour of neutral stability) have been presented at various Mach numbers. In this discussion these previous results will be used to compute the overall amplification in the boundary layer of a disturbance of given frequency as it progresses in the (downstream)  $x$ -direction.

The linearized stability theory is concerned with the time behavior of a disturbance in a boundary layer (more exactly, a parallel flow) that does not vary with time. The disturbance is presented by

$$Q(x, y, t) = Q'(y) \exp[i\alpha(x - ct)], \quad (1)$$

where  $x, y$  are the space coordinates,  $t$ , the time,  $Q$ , any flow quantity,  $\alpha$ , the wave number, and  $c$ , the (complex) wave velocity. All quantities are dimensionless. Therefore, the disturbance is a sine wave in  $x$  that moves in the positive  $x$ -direction with speed  $c_r$ . The time behavior of the disturbance amplitude is represented by the factor  $\exp(\alpha c_i t)$ . This picture can be converted into boundary-layer terms by considering a disturbance in the boundary layer to act at a given  $x$  as if it were in a parallel flow with the same velocity and temperature profiles as the boundary layer possesses at that particular  $x$ . The local rate of amplification with respect to  $x$  can be obtained from the time rate of amplification of the theory by making the transformation

$$t^* = \frac{x^*}{c_r^*}, \quad (2)$$

where an  $*$  refers to a dimensional quantity. Here the phase velocity is used, rather than the more exact group velocity (Ref. 11), for the sake of simplicity.

The derivation of a formula for the amplitude of a disturbance at any  $x$ -distance follows directly from Eq. (1). The amplitude of the disturbance (at some particular value of  $y$ ) at time  $t^*$  is

$$A = |Q'(y)| \exp(\alpha^* c_i^* t^*). \quad (3)$$

The time rate of amplification is

$$\frac{dA}{dt^*} = \alpha^* c_i^* A. \quad (4)$$

The simple form of this result is a consequence of the assumption that all of the quantities in Eq. (3) are independent of time. The fact that this assumption does not correspond directly to a boundary layer leads to considerable difficulties when a comparison is attempted between experimental and theoretical amplifications (Ref. 12). Integration of Eq. (4) yields

$$\frac{A}{A_1} = \exp \int_{t_1}^{t^*} \alpha^* c_i^* dt^*. \quad (5)$$

Eq. (5) gives the growth of the disturbance amplitude from time  $t_1^*$  to time  $t^*$ . Eq. (2) can now be used to convert Eq. (5) from an integration with respect to time to an integration with respect to distance. From the definitions of the dimensionless quantities,

$$c_r = \frac{c_r^*}{U^*}, \quad c_i = \frac{c_i^*}{U^*}, \quad \alpha = \frac{2\pi x^*}{\lambda^*} \frac{1}{R}, \quad (6)$$

where  $U^*$  is the free-stream velocity,  $\lambda^*$  is the wavelength, and  $R$  is the square root of  $R_x$  (the  $x$ -Reynolds number), Eq. (5) can be written

$$\frac{A}{A_1} = \exp \left( 2 \int_{R_1}^R \frac{\alpha c_i}{c_r} dR \right). \quad (7)$$

The right-hand side of Eq. (7) is evaluated by the following simple procedure. From a plot of  $\beta\nu/U^2$  (the dimensionless frequency, with  $\nu$  = the kinematic viscosity) versus Reynolds number, the values of  $R$  at the intersection points of the chosen frequency with the contours of  $c_i$  = constant are read off. Then, from plots of  $\alpha$  versus  $R$  and  $c_r$  versus  $R$ , the corresponding values of  $\alpha$  and  $c_r$  at the intersection points are obtained and  $\alpha c_i/c_r$  can be computed. In this manner, a plot of  $\alpha c_i/c_r$  versus  $R$  can be constructed for several values of  $\beta\nu/U^2$ . Such a plot is shown in Fig. 7 for  $M_1 = 3.8$  and 5.8. The area under a particular curve up to a Reynolds

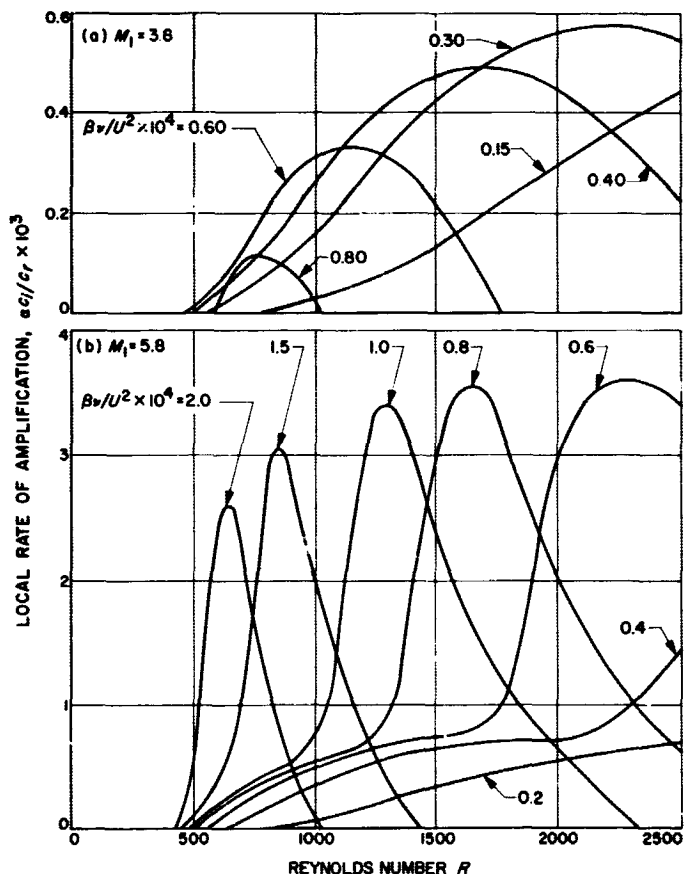


Fig. 7. Local rate of amplification vs Reynolds number for several values of dimensionless frequency at  $M_1 = 3.8$  and 5.8

number  $R$ , when used in Eq. (7), then yields the ratio of the disturbance amplitude at that  $R$  to its amplitude at the lower branch of the neutral curve. If the chosen value of  $R$  is downstream of the upper branch of the neutral-stability curve, then Eq. (7) yields the maximum amplification of the disturbance. The actual amplification at  $R$  would be smaller because of the damping region between the upper branch and  $R$ .

Perhaps the result of greatest interest is the frequency response of the boundary layer at a given Reynolds number, i.e., a curve of amplitude ratio versus frequency at a fixed  $R$ . The frequency-response curve is obtained by measuring the areas under each of the local amplification-rate curves of Fig. 1 (a or b) at a given  $R$ . From this frequency-response curve, it is easy to read off the maximum amplification that can be obtained at the Reynolds number of the curve and also the disturbance frequency that results in the maximum amplification. Four response curves at  $M_1 = 3.8$  and four curves at  $M_1 = 5.8$  are given in Figs. 8 and 9, respectively. For comparison, the two response curves computed by Lees (Ref. 11) at  $M_1 = 0.7$  are given in Fig. 10.

#### Observations concerning the results

(1) The over-all amplification at  $M_1 = 3.8$  is much less than that at  $M_1 = 0.7$ . This difference is a result of

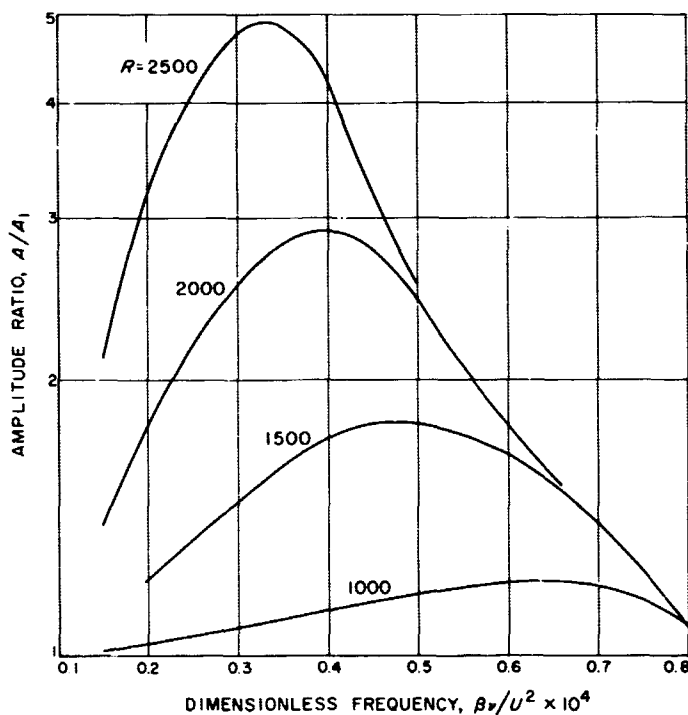


Fig. 8. Amplitude ratio vs dimensionless frequency for four Reynolds numbers at  $M_1 = 3.8$

the much smaller maximum value of  $\alpha c_i$  at  $M_1 = 3.8$  than at  $M_1 = 0.7$ . At the intermediate Mach number of 2.2, computations not presented here show that  $\alpha c_i$  is even smaller than at  $M_1 = 3.8$ , with a maximum amplitude ratio at  $R = 2500$  of only 1.28.

(2) At  $M_1 = 5.8$ , the maximum amplification increases sharply from the levels found at  $M_1 = 3.8$ . This increase results from the presence of a second instability region (Ref. 13), in which the maximum value of  $\alpha c_i$  is larger than in the instability region that corresponds to the region known at lower Mach numbers. For instance, at  $R = 1500$  ( $R_x = 2.25 \times 10^6$ ), the maximum amplification at  $M_1 = 5.8$  is about 20, compared to only 1.8 at  $M_1 = 3.8$  and to 620 at  $M_1 = 0.7$ .

(3) At  $R = 1500$  and with a  $R_x/\text{in.}$  of  $10^5$ , the frequency for maximum amplification is about 4500 cps at  $M_1 = 0.7$ , 24,000 cps at  $M_1 = 3.8$ , and 42,000 cps at  $M_1 = 5.8$ .

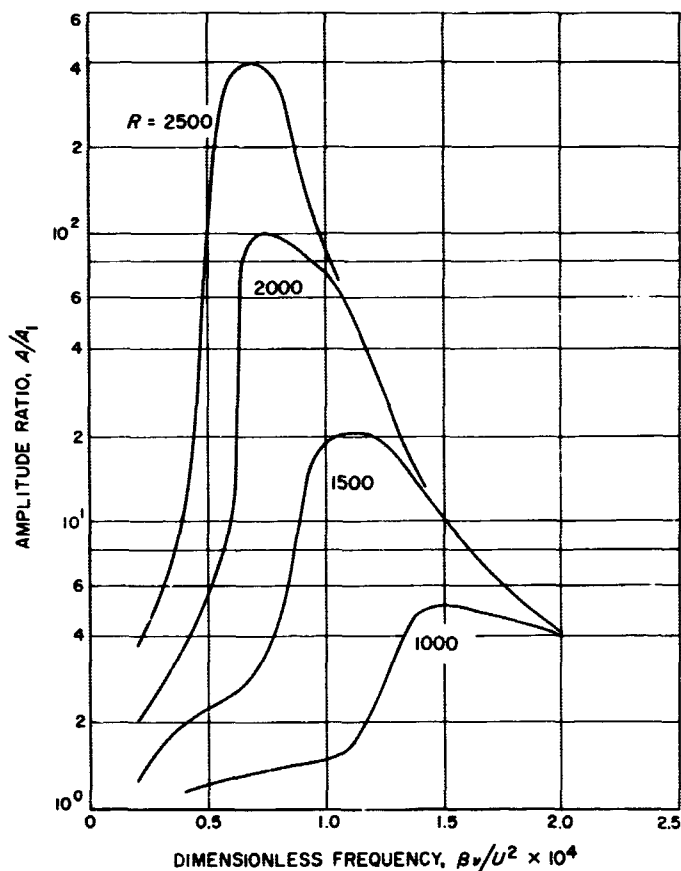


Fig. 9. Amplitude ratio vs dimensionless frequency for four Reynolds numbers at  $M_1 = 5.8$

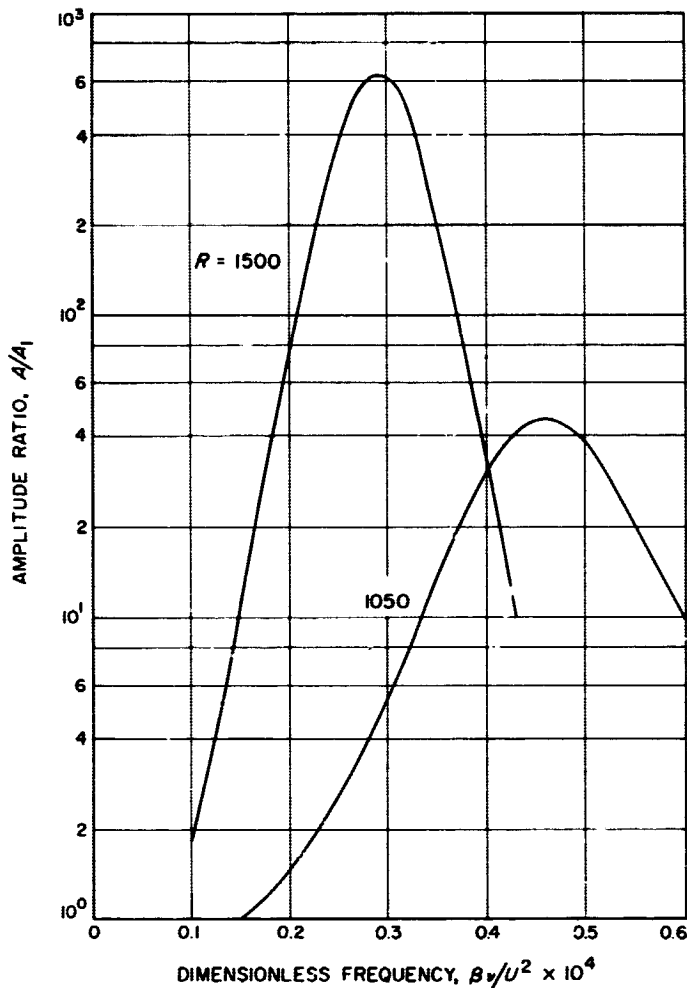


Fig. 10. Amplitude ratio vs dimensionless frequency for two Reynolds numbers at  $M_1 = 0.7$  (Ref. 11)

(4) The above results indicate that the boundary layer is most stable somewhere around  $M_1 = 2.5$ . The instability increases slowly with Mach number until near

$M_1 = 5$  the second instability region becomes of importance and there is a rapid increase in instability.

(5) When assessing the implication of these results for the transition problem, it must be kept in mind that the actual spectrum of the boundary-layer disturbances at a given  $R_x$  depends not only on their magnitude before amplification, but also on their spectrum before amplification. For instance, if at  $M_1 = 5.8$  there is negligible energy present at 42,000 cps compared to that present at lower frequencies, then 42,000 cps loses its importance as a trigger for the transition process even though it is the most highly amplified frequency.

(6) The experimental evidence concerning transition in a supersonic wind tunnel shows that the transition Reynolds number first decreases with increasing Mach number up to about  $M_1 = 3.5$  and then increases. Also, the laminar boundary layer becomes increasingly hard to trip with increasing Mach number. A typical result is that the transition Reynolds number measured on a model with turbulent boundary layers on the tunnel side walls is about  $1.5 \times 10^6$  at  $M_1 = 3.8$ . On the other hand, when the tunnel is run at low pressure so that the tunnel side walls have laminar boundary layers, the  $x$ -Reynolds number of the side-wall boundary layer, as measured from the tunnel throat, is  $25 \times 10^6$ .

(7) All of the results strongly illustrate the fact that almost nothing of a fundamental nature is known about the mechanism of transition at supersonic Mach numbers, the influence of external disturbances such as sound radiation, and the relationship, if any between laminar instability and transition.

## References

1. Sherman, F. S., in *Rarefied Gas Dynamics*, Supplement 2, Vol. II, J. A. Laurmann, Ed., Academic Press, New York, 1963.
2. Wegener, P. P., and Ashkenas, H. I., "Wind Tunnel Measurements of Sphere Drag at Supersonic Speeds and Low Reynolds Numbers," *Journal of Fluid Mechanics*, Vol. 10, Part 4, 1961.
3. Ashkenas, H. I., "Pitot Tube Corrections in Low Density Flows," SPS 37-15, Vol. IV, pp. 123, 124, Jet Propulsion Laboratory, Pasadena, California, June 30, 1962.

## References (Cont'd)

4. Owen, P. L., and Thornhill, C. K., *The Flow in an Axially-Symmetric Supersonic Jet from a Nearly-Sonic Orifice into a Vacuum*, Aeronautical Research Council Reports and Memoranda No. 2616, 1948.
5. Sherman, F. S., *Self-Similar Development of Inviscid Hypersonic Free-Jet Flows*, Lockheed Missiles and Space Company Technical Report: Fluid Mechanics 6-90-63-61, 1963.
6. Mack, L. M., *The Laminar Boundary Layer on a Disk of Finite Radius in a Rotating Flow. Part I: Numerical Integration of the Momentum-Integral Equations and Application of the Results to the Flow in a Vortex Chamber*, Technical Report No. 32-224, Jet Propulsion Laboratory, Pasadena, California, May 20, 1962.
7. Mack, L. M., *The Laminar Boundary Layer on a Disk of Finite Radius in a Rotating Flow. Part II: A Simplified Momentum-Integral Method*, Technical Report No. 32-366, Jet Propulsion Laboratory, Pasadena, California, January 31, 1963.
8. Batchelor, G. K., "Note on a Class of Solutions of the Navier-Stokes Equations Representing Steady Rotationally-Symmetric Flow," *Quarterly Journal of Mechanics and Applied Mathematics*, Vol. 4, p. 29, 1951.
9. Stewartson, K., "On the Flow between Rotating Co-axial Disks," *Proceedings of the Cambridge Philosophical Society*, Vol. 49, p. 333, 1953.
10. Lance, G. N., and Rogers, M. H., "The Axially Symmetric Flow of a Viscous Fluid between Two Infinite Rotating Disks," *Proceedings of the Royal Society, Series A266*, p. 109, 1962.
11. Lees, L., *Instability of Laminar Flows and Transition to Turbulence*, Report No. 7A-7-006, Consolidated Vultee Aircraft Corporation, San Diego, California, 1952.
12. Laufer, J., and Vrebalovich, T., "Stability and Transition of a Supersonic Laminar Boundary Layer on an Insulated Flat Plate," *Journal of Fluid Mechanics*, Vol. 9, Part 2, pp. 257-299, 1960.
13. Mack, L. M., "The Stability of the Compressible Laminar Boundary Layer According to a Direct Numerical Solution, Part IV," *SPS 37-24*, Vol. IV, pp. 271-274, Jet Propulsion Laboratory, Pasadena, California, December 31, 1963.

## XVI. Physics

### A. Rocket Measurement of Nitric Oxide in the Upper Atmosphere<sup>1</sup>

C. A. Barth

The identification of the nitric oxide gamma bands in the ultraviolet dayglow of the Earth was presented in Ref. 1. This discussion gives the results of the analysis of 16 of the spectral scans obtained during the flight of the NASA Aerobee 4.85. The distribution of nitric oxide in the upper atmosphere between 75 and 125 km is obtained from these spectra.

The trajectory of the rocket is presented in Fig. 1. The rocket reached an apogee of 184 km at 230 sec. The arrows on the trajectory show the attitude of the rocket and, hence, the zenith angle of the direction that the ultraviolet spectrometer observed throughout the flight. This angle was 9 deg at 63 sec when observations were begun, 3 deg at 100 sec, 9 deg again at 147 sec, 30 deg at 191 sec, 60 deg at 250 sec, 90 deg at 310 sec, and 120 deg at 373 sec. The Sun was at a 60-deg zenith angle and a 200-deg azimuth angle throughout these observations. The rocket azimuth angle was 170 deg.

<sup>1</sup>These results were presented at the Fifth International Space Science Symposium of COSPAR, Florence, Italy, May 12-20, 1964.

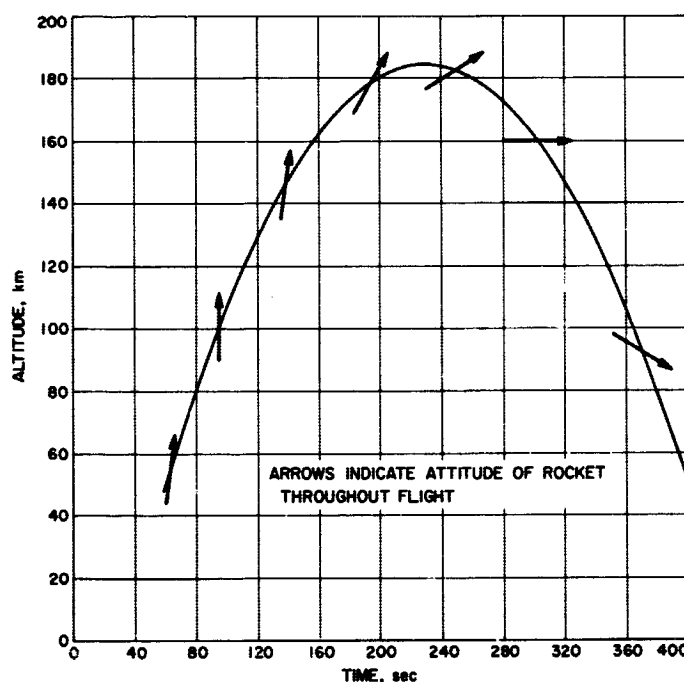


Fig. 1. Rocket trajectory

Five successive spectra obtained between 76 and 114 km during the ascent are given in Fig. 2. The altitude and zenith angle are given for each scan. The portion of the scan that is shown is from 1950 to 2650 Å. The spectra were obtained at 10-Å resolution at a scan rate



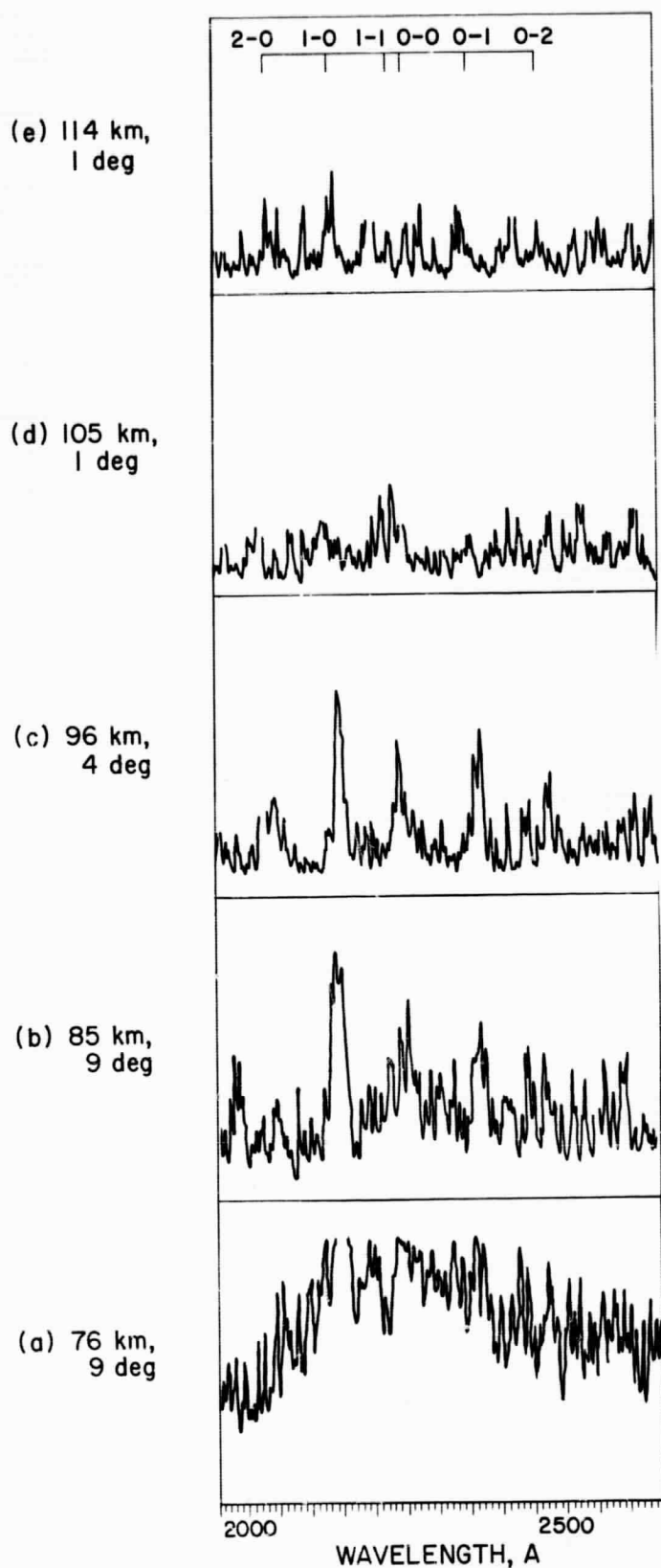


Fig. 2. Ultraviolet spectral scans obtained during the ascent at a 10-Å resolution and a 1000-Å/sec scan rate

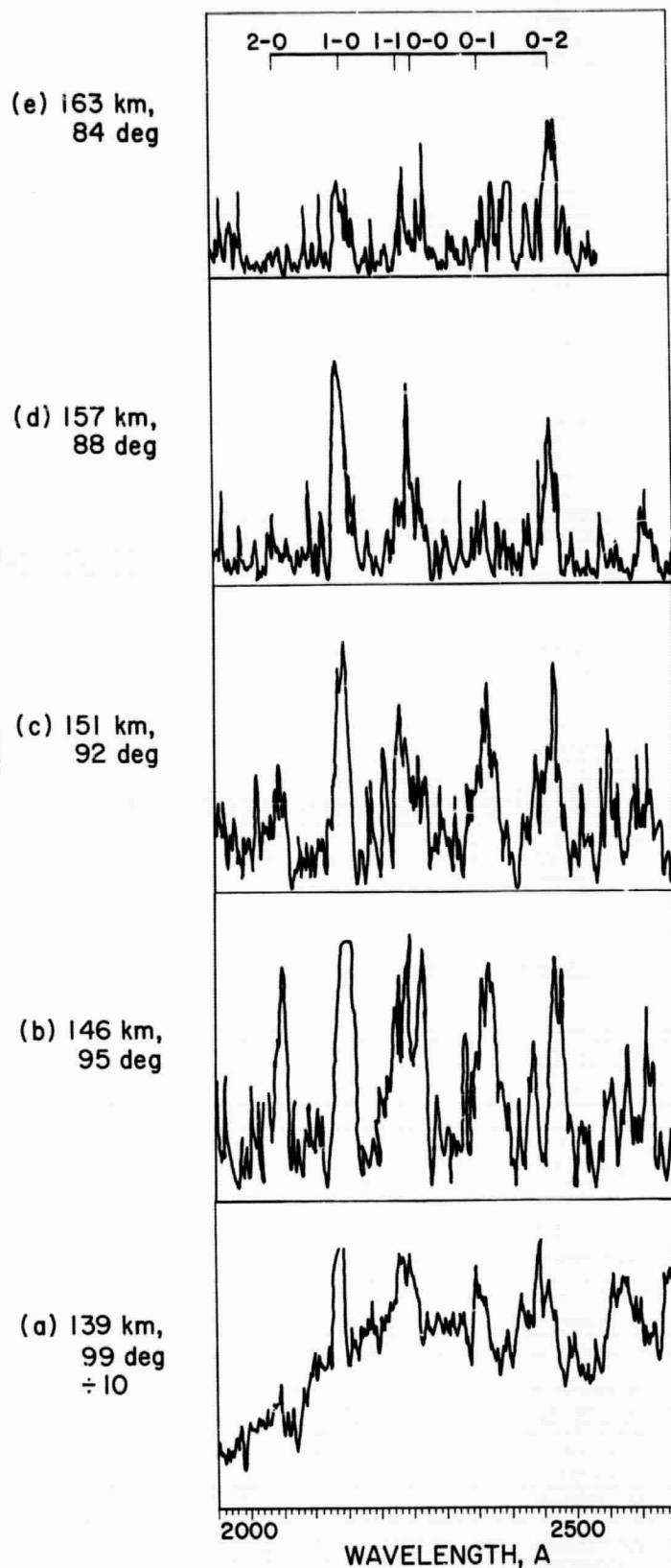


Fig. 3. Ultraviolet spectral scans obtained during observations of the horizon at a 10-Å resolution and a 1000-Å/sec scan rate

of 1000 Å/sec. The positions of the six most prominent gamma bands (Ref. 1)—2-0, 1-0, 1-1, 0-0, 0-1, and 0-2—are indicated in the upper portion of Fig. 2. The 1-0 band at 2150 Å can be seen in all five spectra. The spectrum in Fig. 2(a), which was recorded at 76 km, contains a background of Rayleigh scattering as well as the nitric oxide dayglow. The 1-0 band is off-scale in this high-gain spectrum. (It was properly recorded on a less-sensitive telemetry channel.) The progressive decrease in intensity of the gamma bands with increasing altitude is seen in the remaining spectra of Fig. 2. It is this series of spectra that provides the measurement of the distribution of nitric oxide in the upper atmosphere.

Five successive spectra that were obtained late in the flight during the scan of the horizon are shown in Fig. 3. Once again the spectral resolution is 10 Å and the wavelength interval is 1950 to 2650 Å. The rocket was far above the dayglow layer during these observations, which were begun at a 163-km altitude and completed by the time a 139-km altitude was reached. As the viewing angle of the spectrometer passed through the horizon, the nitric oxide bands appeared and increased greatly in intensity. The intensity viewed toward the horizon was much greater than that viewed in the zenith. The 1-0 band is off-scale in Fig. 3(a) and 3(b), even though the scan in Fig. 3(a) was recorded with a sensitivity that was a factor of 10 less than the sensitivity of the rest of the scans. The spectrum in Fig. 3(a) also shows a contribution from the Rayleigh scattering of the lower atmosphere. It is this series of spectra, taken far above the dayglow layer, that proves that the nitric oxide gamma bands originate in the atmosphere and are not associated with any local contamination of the rocket.

A spectrum obtained at 4-Å resolution and a scan rate of 325 Å/sec is shown in Fig. 4. This spectrum was recorded when the rocket was at an altitude of 141 km and the instrument zenith angle was 97 deg. This high-resolution spectrum, besides displaying the capabilities of the instrumentation, clearly shows the dayglow features to be spectral bands of approximately 15-Å half-width.

The emission rate of the 1-0 gamma band at 2150 Å is given in Fig. 5 as a function of altitude. These data were taken from spectra obtained during the ascent when viewing the zenith. The first five of these were shown in Fig. 2. The emission rate of the 1-0 gamma band is converted into the total amount of nitric oxide above the rocket by dividing by the appropriate g-factor

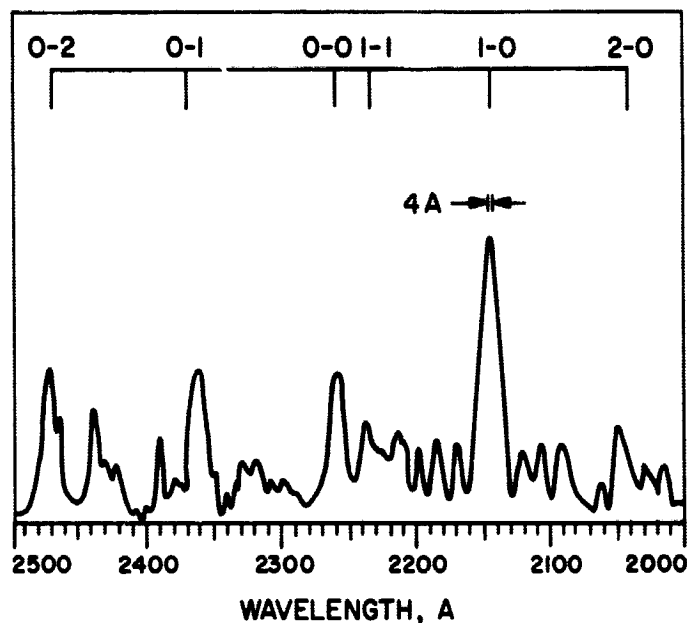


Fig. 4. High-resolution ultraviolet dayglow spectrum obtained at a 4-Å resolution and a 325-Å/sec scan rate

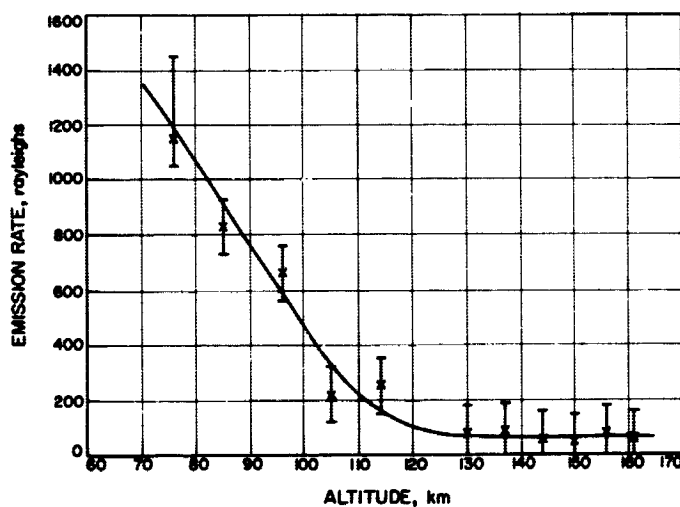


Fig. 5. Emission rate of the 1-0 nitric oxide gamma band as a function of altitude

from Table 1, p. 171, of Ref. 1. The column density of nitric oxide is given in Fig. 6 as a function of altitude. The change of nitric oxide density with height, or the slope of the curve in Fig. 6, gives the volume density of nitric oxide in the upper atmosphere. This final reduction of the flight data is presented in Fig. 7. The density is  $6.2 \times 10^7$  molecules/cm<sup>3</sup> at 75 km; it remains more or less constant until 90 km when it begins to drop off, reaching a value of  $6 \times 10^6$  molecules/cm<sup>3</sup> at 125 km.

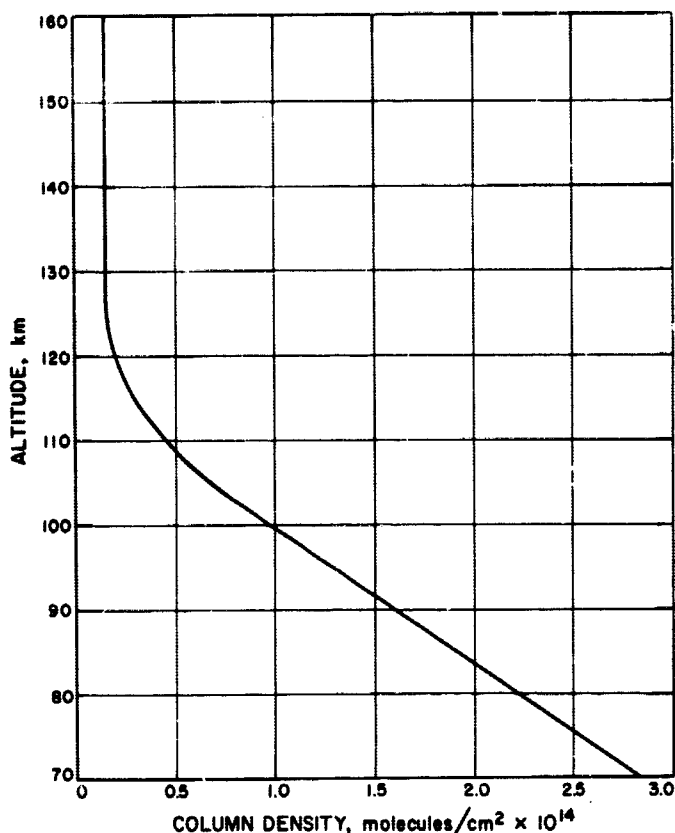


Fig. 6. Overhead column density of nitric oxide in the upper atmosphere as derived from dayglow measurements

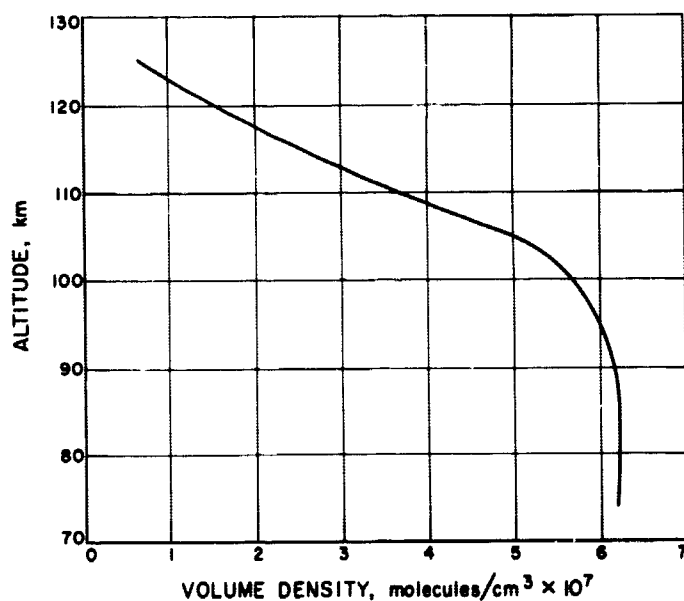


Fig. 7. Distribution of nitric oxide in the upper atmosphere as a function of altitude

## B. Improved Efficiency Calculations for the Fission-Electric Cell

J. L. Shapiro

Previous calculations (Ref. 2) of the efficiency of the fission-electric cell have employed a model assuming mono-energetic fission fragments. It was also assumed that a fragment's electric charge remained constant as the fragment passed through the fuel layer. These approximations have been modified as follows:

It is assumed that each fission releases one light fragment and one heavy fragment. The light fragment has an initial charge in the fuel layer  $Z_l^0 = 21$  and an initial energy  $E_l^0 = 100$  Mev. For the heavy fragment,  $Z_h^0 = 23$  and  $E_h^0 = 65$  Mev. Furthermore, it is assumed that the charge of the light fragment varies as (velocity)<sup>1/4</sup>, while that of the heavy fragment varies as the velocity itself. Although there is little information concerning the ranges in solids for each fragment, on the basis of measurements reported in gases (Ref. 3), it will be assumed that the respective ranges are proportional to the initial energies.

Finally, the linear energy-loss model originally used has been replaced with the assumption that the fission fragment velocity decreases linearly with penetration. Thus, the equations presently being used for calculating efficiency become (compare with Ref. 2, p. 21), for the cylindrical case:

$$\mathcal{E} = \frac{E_0}{E'} \frac{V}{\pi r} \int_0^{\delta_{max}} d\delta \int_0^{\alpha_{max}} d\alpha \int_0^{r_{max}} (1-r)^c \cos^2 \alpha \cos \delta dr,$$

where

$$r_{max} \text{ is the smaller of } \begin{cases} 1 - \left[ \frac{V}{\cos^2 \alpha \left( 1 - \frac{R_1^2}{R_2^2} \sin^2 \delta \right)} \right]^d, \\ \frac{r}{\cos \alpha \cos \delta} \end{cases},$$

$$\cos^2 \alpha_{max} = \frac{V}{1 - \frac{R_1^2}{R_2^2} \sin^2 \delta},$$

$$\sin^2 \delta_{max} \text{ is the smaller of } \begin{cases} \frac{R_2^2}{R_1^2} (1 - V) \\ 1 \end{cases},$$

and  $c = d = 1$  for the heavy fragment, while  $c = 1/2$  and  $d = 1/4$  for the light fragment.

In these equations the voltage  $V$  is in units of  $E_0/e'$ , where  $E_0$  is the initial fragment energy and  $e'$  is the initial fragment charge. Similarly, the fuel layer thickness  $\tau$  is in units of fragment range. Therefore, for a given actual voltage and thickness, separate calculations must now be made for the light and heavy fragments and their results combined to yield the actual efficiency.

Machine computations are now under way which will show the effect of this more realistic efficiency determination.

## C. Local Measurement of Gravitational Fields

H. Wahlquist and F. Estabrook

One useful feature of the dyadic formulation of general relativity (Ref. 4) is its physical interpretability. The dyadic equations are written directly in terms of local, proper 3-vector and dyadic quantities, which are shown to be exactly those quantities usually operationally defined and experimentally measured. The equations are, therefore, immediately applicable to the analysis of gravitational experiments.

As an example of this, we have derived two equations which show how, in principle, 14 components of the Riemann tensor may be locally measured by an arbitrarily moving observer who compares the relative acceleration and precession of two adjacent inertially oriented test particles. To interpret the equations, one may think of an arbitrarily moving laboratory (Earth-based, satellite, or other), small enough so that within it space is Euclidean, equipped with  $x, y, z$  Cartesian axes and equipped everywhere with standard clocks. The clocks cannot, however, be synchronized because of the arbitrary motion of the room.

Consider, then, two inertially oriented test particles or "stable-platforms" in this laboratory. Take one of these to be fixed at the spatial origin and the other to be at  $\mathbf{r}$ . The second may move arbitrarily about the room; i.e.,  $\dot{\mathbf{r}} \neq 0$ . Each of these stable platforms will, in general, *precess* with respect to the Cartesian axes: we denote the angular velocities of precession as  $-\omega$  and  $-\omega_1$ , respectively. Each of these is measured using a standard clock

at the point of measurement. Next, consider two accelerometers, coincident with the stable platforms and similarly measuring accelerations  $\mathbf{a}$  and  $\mathbf{a}_1$ , respectively. The operational equations relating these measurements are shown to be:

$$\mathbf{a}_1 = \mathbf{a}(1 - \mathbf{a} \cdot \mathbf{r}) + \ddot{\mathbf{r}} + 2\omega \times \dot{\mathbf{r}} + \omega \times (\omega \times \mathbf{r}) + \dot{\omega} \times \mathbf{r} + \mathbf{Q} \cdot \mathbf{r} \quad (1)$$

$$-\omega_1 = (-\omega)(1 - \mathbf{a} \cdot \mathbf{r}) + \mathbf{a} \times (\dot{\mathbf{r}} + \omega \times \mathbf{r}) - \mathbf{B} \cdot \mathbf{r} + \mathbf{t} \times \mathbf{r} \quad (2)$$

The first of these is a quasi-Newtonian equation: the usual centrifugal, Coriolis, and angular acceleration terms will be recognized. A special relativistic clock-rate correction factor  $[1 - (\mathbf{a} \cdot \mathbf{r}/c^2)]$ , where  $c^2 = 1$  in our units] is but another manifestation of the Doppler shift predicted by special relativity for accelerating frames and recently verified in local, terrestrial experiments using the Mössbauer effect (compare Ref. 5, p. 411). The term  $\mathbf{Q} \cdot \mathbf{r}$  is the general relativistic term expressing the tidal acceleration effect of the space-time curvature. In contradistinction to all other fields in physics, it is characteristic of the gravitational field that it is observable *not* as a force or torque on single test particles, but rather as a *differential* or tidal effect between two or more adjacent test particles. If the test particles are free ( $\mathbf{a}_1 = \mathbf{a} = 0$ ), the first equation reduces to the equation of geodesic deviation of Synge (Ref. 5).

In the second equation, all terms leading to a difference between the stable platform angular velocities  $-\omega_1$  and  $-\omega$  are of special or general relativistic origin. We again find a clock-rate correction factor. The second relativistic term is the differential Thomas precession. These two terms can be derived from the usual Thomas precession formula in the differential limit, provided care is taken to express all precession rates in terms of local proper times. The terms  $-\mathbf{B} \cdot \mathbf{r}$  and  $\mathbf{t} \times \mathbf{r}$  again express the effects of other components of the Riemann tensor, i.e., the space-time curvature or gravitational field.

According to Einstein's theory,  $\mathbf{t}$  is to be identified with the local density of momentum, as seen in the laboratory  $x, y, z$  system. The quantity  $\mathbf{B}$  is a traceless dyadic involving five local components of the Weyl tensor, while  $\mathbf{Q}$  involves the other five components of the Weyl tensor (in a traceless dyadic  $\mathbf{A}$ ), the local stress tensor  $\mathbf{T}$ , and the local mass density  $\rho$  as follows:

$$\mathbf{Q} = \mathbf{A} + \mathbf{T} + \frac{1}{3}(\rho - 2 \text{tr } \mathbf{T}) \mathbf{I} \quad (3)$$

If the laboratory is sensibly empty,  $t$ ,  $T$ , and  $\rho$  must all be negligible. The gravitational terms will then involve only the dyadics  $A$  and  $B$ , which represent the local gravitational field produced by external matter and stress systems. In flat Minkowski space-time,  $A$  and  $B$  will also vanish. The details of this local interpretation of the Riemann tensor are given elsewhere.<sup>2</sup>

<sup>2</sup>Estabrook, F., and Wahlquist, H., "Dyadic Analysis of Space-Time Congruences" (To be published).

## D. Derivation of the Master Equation with the Wigner-Weisskopf Formalism for Radiation Damping

O. von Roos

The so-called "master equation" (Ref. 6) expresses the time rate of change of the occupation probability  $P_\alpha$  of a state  $|\alpha\rangle$  by means of the following irreversible equation:

$$\dot{P}_\alpha = \sum_\beta [W_{\beta\alpha} P_\beta - W_{\alpha\beta} P_\alpha]. \quad (1)$$

It is our aim to derive this equation from Schrödinger's equation by means of two assumptions: (1) the random phase assumption, and (2) the validity of the Wigner-Weisskopf perturbation treatment. The first assumption means that, in a double sum over probability amplitudes  $\sum_{\alpha,\beta} c_\alpha c_\beta^*$ , only diagonal terms have to be retained; i.e.,

$$\sum_{\alpha,\beta} c_\alpha c_\beta^* = \sum_\alpha |c_\alpha|^2. \quad (2)$$

As to the second assumption, the derivation given below will clarify this assumption amply.

Starting from Schrödinger's equation for a many-body system:

$$(H_0 + H_1)\psi = i\hbar \dot{\psi}, \quad (3)$$

we find, with

$$H_0|\alpha\rangle = E_\alpha|\alpha\rangle \quad (4)$$

$$\psi = \sum_\alpha c_\alpha(t) \exp\left(-\frac{it}{\hbar} E_\alpha\right) |\alpha\rangle, \quad (5)$$

that

$$i\hbar \frac{d}{dt} [c_\beta(t)] = \sum_\alpha c_\alpha(t) \exp(i\omega_{\beta\alpha}) \langle \beta | H_1 | \alpha \rangle. \quad (6)$$

This equation is still exact. We now propose to solve Eq. (6) in first-order perturbation theory, assuming for the expansion coefficients  $c_\alpha$  at  $t = 0$ :

$$c_\alpha(t) \sim c_\alpha(0) \exp\left(-\frac{1}{2} \Gamma t\right) \quad (7)$$

(i.e., the Wigner-Weisskopf hypothesis). By inserting Eq. (7) into Eq. (6), we can integrate immediately:

$$c_\beta(t) = \frac{i}{\hbar} \sum_\alpha c_\alpha(0) \frac{\exp\left[t\left(i\omega_{\beta\alpha} - \frac{1}{2} \Gamma\right)\right] - 1}{i\omega_{\beta\alpha} - \frac{1}{2} \Gamma} \langle \beta | H_1 | \alpha \rangle + c_\beta(0) \exp\left(-\frac{1}{2} \Gamma t\right). \quad (8)$$

Obviously,

$$|c_\beta(t)|^2 = P_\beta(t) \quad (9)$$

is the probability of occupancy of the state  $|\beta\rangle$ . Taking the square of Eq. (8) and applying the random phase assumption, Eq. (2), yields:

$$P_\beta(t) = P_\beta(0) \exp(-\Gamma t) + \frac{1}{\hbar^2} \sum_\alpha P_\alpha(0) \left| \frac{\exp(i\omega_{\beta\alpha}) - 1}{\omega_{\beta\alpha}} \right|^2 \times |\langle \beta | H_1 | \alpha \rangle|^2 \quad (10)$$

In order to determine  $\Gamma$ , we simply insert Eq. (8) into Eq. (6) and obtain:

$$-\frac{i\hbar \Gamma}{2} c_\beta(0) \exp\left(-\frac{1}{2} \Gamma t\right) = \sum_\alpha \frac{1}{i\hbar} \sum_{\alpha'} c_{\alpha'}(0) \frac{\exp\left[t\left(i\omega_{\alpha\alpha'} - \frac{1}{2} \Gamma\right)\right] - 1}{i\omega_{\alpha\alpha'} - \frac{1}{2} \Gamma} \times \langle \alpha | H_1 | \alpha' \rangle \exp(i\omega_{\beta\alpha}) \langle \beta | H_1 | \alpha \rangle. \quad (11)$$

Multiplying Eq. (11) by  $c_\beta^*(0)$  and applying again the random phase assumption yields (neglecting the small quantity  $\Gamma$  in the exponentials):

$$-\frac{i\hbar \Gamma}{2} = \frac{1}{i\hbar} \sum_\alpha \frac{1 - \exp(i\omega_{\beta\alpha})}{-i\omega_{\beta\alpha}} |\langle \alpha | H_1 | \beta \rangle|^2. \quad (12)$$

In the limit

$$\Gamma^{-1} \gg t \gg \omega_{\beta\alpha}^{-1}, \quad (13)$$

i.e., in times large compared to characteristic frequencies of the system but small compared to any times comparable to measurement times (macroscopic times), we have from Eq. (12):

$$\operatorname{Re} \Gamma = \frac{2\pi}{\hbar^2} \sum_{\alpha} \delta(\omega_{\beta\alpha}) |\langle \alpha | H_1 | \beta \rangle|^2, \quad (14)$$

$$\operatorname{Im} \Gamma = 0. \quad (15)$$

Eq. (10) gives, in the same limit:

$$P_{\beta}(t) = P_{\beta}(0) (1 - \Gamma t) + \frac{2\pi}{\hbar^2} \sum_{\alpha} \delta(\omega_{\beta\alpha}) |\langle \beta | H_1 | \alpha \rangle|^2 P_{\alpha}(0). \quad (16)$$

Eq. (16) together with Eq. (15) then yields:

$$\begin{aligned} \frac{P_{\beta}(t) - P_{\beta}(0)}{t} &= \frac{dP_{\beta}(0)}{dt} \\ &= \frac{2\pi}{\hbar^2} \sum_{\alpha} \delta(\omega_{\beta\alpha}) \\ &\quad \times |\langle \beta | H_1 | \alpha \rangle|^2 \{P_{\alpha}(0) - P_{\beta}(0)\}, \end{aligned} \quad (17)$$

which is just Eq. (1) with the well-known expression for the transition probability

$$W_{\beta\alpha} = W_{\alpha\beta} = \frac{2}{\hbar^2} \delta(\omega_{\beta\alpha}) |\langle \beta | H_1 | \alpha \rangle|^2. \quad (18)$$

Since the Wigner-Weisskopf theory is only a special case of the general damping theory by Heitler (Ref. 7), an application of the latter theory is of great interest in nonequilibrium statistical quantum mechanics, a fact which thus far has not received any attention. Work along these lines is in progress.

## E. Internal Automorphisms of a Two-Particle System

J. S. Zmuidzinas

This discussion supplements a previous one on relativistic many-particle states (Ref. 8). As was shown there, the relativistic two-particle states are specified by 12

quantum numbers, 6 of them pertaining to the center-of-mass characteristics of the two-particle system and the other 6 describing the "internal structure" of this system. A typical state is

$$|m s p \lambda; m_1 s_1 \lambda_1 m_2 s_2 \lambda_2\rangle \quad (1)$$

with obvious slight changes in the notation of Ref. 8. As we have shown,

$$\lambda_1 = W^{(1)} \cdot P^{(2)} / p_0 |p|,$$

where  $p_0 = m$  and  $p$  are the total center-of-mass energy and momentum, respectively. It is readily verified that

$$|p| = \frac{1}{2m} \Delta^{\frac{1}{2}}(m_1^2, m_2^2, m^2),$$

$$\Delta(a, b, c) = a^2 + b^2 + c^2 - 2(ab + bc + ca),$$

so that

$$\lambda_1 = 2W^{(1)} \cdot P^{(2)} \Delta^{-\frac{1}{2}}(m_1^2, m_2^2, m^2).$$

The three masses and  $I_n = W^{(1)} \cdot P^{(2)}$  are invariants under all transformations of  $P_{\mu}^{\dagger}$ , and hence so are  $\lambda_1$  and  $\lambda_2$ . The internal quantum numbers  $m_1, s_1, \lambda_1, m_2, s_2$ , and  $\lambda_2$ , being eigenvalues of invariant operators, are not affected by an arbitrary inhomogeneous Lorentz transformation ( $a, \Lambda$ ):

$$\begin{aligned} U(a, \Lambda) |m s p \lambda; m_1 s_1 \lambda_1 m_2 s_2 \lambda_2\rangle \\ = e^{-ia \cdot \Lambda p} \sum_{\lambda'} D_{\lambda' \lambda}^s(R(\Lambda, p)) |m s p' \lambda'; m_1 s_1 \lambda_1 m_2 s_2 \lambda_2\rangle, \end{aligned} \quad (2)$$

where  $p'_{\mu} = \Lambda_{\mu}^{\nu} p_{\nu}$ ,  $R$  is the Wigner rotation operator (Ref. 9), and  $D_{\lambda' \lambda}^s$  are the matrices of the  $(2s+1)$ -dimensional representation of the three-dimensional rotation group. Eq. (2) holds for the case  $m, p_0 > 0$  of physical interest. The natural question arises whether there exist some unitary transformations of the states (1) generated by the operators  $P_{\mu}^{(i)}, M_{\mu\nu}^{(i)}, i = 1, 2$ , which mix the internal variables  $\lambda_1$  and  $\lambda_2$  ( $m_1, s_1, m_2, s_2$  are necessarily fixed within the framework of the Poincaré group) and yet leave the "external" quantum numbers  $m, s, p, \lambda$  unaffected. Clearly, such transformations must be built up from invariants which commute with all  $P_{\mu} = P_{\mu}^{(1)} + P_{\mu}^{(2)}$  and  $M_{\mu\nu} = M_{\mu\nu}^{(1)} + M_{\mu\nu}^{(2)}$  of the external Lie algebra. The set of all such invariants is denoted by

$\mathcal{E}_{int}$ . We have found all of them, omitting details, they are as follows:

$$\begin{aligned}
 \mu_1 &= (P^{(1)})^2, & \mu_2 &= (P^{(2)})^2, \\
 \omega_1 &= (W^{(1)})^2, & \omega_2 &= (W^{(2)})^2, \\
 \mu &= P^{(1)} \cdot P^{(2)}, & \omega &= W^{(1)} \cdot W^{(2)}, \\
 \bar{\lambda}_1 &= W^{(1)} \cdot P^{(2)}, & \bar{\lambda}_2 &= W^{(2)} \cdot P^{(1)}, \\
 \kappa_1 &= W^{(1)} \cdot W, & \kappa_2 &= W^{(2)} \cdot W, \\
 \nu &= (W)^2, & \alpha &= [P^{(1)} P^{(2)} W^{(1)} W^{(2)}], \\
 \beta_1 &= [WPW^{(1)} P^{(1)}], & \beta_2 &= [WPW^{(2)} P^{(2)}],
 \end{aligned} \quad (3)$$

where

$$W_\mu = \frac{1}{2} \epsilon_{\mu\nu\rho\sigma} M^{\nu\rho} P^\sigma,$$

$$[A B C D] = \epsilon_{\mu\nu\rho\sigma} A^\mu B^\nu C^\rho D^\sigma.$$

The fourteen Hermitian internal invariants form a basis  $\mathcal{B}$  of  $\mathcal{E}_{int}$  in the sense that every other element of  $\mathcal{E}_{int}$  can be written as a linear combination of invariants of the form  $x \cdot y$ ,  $(x \cdot y) \cdot z$ ,  $\dots$ , with  $x, y, z, \dots$  in  $\mathcal{B}$ ; the Jordan product  $x \cdot y$  is defined by

$$x \cdot y \equiv \frac{1}{2} (xy + yx) = y \cdot x = (x \cdot y)^*,$$

and it reduces to the ordinary product  $xy = yx$  whenever  $x$  and  $y$  commute. If  $X \in \mathcal{E}_{int}$ , then  $E_X(\xi) = \exp(-i\xi X)$ ,  $\xi$  real, generates a one-parameter (Abelian) group of unitary transformations or automorphisms of the two-particle Hilbert space  $\mathcal{H}^{(2)}$ , the closure of the set of all linear combinations of states (1). The states transform as follows:

$$\begin{aligned}
 E_X(\xi) |m s p \lambda; m_1 s_1 \lambda_1 m_2 s_2 \lambda_2\rangle \\
 = \sum_{\lambda'_1 \lambda'_2} \langle \lambda'_1 \lambda'_2 | E_X(\xi) | \lambda_1 \lambda_2 \rangle |m s p \lambda; m_1 s_1 \lambda'_1 m_2 s_2 \lambda'_2\rangle
 \end{aligned}$$

One would like to characterize all internal automorphisms of  $\mathcal{H}^{(2)}$  in the hope that knowledge of them may lead to an understanding of the dynamics of particles. After all, the S-matrix is a Lorentz-invariant automorphism of  $\mathcal{H}^{(\infty)}$ , the Hilbert space of an infinite number of particles. As a first step in this program, one must know the commutation relations of the basic internal invariants (3). We first note that  $\mu_1, \mu_2, \omega_1, \omega_2, \mu$ , and  $\nu$  commute with themselves and all the other invariants and may therefore be considered constants. The commutators between the remaining eight invariants  $\bar{\lambda}_{1,2}, \kappa_{1,2}, \beta_{1,2}, \alpha$ , and

$\omega$  have been computed, but will not be reproduced here because of their length; they have the general form

$$[X_i, X_j] = i c_{ij}{}^{kl} X_k \cdot X_l \quad (c's \text{ real}), \quad (4)$$

with  $X_i, X_j, \dots$  in  $\mathcal{B}$ . The elements of  $\mathcal{B}$  obviously do not form a Lie algebra. One hopes that, by suitably enlarging the basis  $\mathcal{B}$  with elements of  $\mathcal{E}_{int}$ , one can arrive at a closed system of invariants forming a finite-dimensional Lie algebra of internal symmetries of the two-particle system. This hope is substantiated by the fact that the Jordan product in Eq. (4) is always of the form  $I \cdot X_l = X_l, \bar{\lambda}_1 \cdot X_l$ , or  $\bar{\lambda}_2 \cdot X_l$ , and by the fact that relations exist between certain Jordan products. We are at present investigating this problem.

## F. Two Variational Principles for London's Equation

M. M. Saffren

London (Ref. 10) proposed two equations which, together with Maxwell's equations, should govern the electrodynamics of superconductors. For stationary states of the superconductor, only one of these equations is germane, namely,

$$\nabla \times \mathbf{v}_s(\mathbf{r}) = -\frac{e}{mc} \mathbf{B}(\mathbf{r}), \quad (1)$$

where  $e$ ,  $m$ , and  $\mathbf{v}_s(\mathbf{r})$  denote the charge, mass, and velocity field, respectively, of the superconducting electrons, and  $\mathbf{B}(\mathbf{r})$  is the magnetic field. It is perhaps interesting, though not surprising, that this equation can be made to follow from a variational principle. In this discussion we describe two such principles.

Of course, aside from any intrinsic interest it may have, the casting of London's equation into the form of a variational principle has the value that approximate solutions of London's equation can be defined and the "best" approximation having a prescribed form can be obtained; this can be quite useful when the geometry of the superconductor is sufficiently complicated to forestall an exact solution.



One of the principles is that the magnetic field realized in a superconductor be such as to render the energy of the superconductor a minimum. This energy is comprised of the kinetic energy of the superelectrons and the energy stored in the magnetic field present in the superconductor. The second variational principle, while it is not as easily described, has an important advantage over the first principle. Unlike the first principle, the second includes the fields outside the superconductor as well as those inside; in this sense, the second is more complete. Moreover, we show this principle to be a natural generalization of the well-known variational principle that holds for static magnetic fields when superconductors are absent.

We now proceed to derive the first variational principle. We suppose that superelectrons find themselves in a lattice whose velocity field we denote as  $\mathbf{v}_l(\mathbf{r})$ . We also suppose that the magnetic field at the surface of the superconductor is given. The lattice motion and external field may be thought of as "sources" of the velocity field,  $\mathbf{v}_s(\mathbf{r})$ , of the superconducting electrons.

The essential step is to express the kinetic energy of the superelectrons in terms of the magnetic field present in the superconductor. To do this, we use the Maxwell equation

$$\nabla \times \mathbf{B} = \frac{4\pi}{c} \mathbf{j}, \quad (2)$$

which holds for stationary fields. We now write  $\mathbf{j}$  as

$$e\rho_s(\mathbf{r})\mathbf{v}_s(\mathbf{r}) - e\rho_l(\mathbf{r})\mathbf{v}_l(\mathbf{r}),$$

where  $e$  is the charge on the electron and  $e\rho_s(\mathbf{r})$  and  $e\rho_l(\mathbf{r})$  are the charge densities of the superelectrons and the lattice (here defined as the rest of the solid), respectively. For simplicity we assume

$$\rho_s(\mathbf{r}) = \rho_l(\mathbf{r}) = \rho = \text{constant}.$$

Then, if  $m$  is the electron mass, the kinetic energy takes the form

$$\frac{1}{2} m \int \frac{1}{\rho} (\rho \mathbf{v}_s)^2 d\tau = \frac{1}{8\pi} \int \left\{ \lambda^2 \nabla \times \mathbf{B} + \frac{mc}{\lambda e} \mathbf{v}_l(\mathbf{r}) \right\}^2 d\tau, \quad (3)$$

where  $\lambda^2 = mc^2/4\pi\rho e^2$ . The total energy in the superconductor,  $E$ , is thus

$$\frac{1}{8\pi} \int \left[ \left\{ \lambda^2 \nabla \times \mathbf{B} + \frac{mc}{\lambda e} \mathbf{v}_l(\mathbf{r}) \right\}^2 + \mathbf{B}^2 \right] d\tau. \quad (4)$$

By setting the variation of this expression equal to zero, we now show that we are led to London's equation. The variation is

$$\frac{1}{4\pi} \int \left\{ (\nabla \times \delta \mathbf{B}) \cdot \lambda \left[ \lambda \nabla \times \mathbf{B} + \frac{mc}{\lambda e} \mathbf{v}_l(\mathbf{r}) \right] + \delta \mathbf{B} \cdot \mathbf{B} \right\} d\tau. \quad (5)$$

Integration by parts yields

$$\frac{1}{4\pi} \int \delta \mathbf{B} \cdot \left[ \lambda^2 \nabla \times \nabla \times \mathbf{B} + \frac{mc}{e} \nabla \times \mathbf{v}_l(\mathbf{r}) + \mathbf{B} \right] d\tau + \frac{1}{4\pi} \oint \delta \mathbf{B} \cdot \left[ \hat{\mathbf{n}} \times \lambda^2 \nabla \times \mathbf{B} + \frac{mc}{e} \mathbf{v}_l(\mathbf{r}) \right] d\sigma, \quad (6)$$

where  $\hat{\mathbf{n}}$  is the unit normal. Since  $\delta \mathbf{B}$  in the interior of the superconductor is arbitrary, then in the interior

$$\lambda^2 \nabla \times \nabla \times \mathbf{B} + \frac{mc}{e} \nabla \times \mathbf{v}_l(\mathbf{r}) + \mathbf{B} = 0, \quad (7)$$

if the variation of  $E$  is to vanish.

This is nothing more than

$$\frac{mc}{e} \nabla \times \mathbf{v}_s(\mathbf{r}) = -\mathbf{B}, \quad (8)$$

which, as we see, holds independent of the motion of the lattice.

How we handle the surface terms in the surface integral depends upon how  $\mathbf{B}$  is prescribed at the boundary. If  $\mathbf{B}$  is prescribed at the boundary by requiring it to be continuous with some fixed external field, then, of course,  $\delta \mathbf{B} = 0$  at the boundary. If  $\mathbf{B}$  at the surface is free to take on a value that makes the energy a minimum, then

$$\hat{\mathbf{n}} \times \left[ \lambda^2 \nabla \times \mathbf{B} + \frac{mc}{e} \mathbf{v}_l(\mathbf{r}) \right] = 0 \quad (9)$$

on the surface, since  $\lambda^2 \nabla \times \mathbf{B} + (mc/e) \mathbf{v}_l(\mathbf{r})$  is nothing more than  $(mc/e) \mathbf{v}_s(\mathbf{r})$ . The demand that the energy in the *superconductor alone* be a minimum is seen to require that  $\mathbf{v}_s(\mathbf{r})$  vanish on the surface of the superconductor. However, this condition may give rise to external magnetic fields, and, therefore, the *total energy* of the system comprised of the superconductor and its external fields is not a minimum, even though the energy of the superconductor is. In this variational principle the fields external to the superconductor enter only to supply boundary conditions for  $\mathbf{B}$  on the surface of the superconductor. Because the external field does not enter directly into the principle, we must be sure when we

supply the boundary fields that they do come from some solution of  $\nabla \cdot \mathbf{B} = 0$ ,  $\nabla \times \mathbf{B} = 0$ , which are Maxwell's equations for  $\mathbf{B}$  in free space.

We now set up a variational principle that frees us from this constraint. This principle also determines the field external to the superconductor and so allows us to use approximate instead of exact fields not only inside the superconductor but also outside. This permits us to guess the entire field configuration throughout all space. The principle consists of making stationary, with respect to variations in  $\mathbf{A}$ , the functional

$$\int_{\text{all space}} (\nabla \times \mathbf{A})^2 d\tau + \int_{\text{superconductor}} \left\{ \frac{1}{\lambda^2} \left[ \mathbf{A} + \frac{mc}{e} \mathbf{w}_l(\mathbf{r}) \right]^2 - \left[ \frac{mc}{\lambda e} \mathbf{w}_l(\mathbf{r}) \right]^2 \right\} d\tau, \quad (10)$$

where  $\mathbf{w}_l(\mathbf{r})$  is the transverse part of  $\mathbf{v}_l(\mathbf{r})$ . The conditions on  $\mathbf{A}$  are that it be continuous across boundaries and that, unless there is a magnetic field at infinity,  $\mathbf{A}$  vanishes there. The field  $\mathbf{A}$  is the vector potential; i.e.,  $\mathbf{B} = \nabla \times \mathbf{A}$ . Variation of  $\mathbf{A}$  leads to the requirement that, if  $\mathbf{A}(\mathbf{r})$  makes the functional stationary, then the tangential components of  $\nabla \times \mathbf{A}$  are continuous across boundaries (not surprising), and

$$\nabla \times \nabla \times \mathbf{A} = 0 \text{ in space;} \quad (11a)$$

$$\lambda^2 \nabla \times \nabla \times \mathbf{A} + \frac{mc}{e} \mathbf{w}_l(\mathbf{r}) + \mathbf{A} = 0 \text{ in the superconductor.} \quad (11b)$$

On taking the curl of the last equation, we obtain London's equation.

If there are external sources or a field at infinity, the principle must be modified. When  $\mathbf{B}_\infty$  is the field at infinity, one replaces

$$\int (\nabla \times \mathbf{A})^2 d\tau$$

by

$$\int [(\nabla \times \mathbf{A})^2 - (\mathbf{r} \times \mathbf{B}_\infty) \cdot \nabla \times \mathbf{A}] d\tau.$$

More interesting is the case where there are external transverse currents,  $\mathbf{j}_{ext}$ , which act as external sources of the field; then,

$$\int_{\text{all space}} (\nabla \times \mathbf{A})^2 d\tau$$

must be replaced by

$$\int_{\text{all space}} \left[ (\nabla \times \mathbf{A})^2 - \frac{8\pi \mathbf{j}_{ext} \cdot \mathbf{A}}{c} \right] d\tau.$$

In fact, if we put the current due to the lattice motion into  $\mathbf{j}_{ext}$ , we can write the variational principle quite generally as:

$$\int_{\text{all space}} \left[ (\nabla \times \mathbf{A})^2 - \frac{8\pi \mathbf{j}_{ext} \cdot \mathbf{A}}{c} \right] d\tau + \int_{\text{superconductor}} \frac{1}{\lambda^2} \mathbf{A}^2 d\tau. \quad (12)$$

Actually, in this form the variational principle is easily seen to be proportional to the negative of the total magnetic energy in a system comprised of sources and superconductors. To see this, we write the above expression as

$$\int_{\text{all space}} \left\{ (\nabla \times \mathbf{A})^2 - \frac{8\pi}{c} \left[ \mathbf{j}_{ext} - \frac{c}{8\pi} \frac{\mathbf{A}_{sup}}{\lambda^2} \right] \cdot \mathbf{A} \right\} d\tau \quad (13)$$

where

$$\mathbf{A}_{sup} = \begin{cases} \mathbf{A} & \text{inside the superconductor} \\ 0 & \text{outside the superconductor} \end{cases}$$

However, since in the superconductor

$$\mathbf{j}_s = -\frac{c}{8\pi} \frac{\mathbf{A}}{\lambda^2}, \quad (14)$$

we see that our variational principle is

$$\int \left\{ (\nabla \times \mathbf{A})^2 - \frac{8\pi}{c} \mathbf{j} \cdot \mathbf{A} \right\} d\tau, \quad (15)$$

a well-known expression, except—and this is the essential point—we have replaced  $\mathbf{v}_s$  by  $-(e/mc)\mathbf{A}$ . Thus, part of  $\mathbf{j}$  takes part in the variation, whereas in the usual expression  $\mathbf{j}$  is regarded as fixed and is not varied.

(2)

## References

1. Barth, C. A., "Rocket Measurement of the Nitric Oxide Dayglow," SPS 37-26, Vol. IV, pp. 169-171, Jet Propulsion Laboratory, Pasadena, California, April 30, 1964.
2. Heindl, C. J., *Efficiency of Fission-Electric Cells*, Technical Report No. 32-105, Jet Propulsion Laboratory, Pasadena, California, May 25, 1961.
3. Weinberg, A. M., and Wigner, E. P., *The Physical Theory of Neutron Chain Reactors*, p. 132, The University of Chicago Press, 1958.
4. Estabrook, F., and Wahlquist, H., "Dyadic Formalism and the Problem of Rigidity in Relativity," SPS 37-25, Vol. IV, pp. 265, 266, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.
5. Synge, J. L., *Relativity: The General Theory*, North-Holland Publishing Company, Amsterdam, 1960.
6. Pauli, W., *Probleme der Modernen Physik*, p. 30 (Festschrift zum 60 Geburtstag A. Sommerfelds, herausgegeben von P. Debye), Hirzel, Leipzig, 1928.
7. Heitler, W., *The Quantum Theory of Radiation*, Clarendon Press, Oxford, 1957.
8. Zmuidzinas, J. S., "Relativistic Many-Particle States," SPS 37-26, Vol. IV, pp. 174, 175, Jet Propulsion Laboratory, Pasadena, California, April 30, 1964.
9. Wigner, E. P., *Annals of Mathematics*, Vol. 40, pp. 149-204, 1939.
10. London, Fritz, *Superfluids*, Vol. I, Dover Publications, New York, 1961.

**BLANK PAGE**

## TELECOMMUNICATIONS DIVISION

## XVII. Communications Elements Research

## A. Low Noise Amplifiers

C. T. Stelzried, W. V. T. Rusch, R. Brantner,  
and R. Clauss

## 1. Millimeter Wave Work, C. T. Stelzried and W. V. T. Rusch

*a. Summary.* The objective of the millimeter wave work is to investigate millimeter wave components and techniques to ascertain the future applicability of this frequency range to space communications and tracking. This involves the development of instrumentation for accurate determination of insertion loss, VSWR, power and equivalent noise temperature of passive elements, and gain and bandwidth of active elements at millimeter wavelength.

In order to bring together state-of-the-art millimeter wave circuit elements and evaluate their use in a system, we have built a radio telescope consisting of a 60-in. antenna and a super-heterodyne radiometer. The radio telescope was used to observe the 90-Gc temperature of the Moon during the December 30, 1963 eclipse (Refs. 1 and 2). This experiment was a joint effort by personnel from JPL and the Electrical Engineering Department of the University of Southern California.

*b. Recent work.*

*90-Gc radiotelescope observations.* The 90-Gc radiometer is currently undergoing improvements and modification. Lunar observations were made from the JPL Mesa antenna range during the period April 13 to 27 in order to determine the present measurement capability in the field.

Prior to the observation, the antenna was refocused (Fig. 1) and antenna patterns were made [Fig. 2(a)]

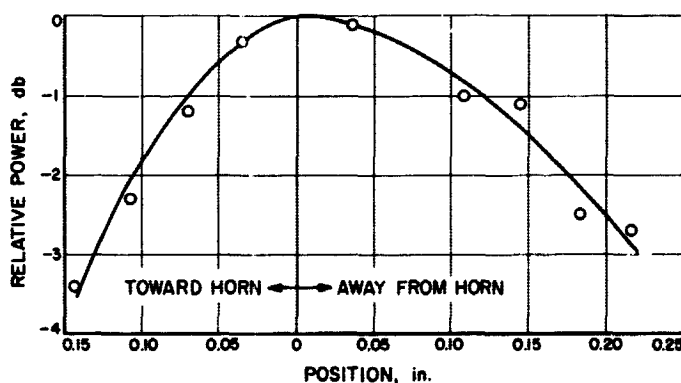


Fig. 1. 90-Gc antenna subreflector focusing

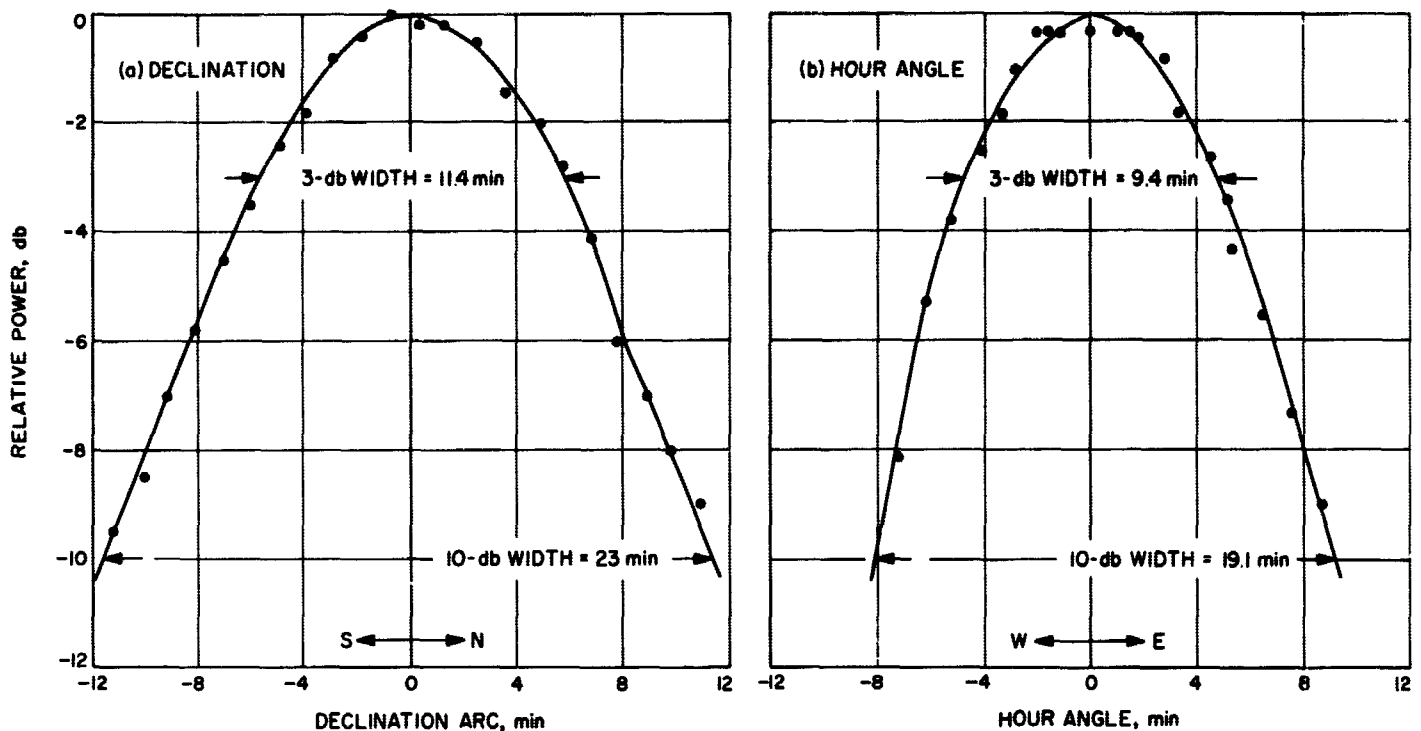
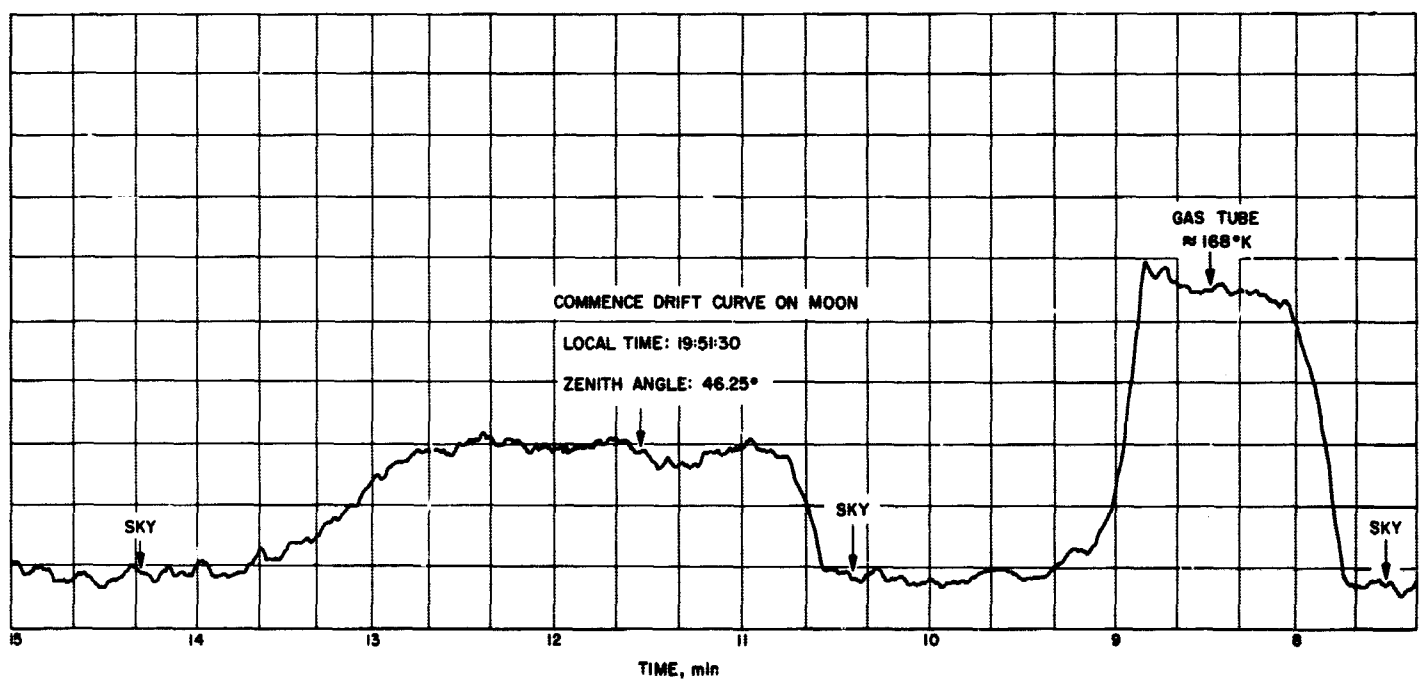


Fig. 2. 90-Gc antenna patterns: (a) declination and (b) hour angle

and (b)]. The patterns were measured on a 0.8-mi-long range using a 10-mw continuous wave (CW) klystron source. Considerable difficulty was experienced with instability while attempting to take patterns using the

radiometer receiver. This was due primarily to klystron frequency instability in conjunction with the narrow bandwidth radiometer (10 Mc). In order to resolve the frequency instability trouble, the klystron was modu-



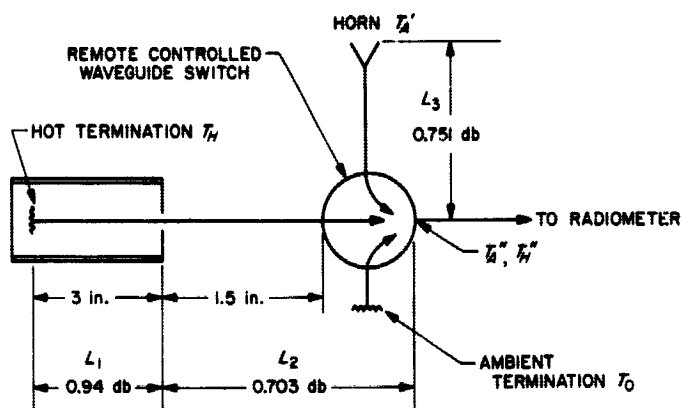


Fig. 3. 90-Gc radiometer waveguide calibrations

lated at 1000 cps, the received signal was video detected, and the 1000-cps signal amplified in a narrow-band amplifier. With this arrangement it was possible to obtain 10 db of dynamic range with better than 0.5-db resolution. A rotary vane attenuator in the transmitter line was used to calibrate the relative power.

The manual waveguide calibration switch has been replaced with a remote controlled waveguide switch, and the waveguide has been recalibrated as shown in Fig. 3. The losses to the hot load are high, because of the unplated stainless steel waveguide (approximately 0.313 db/in. at 90 Gc). Plating techniques are being tested.

When a suitable technique is available the hot load guide will be plated; this will reduce the loss and inaccuracy.

As shown on Fig. 3, the equivalent noise temperatures are denoted by a double prime at the output of the waveguide switch and the antenna temperature by a single prime at the horn input. Using the method of Ref. 3, we have, at the waveguide switch output, the reference temperatures

$$\begin{aligned} (T_H'' - T_0) &= (T_H - T_0) \frac{L_1 - 1}{0.2303 L_1 (\text{db}) L_1 L_2} \\ &= (T_H - T_0) (0.76) \end{aligned} \quad (1)$$

Using this calibration, the measured Moon temperature defined at the waveguide switch output must be multiplied by the loss  $L_3$  (0.751 db) to refer to the horn input.

During a large percentage of the time during the 2 wk of observation, it was either raining or cloudy and overcast. The antenna was not set in the position of a polar mount for these measurements. The measurement technique consisted of optically setting the antenna to the center of the Moon and letting the Moon drift out of the antenna patterns giving  $\frac{1}{2}$  a normal drift curve. A typical curve taken in this manner at 19:44:00 local time on April 16, 1964 is shown in Fig. 4. The sky was cloudy. The radiometer had 10-Mc bandwidth and a recording time constant of 10 sec.

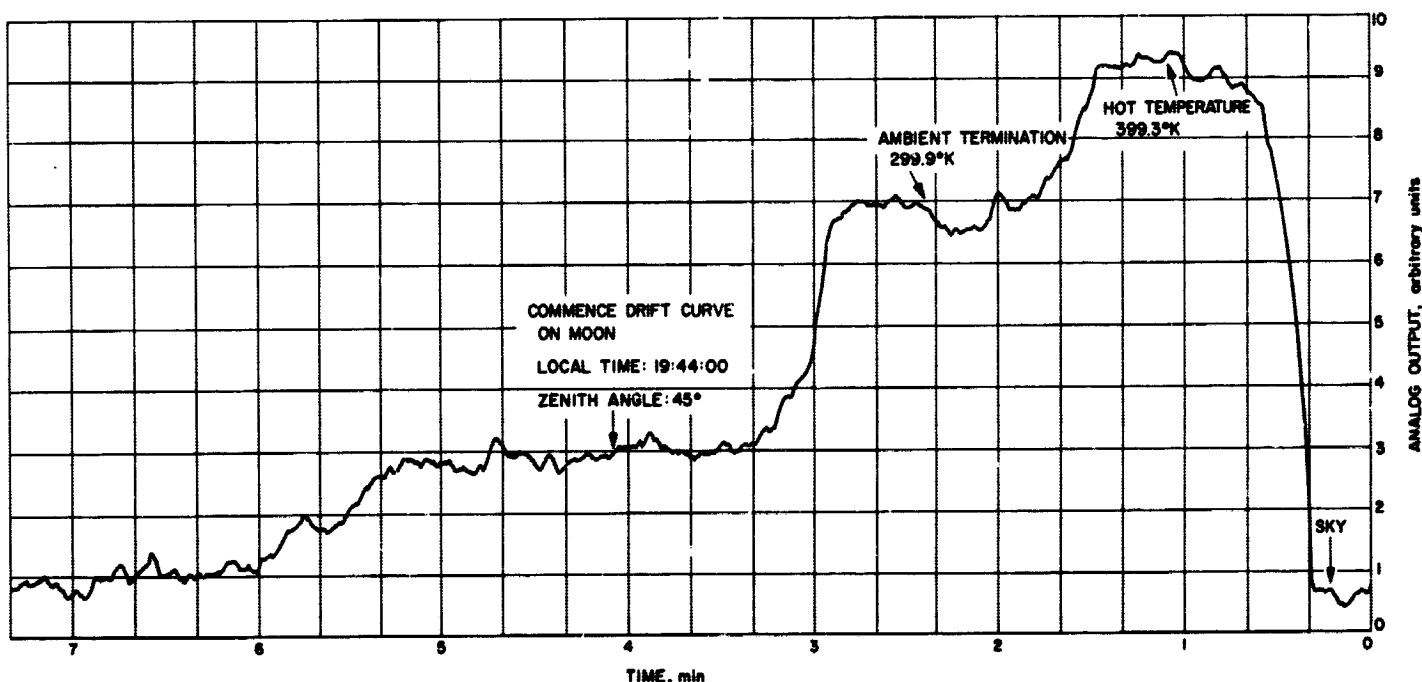


Fig. 4. 90-Gc radiotelescope Moon drift curves (April 16, 1964) from JPL Mesa



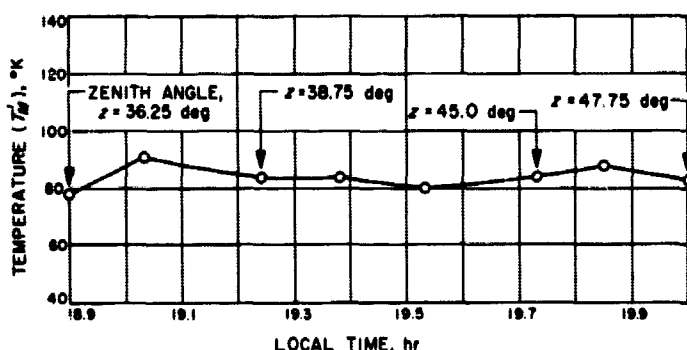


Fig. 5. 90-Gc Moon temperature measured at horn input (April 16, 1964)

Some failures have been experienced with the 90-Gc mixer crystals. With the present unmatched diodes, the receiver equivalent noise temperature is approximately 20,000°K. The theoretical rms stability for a Dicke radiometer with square wave modulation is approximately

$$\Delta T \approx \left[ \frac{2}{(10 \text{ sec})(10 \text{ Mc})} \right]^{1/2} 20,000^\circ\text{K} = 3^\circ\text{K} \quad (2)$$

The reduced Moon temperature data calibrated to the horn input is shown in Fig. 5 for eight drift curves. The average temperature over the measurement period is 84.5°K with a standard deviation of 6.5°K.

## 2. Solid State Circuits, R. Brantner

**a. Summary.** An intermediate frequency amplifier using field effect transistors (FET) is being studied for possible application as a low noise amplifier where space and power requirements preclude use of vacuum tube amplifiers. Experimental models of a 10-Mc amplifier have been built and tested.

**b. Recent work.** As reported, a first model of the amplifier was built and tested. Results indicated the need for much better shielding and inter-stage signal isolation. Accordingly, a carefully shielded unit was constructed incorporating precision mechanical construction for adequate shielding and very efficient RF filtering in the power distribution lines. Using this unit, careful measurements indicate that part of the very high gain previously reported can be attributed to regeneration between stages. In the new model, the amplifier proper has been set for a gain of approximately 30 db and a bandwidth of approximately 3.5 Mc. The input matching network was then adjusted for a bandwidth of 1.0 Mc; at this setting, the gain was 52 db and the noise figure (NF) was 6.0 db.

Attempts have been made to operate the unit at varying supply voltages but at B+ voltages in the 8- to 12-v range, both the bandwidth and gain decreased substantially. Consequently, tests are being made using a power input of 16 v at about 40 ma. Under these conditions, all RCA Type TA 2330 FETs now on hand have been tested in the input stage for over-all NF and first-stage gain. The results are as follows:

Characteristics	Minimum	Average	Maximum
Over-all NF, db	5.9	8.5	13.6
First-stage gain, db	-0.7	5.8	7.8

The majority of the FETs on hand have a gain between 4.5 and 7.0 db, with over-all NF between 6.6 and 10.1 db. A significant point is that many FETs have the same gain but produce as much as 2.8- to 3.0-db variation in the NF. This indicates the necessity of selecting FETs for both gain and NF until units of far more uniform characteristics are available. All units gave essentially the same bandwidth of 1.0 Mc.

Since a gain of 30 db would be adequate for use in a low noise preamplifier, the main difficulty, at present, is in broad-banding the input matching network to obtain a usable bandwidth with a good NF. Work is proceeding along these lines. Also, studies are in progress which may make possible greater amplifier bandwidth, with no sacrifice of gain, by using modified interstage coupling networks. All tests to date have been performed using a 50-ohm source and 50-ohm load. Past experience with IF amplifiers indicates that broad-band inputs are far easier to achieve with the 300- to 400-ohm source impedance of a diode mixer. This is another area of investigation in the improvement of the FET IF amplifier.

## 3. Microwave Noise Source, C. T. Stelzried and R. Claus

**a. Summary.** The principal requirements on a gas tube noise source for calibrating a microwave receiving system are: low insertion loss, good short- and long-term stability of the injected noise, long life, and ease of operation. Present techniques used at JPL utilize a noise tube on the side arm of a directional coupler to inject a noise pulse of the desired amplitude. The excess noise added to the system by the unfired source adds a fair amount of noise (from 3 to 0.75°K, typically) which leads to very large decoupling factors, up to 26 db. The net result is a noise pulse which is sometimes too small for proper system calibration. Short-term instability of the

excess noise from commercial microwave noise sources has been measured (Ref. 4). Short-term instability can be traced to changes in environmental temperature, mechanical vibration, and voltage supply. Electrode and gas impurity and previous history such as electrical overloads contribute to long-term instability. In an operational configuration, long cable runs between the noise source and its power supply are sometimes necessary. As a result, difficulty has been experienced with the high-voltage pulsing required to excite the noise source. Some noise sources have had a notable short life under operational conditions.

A commercial neon bulb Type NE-2H has been used to fabricate an experimental noise source. Preliminary measurements indicate a low insertion loss and a possible application in future receiving systems.

**b. Recent work.** A commercial neon bulb Type NE-2H has been used to fabricate an experimental WR 430 waveguide noise source (Figs. 6 and 7). In this design, the tube is inserted directly in the waveguide. The small tube has the advantage that a high-voltage pulse is not needed for excitation due to the proximity of the electrodes. Laboratory measurements were made at 3 ma of current using a 180-v battery with a series dropping resistor. With the bulb inserted 0.33 in. into the center of the waveguide E-field, the excess noise at 2295 Mc was approximately  $20^\circ\text{K}$ —a useful value for use in a very low temperature system. The VSWR was 1.02, and the insertion loss with the tube not fired was 0.0066 db (which would contribute less than  $0.5^\circ\text{K}$  to a receiving system noise temperature). The insertion loss without the tube installed was about 0.005 db. Firing the tube resulted in an increase in the insertion loss of 0.007 db. The insertion loss  $L$  through a gas relates the equivalent noise temperature  $T_g$  of the gas to the equivalent excess noise temperature  $T_e$  by

$$T_e = T_g \left( 1 - \frac{1}{L} \right) \approx T_g \frac{L(\text{db})}{4.343} \quad (1)$$

This indicates an approximate equivalent noise temperature for the gas in a NE-2H tube with a 3-ma current of

$$T_g \approx 20^\circ\text{K} \frac{4.343}{0.007} = 6200^\circ\text{K}$$

The excess noise in this arrangement was measured to be almost directly related to the tube current, so that a well regulated power supply is needed. A 1% excess noise stability therefore requires better than 1% power supply

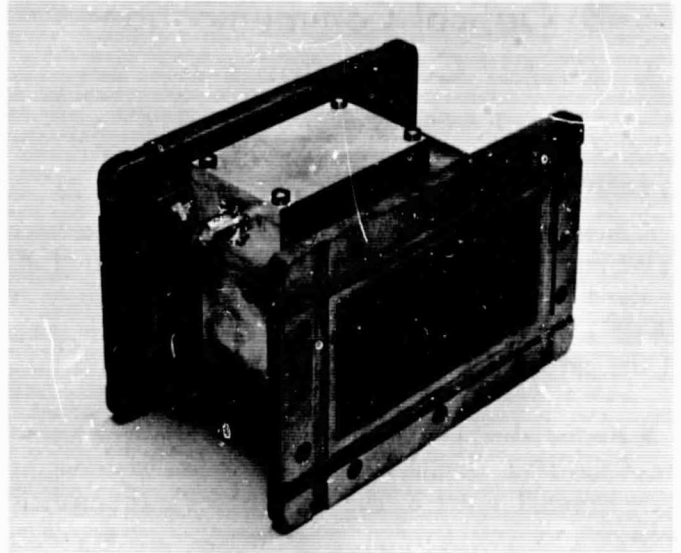


Fig. 6. Experimental WR 430 waveguide noise source showing location of noise source

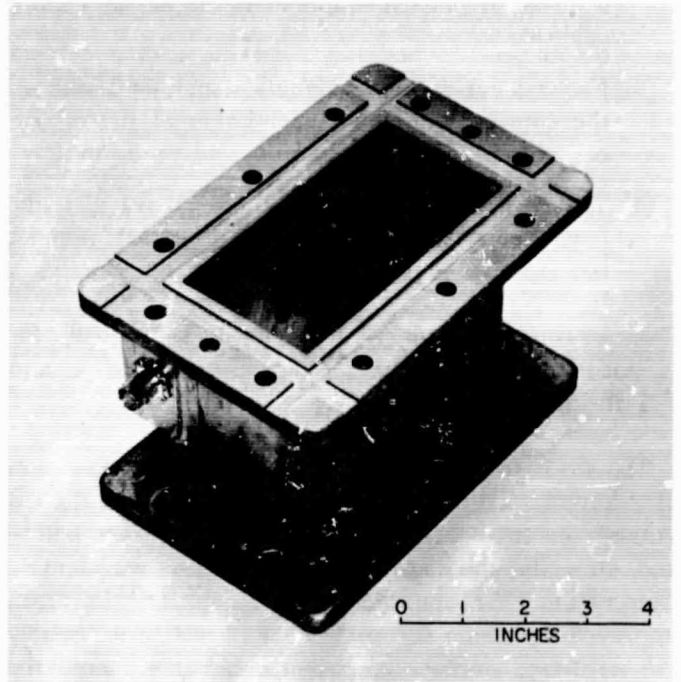


Fig. 7. External view of experimental WR 430 waveguide noise source

current stability. Commercial power supplies are available with stabilities better than 0.01%.

Future work on these noise sources will include measurements of leakage and output stability. Also, designs to minimize insertion loss and VSWR will be studied.

## B. Optical Communications Components

J. Siddoway and J. Maserjian

### 1. Lasers, J. Siddoway

**a. Summary.** A modification of the traveling wave laser (TWL) amplifier has increased the gain by about 0.7 db at the 1.15- $\mu$  wavelength. The amplifier is still lossy at 0.6328  $\mu$ , although considerable improvement in the gain of the gas is observed.

**b. Recent work.** Performance characteristics and analysis of the TWL amplifier for three spectral wavelengths were reported in Ref. 5. The major points in the analysis were:

- (1) The gain of the gas in a He-Ne system depends upon the radius of the tube ( $\approx 1/r$ ).
- (2) For very high-gain wavelengths (such as 3.39  $\mu$ ), a linear single-pass amplifier is simpler and more efficient.
- (3) For very low-gain wavelengths (such as 0.6328  $\mu$ ), the gain is limited by the number of passes one can obtain with the optics.
- (4) The TWL amplifier is best suited for wavelengths (e.g., 1.15  $\mu$ ) where one can obtain several percent net gain per pass.

Gain measurements reported previously were 40 db for 12 passes at 3.39  $\mu$ , 7 db for 56 passes at 1.15  $\mu$ , and -3 db for 40 passes at 0.6328  $\mu$ .

The TWL amplifier has been modified by the insertion of a longitudinal spacer to decrease the narrow dimension of the tube. Fig. 8 shows the gain curves before and after the modification for the 1.15- $\mu$  wavelength. Gain of the TWL amplifier at this wavelength has been improved by about 0.7 db for 44 passes. This is now the maximum number of passes obtainable before the TWL amplifier begins to oscillate due to diffraction coupling. The amplifier is still lossy at the 0.6328- $\mu$  wavelength, although the measured net loss is now  $\approx 0.5$  db for 48 passes versus 3-db loss for 40 passes before modification. Performance characteristics for the 3.39- $\mu$  wavelength, since modification of the amplifier, are not yet conclusive. The narrow dimension of the TWL amplifier is now comparable to the diameter of the laser oscillator tubes; further improvement in gain characteristics may result from the use of bundles of small bore tubes as discussed in Ref. 5.

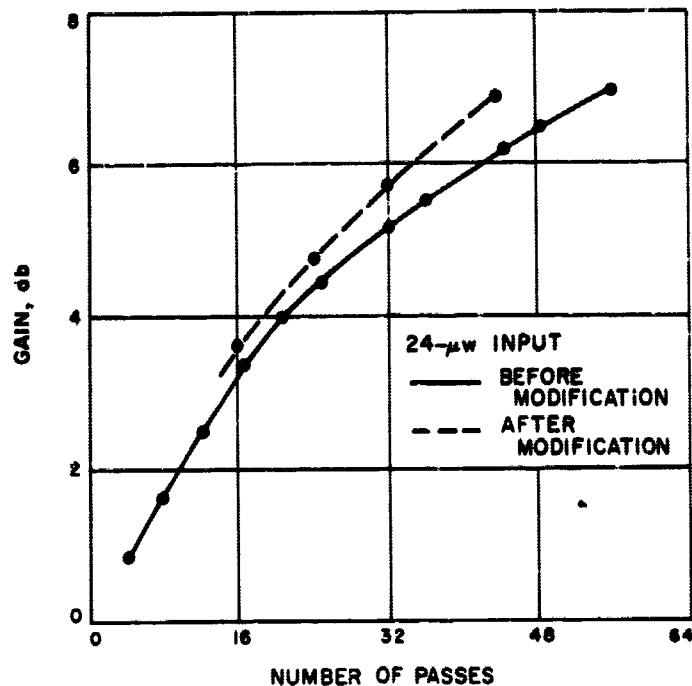


Fig. 8. Gain versus number of passes for 1.15- $\mu$  wavelength in He-Ne

### 2. Superconducting Bolometers, J. Maserjian

**a. Summary.** Work on a superconducting bolometer utilizing a vacuum deposited film of tantalum has been described in previous reports (Ref. 6). In order to make a fair appraisal of the feasibility of such a device for infrared detection, one should consider only an optimized design. The design theory depends upon the thermal, optical, and electrical properties of the device. The thermal and optical properties can be treated in a general way independent of the specific electrical properties and will be the subject of this report. It is anticipated that electrical measurements in progress will provide the necessary information for completing the device evaluation in a subsequent report.

#### b. Recent work.

**Thermal considerations in the design of superconducting bolometers.** The traditional design of bolometers attempts to achieve a high degree of thermal isolation for the bolometer element, as might be obtained with only radiation exchange of energy. This provides the maximum sensitivity of the bolometer, but only at extremely low modulation frequencies of the radiation (typically much less than 1 cps). For reasonable frequencies essential for communication applications, the

thermal transfer may in general be described as a solution of the thermal diffusion equation satisfying a given geometry. That is,

$$\frac{\partial T}{\partial t} = \kappa \nabla^2 T \quad (1)$$

It is tacitly assumed that thermal convection is avoided, since it will contribute to excessive noise and reduce the thermal response. The usual treatment is to assume parallel heat flow in which the thermal diffusion length is short compared with the length of the conducting medium attached to a heat sink so that only a fixed thermal conductance  $K$  need be considered (Ref. 7). In this report, the general solution is presented for a harmonic response assuming parallel heat flow into a semi-infinite medium. This solution predicts the optimum conditions required for maximum response at a given frequency.

The assumption of parallel heat flow follows from a requirement believed essential for superconducting bolometers and is not merely a mathematical simplification. It is essential that the superconductor be at a uniform temperature (isothermal condition) in order to avoid movement of boundaries between the normal and superconducting state along temperature gradients. Indeed the nonisothermal case has been the subject of a special kind of superconducting bolometer (Ref. 8). However, because of the instabilities believed inherent in boundary movement and the associated excess noise (Refs. 9, 10), the isothermal requirement appears most reasonable. In the situation where a superconducting film is deposited on a flat surface, the isothermal condition at the film is equivalent to a requirement of parallel heat flow into the substrate medium. The geometry may then be as shown in Fig. 9 where the superconductor film of thickness  $l_1$  [region (1)] is at the end of a long cylinder [region (2)] of area  $A$  attached to a heat sink. The cylinder, if placed in a vacuum, would experience negligible

radiation loss compared with heat conducted along the cylinder.

If the radiation absorbed at the film is  $Q_0 + Q \sin \omega t$  and the film is maintained at some temperature in the superconducting transition region  $T_1$  by means of a heater bias (Ref. 6), the average power  $Q_0$  may be ignored, incorporating it into the heater bias. We then solve for the harmonic steady-state response at the surface. It is assumed that  $l_1$  is much shorter than the diffusion length in region (1); that is,

$$l_1 < (\kappa_1/\omega)^{1/2}, \quad (2)$$

where  $\kappa_1 = k_1/\rho_1 c_1$ , and  $c, \rho, k$  are the specific heat, density, and conductivity, the index referring to region (1), so that the entire film may be considered at a uniform temperature  $T_1 + \Delta T_1$ . Applying the boundary condition  $Q \sin \omega t = Q_1 + Q_2$  (where  $Q_1, Q_2$  are the harmonic components of heat flux flowing into regions (1) and (2), respectively) to the steady-state solution of Eq. (1) given for region (2) alone (Ref. 11), we obtain,

$$\Delta T_1 = \Delta T \cos(\omega t - \phi) \quad (3)$$

where

$$\phi = \tan^{-1} [1 + (2\omega/\omega_0)^{1/2}]$$

$$\Delta T = \frac{Q/A}{C_1 \omega [(1 + 2\omega_0/\omega)^{1/2} + (\omega_0/\omega)]^{1/2}} \quad (4)$$

$$\omega_0 = \frac{k_2 c_2 \rho_2}{C_1^2} \quad (5)$$

$C_1 = c_1 \rho_1 l_1$ , the heat capacity per unit area of region (1).

The amplitude of the harmonic temperature response  $\Delta T$  described by Eq. (4) is plotted in Fig. 10 for an arbitrary  $Q/A, C_1, \omega_0$ . (The solid curve represents response of long rod; dotted curves (a) and (b) represent

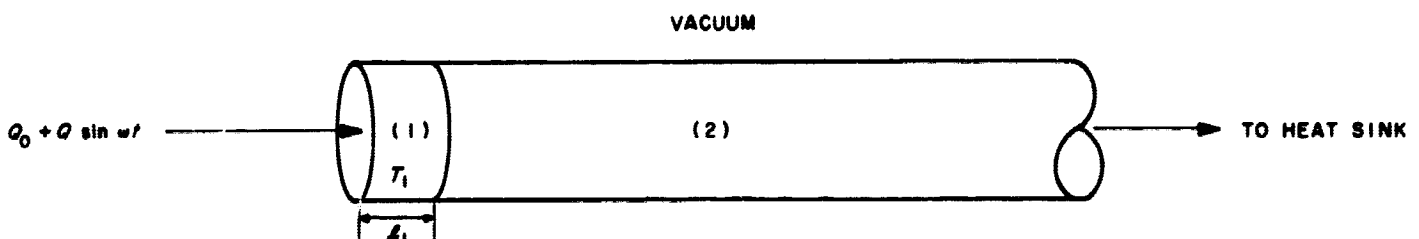


Fig. 9. Bolometer geometry assumed in thermal calculations

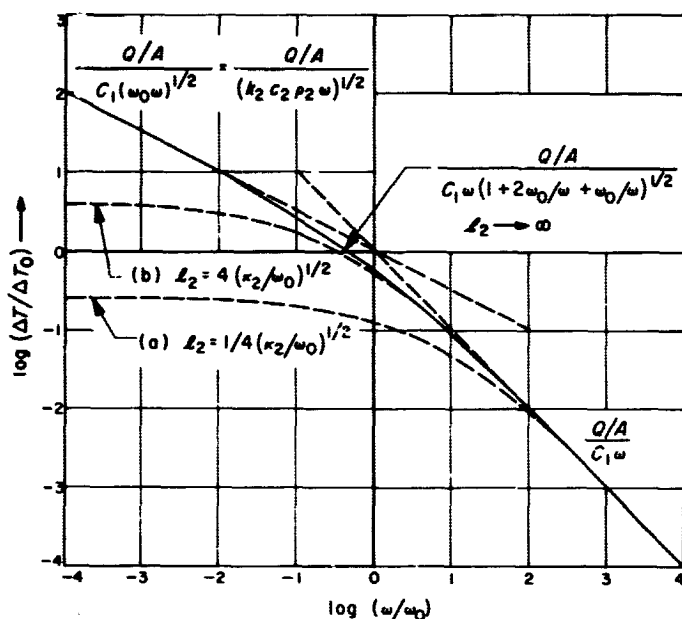


Fig. 10. Temperature response of bolometer

response of rods with length  $l_2$  indicated.) The asymptotic forms are

$$\Delta T = \frac{Q/A}{C_1(\omega_0\omega)^{1/2}} = \frac{Q/A}{(k_2C_2\rho_2\omega)^{1/2}}, \quad \omega \ll \omega_0 \quad (6)$$

$$\Delta T = \frac{Q/A}{C_1\omega}, \quad \omega \gg \omega_0 \quad (7)$$

Eq. (6) is the limiting case where the heat capacity of the film region (1) is negligible compared with diffusion into region (2). Eq. (7) is the limit where only the heat capacity of the film is important. The transition frequency  $\omega_0$  represents the condition where the energy stored in the film equals the energy stored in region (2).

Clearly the temperature response is always higher at lower frequencies. However, let us assume that we desire to operate near some frequency which we call  $\omega_1$ . Then the best we can do is to use the minimum value of  $k_2C_2\rho_2$  applicable to the bolometer and insist that  $\omega_0 \geq \omega_1$ . From Eq. (5), the latter condition requires that  $C_1 \leq (k_2C_2\rho_2/\omega_1)^{1/2}$  which is easily met with thin films at moderate frequencies. If we assume  $C_1$  to be the minimum value applicable to the bolometer, then  $\omega_0$  represents a practical upper frequency limit on the bolometer without making an excessive sacrifice in response.

The solid curve shown in Fig. 10, if considered to represent the minimum applicable values of  $C_1$  and  $k_2C_2\rho_2$ , will envelop all possible response curves based on different parameters. Most significantly, this includes any

response with a short conducting medium, which is the case usually considered. This case is illustrated by the dotted curves (a) and (b) where the same parameters are assumed except that the length  $l_2$  of region (2) is shortened. In curve (a),  $l_2$  equals one-fourth the diffusion length at  $\omega_0$ , i.e.,

$$l_2 = \frac{1}{4}(\kappa_2/\omega_0)^{1/2} = \frac{C_1}{4\rho_2C_2},$$

and in curve (b),  $l_2$  equals four times the diffusion length at  $\omega_0$ , i.e.,

$$l_2 = 4(\kappa_2/\omega_0)^{1/2} = \frac{4C_1}{\rho_2C_2}.$$

The plateau is given in each case by  $Q/A/k_2/l_2$ . Clearly the optimum response can only be achieved with a conducting rod a few times longer than the diffusion length at the operating frequency. An additional advantage of a long rod is the greater isolation of the bolometer from the thermal fluctuations of the heat sink (helium bath). This will be discussed further in a subsequent report which will consider bolometer noise.

An upper limit on the rod length  $l_2$  is determined by the gradient that can be supported between the superconductor and heat sink. This is given by

$$(l_2)_{\max} = \frac{k_2(T_1 - T_0)}{Q_0/A}$$

where  $T_0$  is the bath temperature and  $Q_0$  is the average absorbed radiation. In practice, at low intensity levels, this limit would impose no problem.

It is of interest to calculate values of  $f_0 = \omega_0/2\pi$  for some typical parameters. The parameters at 4.2°K along with the result are given below:

Conducting medium	$k_2C_2\rho_2 \frac{\text{Joules}^2}{\text{sec-cm}^2 \text{ deg}^2}$	$f_0$ , cps
Sapphire	$1 \times 10^{-4}$ (estimated)	$6 \times 10^9$
Glass	$5 \times 10^{-7}$	$3 \times 10^7$

The value used for  $C_1$  is  $5 \times 10^{-8}$  joules/deg cm<sup>2</sup> for a 1000 Å film of tantalum. The values of  $f_0$  are seen to be quite large and would not be exceeded by any reasonable operating frequency. This permits us to increase  $C_1$  if necessary without decreasing the response at lower frequencies. This is useful in considerations for optimizing absorption of the radiation described briefly below. Glass was chosen as a possibility because it has one of the

lowest values of  $k_2 c_2 \rho_2$  at 4.2°K and may be applicable as a substrate for the tantalum film. However, the use of ultra-high vacuum techniques would be mandatory in this case and is presently being investigated. Sapphire has been used in previous samples because it permits use of a conventional vacuum system in the preparation of the films (Ref. 6). However, the thermal response of sapphire is more than one order of magnitude less than that of glass ( $\omega < \omega_0$ ). An additional incentive for investigating the use of ultra-high vacuum techniques and glass substrates is based on expected improvements in the electrical performance of the bolometer.

*Optical considerations in the design of superconducting bolometers.* No mention has been made of the area  $A$ . Clearly, the smaller  $A$  can be made for a given  $Q$ , the larger the response. A practical lower limit is considered to be the order of 1-mm diameter. This is a reasonable limit when one considers problems involved in focusing of the incident radiation, as well as the physical size requirements on the electrical resistance of the element.

The radiation impinging on the bolometer should ideally be completely absorbed. Interference methods for accomplishing this have been described previously (Ref. 12). Although the discussion in Ref. 12 describes only partial interference from an anodized film, the method can be carried to the limit of complete absorption by suitably adjusting the surface reflectance at the oxide-air interface. This can be done, for example, by coating the oxide with a partially transparent film of metal. The interference method will increase the total value of  $C_1$  used in the previous analysis; but as pointed out this will not matter until the transition frequency  $f_0$  approaches the operating frequency. This is extremely unlikely in this case, since the change in  $C_1$  would at most be only about a factor of two.

## C. Antennas for Space Communications

P. D. Potter

### 1. Antenna Feed Research: Spherical Wave Functions

*a. Summary.* In the previous reportings, the theory of spherical waves was reviewed and applied to paraboloidal antenna feed systems. Members of a class of physically

realizable feed system radiation patterns were expanded as finite sums of spherical wave functions; this class of patterns has the property of resulting in azimuthally symmetrical aperture illumination with no cross polarization. The maximum realizable order  $N$  of the wave functions and the corresponding maximum achievable paraboloid performance were presented in terms of the paraboloid diameter in wavelengths.

The problem considered in this reporting is that of utilizing the spherical wave formulation to provide an analytical synthesis technique for nonoptical Cassegrain subreflectors with maximum possible performance. The synthesis formulation is presented, and preliminary numerical results are compared with a previous empirically derived, nonoptical Cassegrain subreflector. The method presented is a rigorous boundary value solution to the problem.

*b. Recent results.* Fig. 11 shows the assumed geometry of the Cassegrain-type paraboloidal antenna feed system. For the antenna system to operate efficiently, the wave from the feedhorn must be transformed by the subreflector into a wave which is essentially spherical in the vicinity of the paraboloid. It is generally (and incorrectly) assumed that, because the desired wave scattered from the subreflector is essentially spherical at the paraboloid surface, it must be spherical at the subreflector surface. This is, in fact, only true in the limit of vanishing wavelength. In this reporting, a general equation for the desired surface is obtained; for the case of a spherical wave from the feedhorn and vanishing wavelength, the surface equation reduces to that of a hyperboloid.

The total vector field,<sup>1</sup>  $E_T$ , in the vicinity of the subreflector is the sum of the incident field,  $E_H$ , from the feedhorn and the scattered field,  $E_s$ .

$$E_T = E_H + E_s \quad (1)$$

A general horn pattern,  $E_H$ , is assumed as follows:

$$E_H \triangleq - \frac{[F_{H\gamma}(\gamma) \sin \xi a_\gamma + F_{H\epsilon}(\gamma) \cos \xi a_\epsilon]}{kr} \times \exp \left\{ -jk \left[ r - 2a - \frac{\delta_H(\gamma)}{k} \right] \right\} \quad (2)$$

<sup>1</sup>The harmonic time dependence,  $e^{j\omega t}$ , is suppressed throughout this reporting.

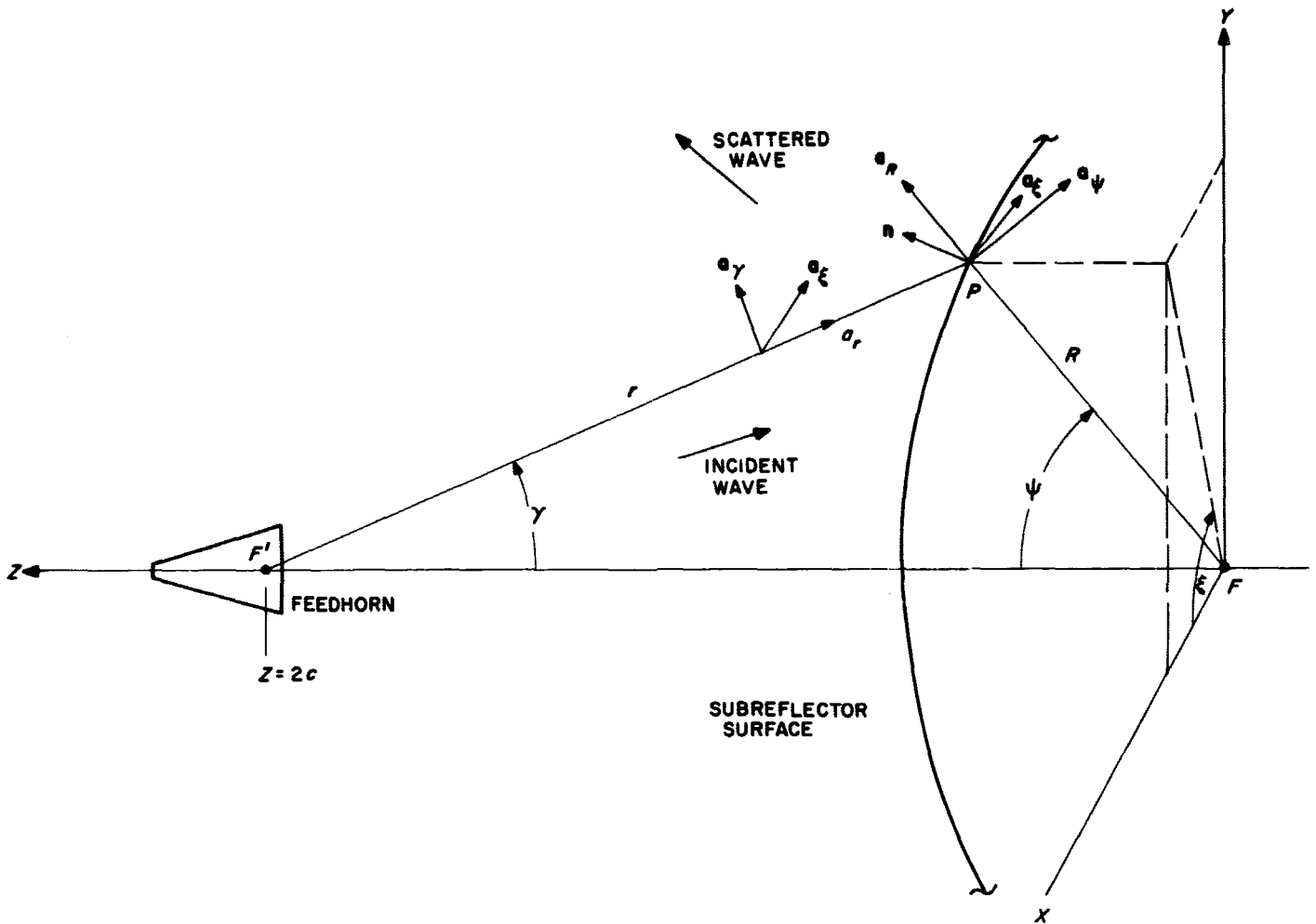


Fig. 11. Feed system geometry

where

$F_{HY}(\gamma)$  = feedhorn polar radiation pattern

$F_{H\xi}(\gamma)$  = feedhorn azimuthal radiation pattern

$k$  = free space propagation constant

$a$  = a geometrical constant

$\delta_H(\gamma)$  = feedhorn phase pattern, assumed to be azimuthally symmetric

The scattered field,  $E_s$ , is now set equal to the sum of spherical wave functions. The coefficients of these waves,  $a_{TE_n}$ , have been previously determined (Ref. 13) by fitting the sum to a desired far-field feed system radiation pattern. Because of the orthogonality of the waves, the scattered field,  $E_s$ , may be uniquely expressed in all regions of space (except the origin) in terms of its far-field value. Thus  $E_s$ , in the region of the subreflector, is a known quantity.

$$E_s \triangleq F(R, \psi) (\sin \xi a_\psi + \cos \xi a_\xi) \quad (3)$$

where

$$F(R, \psi) = jZ_0 \sum_{n=1}^{n=N} a_{TE_n} h_n(\rho) \left[ \frac{P_n^1(\psi)}{S_{11}\psi} + \frac{dP_n^1(\psi)}{d\psi} \right] \quad (4)$$

where

$$Z_0 = 120\pi \text{ ohms}$$

$a_{TE_n}$  = wave coefficient

$$\rho = kR$$

$h_n(\rho)$  = spherical Hankel function

$P_n^1(\psi)$  = associated Legendre polynomial of the first kind

Eq. (3) is based on Eqs. (5) and (6) of Ref. 13, p. 198, which show that a symmetric far-field remains symmetric in the region  $\rho \gg 2n$ . This region turns out to be physically reasonable for the subreflector.

Eq. (4) may be expressed as a quasi-spherical wave as follows:

$$F(R, \psi) \triangleq \frac{F_s(R, \psi)}{\rho} e^{-j[\rho - \delta(R, \psi)]} \quad (5)$$

The unit vectors in Eqs. (2) and (3) are geometrically related as follows:

$$\mathbf{a}_\gamma = \mathbf{a}_R \sin(\psi + \gamma) + \mathbf{a}_\psi \cos(\psi + \gamma) \quad (6)$$

Combining Eqs. (1), (2), (3), (5), and (6),

$$\begin{aligned} \mathbf{E}_T = & -\exp \left\{ -jk \left[ r - 2a - \frac{\delta_H(\gamma)}{k} \right] \right\} \\ & \times \frac{F_{H\gamma}(\gamma)}{kr} \sin(\psi + \gamma) \sin \xi_{a_R} \\ & + \left\{ \frac{F_s(R, \psi)}{\rho} e^{-j[\rho - \delta(R, \psi)]} - \frac{F_{H\gamma}(\gamma)}{kr} \right\} \\ & \times \exp \left\{ -jk \left[ r - 2a - \frac{\delta_H(\gamma)}{k} \right] \right\} \left[ \cos(\psi + \gamma) \right] \\ & \times \sin \xi_{a_\psi} + \left\{ \frac{F_s(R, \psi)}{\rho} e^{-j[\rho - \delta(R, \psi)]} - \frac{F_{H\gamma}(\gamma)}{kr} \right\} \\ & \times \exp \left\{ -jk \left[ r - 2a - \frac{\delta_H(\gamma)}{k} \right] \right\} \cos \xi_{a_\xi} \quad (7) \end{aligned}$$

The boundary condition at the subreflector surface requires that the surface normal,  $\mathbf{n}$ , have no component perpendicular to the total field,  $\mathbf{E}_T$ . Thus,

$$\mathbf{n} \times \mathbf{E}_T = 0 \quad (8)$$

The restriction is now imposed that the subreflector be a surface of revolution. This restriction considerably simplifies the problem and results in a polarization diverse solution. Since, under this condition  $\mathbf{n}$  has no azimuthal component,  $\mathbf{E}_T$  must have no azimuthal component. From Eq. (7), two equations result:

$$F_{H\xi}(\gamma) = \frac{r}{R} F_s(R, \psi) = \frac{\sin \psi}{\sin \gamma} F_s(R, \psi) \quad (9a)$$

$$R - \frac{\delta(R, \psi)}{k} = r - 2a - \frac{\delta_H(\gamma)}{k} \quad (9b)$$

These two equations, together with Eqs. (4) and (5), determine both the reflector surface and the required horn azimuthal pattern. From geometrical considerations,  $r$  may be eliminated from Eq. (9b), and the result is the polar equation of the required subreflector surface:

$$R = \frac{c^2 - \left[ a - \frac{\lambda \delta(R, \psi) - \lambda \delta_H(\gamma)}{4\pi} \right]^2}{c \cos \psi + a - \frac{\lambda \delta(R, \psi) - \lambda \delta_H(\gamma)}{4\pi}} \quad (10)$$

In the limit of zero wavelength, Eq. (10) reduces to the polar equation of an hyperboloid:

$$\lim_{\lambda \rightarrow 0} R = \frac{c^2 - a^2}{c \cos \psi + a} \quad (11)$$

The polar component,  $F_{H\gamma}(\gamma)$ , of the feedhorn may be determined from the boundary condition (8). The result is:

$$\frac{F_{H\xi}(\gamma)}{F_{H\gamma}(\gamma)} = \cos(\psi + \gamma) + \frac{1}{R} \sin(\psi + \gamma) \frac{dR}{d\psi} \quad (12)$$

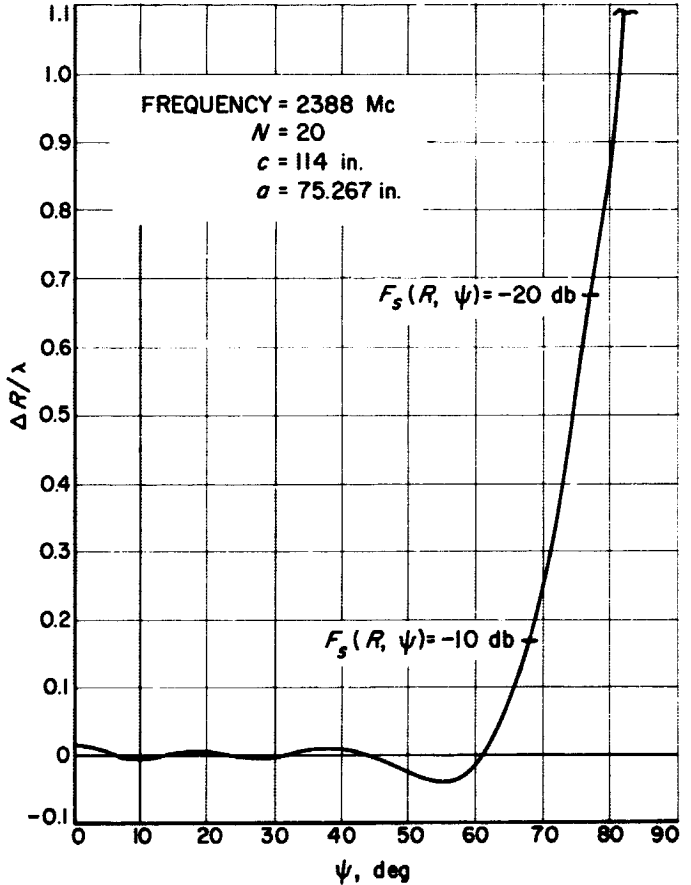


Fig. 12. Deviation of synthesized subreflector from an hyperboloid



where  $dR/d\psi$  is evaluated on the surface.

The two polar angles are related geometrically as follows:

$$\gamma = \sin^{-1} \frac{\sin \psi}{\left[ 1 + 4 \left( \frac{c}{R} \right)^2 - 4 \left( \frac{c}{R} \right) \cos \psi \right]^{1/2}} \quad (13)$$

An abbreviated machine program for evaluating Eqs. (10), (5), and (9a) has been used to synthesize an  $N = 20$  subreflector for the 85-ft Goldstone antenna operating at 2388 Mc. Fig. 12 shows the deviation of the subreflector,  $\Delta R/\lambda$ , in wavelengths, from the optical case. The surface which was empirically designed for use at 2388 Mc in the planetary radar system consists of an hyperboloid with a conical beamshaping flange (Ref. 14), as shown in Fig. 13. Fig. 14 shows the difference between the optimum (calculated) hyperboloid-flange combination and the computed reflector surface. It can be seen that this simple system is an almost perfect fit to the ideal

surface. Unfortunately, this case of  $\alpha = 21.5$  deg was not experimentally evaluated when the 2388-Mc system was designed.

Fig. 15 shows the scattered field function  $F_s(R, \psi)$  at the ideal reflector surface, the far scattered field, and the ideal far field. Although the formulation considered in

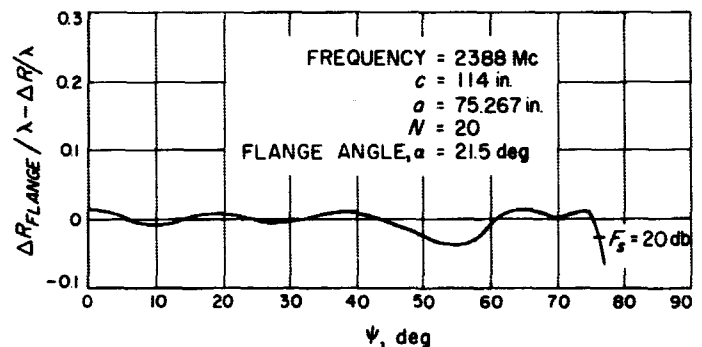


Fig. 14. Deviation of synthesized subreflector from an hyperboloid-flange

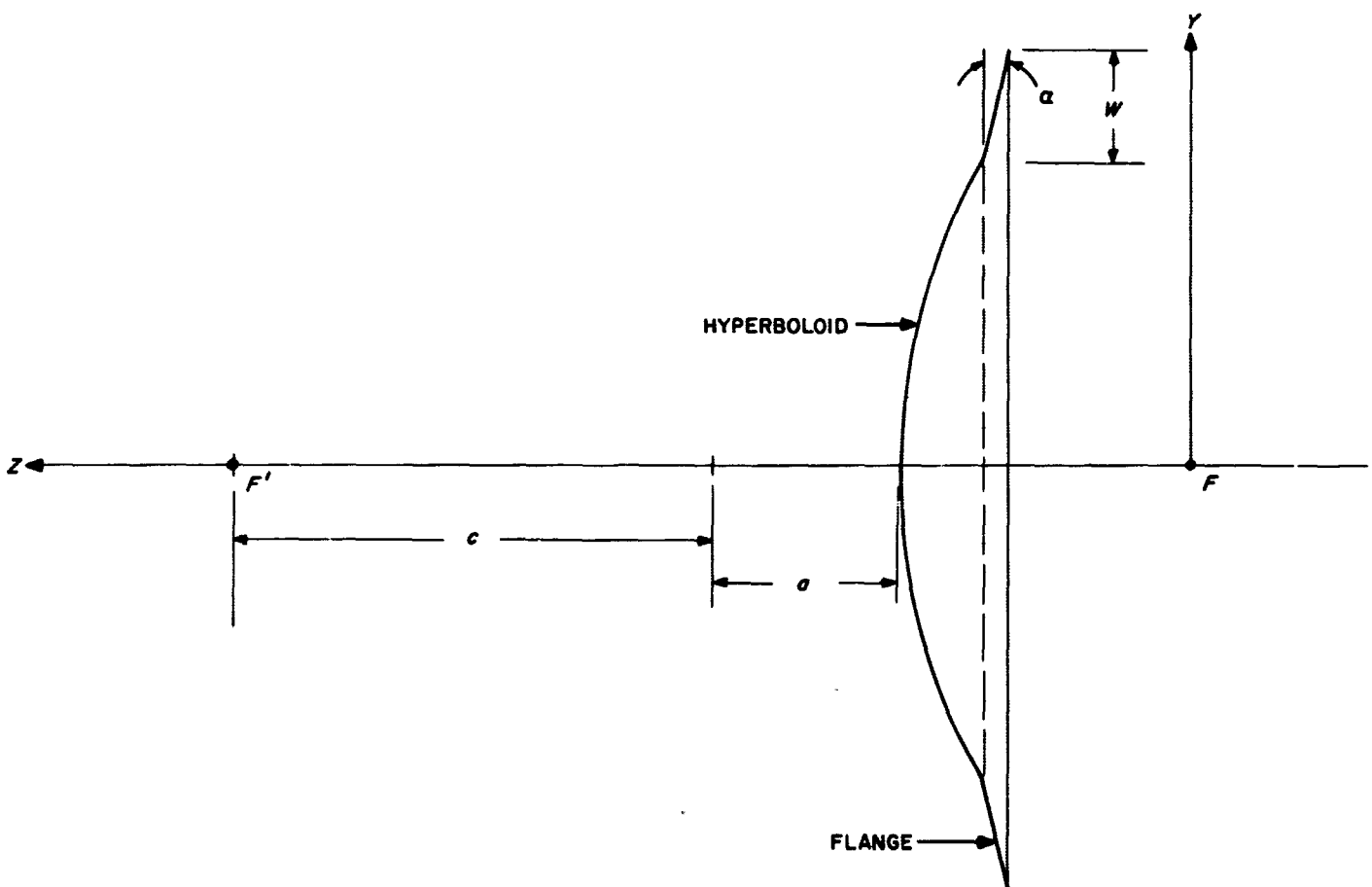


Fig. 13. Beamshaping flange geometry

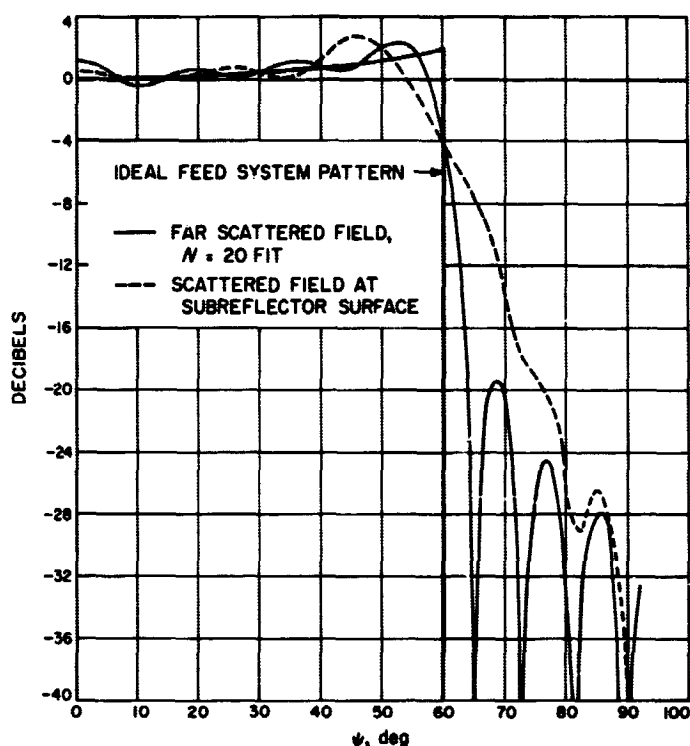


Fig. 15. Scattered fields

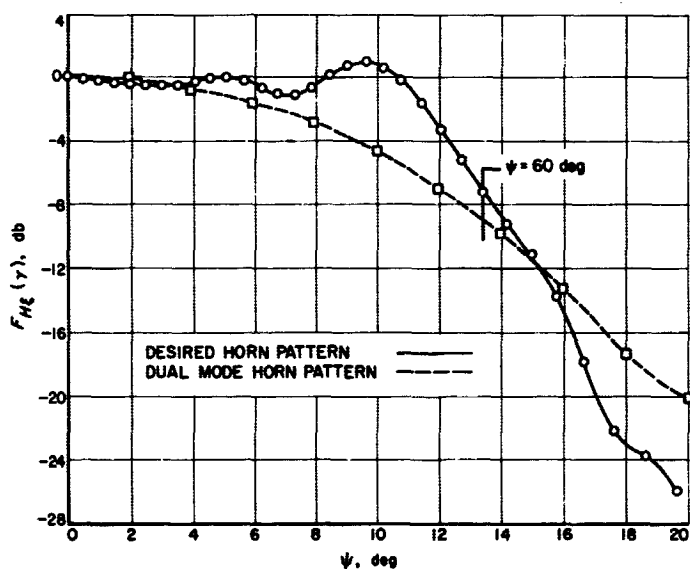


Fig. 16. Synthesized horn radiation pattern

this reporting generates a subreflector surface of infinite extent, it is clear from Fig. 15 that the fields are extremely small in the outer region of the surface. It is therefore felt that the ideal surface may be truncated without significant reduction of performance. The effect of truncation will be investigated on the Arbitrary Reflector of Revolution Scattering Computer Program (Ref. 15).

The required azimuthal horn radiation pattern is shown in Fig. 16, together with the dual-mode horn (Ref. 16) pattern which was used with the 2388-Mc experimental system. It can be seen that the required pattern is qualitatively a sector beam which, in principle, may be physically realized by a suitable multimode feedhorn.

A request has been submitted for a comprehensive computer program to synthesize subreflectors and the corresponding horn patterns. It is expected that this program will be available in 2 to 4 mo.

## D. RF Techniques

T. Otoshi and D. White

### 1. H-Band Rotary Vane Attenuator,

T. Otoshi and D. White

**a. Summary.** This report presents results of preliminary calibrations made at 8448 Mc on a precision H-band rotary vane attenuator. Accurate measurements were made of attenuation values as low as 0.02 db and as high as 40 db. This rotary vane attenuator will serve as a JPL interlaboratory attenuation standard traceable to the National Bureau of Standards and will permit intercomparison of the accuracies of various 8448-Mc insertion loss test sets used at JPL. This attenuator should also be useful in the calibration of standard noise sources for the experimental 8448-Mc antenna system at Goldstone (Venus site).

#### b. Recent work.

**Theory of operation.** A rotary vane attenuator essentially consists of two end sections which are rectangular-to-circular waveguide transitions, and a center section which is a circular waveguide free to rotate between the end sections. A thin resistive card is placed across each section of waveguide to absorb tangential electric fields. If there are no discontinuities and if the resistive elements completely absorb the tangential components of the electrical field, then the attenuation in decibels is given by:

$$A = -20 \log_{10} \cos^2 \theta$$

where  $\theta$  is the angular displacement of the card in the rotor with respect to the cards in the stator. The accuracy to which this angular displacement can be determined is usually the principal factor in limiting the over-all accuracy of such an attenuator.

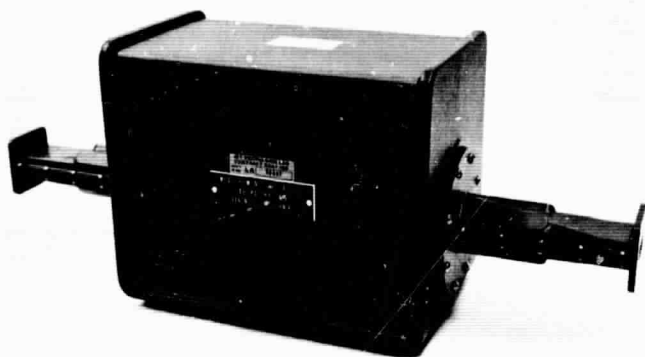


Fig. 17. H-band rotary vane attenuator

*Design description.* A photograph of the calibrated attenuator may be seen in Fig. 17. This attenuator is a Hewlett-Packard H382A model which was modified to permit readout of the rotary vane angle position in degrees, minutes, and seconds. The readout exhibits a range of 0 to 90 deg with a minimum resolution of 1 sec of arc. The maximum backlash between the position of the center vane and the "seconds dial" on the readout mechanism is less than 10 sec of arc. If the drive shaft is turned in the same direction, the resetability of the vane angle is believed to be within  $\pm 2$  sec of arc.

Fig. 18 is a photograph of the gear assembly of this precision attenuator. The circular waveguide rotor is

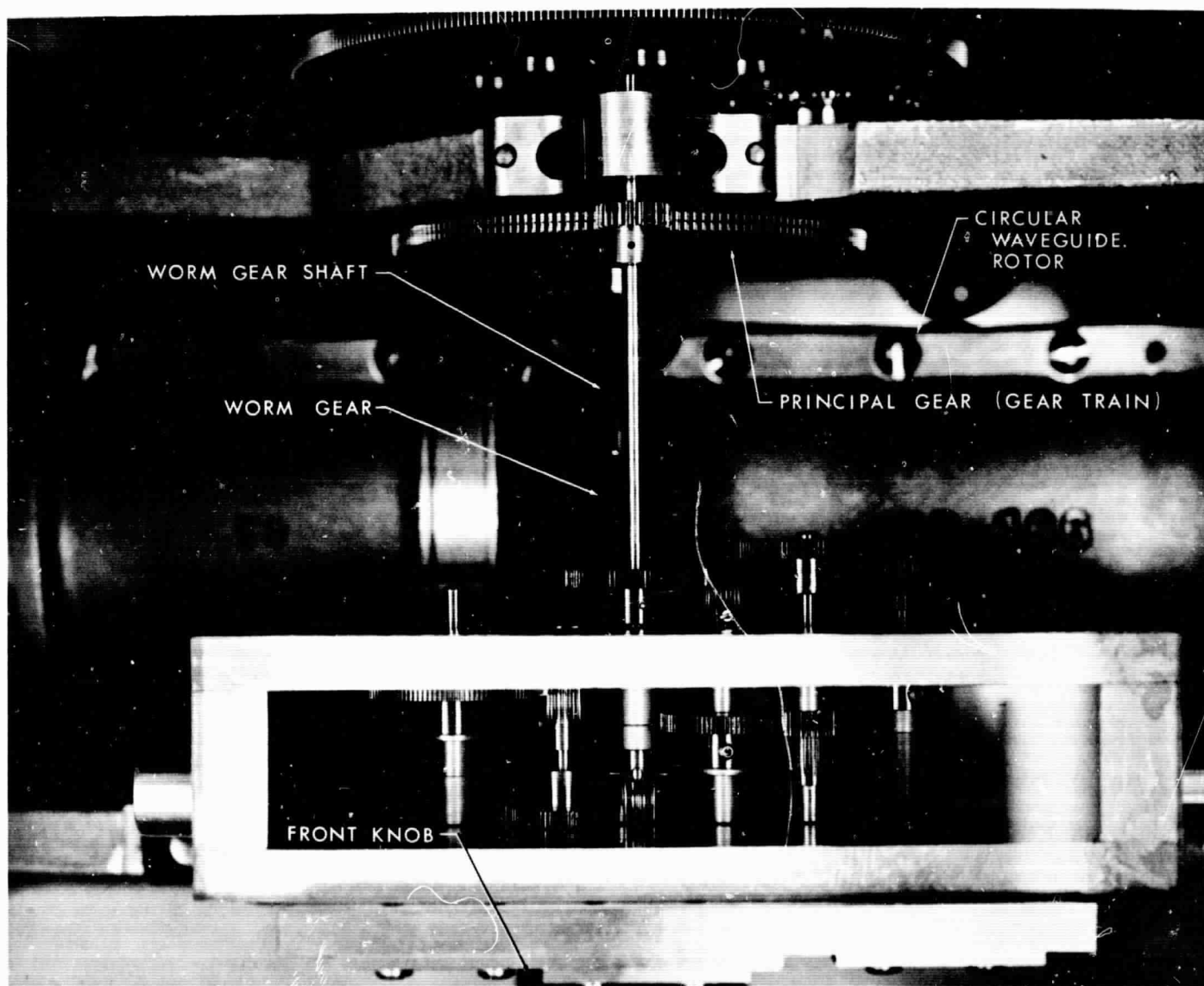


Fig. 18. Top view of gear drive assembly (attenuator cover off)

driven by a Hewlett-Packard worm gear. Also mounted on the worm gear shaft is the principal gear of a gear train which drives the readout dials. All gears belonging to the gear train for the readout dials are Class Precision 3 gears. The gear train and readout system were designed by the Measurement Specialties Laboratory, Inc., in Van Nuys, California.

#### *Preliminary tests.*

(1) *0-deg vane angle setting.* If the center vane is not correctly aligned relative to the end vanes at the indicated 0-deg position, then a systematic angular bias error will occur at other vane angle settings. One method for determining this systematic error was described by Wilbur Larson of the National Bureau of Standards (Ref. 17). At various dial settings, deviations of measured and theoretical attenuations are converted to estimated vane angle errors. The average of these estimated vane angle errors is used to make the correc-

tion necessary to establish the new 0-deg dial setting. Using this method, good agreement between measured and theoretical attenuations was obtained.

(2) *Measurement system.* The ac ratio transformer technique, described in Ref. 18, was used to calibrate the rotary vane attenuator. Detection is accomplished by means of a dual channel bolometer system. Attenuation power ratios are compared on a precise ac ratio transformer. The ac ratio transformer test set and H-band insertion loss measurement system may be seen in Fig. 19. The source and load assemblies of this measurement system were matched to VSWRs of less than 1.02 at 8448 Mc.

As stated in Ref. 18, the principal limitations on the accuracy of this measurement technique for insertion loss measurements are the deviations from square law response of the bolometers used. A bolometer is said to have square law characteristics if its detected ac output

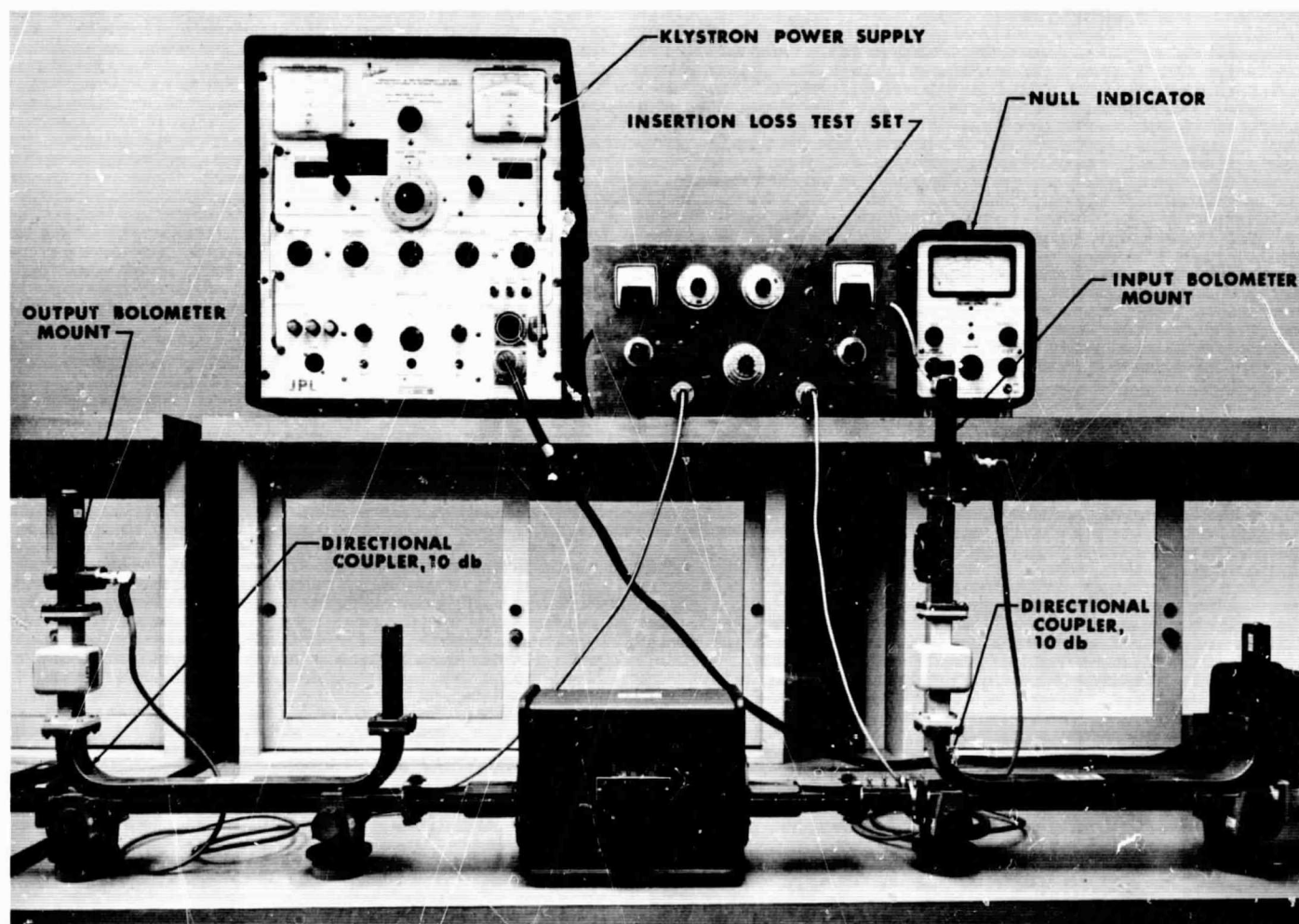


Fig. 19. H-band rotary vane attenuator and measurement system

voltage is proportional to the microwave input power. The bolometers and bolometer mounts used for the measurement system were the Sperry 821 barretters and Hewlett-Packard Model H485B detector mounts, respectively. These barretters are rated to operate at 200-ohm resistance with an 8.75-ma dc current applied.

Bolometer deviation from square law characteristics (also referred to as bolometer nonlinearity characteristics)

can be determined approximately by experimentally measuring fixed attenuation differences, each time with the bolometer operating at a new RF power level. The following describes the procedure used to determine bolometer errors of the measurement system.

With the vane angle set at the 0-deg vane angle setting, the RF power levels into the input and output bolometers were adjusted to 0.2-mw average power. The

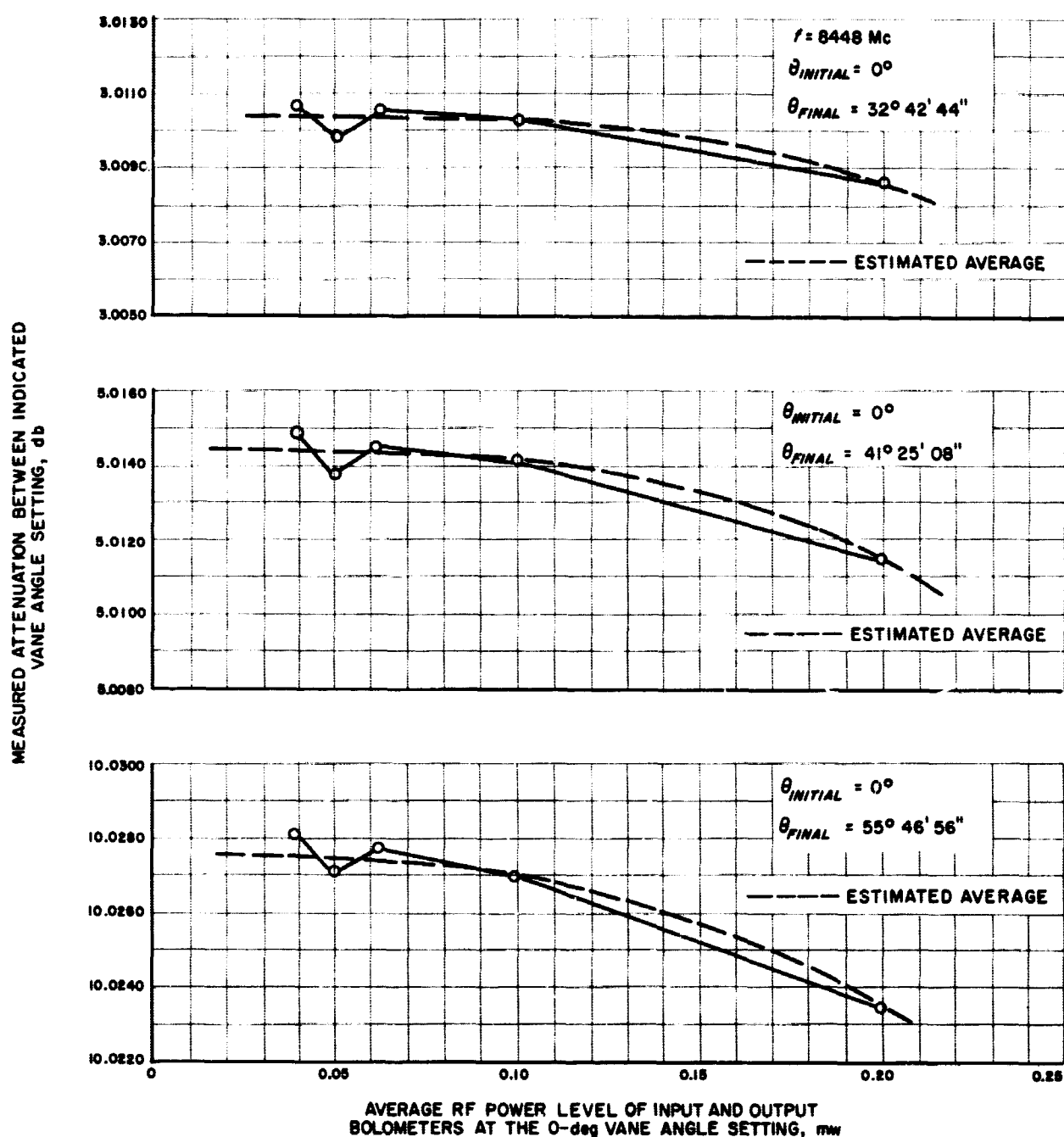


Fig. 20. Measured attenuation dependence on RF operating levels of bolometers

attenuation differences relative to the 0-deg setting were then measured at the nominal 3-, 5- and 10-db vane angle settings. This procedure was then repeated at other RF power levels. The range of RF power levels into the bolometers for these tests was between 0.2- and 0.04-mw average power (or 0.4- to 0.08-mw peak power since the klystron was square wave modulated). Fig. 20 shows the results of the bolometer tests. For the bolometer pair tested, it appears that if the bolometers are operated at RF power levels less than 0.2-mw peak power (0.1-mw average power), then the deviations in measured attenuations, for a 10-db attenuation measurement, should be less than 0.005 db. At these RF power levels and operating conditions, the bolometers should very nearly be operating as square law detectors (Ref. 19).

For attenuation measurements greater than a 10-db step, resolution and precision of the ac ratio transformer becomes somewhat critical. A loss of  $\pm 1$  dial division of resolution on the particular ac ratio transformer being used (ESI Model DT-45) would result in errors of approximately  $\pm 0.0005$  db,  $\pm 0.005$  db, and  $\pm 0.05$  db for 10-, 20-, and 30-db step attenuation measurements, respectively.

#### Calibration results.

(1) *VSWR versus vane angle.* Fig. 21 shows the VSWR of the rotary vane attenuator measured as a function of

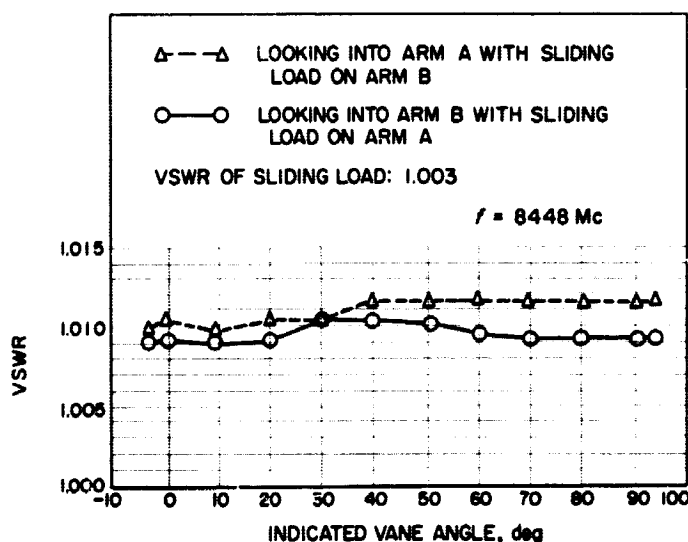
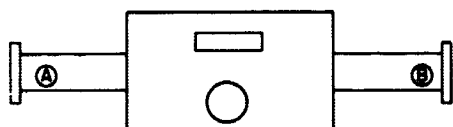


Fig. 21. VSWR versus indicated vane angle setting

vane angle. The maximum VSWR measured was 1.012, and the maximum VSWR difference was less than a factor of 1.004. For this measurement a tuned waveguide reflectometer system operating at 8448 Mc was used. The accuracy of this VSWR measurement is estimated to be within a factor of 1.005.

(2) *Attenuation versus vane angle.* For the calibration of the rotary vane attenuator, the measurement method used was to measure attenuation differences between indicated and reference vane angles. The maximum attenuation difference measured in one step with the ac ratio transformer test set was 20 db. At the reference vane angle settings, the RF power levels into the input and output bolometers were set at approximately 0.06 mw (average power).

Table 1 gives a comparison of measured and theoretical attenuation differences. The resetability of the rotary vane angle at a particular setting may be seen from the repeatability of the measured data.

Table 2 shows attenuation differences using the 0-deg setting as the reference position for all readings. This table was compiled from the results given in Table 1. It may be seen that the theoretical and average measured attenuation values agree to within  $\pm 0.005$  db for attenuations up to 6 db and  $\pm 0.015$  db for attenuations up to 20 db. At the nominal 40-db setting, the average measured deviation was  $+0.066$  db.

#### Error analysis.

(1) *Calibration errors.* The accuracy of the calibration method is dependent upon the resetability of the rotary vane to within  $\pm 2$  sec of arc. Other measurement errors would be bolometer nonlinearity error, external mismatch error, and uncertainties in the accuracy of the insertion loss test set. Accounting for cumulative errors of the calibration method, it is estimated that the accuracies of values given in Table 2 are better than  $\pm 0.02$  db at 10 db,  $\pm 0.04$  db at 20 db,  $\pm 0.07$  db at 30 db, and  $\pm 0.10$  db at 40 db.

(2) *Attenuator errors.* Four principal errors causing deviations from theoretical attenuations would be transmission error, alignment error, internal mismatch error, and angular readout error (Ref. 20).

The transmission error is an error caused by incomplete attenuation of the tangential electric field components by the resistive cards. For most rotary vane

Table 1. Theoretical and measured 8448-Mc attenuation differences between indicated and reference vane angle settings

Part A. Attenuation differences between indicated and 0-db reference vane angle settings								Part B. Attenuation differences between indicated and 10-db reference vane angle settings							
Indicated vane angle setting			Theoretical attenuation relative to the 0-db setting, db	Measured values, db				Indicated vane angle setting			Theoretical attenuation relative to the 0-db setting, db	Measured values, db			
deg	min	sec		1	2	3	Average	deg	min	sec		1	2	3	Average
0	0	0	0	Reference setting				55	46	56	10.0	Reference setting			
02	44	56	0.02	0.0196	0.0195	0.0201	0.0197	59	55	17	12.0	1.9972	1.9964	1.9967	1.9968
03	53	12	0.04	0.0367	0.0381	0.0379	0.0382	63	28	08	14.0	3.9910	3.9917	3.9924	3.9917
04	45	33	0.06	0.0581	0.0577	0.0581	0.0580	66	32	24	16.0	5.9762	5.9752	5.9775	5.9763
05	29	40	0.08	0.0779	0.0775	0.0777	0.0777	69	13	05	18.0	7.9846	7.9843	7.9887	7.9859
06	08	31	0.10	0.0985	0.0983	0.0983	0.0984	71	33	54	20.0	9.9870	9.9887	9.9898	9.9885
08	40	39	0.20	0.1984	0.1978	0.1985	0.1982	Part C. Attenuation differences between indicated and 20-db reference vane angle settings							
12	14	54	0.40	0.3997	0.3986	0.3998	0.3994								
14	58	20	0.60	0.6019	0.6008	0.6017	0.6015	71	33	54	20.0	Reference setting			
17	15	19	0.80	0.8036	0.8018	0.8024	0.8026	73	37	48	22.0	2.0199	2.0190	2.0191	2.0193
19	15	18	1.0	1.0034	1.0020	1.0038	1.0031	75	27	08	24.0	4.0352	4.0323	4.0319	4.0331
26	58	09	2.0	2.0005	2.0000	2.0006	2.0004	77	03	48	26.0	6.0448	6.0438	6.0420	6.0435
32	42	44	3.0	3.0020	3.0015	3.0031	3.0022	78	29	27	28.0	8.0453	8.0442	8.0439	8.0445
37	24	29	4.0	3.9971	3.9961	3.9982	3.9971	79	45	24	30.0	10.0461	10.0470	10.0435	10.0455
41	25	08	5.0	4.9985	4.9968	4.9992	4.9982	82	20	12	35.0	15.0489	15.0481	15.0449	15.0473
44	55	55	6.0	6.0107	6.0096	6.0112	6.0105	84	15	39	40.0	20.0663	20.0676	20.0676	20.0672
48	03	38	7.0	7.0131	7.0128	7.0147	7.0135								
50	52	45	8.0	8.0099	8.0088	8.0099	8.0095								
53	26	25	9.0	9.0146	9.0141	9.0158	9.0148								
55	46	56	10.0	10.0108	10.0108	10.0105	10.0107								

attenuators, this error is usually not significant for attenuations below 40 db. One method of evaluating transmission error is to measure the attenuation of the resistive card in the rotor when the center vane is set at 90 deg and, using this measured value, to calculate the transmission error (Ref. 20). The attenuation at 90 deg was not measured for this rotary vane attenuator because a suitable receiver required for this measurement was not available.

The effects of internal mismatch error and alignment error were not considered to be serious and were not calculated.

The major part of the deviation between the theoretical and measured attenuations shown in Table 2 is believed to be principally due to angular readout errors. This belief is based on further analysis made of the deviations between measured and theoretical attenuations. If the db deviations are converted into estimated readout errors in degrees (using the method described in Ref. 17) and plotted against the indicated vane angle, it may be seen that a periodicity exists in the plotted curve (Fig. 22). The variations in the readout error appear to repeat every 29 or 30 deg of indicated vane angle.

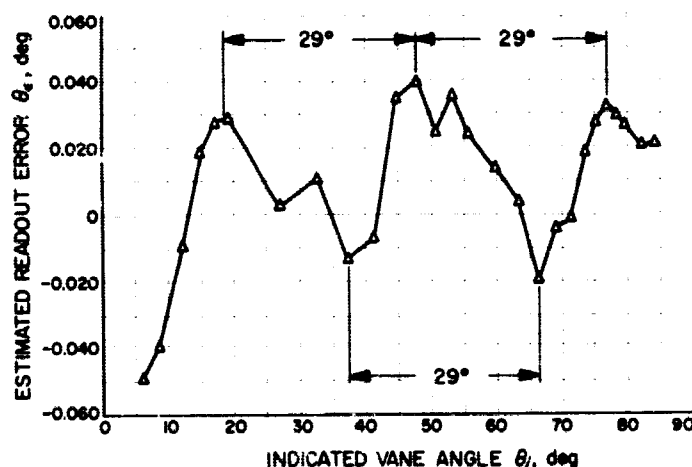


Fig. 22. Estimated readout error of the H-band rotary vane attenuator

It was later discovered that the worm gear shaft is designed to turn one complete revolution for 30-deg rotation of the rotor. Since both the worm gear (which turns the circular waveguide rotor) and the principal gear of the readout gear train are mounted on the worm gear shaft, it is reasonable to conclude that the readout error is a function of worm gear shaft rotation. The major part

**Table 2. Calibration of the H-band rotary vane attenuator from 0 to 40 db at 8448 Mc<sup>a</sup>**

1 Indicated vane angle <sup>b</sup> $\theta_i$ , deg	2 Theoretical attenuation correspond- ing to indicated vane angle $A_i$ , db	3 Average measured attenuation $A_m$ , db	4 Average measured deviation ( $A_m - A_i$ ), db	5 Estimated readout error <sup>c</sup> $\theta_e$ , deg
2.749	0.02	0.0197	-0.0003	-0.021
3.887	0.04	0.0382	-0.0018	-0.088
4.759	0.06	0.0580	-0.0020	-0.080
5.495	0.08	0.0777	-0.0023	-0.078
6.142	0.10	0.0984	-0.0016	-0.049
8.678	0.20	0.1982	-0.0018	-0.039
12.248	0.40	0.3994	-0.0006	-0.009
14.972	0.60	0.6015	+0.0015	+0.019
17.255	0.80	0.8026	+0.0026	+0.028
19.255	1.0	1.0031	+0.0031	+0.029
26.969	2.0	2.0004	+0.0004	+0.003
32.712	3.0	3.0022	+0.0022	+0.011
37.408	4.0	3.9971	-0.0029	-0.013
41.419	5.0	4.9982	-0.0018	-0.007
44.932	6.0	6.0105	+0.0105	+0.025
48.061	7.0	7.0135	+0.0135	+0.040
50.879	8.0	8.0095	+0.0095	+0.025
53.440	9.0	9.0148	+0.0148	+0.036
55.782	10.0	10.0107	+0.0107	+0.024
59.921	12.0	12.0075	+0.0075	+0.014
63.469	14.0	14.0024	+0.0024	+0.004
66.540	16.0	15.9870	-0.0130	-0.019
69.218	18.0	17.9966	-0.0034	-0.004
71.565	20.0	19.9992	-0.0008	-0.001
73.630	22.0	22.0185	+0.0185	+0.019
75.452	24.0	24.0323	+0.0323	+0.028
77.063	26.0	26.0427	+0.0427	+0.033
78.491	28.0	28.0437	+0.0437	+0.030
79.757	30.0	30.0447	+0.0447	+0.027
82.337	35.0	35.0465	+0.0465	+0.021
84.261	40.0	40.0664	+0.0664	+0.022

<sup>a</sup>Measured data was compiled from Table 1.

<sup>b</sup>Indicated vane angle settings of Table 1 converted to degrees.

<sup>c</sup>See NBS Technical Note 177 (Ref. 17).

of the angular readout error is thought to be caused by the worm gear because it was not designed to be used for ultra-precise positioning of the rotary vane.

The curve shown on Fig. 22 should be useful for predicting the attenuations which would be measured at indicated vane angle settings not calibrated. The measured attenuation values closely follow the relationship:

$$A_m = -40 \log_{10} \cos(\theta_i + \theta_e)$$

where

$\theta_i$  = indicated vane angle, deg

$\theta_e$  = estimated readout error, deg

In order to establish this attenuator as a reliable attenuation standard, the attenuator should be calibrated independently, preferably using another accurate measurement technique. An arrangement has been made with the National Bureau of Standards to calibrate this precision attenuator using their modulated sub-carrier technique of measuring microwave attenuation.

## 2. AC Ratio Transformer Insertion Loss Set, D. White

**a. Summary.** This report discusses recent tests and modifications made on the ac ratio transformer insertion loss set described in Ref. 18. Two major changes made were replacement of the type of bridge transformer used and removal of the shield balance circuit.

**b. Recent work.** Since the development of the S-band rotary vane attenuator described in Refs. 21 and 22, it has been possible to resolve very small differences between the measured values of insertion loss obtained from two ac ratio transformer insertion loss sets of the type described in Ref. 18.

The precision rotary vane attenuator was used to compare the two insertion loss sets operating at 2388 Mc. Measured values were obtained at indicated attenuator settings of 1, 5, 10, and 20 db, respectively. As shown in Table 3-A, there was a small but definite difference in the measured values obtained from the two insertion loss sets. The cause of this discrepancy was traced to the output bridge transformers of the insertion loss test

**Table 3. Comparison of the two ac ratio transformer insertion loss test sets**

Part A. Before modification		
Indicated rotary vane attenuator setting, db	Average measured loss using Unit No. 1, db	Average measured loss using Unit No. 2, db
1	1.0018	1.0002
5	5.0057	5.0002
10	10.0130	10.0074
20	20.0453	20.0364
Part B. After modification		
1	0.9993	0.9990
5	4.9995	4.9993
10	10.0073	10.0065
20	20.0411	20.0389

Notes: Due to a small error in the readout system of the rotary vane attenuator, the indicated rotary vane attenuator setting yields an insertion loss value that is slightly high (see Ref. 23 for details) and becomes especially noticeable for values above 5 db.



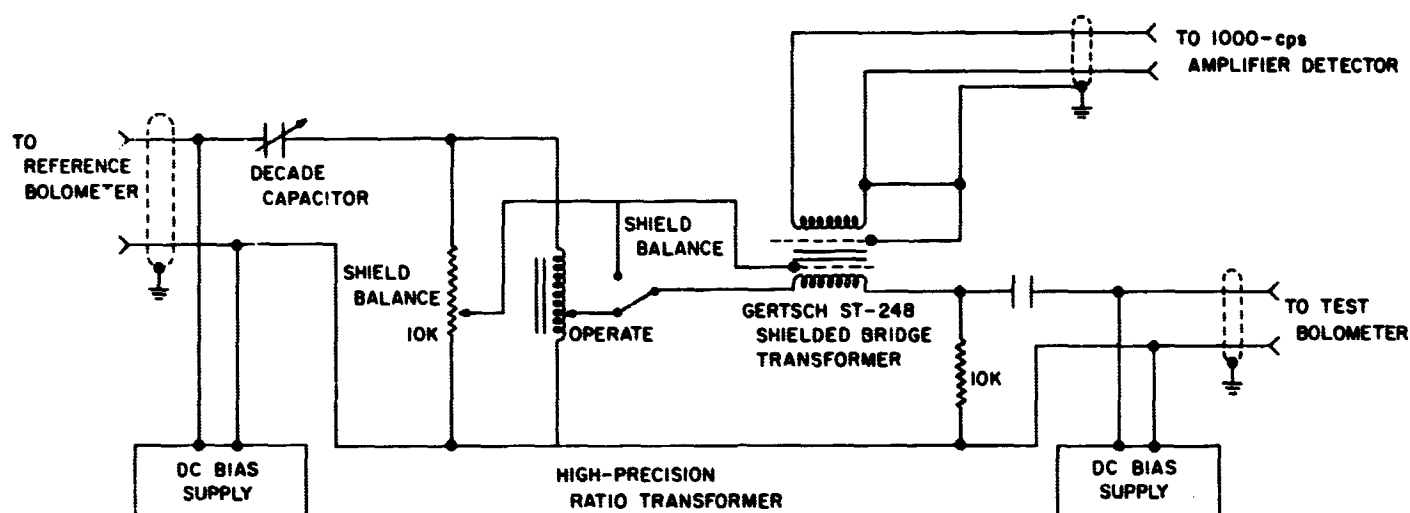


Fig. 23. Diagram of ratio transformer insertion loss set before modification

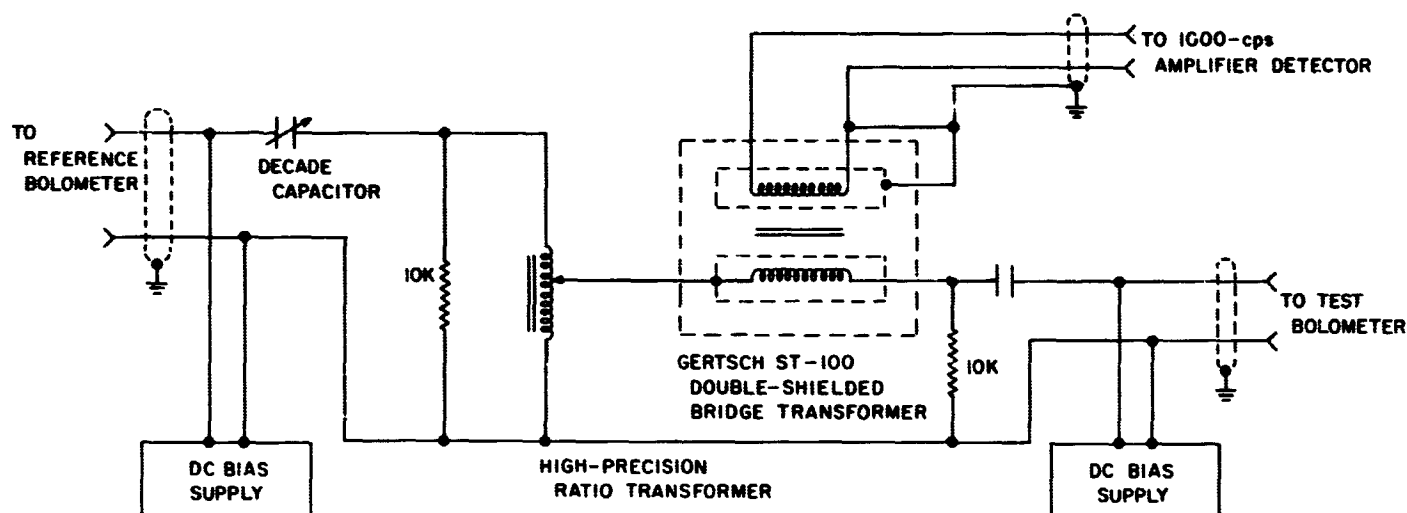


Fig. 24. Diagram of ratio transformer insertion loss set after modification

sets. In construction of the two units, the output leads of the bridge transformers (ST-248, as shown in Fig. 23) were arbitrarily reversed with respect to each other. It was found that, when the bridge transformers of the

**Table 4. Comparison of the ac ratio transformer insertion loss test set with and without shield balance**

Indicated rotary vane attenuator setting, db	Unit No. 2: average measured loss using shield balance operation, db	Unit No. 2: average measured loss with shield tied to wiper of ratio transformer, db
5	4.9993	4.9997
10	10.0065	10.0050
20	20.0389	20.0394

two units were hooked up identically, the measured insertion loss values compared almost identically between units. It was finally determined that the ST-248, being a single shielded transformer, is subject to capacitive ground loops that become apparent when working with low-level voltages. When the ST-248 transformers were replaced with double-shielded transformers, the ground loop condition was eliminated and the two insertion loss sets agreed favorably regardless of the polarity of the output leads of the bridge transformer. (See Table 3-B and Fig. 24.)

A significant improvement in the operation of the insertion loss set was also made possible when the double-shielded bridge transformer was installed. When the shield-to-winding capacitance ground loop was elimi-

nated, it was found that the wiper arm of the ratio transformer could be tied directly to the primary shield of the bridge transformer without any noticeable loading effects (Table 4). It was now possible to replace the shield balance potentiometer with a fixed 10K resistor and elimi-

nate a step in the operation. The shield balance potentiometer was used to drive the shield of the bridge transformer to the same potential as the primary winding of the bridge transformer with a minimum loading on the reference signal (Figs. 23 and 24).

## References

1. Stelzried, C. T., and Rusch, W. V. T., "Millimeter Wave Work," SPS 37-25, Vol. IV, pp. 113-117, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.
2. Stelzried, C. T., and Rusch, W. V. T., "Millimeter Wave Work," SPS 37-26, Vol. IV, pp. 181-189, Jet Propulsion Laboratory, Pasadena, California, April 10, 1964.
3. Stelzried, C. T., "Temperature Calibration of Microwave Termination," SPS 37-26, Vol. IV, Table 2, p. 190, Jet Propulsion Laboratory, Pasadena, California, April 30, 1964.
4. Stelzried, C. T., "Low Noise Amplifiers," SPS 37-18, Vol. IV, Jet Propulsion Laboratory, Pasadena, California, December 31, 1962.
5. Siddoway, J. C., "Lasers," SPS 37-26, Vol. IV, pp. 194 to 195, Jet Propulsion Laboratory, Pasadena, California, April 30, 1964.
6. Erpenbach, H., Wells, W. H., Finnie, C., "Superconducting Bolometer," SPS 37-21, 23, 24, 25, Vol. IV, 1963; SPS 37-26, Vol. IV, 1964.
7. Milton, R. M., *Chemical Review*, No. 39, p. 419, 1946.
8. Franzen, W., *Journal of the Optical Society of America*, No. 53, p. 596, 1963.
9. Baird, D. C., *Canadian Journal of Physics*, No. 37, p. 129, 1959.
10. Lalevic, B., *Journal of Applied Physics*, No. 31, p. 1234, 1960.
11. Carslaw, H. S., and Jaeger, J. C., "Conduction of Heat in Solids," p. 75, Oxford, 1940.
12. Erpenbach, H., and Maserjian, J., "Superconducting Bolometer," SPS 37-22, Vol. IV, pp. 182-184, August 31, 1963.
13. Potter, P. D., "Antenna Feed Research; Spherical Wave Functions," SPS 37-26, Vol. IV, pp. 197-200, Jet Propulsion Laboratory, Pasadena, California, April 30, 1964.
14. Potter, P. D., "A Simple Beamshaping Device for Cassegrain Antennas," Technical Report No. 32-214, Jet Propulsion Laboratory, Pasadena, California, January 31, 1962.
15. Rusch, W. V. T., "Scattering of a Spherical Wave by an Arbitrary Truncated Surface of Revolution," Technical Report No. 32-434, Jet Propulsion Laboratory, Pasadena, California, May 27, 1963.

## References (Cont'd)

16. Potter, P. D., "A New Horn Antenna with Suppressed Sidelobes and Equal Beamwidths," *The Microwave Journal*, Vol. VI, No. 6, pp. 71-78, June 1963.
17. Larson, W., "Table of Attenuation Error as a Function of Vane-Angle Error for Rotary Vane Attenuators," NBS Technical Note 177, National Bureau of Standards, May 20, 1963.
18. Schuster, D., and Finnie, C., "AC Ratio Transformer Technique," SPS 37-22, Vol. IV, pp. 189-196, Jet Propulsion Laboratory, Pasadena, California, August 31, 1963.
19. Sorger, G. U., and Weinschel, B. O., "Comparison of Deviations from Square Law for RF Crystal Diodes and Barretters," *IRE Transactions on Instrumentation*, Vol. I-8, pp. 103-111, December 1959.
20. James, A. V., "A High-Accuracy Microwave Attenuation Standard for Use in Primary Calibration Laboratories," *IRE Transactions on Instrumentation*, Vol. I-11, pp. 285-290, December 1962.
21. Otoshi, T., and White, D., "S-Band Rotary Vane Attenuator," SPS 37-24, Vol. IV, pp. 156-161, Jet Propulsion Laboratory, Pasadena, California, December 31, 1963.
22. Otoshi, T., "S-Band Rotary Vane Attenuator," SPS 37-25, Vol. IV, pp. 128-132, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.

## XVIII. Communication System Development

### A. Signal-to-Noise Monitoring

C. E. Gilchrist

#### 1. Statistical Properties

*a. Résumé.* Two systems for determining signal-to-noise ( $S/N$ ) from measurements are analyzed for efficiency (minimum number of samples or equivalently minimum time) for inclusion in the DSIF for advanced planetary communications systems. Our system is based on the  $S/N$  estimator from a linear system. The probability density function is determined along with confidence intervals. The second system is based on a decision device which determines the bit error rate estimator. The probability density function is determined along with confidence intervals. Charts of confidence intervals for various numbers of samples are given for both systems. The conclusion is that the  $S/N$  estimator method is by far the most efficient.

*b. Recent work.* The communications system performance monitor function during actual spacecraft missions is an extremely desirable one in that it lends confidence to real-time spacecraft operation decisions based on the communication system. It also allows some correction of detected system degradation during spacecraft operations.

With this general preamble, we shall restrict this study to a particular facet of the problem; that is, we shall study a system for monitoring the signal-to-noise ratio at various points in the receiving system.

There are two major points in the system which need to be monitored; these are:

- (1) Output of the receiver itself in regard to carrier-power-to-noise-power.
- (2) Output of the data system or the data subcarrier-power-to-noise-power.

The first of these can be monitored reasonably well through existing hardware at the DSIF. In particular, elaborate calibrations are available for both received carrier power (automatic gain control) and the noise power (system noise temperature). Unfortunately, these have nonlinear calibrations, and the measurements are operationally difficult to present in engineering units at the necessary operations locations. For the second, nothing exists, but may only be derived indirectly through calculations involving the modulation index. A degrading modulation index is basically the parameter which we would like to monitor through the data channel signal-to-noise measurement in combination with the RF carrier-power-to-noise measurement.

Proposed systems for monitoring the data and carrier channel signal to noise should have the following features:

- (1) It should be readily and accurately analyzable.
- (2) It should be efficient (minimum number of samples and minimum time for processing).
- (3) It should not require manual operations.

Requirement (1) restricts one to systems which are linear and Gaussian, since they are the easiest to analyze and have the most information available. Requirement (2) will be the main concern in the analysis which is to follow. The basic problem is how many samples are required to obtain a certain confidence in our answer. Requirement (3) dictates that the processing should probably be digital, but should not be so elaborate that it could not be done at the DSIF with a small special-purpose processor. Since the Digital Instrumentation System (DIS)

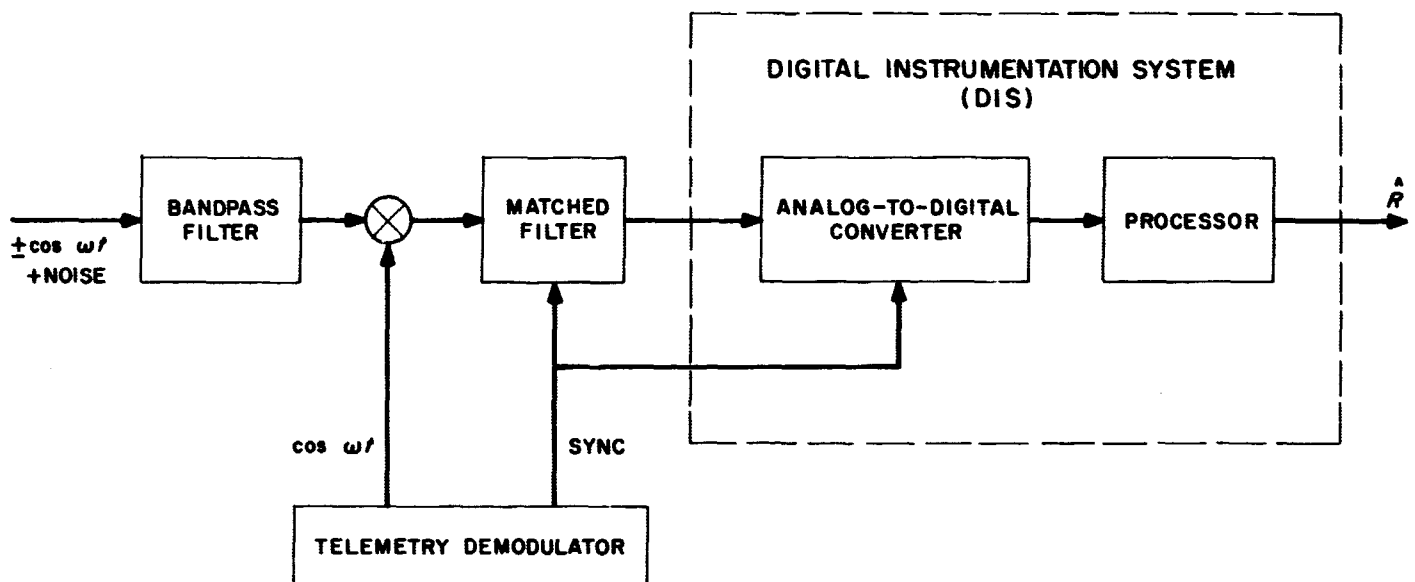


Fig. 1. Subcarrier signal-to-noise estimator system

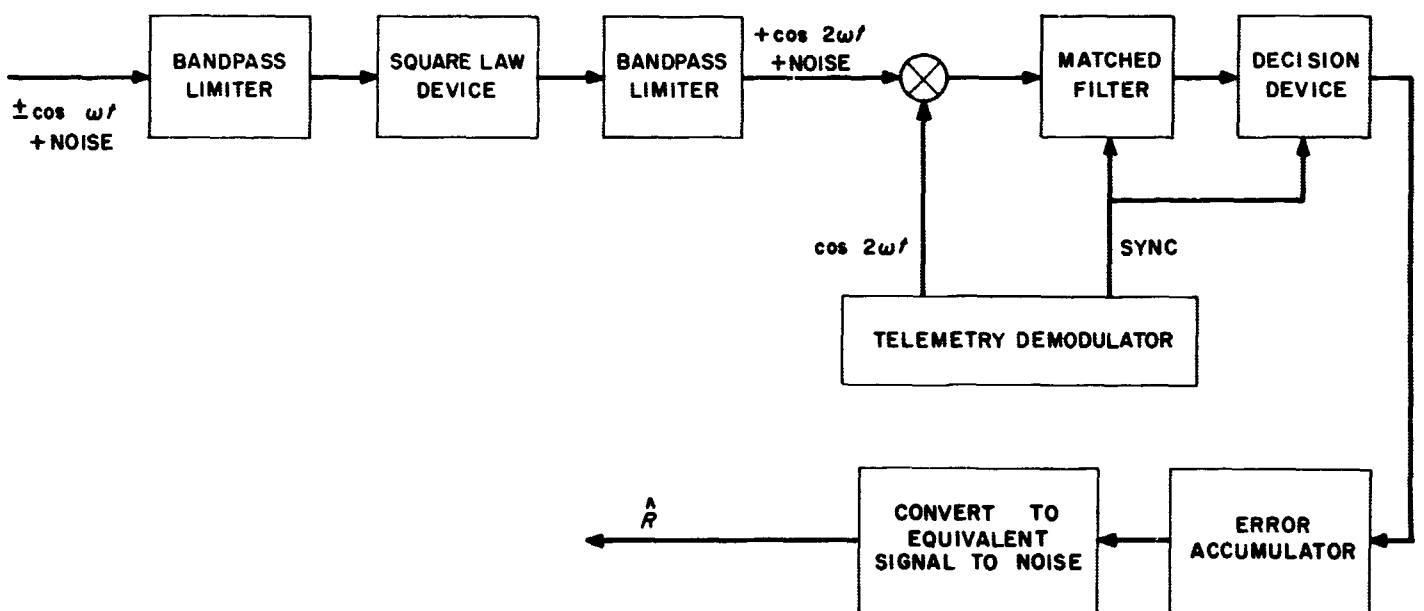


Fig. 2. Subcarrier bit error estimator system

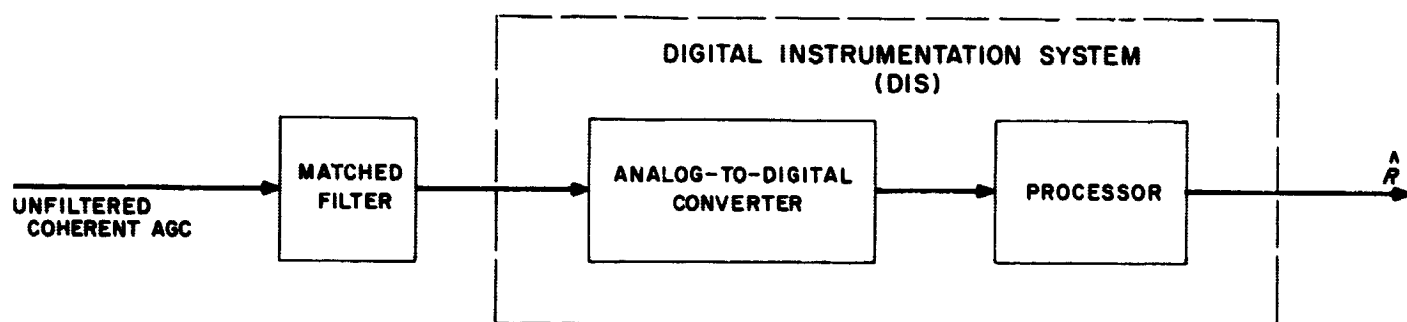


Fig. 3. Carrier signal-to-noise estimator system

will be operational at all DSIF Stations in CY 1965, this offers no difficulty.

A system, which at this time appears to satisfy these requirements best in relation to some of the fixed system constraints (unenumerated except that future telemetry systems will largely be digital), is shown in Figs. 1 (an alternate proposal to Fig. 2), 2 (an alternate proposal to Fig. 1), and 3; all systems, however, will be analyzed. Because the data is plus or minus in the digital sub-carrier channel, the output is not Gaussian. This difficulty is easily circumvented if the range of values of interest is restricted. First of all, the specification on the allowable bit error rate of the system has been set at  $5 \times 10^{-3}$  which means that  $\mu/\sigma$ , where  $\mu$  is the average signal and  $\sigma$  is the average noise value, is  $> 2.58 = 8.23$  db at the matched filter output. Fig. 4 shows what this means in the output probability density function. What is proposed is that the samples of the system are taken without regard to sign yielding the probability density function shown in Fig. 5, where the dotted line shows the Gaussian distribution which has the same area to be used as an approximation in the analysis to follow. Consideration of the above indicates that  $\mu$  will be higher and  $\sigma$  lower, thus giving an increased answer for  $\mu/\sigma$ ; but this effect should be small even for the worst case. Because of this problem, the system was chosen as outlined so that the  $\mu/\sigma$  could be as large as possible.

In the system outlined, the basic problem is to make a statistical estimate of the signal-to-noise ratio from  $n$  measurements  $\{X_1, X_2, \dots, X_n\}$  of the output universe. Since the universe average  $\mu$  is the signal and the universe standard deviation  $\sigma$  is the noise, the estimate of the signal to noise is the ratio of the estimates of  $\mu$  and  $\sigma$  or  $\hat{\mu}$  and  $\hat{\sigma}$  (estimators). These estimators are random variables which depend on the measurements  $\{X_1, X_2, \dots, X_n\}$ . The signal-to-noise ratio is then a functional relationship between two random variables and is itself a random variable.

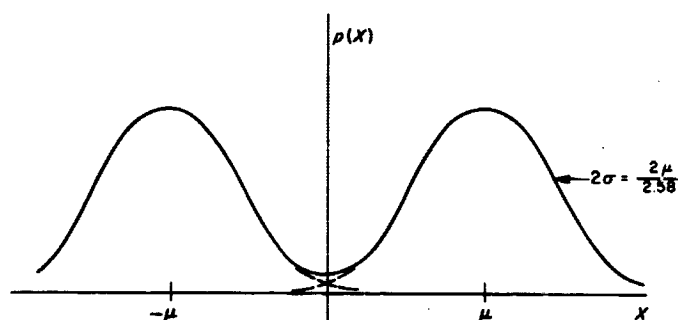


Fig. 4. Worst case probability density function for matched filter output

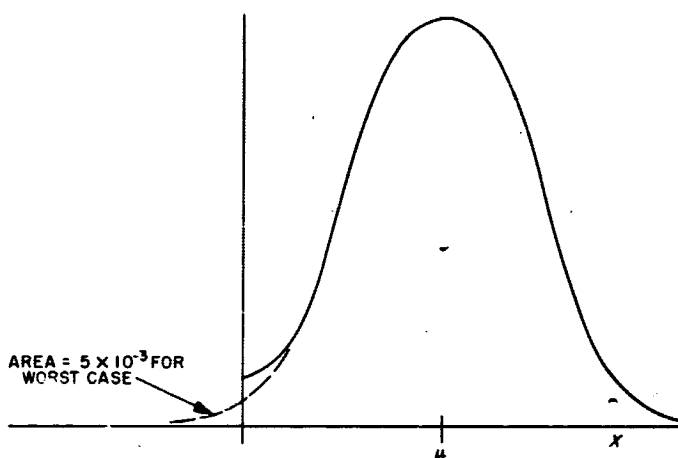


Fig. 5. Probability density function of worst case without regard to sign

There are a number of different estimators of  $\mu$  and  $\sigma$  covering different circumstances and efficiencies, but for  $\mu$  we shall use the estimator

$$\hat{\mu} = \frac{\sum_{i=1}^n X_i}{n}$$

and for  $\sigma$  we shall use

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \hat{\mu})^2}{n-1}$$

which is an unbiased estimator depending upon the estimate  $\hat{\mu}$  of  $\mu$ , where we might have used any one of the following:

$$\hat{\sigma}_1^2 = \frac{\sum_{i=1}^n (X_i - \hat{\mu})^2}{n} \quad (\text{a biased estimator})$$

$$\hat{\sigma}_2^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n-1}$$

$$\hat{\sigma}_3^2 = \frac{\sum_{i=1}^n (X_i - \mu)^2}{n} \quad (\text{a biased estimator})$$

(By an unbiased estimator, we mean that the ensemble average of the estimate is equal to parameter that is being estimated. Biased, of course, means the opposite.)

The measurements  $\{X_1, X_2, \dots, X_n\}$  are made from a universe which we have assumed to be Gaussian and, because of the nature of the matched filter, we can assume that the  $X_i$  are independent.  $\hat{\mu}$  then is functionally related to the  $X_i$  by the previous relation, and thus will have a probability density function which we can calculate. Proceeding then, because of the independence, the multivariate distribution is then (Ref. 1)

$$p(X_1, X_2, \dots, X_n) = \prod_{i=1}^n \frac{1}{(2\pi)^{1/2} \sigma} \exp \left[ -\frac{(X_i - \mu)^2}{2\sigma^2} \right]$$

Now, making the transformation

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i$$

it can be shown that

$$p(\hat{\mu}) = \left( \frac{n}{2\pi} \right)^{1/2} \frac{1}{\sigma} \exp \left[ -\frac{n(\hat{\mu} - \mu)^2}{2\sigma^2} \right]$$

$\hat{\mu}$  then is a random variable with the above probability density function; we can not say with certainty what  $\mu$  is at all. We can now calculate to see if  $\hat{\mu}$  is unbiased.

By inspection of  $p(\hat{\mu})$  it is obvious that

$$E[\hat{\mu}] = \mu;$$

Therefore, it is unbiased.

So far we have examined only a small part of the problem. What we are ultimately going to do is take a ratio of  $\hat{\mu}$  and  $\hat{\sigma}$

$$\hat{R} = \frac{\hat{\mu}}{\hat{\sigma}}$$

where  $\hat{R}$  is an estimator for  $\mu/\sigma = R$  and what we need is the joint distribution  $p(\hat{\mu}, \hat{\sigma})$ . In the calculation for  $\hat{\sigma}$ , we had several choices for the estimator and

$$\hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \hat{\mu})^2$$

was chosen.  $\hat{\mu}$  was chosen in this calculation because we would not know  $\mu$  with any certainty. The distribution which we get for  $p(\hat{R})$  will involve both  $\mu$  and  $\sigma$  which are both unknown. Now it can be shown that (Ref. 2)

$$p(\hat{\mu}, \hat{\sigma}) = \frac{2 \left( \frac{n}{2\pi} \right)^{1/2} \left( \frac{n-1}{2} \right)^{(n-1)/2}}{\Gamma \left( \frac{n-1}{2} \right) \sigma^n} (\hat{\sigma})^{n-2} \times \exp \left[ -\frac{n(\hat{\mu} - \mu)^2 + (n-1)\hat{\sigma}^2}{2\sigma^2} \right]$$

which is the sought-after joint probability density function for  $\hat{\mu}$  and  $\hat{\sigma}$ . Notice that  $p(\hat{\mu}, \hat{\sigma}) = p(\hat{\mu})p(\hat{\sigma})$  which shows the independence of  $\hat{\mu}$  and  $\hat{\sigma}$ . A word of caution should be injected here. The probability density function for  $\hat{\sigma}$  is not unbiased while the probability density function for  $\hat{\sigma}^2$  is. We shall get the square for our ultimate distribution which will eliminate this problem; but we should still check for bias in our estimator  $\hat{R}^2 = \hat{\mu}^2/\hat{\sigma}^2$ .

The variable was  $\hat{R} = \hat{\mu}/\hat{\sigma}$  where  $\hat{\sigma} > 0$  so that certainly the ratio exists and the range for  $R$  is  $-\infty < R < \infty$ .  $R$  can be interpreted as a line emanating from the origin in the positive  $\hat{\sigma}$  direction.

To get  $p(\hat{R})$  we must calculate  $(\partial/\partial\hat{R})P(\hat{R} \leq \hat{R})$ . Now

$$P(\hat{R} < \hat{R}) = \int_0^\infty \int_{-\infty}^{\hat{\mu} - \hat{R}\hat{\sigma}} p(\hat{\mu}, \hat{\sigma}) d\hat{\mu} d\hat{\sigma}$$

which is the sector between the  $-\hat{\mu}$  line and the  $\hat{R}$  line as shown in Fig. 6. Use the following identity for calculating the derivative under the integral:

$$\begin{aligned} \frac{d}{dc} \int_a^b f(x, c) dx &= \int_a^b \frac{\partial}{\partial c} f(x, c) dx \\ &+ f(b, c) \frac{\partial b}{\partial c} - f(a, c) \frac{\partial a}{\partial c} \end{aligned}$$

so that

$$\begin{aligned} p(\hat{R}) &= \frac{\partial}{\partial \hat{R}} P(\hat{R} < \hat{R}) \\ &= \int_0^\infty \frac{\partial}{\partial \hat{R}} \int_{-\infty}^{\hat{\mu} - \hat{R}\hat{\sigma}} p(\hat{\mu}, \hat{\sigma}) d\hat{\mu} d\hat{\sigma} \\ &= \int_0^\infty [0 + p(\hat{\mu} = \hat{R}\hat{\sigma}, \hat{\sigma}) \hat{\sigma} - 0] d\hat{\sigma} \\ &= \int_0^\infty p(\hat{\mu} = \hat{R}\hat{\sigma}, \hat{\sigma}) \hat{\sigma} d\hat{\sigma} \end{aligned}$$

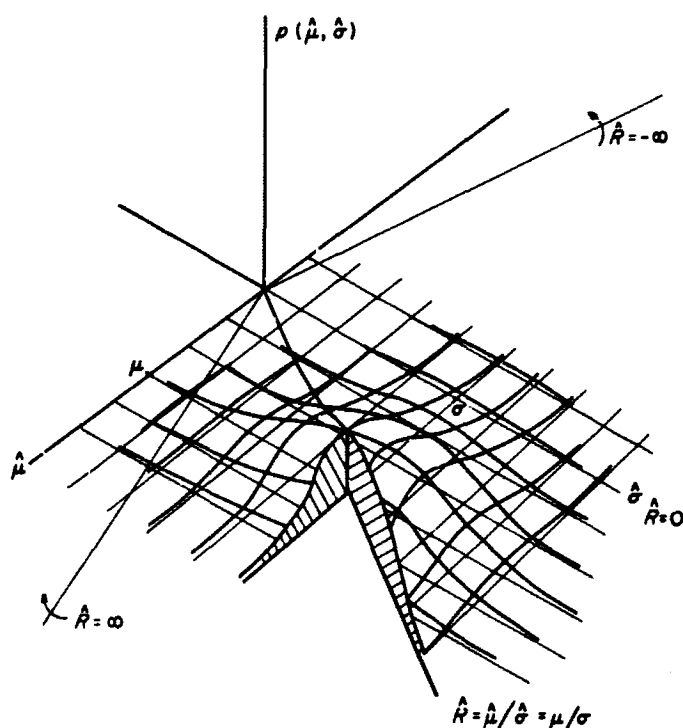


Fig. 6. Joint probability density function  $p(\hat{\mu}, \hat{\sigma})$

Therefore

$$\begin{aligned} p(\hat{R}) &= \frac{2 \left( \frac{n}{2\pi} \right)^{1/2} \left( \frac{n-1}{2} \right)^{(n-1)/2}}{\Gamma \left( \frac{n-1}{2} \right) \sigma^n} \\ &\times \int_0^\infty \hat{\sigma}^{n-1} \exp \left[ - \frac{n(\hat{R}\hat{\sigma} - \mu)^2 + (n-1)\hat{\sigma}^2}{2\sigma^2} \right] d\hat{\sigma} \end{aligned}$$

Expanding the exponent, we obtain

$$\begin{aligned} \frac{n(\hat{R}\hat{\sigma} - \mu)^2 + (n-1)\hat{\sigma}^2}{2\sigma^2} &= \frac{n\hat{R}^2\hat{\sigma}^2 - 2n\hat{R}\hat{\mu}\hat{\sigma} + n\mu^2 + (n-1)\hat{\sigma}^2}{2\sigma^2} \\ &= \frac{\hat{\sigma}^2 (n(\hat{R}^2 + 1) - 1) - 2n\hat{R}\hat{\mu}\hat{\sigma} + n\mu^2}{2\sigma^2} \end{aligned}$$

Now

$$\begin{aligned} p(\hat{R}) &= \frac{2 \left( \frac{n}{2\pi} \right)^{1/2} \left( \frac{n-1}{2} \right)^{(n-1)/2}}{\Gamma \left( \frac{n-1}{2} \right) \sigma^n} e^{-n\mu^2/2\sigma^2} \\ &\times \int_0^\infty \hat{\sigma}^{n-1} \exp \left[ - \frac{\sigma^2 (n(\hat{R}^2 + 1) - 1) - 2n\hat{R}\hat{\mu}\hat{\sigma}}{2\sigma^2} \right] d\hat{\sigma} \end{aligned}$$

which is a form of the non-central student  $t$  distribution (Refs. 8, 9), but (Ref. 3)

$$\begin{aligned} \int_0^\infty z^k e^{-bz - az^2/2} dz &= \\ a^{-(k+1)/2} 2^{(k-1)/2} &\left[ \Gamma \left( \frac{k+1}{2} \right) {}_1F_1 \left( \frac{k+1}{2}, \frac{1}{2}; \frac{b^2}{2a} \right) \right. \\ &\left. - b \left( \frac{2}{a} \right)^{1/2} \Gamma \left( \frac{k}{2} + 1 \right) {}_1F_1 \left( \frac{k}{2} + 1, \frac{3}{2}; \frac{b^2}{2a} \right) \right] \end{aligned}$$

where

$$k = n - 1$$

$$b = -2nR\mu/2\sigma^2$$

$$a = (n(\hat{R}^2 + 1) - 1)/\sigma^2$$



Therefore

$$p(\hat{R}) = \frac{\left(\frac{n}{\pi}\right)^{1/2} (n-1)^{(n-1)/2}}{\Gamma\left(\frac{n-1}{2}\right) (n(\hat{R}^2 + 1) - 1)^{n/2}} e^{-n\mu^2/2\sigma^2} \\ \times \left[ \Gamma\left(\frac{n}{2}\right) {}_1F_1\left(\frac{n}{2}, \frac{1}{2}; \frac{\mu^2 n^2 \hat{R}^2}{2\sigma^2 (n(\hat{R}^2 + 1) - 1)}\right) \right. \\ \left. + \frac{\mu}{\sigma} \frac{n\hat{R}(2)^{1/2}}{(n(\hat{R}^2 + 1) - 1)^{1/2}} \Gamma\left(\frac{n+1}{2}\right) \right. \\ \left. \times {}_1F_1\left(\frac{n+1}{2}, \frac{3}{2}; \frac{\mu^2 n^2 \hat{R}^2}{2\sigma^2 (n(\hat{R}^2 + 1) - 1)}\right) \right]$$

Because we are really interested in  $\hat{R}^2$  and not  $\hat{R}$ , we now make a functional transformation. Since

$$\hat{R}^2 = \hat{R}^2$$

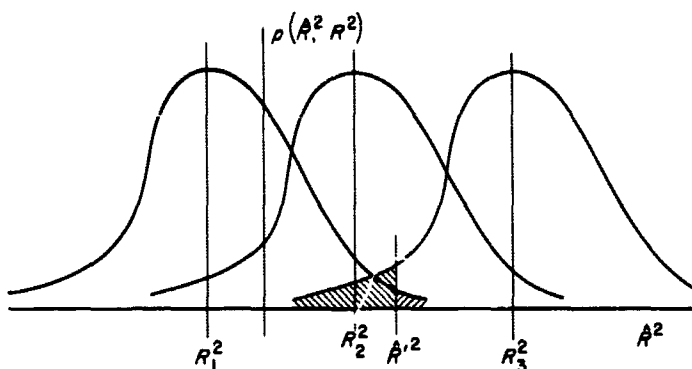
$$d\hat{R}^2 = \frac{d\hat{R}^2}{2(\hat{R}^2)^{1/2}} = \frac{d\hat{R}^2}{2(\hat{R}^2)^{1/2}}$$

Then, provided that the probability that  $\hat{R} < 0$  is very small,

$$p(\hat{R}^2) = \frac{\left(\frac{n}{\pi}\right)^{1/2} (n-1)^{(n-1)/2}}{2\Gamma\left(\frac{n-1}{2}\right) (n(\hat{R}^2 + 1) - 1)^{n/2} (\hat{R}^2)^{1/2}} e^{-n\mu^2/2\sigma^2} \\ \times \left[ \Gamma\left(\frac{n}{2}\right) {}_1F_1\left(\frac{n}{2}, \frac{1}{2}; \frac{\mu^2}{2\sigma^2} \frac{n^2 \hat{R}^2}{(n(\hat{R}^2 + 1) - 1)}\right) \right. \\ \left. + \frac{\mu}{\sigma} \frac{n(\hat{R}^2)^{1/2} (2)^{1/2}}{(n(\hat{R}^2 + 1) - 1)^{1/2}} \Gamma\left(\frac{n+1}{2}\right) \right. \\ \left. \times {}_1F_1\left(\frac{n+1}{2}, \frac{3}{2}; \frac{\mu^2}{2\sigma^2} \frac{n^2 \hat{R}^2}{(n(\hat{R}^2 + 1) - 1)}\right) \right]$$

This is the relation for  $p(\hat{R}^2)$  which we needed.

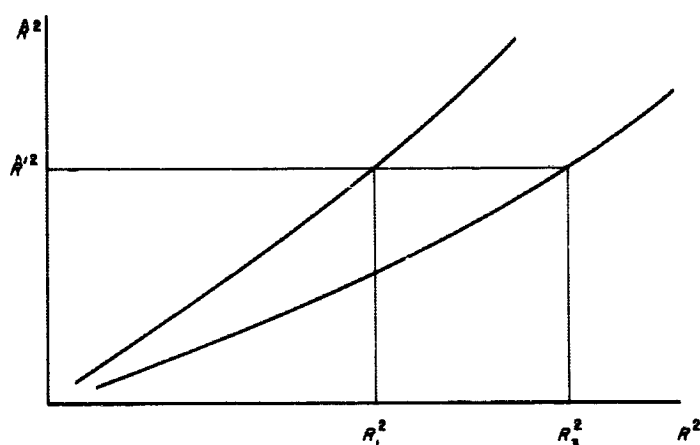
The function that we have obtained now for  $p(\hat{R}^2)$  is in unknown terms  $R^2 = \mu^2/\sigma^2$ , which is the parameter we are trying to estimate. Let us now write  $p(\hat{R}^2)$  as  $p(\hat{R}^2, R^2)$  mentioned before; there is really nothing we can say about  $\hat{R}^2$  for certain. The usual way of handling this problem is through confidence intervals. If  $R^2$  is known, we have probability function for various  $R^2$  as shown in the sketch below:



If experimentally we obtain a value  $\hat{R}^2$ , we can say that if  $R_1$  is the true parameter, then  $\hat{R}^2$  will not exceed  $\hat{R}_1^2$  with probability greater than a specified amount ( $P_1$ ). We can also say that if  $R_3$  is the true parameter then  $\hat{R}^2$  will exceed  $\hat{R}_3^2$  with probability greater than a specified amount ( $P_3$ ). These two specified probabilities need not be equal. We can say that the range  $R_1^2$  to  $R_3^2$  (called the confidence interval) will cover the true value  $R^2$  with a probability of  $(1 - P_1 - P_3)$ . It is obvious that  $P_1$  and  $P_3$  can be chosen in an infinite number of ways. What we try to do is select  $P_1$  and  $P_3$  so that  $R_3^2 - R_1^2$  is the shortest length. If  $p(\hat{R}^2, R)$  is symmetrical, then  $P_1 = P_3$  yields this circumstance; but in the unsymmetrical case, it is difficult in general to obtain the best values for  $P_1$  and  $P_3$ . In practice,  $P_1$  and  $P_3$  are chosen equal because of this difficulty, but the increased range of  $R_1^2$  to  $R_3^2$  is not affected appreciably in the majority of cases. The problem now is one of determining the following integrals for  $R_1$  and  $R_3$ .

$$\int_{-\infty}^{\hat{R}_1^2} p(\hat{R}^2, R_1) d\hat{R}^2 = P_1 = P_2 = \int_{\hat{R}_3^2}^{\infty} p(\hat{R}^2, R_3) d\hat{R}^2$$

For any particular  $\hat{R}^2$  we can plot the values  $R_1^2$  and  $R_3^2$  as shown in the second sketch:



From this sketch then, if we obtain  $\hat{R}^2$  experimentally, we can read the confidence interval by drawing a horizontal line to intersect the curves and project the intersection to the abscissa to obtain the confidence interval  $R_1^2$  to  $R_3^2$ .

In order to reduce the range of the confidence interval, the only parameter we have control of is the number of samples. Obviously, as  $n$  is increased, the confidence interval decreases, and we have more assurance of obtaining a better answer. One of the main problems outlined in the beginning of this treatise was that of determining the number of samples. This is extremely important to the operation because of time required to gather a large number of samples, because the parameters could change over such a large time, and because of the amount of equipment required to process a large number of samples. Figs. 7 through 9 show the confidence intervals for various numbers of samples.

So far, only one system has been analyzed which gives us no measure of the efficiency (minimum number of samples for a given confidence interval) of that system. One other system which has been suggested (Fig. 2) to measure signal-to-noise ratio is that of measuring the bit error rate which is functionally related to signal-to-noise ratio (Refs. 10 and 11). This device essentially uses the probability function shown as the dashed line in Fig. 5 and makes a decision whether the function is  $X > 0$  or  $X < 0$ . The relation between the signal-to-noise ratio is

$$q^* = \frac{1}{2} \left( 1 - \operatorname{erf} \left( \frac{S}{N} \right)^{1/2} \right)$$

and the samples  $\{X_1, X_2, \dots, X_n\}$  will be discretely distributed with parameter  $q$  (Ref. 1)

$$p(x; q) = q^x (1 - q)^{1-x} \quad x = 0, 1$$

The maximum likelihood estimator for  $q$  is

$$\hat{q} = \frac{y}{n}$$

where

$$y = \sum_{i=1}^n x_i$$

which can have the possible values  $0, 1, 2, \dots, n$ . Now

$$p(X_1, X_2, X_3, \dots, X_n; q) = q^{\sum_{i=1}^n x_i} (1 - q)^{n - \sum_{i=1}^n x_i}$$

The only possible values for  $\hat{q}$  are  $0, 1/n, 2/n, \dots, 1$ ; and the probability density function for  $\hat{q}$  is

$$p\left(\hat{q} = \frac{j}{n}, q\right) = \binom{n}{j} q^j (1 - q)^{n-j}$$

$$\frac{j}{n} = 0, \frac{1}{n}, \frac{2}{n}, \dots, 1$$

In an analogous manner to that of the continuous case above for the confidence intervals when we experimentally get a  $\hat{q}' = k/n$ , we find a  $q_1$  and a  $q_3$  from the following

$$\sum_{v=0}^k \binom{n}{v} q_1^v (1 - q_1)^{n-v} = P_1$$

$$= \sum_{v=k}^n \binom{n}{v} q_3^v (1 - q_3)^{n-v} = P_2$$

For small  $n$ , hand trials can be used for the solutions  $q_1$  and  $q_3$ , but for intermediate values of  $n$ , the incomplete beta function can be used. The cumulative beta distribution is

$$F(X; \alpha, \beta) = \frac{(\alpha + \beta + 1)!}{\alpha! \beta!} \int_0^X t^\alpha (1 - t)^\beta dt$$

which, with integration by parts, yields

$$F(X; \alpha, \beta) = - \sum_{i=0}^{\alpha} \binom{\alpha + \beta + 1}{i} X^i (1 - X)^{\alpha + \beta + 1 - i} + 1$$

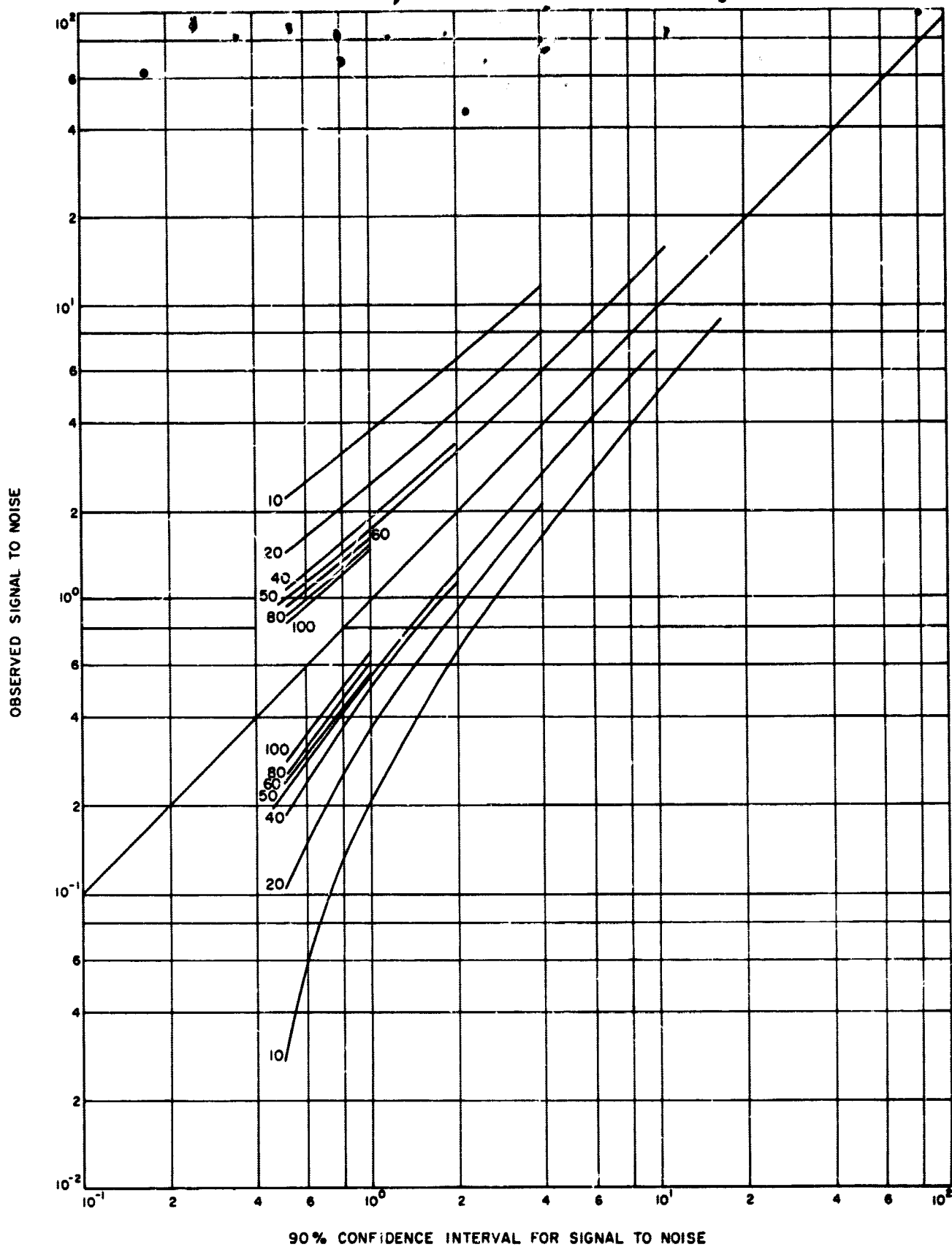


Fig. 7. 90% confidence intervals for S/N

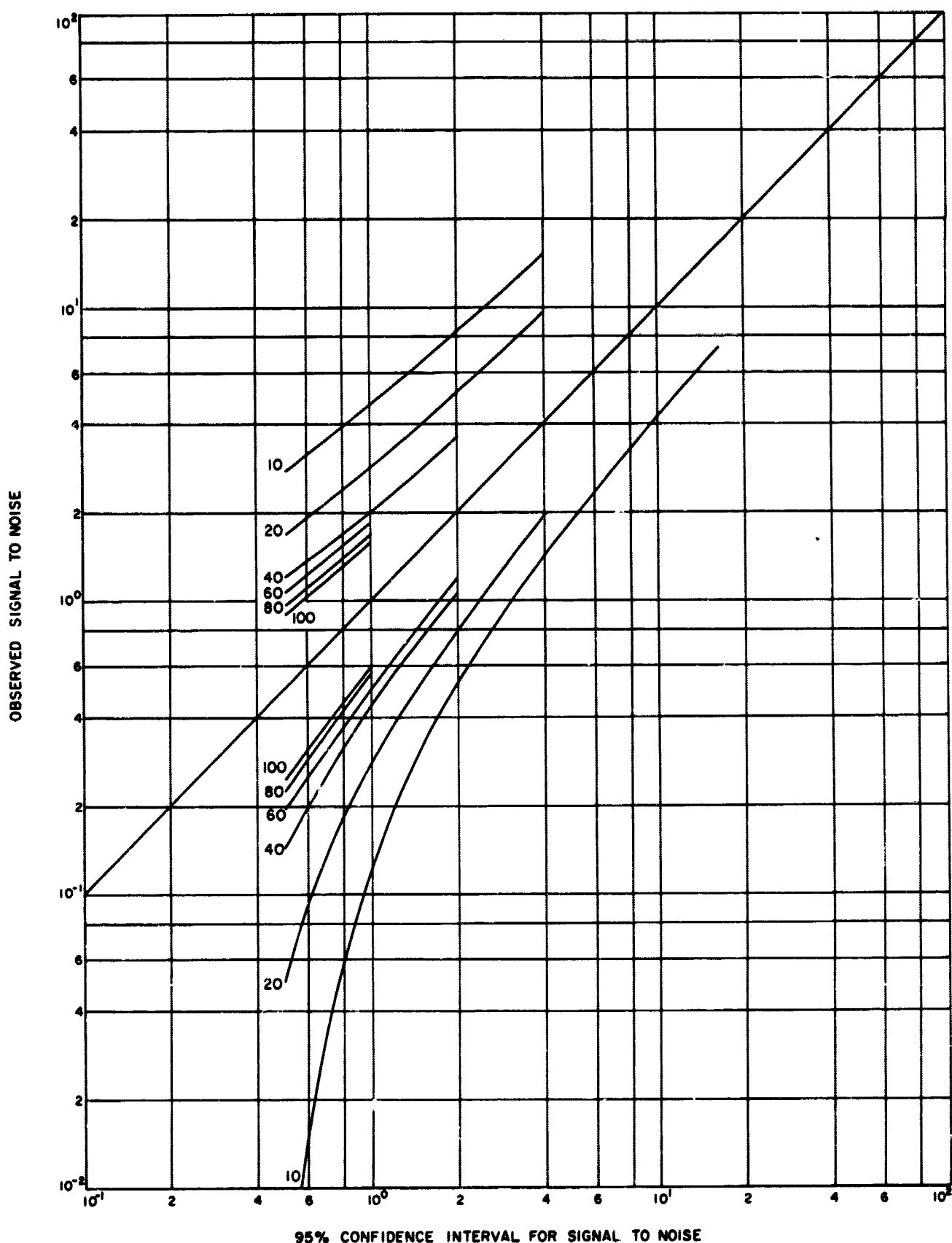


Fig. 8. 95% confidence intervals for S/N

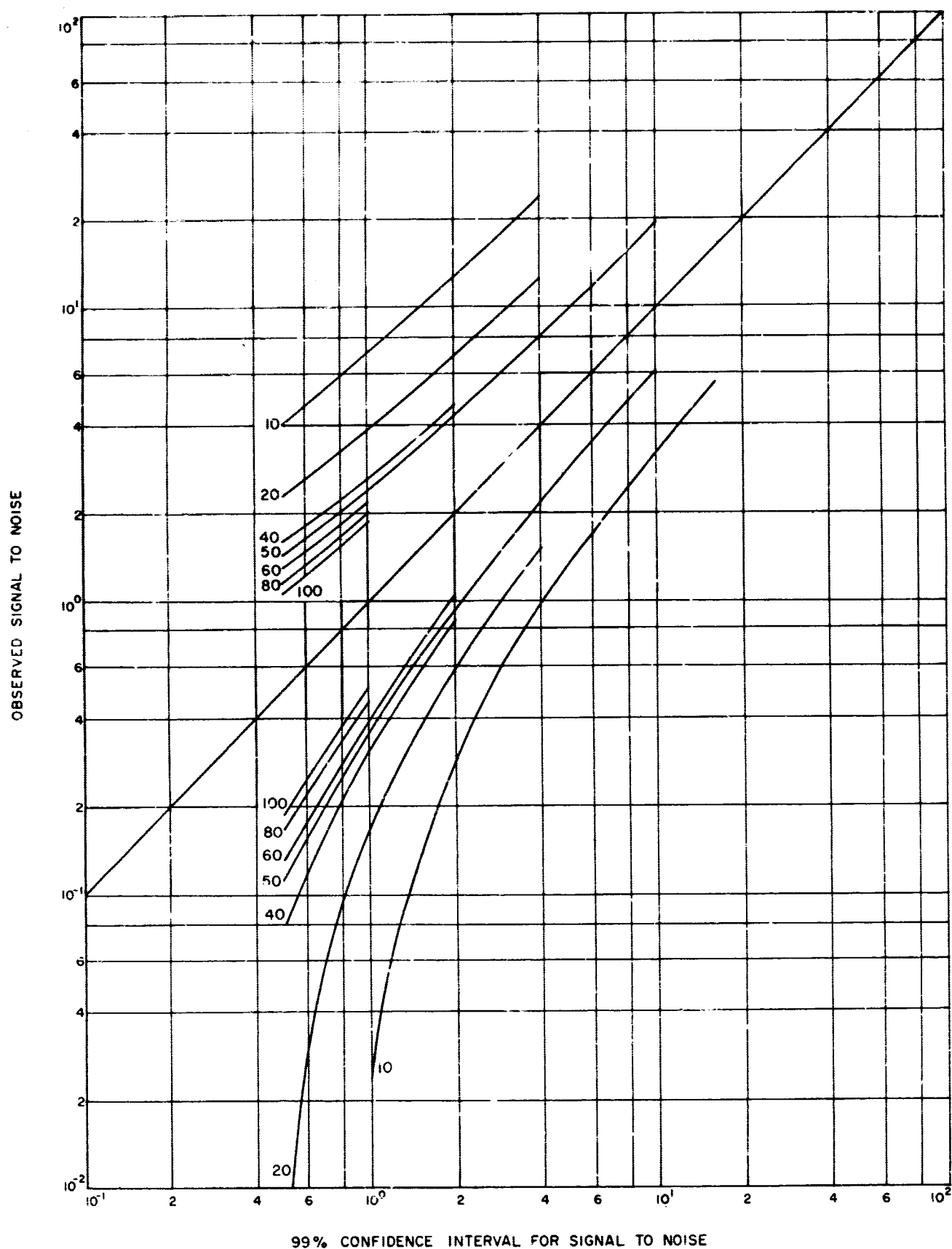


Fig. 9. 99% confidence intervals for S/N

Therefore

$$\sum_{y=0}^k \binom{n}{y} q_1^y (1 - q_1)^{n-y} = 1 - F(q_1; k, n - k - 1) = P_1$$

and

$$\sum_{y=k}^n \binom{n}{y} q_3^y (1 - q_3)^{n-y} = F(q_3; k - 1, n - k) = P_3$$

The incomplete beta function may be found in Pearson's tables (Ref. 7 or Refs. 4 and 5). For values beyond these

tables, approximations must be made. For  $n$  large and  $q$  around 0.5, the normal approximation to the binomial should be made

$$F(q_3; k - 1, n - k) = \frac{1}{(2\pi)^{1/2}} \int_k^x \frac{e^{-t^2/2}}{\sqrt{nq_3(1 - q_3)}} dt$$

When  $q$  is small, the Poisson approximation should be made (Ref. 12)

$$F(q_3; k - 1, n - k) = \sum_{x=k}^{\infty} \frac{(nq)^x e^{-nq}}{x!}$$

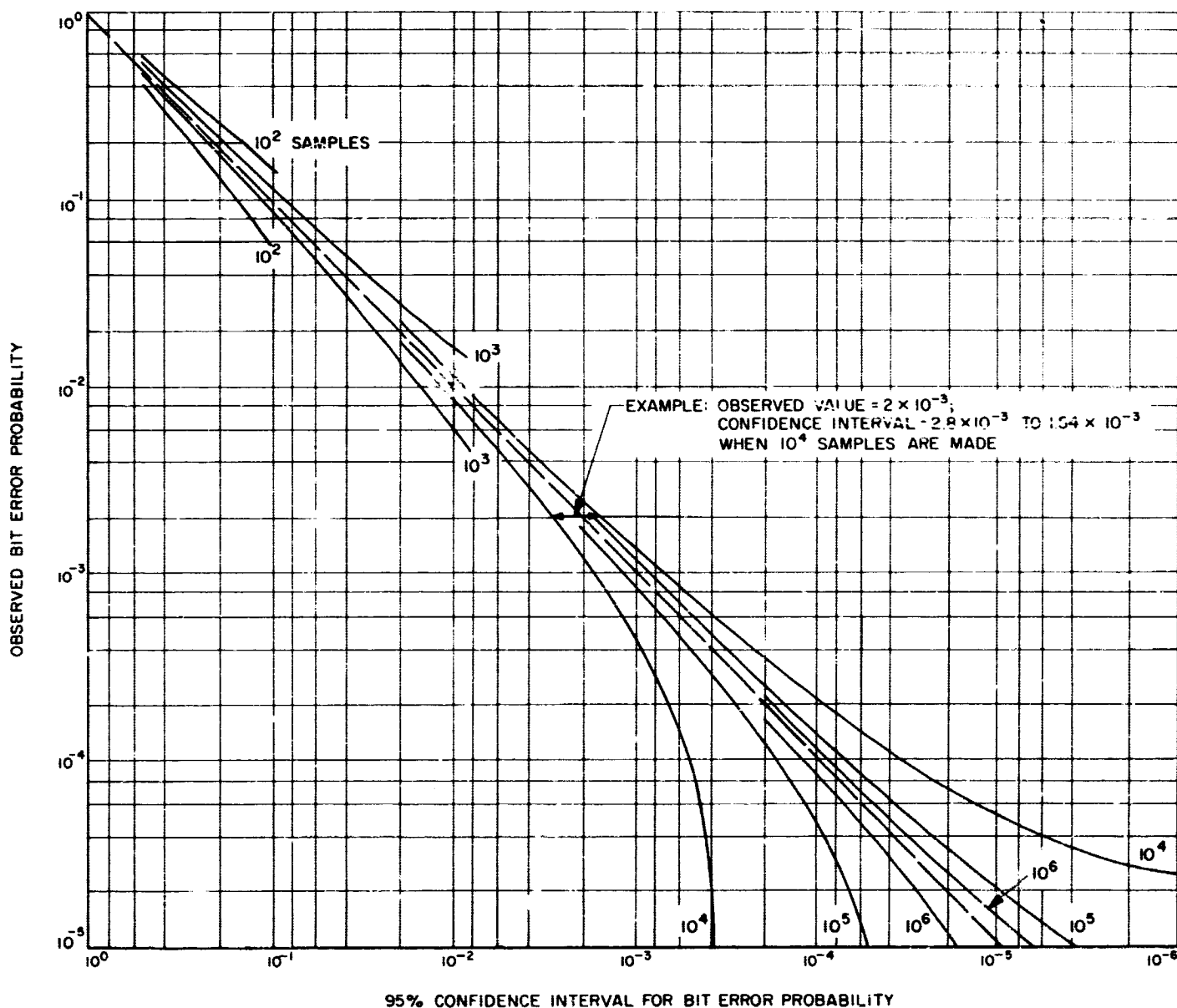


Fig. 10. 90% confidence intervals for bit error rates

Figs. 10 through 15 show the confidence intervals corresponding to the bit error rates.

Conclusions as to the efficiency (number of samples required) of the systems depend largely upon the region of  $S/N$  that is experienced. At high  $ST/N/B = 10$  (corresponding to a bit error rate of about  $10^{-5}$ ), about  $10^5$  samples are needed by the bit error estimator method to have a 90% confidence interval less than  $\pm 1$  db, while

the  $S/N$  estimator method would require only about  $10^3$  (confidence interval proportional to  $1/n^{1/2}$  when  $n$  is above 50). At  $ST/N/B = 1.0$  (corresponding to a bit error rate of about  $10^{-1}$ ), the two methods are comparable. Since the higher  $S/N$  will be prevalent during most of any mission and since  $10^5$  samples would require 3½ hr at 8½ bits/sec (*Mariner* bit rates) while  $10^3$  samples would require 6 min, it would seem that the  $S/N$  estimator method would be the most practical system.

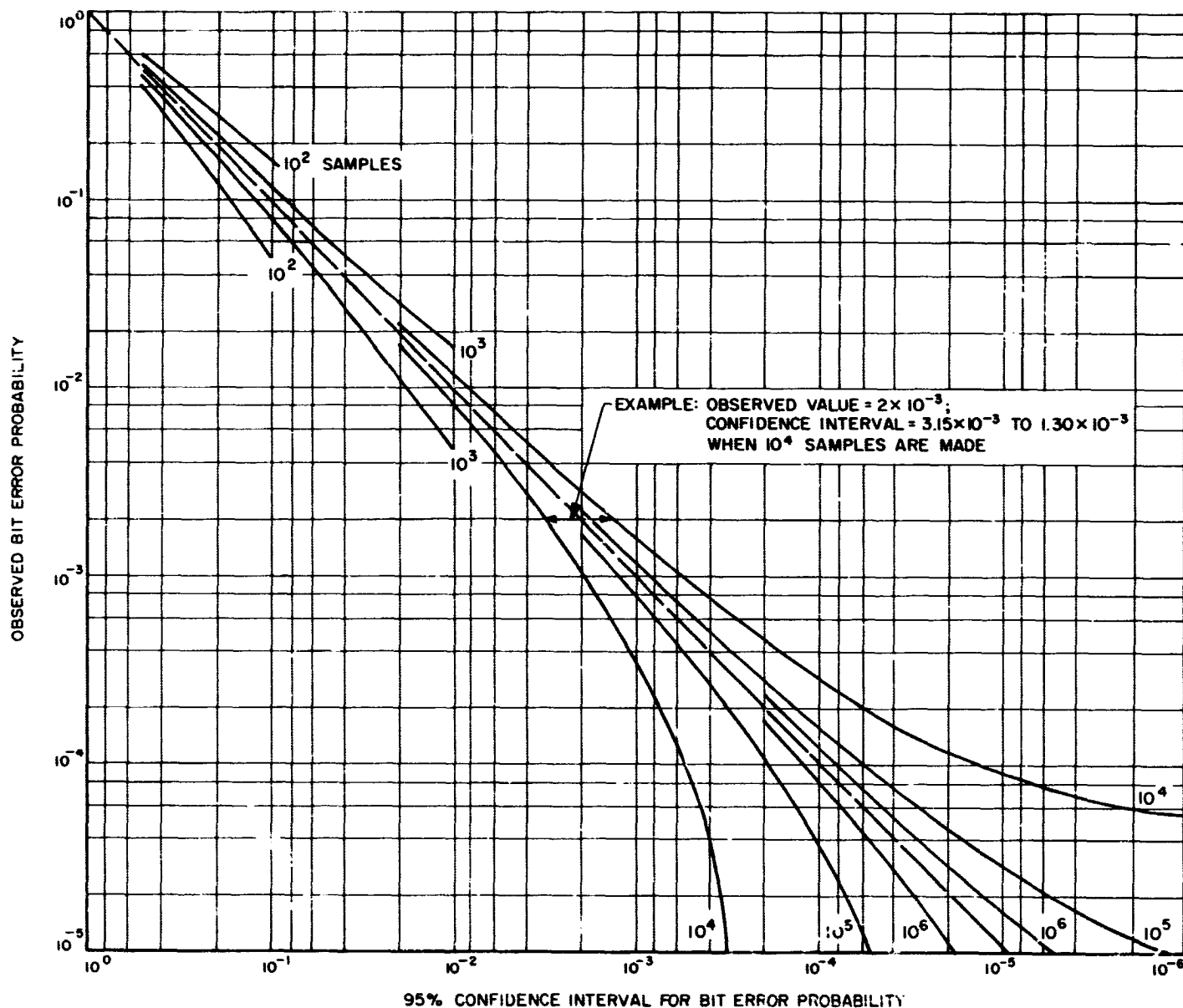


Fig. 11. 95% confidence intervals for bit error rates

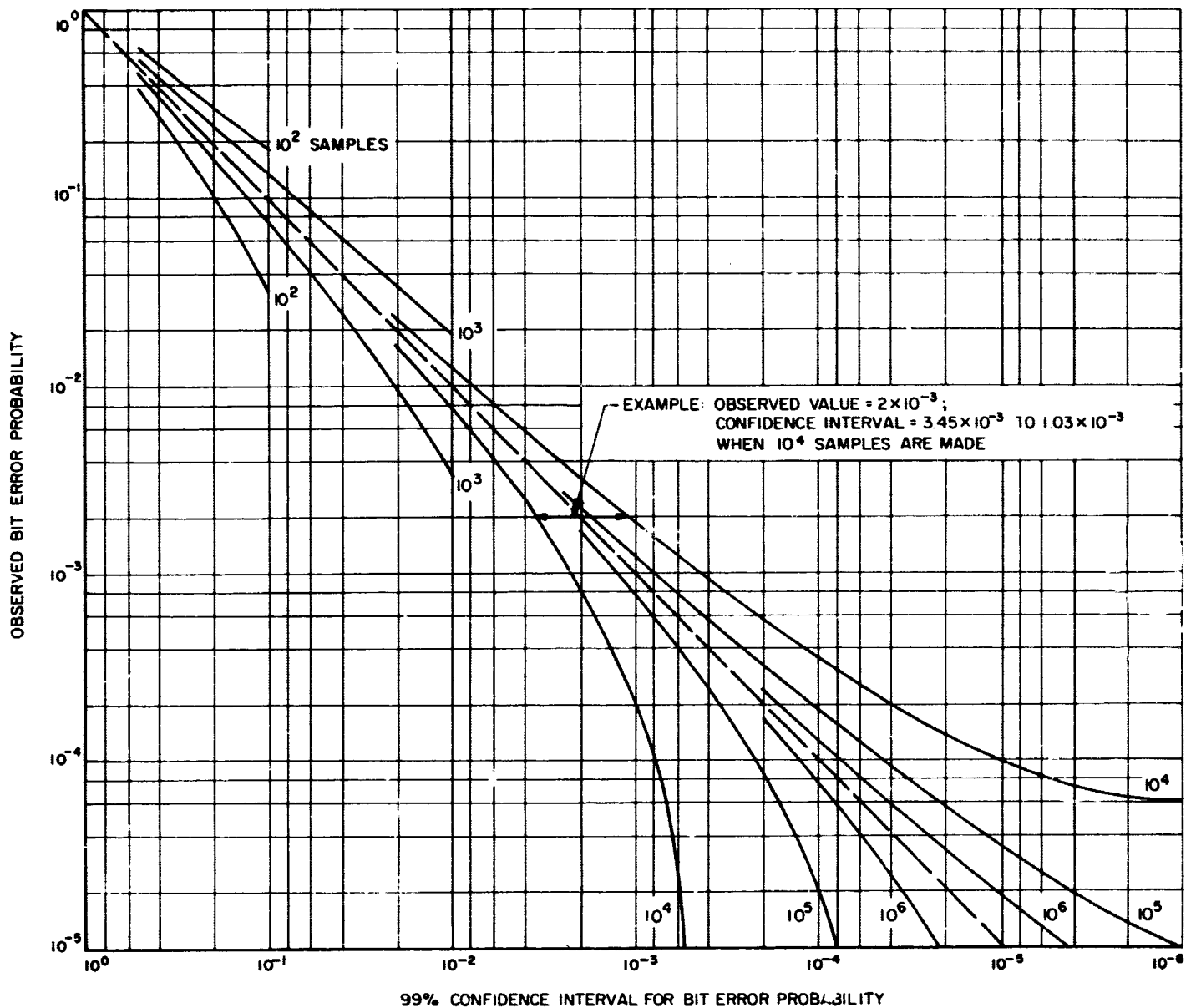
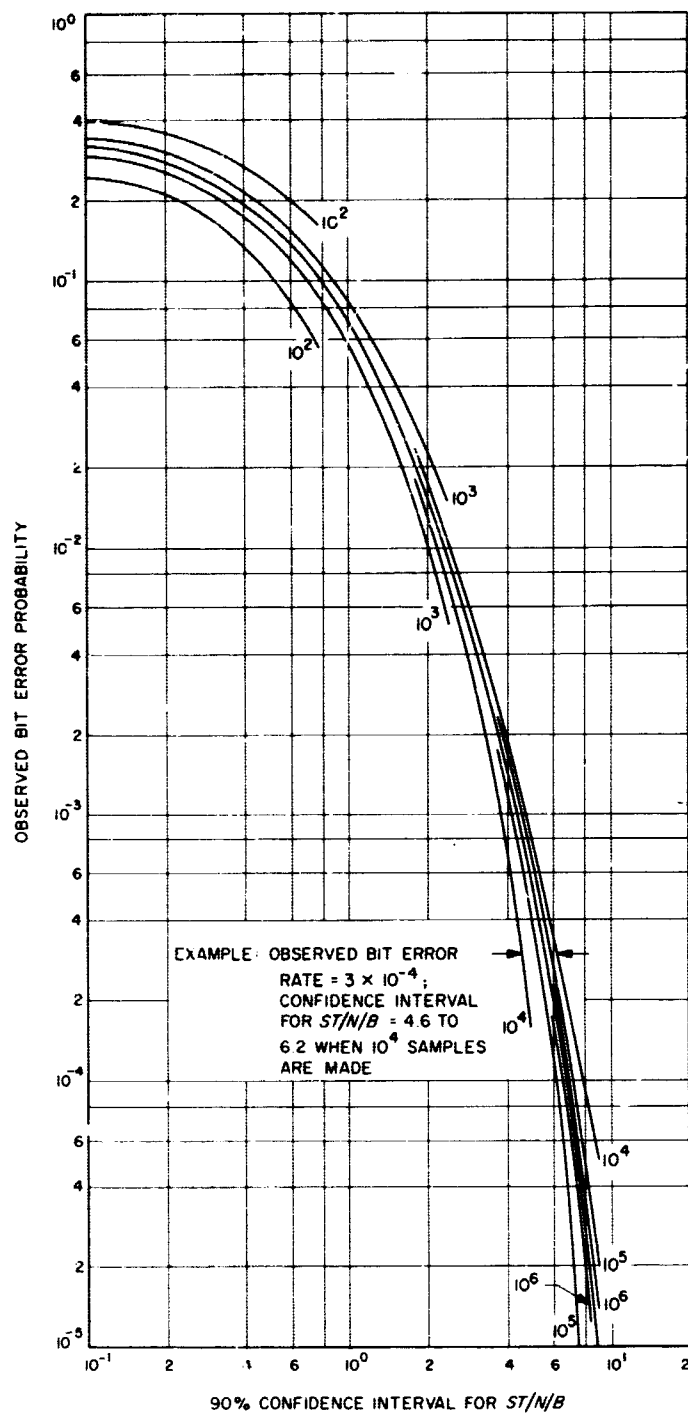
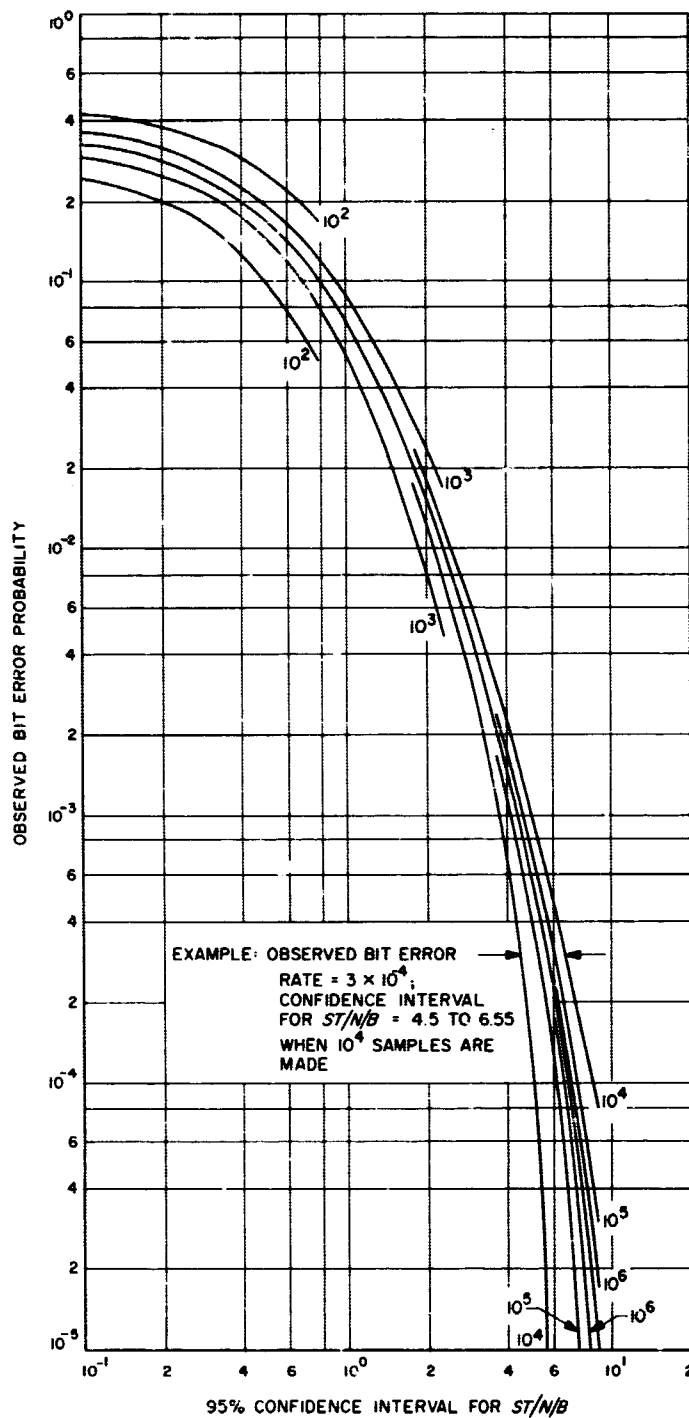


Fig. 12. 99% confidence intervals for bit error rates



Fig. 13. 90% confidence intervals for  $ST/N/B$ Fig. 14. 95% confidence intervals for  $ST/N/B$

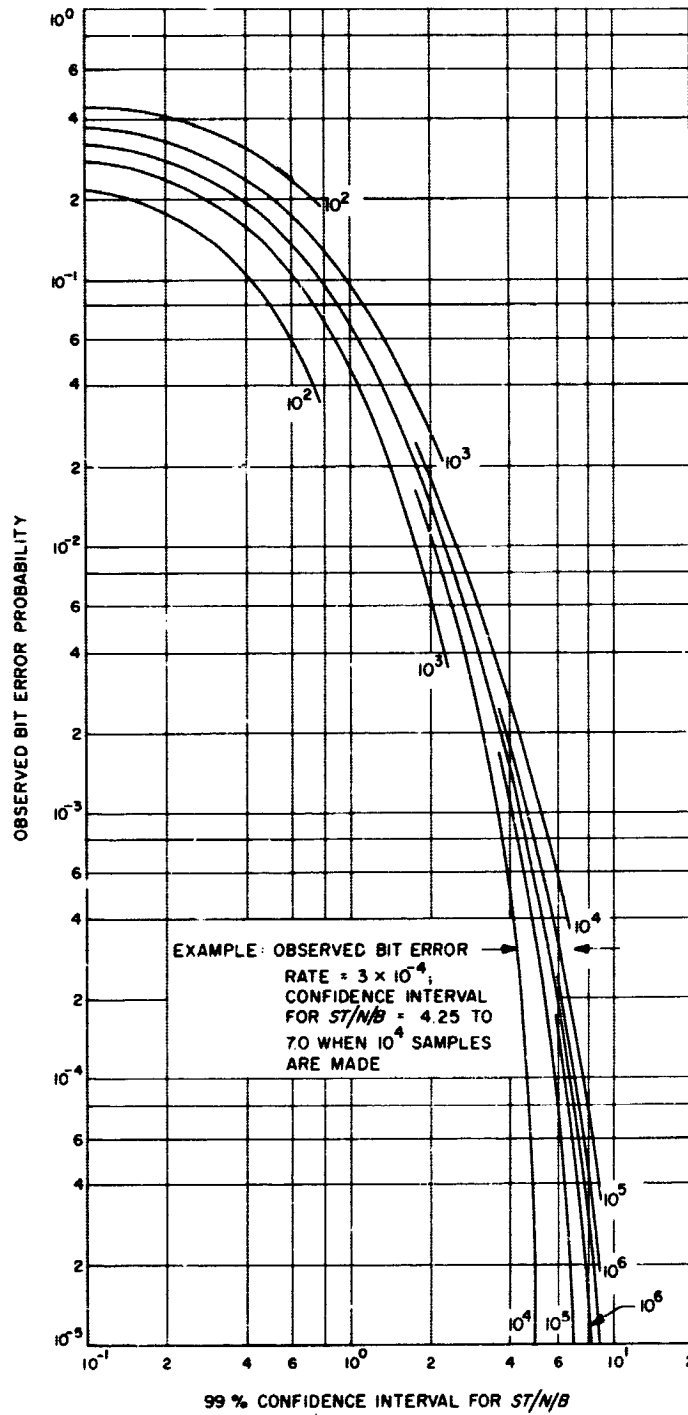


Fig. 15. 99% confidence intervals for  $ST/N/B$

## References

1. Mood, A. M., "Introduction to the Theory of Statistics," McGraw-Hill, 1950.
2. Anderson, R. L., and Bancroft, T. A., "Statistical Theory in Research," McGraw-Hill, 1952.
3. Middleton, D., "An Introduction to Statistical Communication Theory," McGraw-Hill, 1960.
4. Pearson, E. S., and Hartley, H. O., "Biometrika Tables for Statisticians, Vol. I," Cambridge, 1958.
5. Harvard Staff, "Tables for the Cumulative Binomial Probability Distribution, Annals 35," Harvard University Press, 1955.
6. Quenouille, M. H., "Fundamentals of Statistical Reasoning," Hafner, 1958.
7. Pearson, K., "Tables of the Incomplete Beta Function," Cambridge, 1934.
8. Johnson, N. L., and Welch, B. L., "Application of Non-central  $t$ -Distribution," *Biometrika*, Vol. 31, 1940.
9. Resnikoff, J., and Lieberman, G. J., "Tables of the Non-central  $t$ -Distribution," Stanford University Press, 1959.
10. Viterbi, A. J., "On Coded Phase-Coherent Communications," Technical Report No. 32-25, Jet Propulsion Laboratory, Pasadena, California, August 15, 1960.
11. Becke, H. D., and Lawten, J. G., "Theoretical Comparison of Binary Data Transmission Systems," Report No. CA-1172-5-1, Cornell Aeronautical Laboratory, Inc., Buffalo 21, New York, May 1958.
12. Defense Systems Dept., General Electric Co., "Tables of the Individual and Cumulative Errors of Poisson Distribution," Van Nostrand Co., Inc., 1962.

## XIX. Communications Systems Research

### A. Coding Theory

R. C. Titsworth, G. Solomon, and E. C. Posner

#### 1. Random Numbers Generated by Linear Recurrence Modulo 2, R. C. Titsworth

*a. Introduction.* Many situations arise in various fields of interest for which the mathematical model utilizes a random sequence of numbers, events, or both. In many of these applications it is often extremely advantageous to generate by some deterministic means a sequence which appears to be random, even if, upon closer and longer observation, certain regularities become evident. For example, electronic computer programs for generating random numbers to be used in Monte Carlo experiments have proved extremely useful. This article describes a random number generator of this type with several outstanding properties. The numbers are generated by modulo 2 linear recurrence techniques long used to generate binary codes for communications.

*b. Maximal length linear recurrence modulo 2.* Let  $a = \{a_k\}$  be the sequence of 0's and 1's generated by the linear recursion relation

$$a_k = c_1 a_{k-1} + c_2 a_{k-2} + \cdots + c_n a_{k-n} \pmod{2}$$

for any given set of integers  $c_i$  ( $i = 1, 2, \dots, n$ ), each having the value 0 or 1. We, of course, require  $c_n = 1$ , and say that the sequence has degree  $n$ .

The period  $p$  of such a sequence is clearly not greater than  $2^n - 1$ . The necessary and sufficient condition that  $p = 2^n - 1$  is that the polynomial

$$f(x) = 1 + c_1 x + c_2 x^2 + \cdots + x^n$$

be primitive mod 2 (Ref. 1).

We shall assume in the remainder of this article that  $f(x)$  is a primitive  $n$ th degree polynomial mod 2; the sequence  $a$  is then a *maximal length linearly recurring sequence mod 2*. These sequences have been studied and used as codes in communications and information theoretic studies (Refs. 2, 3, 4). The properties of interest to us at present are the following:

$$(1) \sum_{k=1}^p a_k = \frac{p+1}{2} = 2^{n-1}.$$

(2) For every distinct set of  $(0, 1)$  integers  $b_1, b_2, \dots, b_n$ , not all zero, there exists a unique integer  $v$  ( $0 \leq v < p - 1$ ), independent of  $k$ , such that

$$b_1 a_{k-1} + b_2 a_{k-2} + \cdots + b_n a_{k-n} = a_{k+v} \pmod{2}.$$

This is often referred to as the "cycle-and-add" property.

- (3) Every non-zero  $(0, 1)$  binary  $n$ -vector  $(e_1, e_2, \dots, e_n)$  occurs exactly once per period as  $n$  consecutive binary digits in  $a$ .

We shall, in what follows, find it convenient to use a slightly different version of the sequence  $a$ . Let us define

$$\alpha_k = (-1)^{a_k}$$

Under this transformation, we see that if  $a_k$  takes on the values 0 and 1, then  $\alpha_k$  takes the values of +1 and -1, respectively. The properties (1), (2), and (3) are then transformed into

$$(1') \sum_{k=1}^p \alpha_k = -1.$$

- (2') For every distinct set of  $(0, 1)$  integers  $b_1, \dots, b_n$ , not all zero, there exists a unique integer  $v$  ( $0 \leq v < p-1$ ) such that

$$\alpha_{k-1}^{b_1} \alpha_{k-2}^{b_2} \dots \alpha_{k-n}^{b_n} = \alpha_{k+v}.$$

- (3') Every  $\pm 1$  binary  $n$ -vector  $(\epsilon_1, \epsilon_2, \dots, \epsilon_n)$  *except* the all-ones vector occurs exactly once per period as  $n$  consecutive elements in  $\alpha$ .

**c. The Boolean transform.** Let  $g(x)$  be a  $\pm 1$ -valued Boolean function of  $(0, 1)$  variables  $x_1, x_2, \dots, x_n$ . For any  $s = (s_1, s_2, \dots, s_n)$ ,  $s_i = 0$  or 1, define

$$\phi(s, x) = 2^{-n/2} (-1)^{s_1 x_1 + \dots + s_n x_n}$$

These  $2^n$  functions of  $x$ , the Rademacher-Walsh functions (Ref. 5), form an orthonormal basis for  $2^n$ -space. Relative to this basis  $g(x)$  has components  $G(s)$  given by

$$G(s) = 2^{-n/2} \sum_x g(x) \phi(s, x).$$

That is,  $G(s)$  is the projection of  $g(x)$  on  $\phi(s, x)$ , normalized so that

$$\sum_s G^2(s) = 1.$$

Similarly, we have

$$g(x) = 2^{n/2} \sum_s G(s) \phi(s, x).$$

Consider the effect of setting  $x_i = a_{k-i}$  in  $g(x)$ . As a function of  $k$ , a binary  $\pm 1$  sequence  $\{\gamma_k\} = \gamma$  is generated:

$$\begin{aligned} \gamma_k &= \sum_s G(s) (-1)^{s_1 a_{k-1} + \dots + s_n a_{k-n}} \\ &= \sum_s G(s) \alpha_{k-1}^{s_1} \alpha_{k-2}^{s_2} \dots \alpha_{k-n}^{s_n}. \end{aligned}$$

Note in the above equation that the coefficients  $G(s)$  are determined by the logic alone, and that, by (2'), we have fourth property basic to our analysis:

$$(4) \gamma_k = G(0) + \sum_{s \neq 0} G(s) \alpha_{k+v(s)}$$

where the mapping  $v(s)$  of all binary non-zero  $n$ -vectors onto  $\{0, 1, 2, \dots, p-1\}$  is one-to-one.

**d. Random number generation.** Let  $a = \{a_k\}$  be the  $(0, 1)$  sequence generated by an  $n$ th degree maximal-length linear recurrence modulo 2, as described previously, and define a set of numbers of the form

$$y_k = 0.a_{qk+r-1} a_{qk+r-2} \dots a_{qk+r-n} \quad (\text{base } 2),$$

where  $r$  is a randomly chosen integer,  $0 \leq r < 2^n - 1$ . That is,  $y_k$  is the binary expansion of a number whose binary digits are  $n$  consecutive digits in  $a$ ,  $y_k$  spaced  $q$  digits apart. For reasons essential to the analysis, we restrict  $q \geq n$ , and  $(q, 2^n - 1) = 1$ .

We can also write

$$y_k = \sum_{t=1}^n 2^{-t} a_{qk+r-t}.$$

These all lie in the interval  $0 < y_k < 1$ . Because of condition (2), the randomness of the choice of  $r$  is equivalent to the statement that the initial value  $y_0$  is a random choice.

**e. Analysis of the generator.** We shall find it convenient to work with a transformed set of numbers  $w_k$  rather than the  $y_k$ . Specifically, let  $\alpha = \{\alpha_k\}$  be the  $\pm 1$  sequence corresponding to  $a = \{a_k\}$ , and define

$$w_k = \sum_{t=1}^n 2^{-t} \alpha_{qk+r-t}.$$

We see that  $y_k$  and  $w_k$  are related by

$$w_k = 1 - 2^n - 2y_k.$$

There is thus an easy translation between  $w_k$  and  $y_k$ .

By property (3), it is recognized that every one of the possible binary expansions of a number occurs exactly once per period, except that giving  $y_k = 0$ . Hence, for  $y_k$

$$2^{-n} \leq y_k \leq 1 - 2^{-n}$$

and for  $w_k$

$$-1 + 2^{-n} \leq w_k \leq 1 - 3 \cdot 2^{-n}.$$

We generally may assume, merely from the application to which we wish to suit the numbers, that  $n$  is moderately large, so that the numbers  $y_n$  and  $w_n$  are extremely numerous. For example, if  $n = 35$ , there are  $3.43 \times 10^{10}$  of them. We wish to consider only a portion of the total number of these, say  $N$  of them, and to discover, for moderately large  $N$ , how these are distributed.

*f. Correlation properties.* The mean value of  $w_k$  is easily found

$$\begin{aligned} E(w_k) &= \frac{1}{p} \sum_{r=0}^{p-1} w_k = \frac{1}{p} \sum_{t=1}^n 2^{-t} \sum_{r=0}^{p-1} \alpha_{qk+r-t} \\ &= -2^{-n} \end{aligned}$$

a number very nearly equal to zero. This produces  $E(y_k) = 1/2$ .

Define the sample autocorrelation function  $\hat{R}(m)$  of  $w_k$  by

$$\hat{R}(m) = \frac{1}{N} \sum_{k=1}^N w_k w_{k+m}$$

The expected value of  $\hat{R}(m)$  is the true autocorrelation  $R(m)$  function of the process

$$R(m) = E[\hat{R}(m)]$$

and the value  $R(0)$  is the variance  $\sigma^2$  of the process  $w_k$ .

$$R(0) = \frac{1}{p} \sum_{r=0}^{p-1} w_k^2 = \frac{1}{p} \sum_{t=1}^n \sum_{u=1}^n 2^{-(t+u)} \sum_{r=0}^{p-1} \alpha_{qk+r-t} \alpha_{qk+r-u}.$$

The last sum is  $-1$  if  $t \neq u$  and  $p$  if  $t = u$ , by (2'). Hence

$$R(0) = \frac{1}{3} - \left( \frac{2}{3} - 2^{-n} \right) 2^{-n}.$$

This shows that  $w_k$  has essentially the same variance as a uniformly distributed process.

Now consider  $\hat{R}(m)$ ,  $m \neq 0$ . First, its mean value is

$$\begin{aligned} R(m) &= E[\hat{R}(m)] \\ &= \frac{1}{pN} \sum_{k=1}^N \sum_{t=1}^n \sum_{u=1}^n 2^{-(t+u)} \sum_{r=0}^{p-1} \alpha_{qk+r-t} \alpha_{q(k+m)+r-u} \\ &= \frac{1}{p} \sum_{t=1}^n \sum_{u=1}^n 2^{-(t+u)} \sum_{r=0}^{p-1} \alpha_r \alpha_{r+qm+t-u}. \end{aligned}$$

The last sum is again  $-1$  unless  $qm + t - u$  is a multiple of  $p$ . Obviously,

$$qm - n + 1 \leq qm + t - u \leq qm + n - 1.$$

Hence, if  $q \geq n$  and  $m \leq (p-n)/q$ , we see that

$$0 < qm + t - u < p$$

so  $qm + t - u$  can never be a multiple of  $p$ . These conditions we shall now assume to be in effect as one of our hypotheses. The mean value of  $\hat{R}(m)$  is then

$$\begin{aligned} R(m) &= -\frac{1}{p} (1 - 2^{-n})^2 \\ &= -\frac{p}{(p+1)^2} = 2^{-2n} - 2^{-n}. \end{aligned}$$

The mean behavior of the process shows essentially no correlation between  $w_k$  and  $w_{k+m}$  for any  $m$  ( $m = 1, 2, \dots, [(p-n)/q]$ ).

The sample autocorrelation function is itself a random process, and its mean-squared value for  $m \neq 0$  is

$$E[\hat{R}^2(m)] = \sum_{t=1}^n \sum_{u=1}^n \sum_{i=1}^n \sum_{j=1}^n 2^{-(t+u+i+j)} \mu_{tuis}$$

where  $\mu_{tuis}$  is defined by

$$\mu_{tuis} = \frac{1}{pN^2} \sum_{k=1}^N \sum_{l=1}^N \sum_{r=0}^{p-1} \alpha_{r+kq-t} \alpha_{r+(k+m)q-u} \alpha_{r+lq-i} \alpha_{r+(l+m)q-j}.$$

Now since we have restricted  $q \geq n$  and  $1 \leq m \leq (p-n)/q$ , there exist  $v_1$  and  $v_2$  such that

$$\alpha_{r+kq-t} \alpha_{r+kq+km-u} = \alpha_{r+v_1}$$

$$\alpha_{r+lq-i} \alpha_{r+lq+lm-j} = \alpha_{r+v_2}.$$

For fixed values of  $t$ ,  $u$ ,  $i$ , and  $j$ , there is at most one value of  $l$  for each  $k$  such that  $v_i = v_j$ . Hence

$$\mu_{t(i)j} \leq \frac{1}{pN^2} [N(p+1) - N^2]$$

produces the result

$$E[\hat{R}^2(m)] \leq (1 - 2^{-n})^2 \left( \frac{p+1}{pN} - \frac{1}{p} \right) \quad (m \neq 0)$$

and the value of the variance of  $\hat{R}(m)$  is likewise bounded

$$\begin{aligned} \text{var}[\hat{R}(m)] &\leq (1 - 2^{-n})^2 \left( \frac{p+1}{pN} - \frac{1}{p} - \frac{1}{p^2} \right) \\ &< \frac{1}{N} \left( 1 + \frac{1}{p} \right). \end{aligned}$$

This indicates that the deviation of the sample autocorrelation function from its mean value is very small, and decreases proportionately as  $N$  increases.

**g. The distribution properties.** We have shown that  $w_k$  (and consequently,  $y_k$ ) has essentially the same mean and variance as a uniform distribution. Now consider actual distributions of  $N$  values of  $y_k$  on  $(0, 1)$ . To do this, we consider an arbitrary interval of length  $L < 1$ , and observe what percentage of the  $N$  values of  $y_k$  lie in this range.

Since we are considering binary expansions of numbers, intervals of width  $2^{-d}$  are most conveniently considered, and these will surely be sufficient to our needs. To do this, we consider the first  $d$  positions of the vectors representing  $y_k$  for  $k = 1, 2, \dots, N$ , and count the number of these having a specified pattern. This is equivalent to forming a Boolean function on the first  $d$  positions of  $y_k$ , whose value is, say  $-1$ , if  $y_k$  has this initial pattern and  $+1$  otherwise.

More specifically, let  $(e_1, e_2, \dots, e_d)$  be the initial pattern of ones and zeros we seek as a prefix to  $y_k$ . Then define the  $(\pm 1)$  Boolean function  $g(x)$  by

$$g(x) = \begin{cases} -1 & \text{if } x_1 = e_1, x_2 = e_2, \dots, x_d = e_d \\ 1 & \text{otherwise.} \end{cases}$$

The relative number of times  $\hat{T}$  that a number  $y_k$  falls in the specified interval is

$$\hat{T} = -\frac{1}{2N} \left[ N - \sum_{k=1}^N \gamma_k \right]$$

where  $\gamma_k$  has been previously defined; its value is

$$\gamma_k = G(0) + \sum_{s \neq 0} G(s) \alpha_{kq+r+v}(s)$$

by the Boolean transform. The expected value of  $\hat{T}$  is

$$T = E[\hat{T}] + \frac{1}{p} \sum_{r=0}^{p-1} \hat{T}$$

$$\begin{aligned} T &= \frac{1}{2} \left[ 1 - G(0) + \frac{1}{p} \sum_{s \neq 0} G(s) \right] \\ &= \frac{1}{2} \left[ 1 - \left( \frac{p+1}{p} \right) G(0) + \frac{1}{p} \sum_{s \neq 0} G(s) \right] \end{aligned}$$

But it is easy to see from its definition that

$$g(0) = \sum_{s \neq 0} G(s)$$

and that

$$\begin{aligned} G(0) &= 2^{-n} \sum_x g(x) = 2^{-n} (2^n - 2 \cdot 2^{n-d}) \\ &= 1 - 2^{-d+1} \end{aligned}$$

Hence, we have

$$\begin{aligned} T &= \frac{1}{2} \left[ 1 - \left( 1 + \frac{1}{p} \right) (1 - 2^{-d+1}) + \frac{g(0)}{p} \right] \\ &= \left[ \left( 1 + \frac{1}{p} \right) 2^{-d} + \frac{g(0) - 1}{2p} \right]. \end{aligned}$$

Thus, the  $y_k$  are equidistributed in the mean.

The variance about this mean can also be bounded. First, however, we compute

$$\begin{aligned} \frac{1}{p} \sum_{r=0}^{p-1} \sum_{k=1}^N \sum_{l=1}^N \gamma_k \gamma_l &= \sum_{k=1}^N \sum_{l=1}^N \sum_s \sum_u G(s) G(u) \frac{1}{p} \\ &\quad \times \sum_{r=1}^{p-1} \alpha_{r-1}^{s_1} \cdots \alpha_{r-n}^{s_n} \alpha_{r+t-1}^{u_1} \cdots \alpha_{r+t-n}^{u_n} \\ &\quad (t = q(l-k)) \end{aligned}$$

If  $s \neq 0$ , and  $u \neq 0$ , then there exist integers  $v_1$  and  $v_2$  such that

$$\begin{aligned} \alpha_{r-1}^{s_1} \cdots \alpha_{r-n}^{s_n} &= \alpha_{r+v_1} \\ \alpha_{r+t-1}^{u_1} \cdots \alpha_{r+t-n}^{u_n} &= \alpha_{r+v_2} \end{aligned}$$

and for each  $k$  there is at most one  $l$  such that  $v_l = v_k$ . Using this fact and the Schwartz inequality, we see

$$\frac{1}{p} \sum_{r=0}^{r-1} \sum_{k=1}^N \sum_{l=1}^N \gamma_k \gamma_l \leq N^2 \left\{ G^2(0) \left( 1 + \frac{1}{p} \right) - \frac{1}{p} \right\} + N \left( 1 + \frac{1}{p} \right)$$

This calculation then places a bound on the variance of  $\hat{T}$ ,

$$\begin{aligned} \text{var } \hat{T} &= \frac{1}{4} E \left\{ \frac{1}{N} \sum_{k=1}^N \gamma_k - \left( 1 + \frac{1}{p} \right) G(0) + \frac{1}{p} g(0) \right\}^2 \\ &\leq \frac{1}{4} \left\{ - [1 + G^2(0)] \frac{1}{p} \left( 1 + \frac{1}{p} \right) \right. \\ &\quad \left. + \frac{2}{p} \left( 1 + \frac{1}{p} \right) g(0) G(0) + \frac{1}{N} \left( 1 + \frac{1}{p} \right) \right\}. \end{aligned}$$

If the negative terms are omitted, the inequality is stronger,

$$\begin{aligned} \text{var } \hat{T} &< \frac{1}{4} \left( 1 + \frac{1}{p} \right) \left\{ \frac{1}{N} + \frac{2g(0)(1-2^{-d+1})}{p} \right\} \\ &< \frac{1}{4} \left( 1 + \frac{1}{p} \right) \left( \frac{1}{N} + \frac{2}{p} \right) \end{aligned}$$

and again, the deviation from expected behavior decreases as  $N$  grows larger, independently of  $e$ .

**h. Summary.** The conclusions reached by this analysis are stated in the following:

**Theorem.** If  $\{a_k\}$  is a  $(0, 1)$  binary sequence generated by an  $n$ th degree maximal length linear recursion relation modulo 2, if for  $(q, 2^n - 1) = 1$ , and  $q \geq n$ ,  $y_k = 0.a_{kq-1}a_{kq-2} \dots a_{kq-n}$  is the binary expansion of a real positive number in the interval  $(0, 1)$ , and if  $w_k$  is a real number in the interval  $(-1, +1)$  related to  $y_k$  by  $w_k = 1 - 2y_k - 2^{-n}$ , then averaged over all possible (assumed equally likely) initial values  $y_0$  (or  $w_0$ ), then

- (1) The mean value  $\mu$  of the sequence  $w_k$  is

$$\mu = 2^{-n}$$

and variance  $\sigma^2$

$$\sigma^2 = \frac{1}{3} - \left( \frac{2}{3} - 2^{-n} \right) 2^{-n}$$

- (2) The sample autocorrelation function, defined by

$$\hat{R}(m) = \frac{1}{N} \sum_{k=1}^N w_k w_{k+m}$$

has as its mean value  $R(m)$ , given by

$$R(m) = 2^{-2n} - 2^{-n}$$

for  $m = 1, 2, \dots, [(2^n - 1 - m)/q]$ . The variance of  $\hat{R}(m)$  about  $R(m)$  is bounded by

$$\text{var } \hat{R}(m) < \frac{1}{N} \left[ 1 + \frac{1}{(2^n - 1)} \right]$$

- (3) The relative number of times  $\hat{T}$  that  $y_k$  falls in the interval for which the first  $d$  positions of the binary expansion are fixed, i.e., a neighborhood of length  $2^{-d}$  in the interval  $(0, 1)$ , has mean

$$\begin{aligned} T = E[\hat{T}] &= 2^{-d} \left[ 1 + \frac{1}{(2^n - 1)} \right] \\ &+ \frac{1}{2} [g(0) - 1] \left( \frac{1}{2^n - 1} \right) \end{aligned}$$

for any number  $N$  of points  $y_k$ . The variance of  $T$  is bounded by

$$\text{var } [\hat{T}] < \frac{1}{4} \left[ 1 + \frac{1}{(2^n - 1)} \right] \left[ \frac{1}{N} + \frac{2}{(2^n - 1)} \right]$$

**i. Primitive polynomials.** In order to implement the generator, it is necessary to find a primitive polynomial  $f(x)$  over  $GF(2)$ . These have been tabulated up through degree 34 and appear in Peterson (Ref. 6); others appear in Watson (Ref. 7). The form easiest to implement is that in which the recursion relation has fewest terms. Golomb, et al. (Ref. 8) have found primitive trinomials for most degrees through degree 33. A degree 35 polynomial  $f(x) = x^{35} + x^2 + 1$  is very useful for generating numbers on an IBM 7094, whose numerical register contains 35 digits. The period  $p = 2^{35} - 1$  is relatively prime to 35, and so  $q$  may be set equal to 35. Preliminary experimental results indicate that the bounds given here are indeed valid for arbitrary sample sequences  $y_k$ . These tests and experimental properties will be discussed more fully in a subsequent article.



## 2. A New Decoding Procedure for Bose-Chaudhuri Codes, G. Solomon

**a. Introduction.** The decoding procedures presented by Peterson (Ref. 9, Sects. 9.4 and 9.5) for Bose-Chaudhuri (B-C) codes involve matrix inversion, determination of linear independence, and the finding, by trial and error, of roots of certain polynomials. While the procedure outlined is algebraic in nature, it does not take full advantage of the structure of these B-C codes nor does it use the full power of the later Mattson-Solomon polynomial representation and new treatment of the subject (Ref. 10). We present here a complete algorithm for the correct decoding of these codes using the full capabilities of the Bose-Chaudhuri codes. This algorithm is highly algebraic in nature and involves a series of computations of coefficients and substitution of these coefficients in a set of formulae. The end result is the decoded word. This decoding procedure will be used on Deep Space Network interstation teletype links for command verification.

**b.  $(2^k - 1, m)$  Bose-Chaudhuri cyclic code.** We treat for simplicity of discussion only the special case in which the recursion polynomial for the code is primitive, i.e., has one of its roots primitive of degree  $k$ . Thus, let  $A$  be a B-C code of length  $2^k - 1$  and dimension  $m$ . Then following Ref. 10, there is a polynomial  $f(x)$  dividing  $x^{2^k-1} + 1$  of degree  $m$  over the field  $F$  of two elements which serves as the recursion rule via the shift register for the code. For example,

$$f(x) = \sum_{i=0}^m u_i x^i, \quad u_i \in F,$$

can be written as the product of  $r$  irreducible polynomials over  $F$ :

$$f = \prod_{i=1}^r f_{i_1}(x),$$

where  $f_{i_1}$  denotes the irreducible polynomial over  $F$  with  $\beta^{i_1}$  as a root.  $\beta^{i_1}$  is the smallest integral power of  $\beta$ , which is a root of  $f_{i_1}$ , where  $\beta$  is a fixed primitive root of  $f(x)$ . If  $A$  is  $e$ -error-correcting, then by Ref. 10, we know that  $f(x)$  does not contain the first  $2e + 1$  negative powers of  $\beta$  as roots. But here " $e$ -error-correcting" is taken to mean that the  $e$  errors are corrected because of the B-C theory; extra errors that are correctable are not correctable by this (or Peterson's) procedure.

Corresponding to this fixed choice of  $\beta$  and a correspondence of the coordinates of the code vector with the powers of  $\beta$  for each code word  $a = (a_i)$  in  $A$ , there

is a polynomial  $g_a$  of degree at most  $2^k - 1 - (2e + 1)$  of the following form:

$$g_a(x) = \sum c_i x^i, \quad c_i \in GF(2^k)$$

$$c_{2^i} = c_i^2,$$

such that  $g_a(\beta^i) = a_i$  (Ref. 11). The coefficients of these  $g_a$  are given by the Reed inversion formula, in terms of the coordinates and the  $(2^k - 1)$  roots of unity:

$$c_l = \sum_{i=0}^{2^k-2} a_i \beta^{-il}.$$

We note that a necessary and sufficient condition for a received word to be a code word is that  $c_l = 0$  for all positions  $l$  such that  $\beta$  is not a root of the generating polynomial  $f(x)$ . Call this set of positions  $S$ . Thus, since  $c_{2^i} = c_i^2$ , there are  $r + e$  independent coefficients to compute, of which  $e$  at least must be zero. Using the formula (or an alternate method to be presented later), the  $c_i$  are computed. Using certain of the  $c$ 's, an algorithm can be devised, with the aid of the Newton formula relating power sums and symmetric functions of elements, which will yield the correct word.

**c. Structure of the  $c$ 's.** Let us examine the  $c$ 's. For clarity, we relabel  $c_{2^k-1-i} = c_{-i}$  as  $d_i$  for all those  $c$  which should be zero in the event of a correctly received word. Thus, the rule remains that  $d_i$  be zero for all  $i$ . There are at least  $e$  independent such  $d_i$  if we are to assume the B-C bound gives us minimum distance  $2e + 1$ . In particular,  $d_i, i = 1, \dots, 2e$  must be zero.

To see where we are headed, let us assume one error has been made in the  $j$ th position. Then we have,

$$d_1 = \beta^j, \quad d_3 = \beta^{3j}; \quad c'_1 = c_1 + \beta^{-j}.$$

We note  $c_1 = c'_1 + (1/d_1)$ , where  $c_1$  is the correct coefficient.

We note simply that  $d_3 = d_1^3, d_5 = d_1^5, d_l = d_1^l, l \neq 3, 5$ . If we add the reciprocal of  $d_1$  to  $c_1$  and the appropriate powers of the reciprocal to the other  $c$ , we end up with a correct set of  $c$ 's which give the correct word.

Let us now assume that  $t$  errors,  $t \leq e$ , have been made. If  $\gamma_1, \gamma_2, \dots, \gamma_t$  are the powers of  $\beta$  corresponding to the coordinate positions at which an error has

been made, then  $d_m$  is the sum of the  $m$ th powers of  $\gamma_1, \dots, \gamma_t$ :

$$d_m = \sum_{i=1}^t \gamma_i^m,$$

and a correct  $c_i$  can be obtained by just adding to the computed  $c_i'$  the  $l$ th powers of the reciprocals of the  $\gamma_i$ . Since we can compute the  $d_i$  directly and  $d_i$  are power sums of the error positions, all we need to do is relate the power sums of the reciprocals of the error positions to these and we are done.

**d. Use of symmetric functions.** We introduce the symmetric functions of the above  $\gamma_i$ . For  $t$  elements, define  $\sigma(t, i)$  to be the sum of  $t$  elements multiplied  $i$  at a time. We note that  $\sigma(t, 1)$  is just the sum of these elements one at a time and is equal to  $d_1$ ,  $\sigma(t, t)$  is the product of all of them, and  $\sigma(t, j)$  for  $j > t$  is 0. We can compute the  $\sigma$  from the  $d_i$  via the Newton formulae relating the power sums to the symmetric functions. Thus, the  $\sigma$  are given directly from the computed  $d_i$  (or at least the first  $e$  of them). We need only now note that the  $\gamma_i^{-1}, i = 1, \dots, t$ , have a set of symmetric functions easily obtainable from the ones just computed. For, if  $\gamma_1, \gamma_2, \gamma_t$  are the roots of a polynomial with coefficients given by the symmetric functions

$$\sum_{i=0}^t \sigma(t, i) x^{t-i},$$

the reciprocal polynomial with the  $\gamma_1^{-1}, \gamma_2^{-1}, \dots, \gamma_t^{-1}$  as the roots, is given by

$$\sum_i \frac{\sigma(t, i)}{\sigma(t, t)} x^i.$$

Thus, the symmetric functions  $\tau(t, i)$  for these is just  $\sigma(t, t-i)/\sigma(t, t)$ .

Now to obtain the power sums of these elements, we simply use the Newton formulae again, but this time substituting in the  $\tau$ . Once these are computed, we simply add the results obtained to the already calculated  $c$ 's, and we are done. This need only be done for the independent set of  $c$ 's.

From our representation we have ensured that the first  $2e + 1$  power sums are obtainable, and the ordinary symmetric functions can be computed from these (Ref. 9, p. 176). In the computation of the symmetric functions, the exact number  $t$  of errors made is determined and this is used for the reciprocal formulation.

In Peterson's procedures, the exact error location positions are determined from the polynomial of the symmetric functions. This can only be done by the trial and error method and involves substituting all  $2^k - 1$  values in the polynomial.

Even though the calculation given by the Reed formula seems lengthy and involves finite field operations, it is not necessary to actually compute the  $d$ 's this way, as will be shown. We need not even use the polynomials for finding the error vector.

#### *e. Mechanizations of the computations for decoding.*

Let  $A$  be a  $(2^k - 1, m)$  code of the type described above. To affect algebraic decoding, we must mechanize the field operations of  $GF(2^k)$ . Let  $f(x) = f_1(x)$  be the primitive polynomial of degree  $k$ , with  $\beta$  as root. We set up an isomorphism between binary  $k$ -tuples and  $GF(2^k)$  in the following manner. Consider the polynomial  $x + x^2 + x^4 + \dots + x^{2^{k-1}}$ . Form the vector  $B = (b_i)$ ,  $i = 0, 1, 2, \dots, 2^k - 1$ , where  $b_i$  is the value of the polynomial for  $x = \beta^i$ . To the element 1 in  $GF(2^k)$  we correspond the initial  $k$ -tuple of this vector, i.e.,  $(b_0, b_1, b_2, \dots, b_{k-1})$ . To  $\beta^i$  we assign the  $k$ -tuple  $(b_i, b_{i+1}, \dots, b_{i+k-1})$ . Addition of the  $k$ -tuples is the usual point wise addition of coordinates, and multiplication by a power  $m$  of  $\beta$  is the  $k$ -tuple obtained by a shift of  $m$  positions (to the right) along the sequence. This sets up an algebraic isomorphism between the binary  $k$ -tuples and the field  $GF(2^k)$  (Ref. 12). Any initial non-zero  $k$ -tuple would serve, but this particular choice also sets up a direct correspondence between  $GF(2^k)$  and the  $k$  information bits of a code word belonging to the maximal length shift register code generated by  $f_1$ . For example, if  $c \in GF(2^k)$ , then our representation both as a  $k$ -tuple and as the values of the polynomial

$$\sum_{i=0}^{k-1} c 2^i x^{2^i}$$

at the  $\beta^i$  for  $i = 0, 1, 2, \dots, k-1$ , are exactly the same. This will be our field  $GF(2^k)$ , and on these  $k$ -tuples, subject to these rules and field operations, we shall do our computations.

**f. Calculation of the  $d$ 's.** If  $e$  errors are made in transmission, we need compute  $d_1, d_2, \dots, d_e$ , and plug them into the Newton formulae. Since the  $d_i$  are  $k$ -tuples, we shall need  $ek$  operations on the received vector  $a = (a_i)$  and on the above defined and stored  $B$ . (In reality, only the first  $k$  bits of  $B$  need be stored and the rest generated by a shift register device.) Let us compute  $d_1$ . From the Reed formula,  $d_1 = \sum a_i \beta_i$ . By the

above representation, this corresponds to the  $k$ -tuple  $(\sum a_i b_i, \sum a_i b_{i+1}, \sum a_i b_{i+2}, \dots, \sum a_i b_{i+k-1})$ . Defining the dot product  $a \cdot B = \sum a_i b_i$ , and  $T^j B$  as the  $j$ th shift of  $B$  to the right, then  $d_i$  is given by  $(a \cdot B, a \cdot TB, a \cdot T^2 B, \dots, a \cdot T^{k-1} B)$ , a set of *anding* and summing operations. Mechanically, one may envision the received vector  $B$  as it is shifted along the shift register.

To compute  $d_m$ , we note that  $d_m = \sum a_i \beta^{im}$ , which corresponds to the  $k$ -tuple

$$(\sum a_i b_{im}, \sum a_i b_{im+1}, \sum a_i b_{im+2}, \dots, \sum a_i b_{im+k-1}).$$

Thus, if we speed up the shift register of  $B$  by a factor of  $m$ , and take the received vector  $a = (a_i)$  and *and* it with the speeded up  $B$ , we obtain the second coordinate, and so on. Thus, for  $e$  errors, we have a total  $ek$  such *anding and summing* operations.

**g. The (63, 18) Bose-Chaudhuri code.** We apply the results of this report to the (63, 18) Bose-Chaudhuri code considered in Posner-Solomon (SPS 37-26, Vol. IV), which will be used in DSN interstation telemetry. This new decoding procedure simplifies the one presented in SPS 37-26 which otherwise would have been used. The code is a 10-error-correcting code generated by the recursion polynomial

$$\begin{aligned} f(x) &= x^{18} + x^{16} + x^{11} + x^9 + x^8 + x^7 + x^5 + x^3 + x^2 + 1 \\ &= (x^6 + x + 1)(x^6 + x^5 + x^2 + x + 1) \\ &\quad \times (x^3 + 1)(x^3 + x^2 + 1). \end{aligned}$$

For our decoding computations, choose  $\beta$  a root of  $x^6 + x + 1$ . Then to the code word  $a$  is associated the polynomial

$$\begin{aligned} c_0 + c_1 x + c_2 x^2 + \dots + c_{32} x^{32} + c_3 x^5 + c_5 x^{10} + \dots \\ + c_5 x^{34} + c_9 x^9 + c_9 x^{18} + c_9 x^{36} + c_{21} x^{21} + c_{21} x^{42}. \end{aligned}$$

The Reed formula gives the values of  $c$ 's and  $d$ 's for each code word  $a = (a_i)$ , where the computations are performed in  $GF(2^6)$ . For a code word to be received correctly, the  $d$ 's must all be 0 ( $d_1 = d_3 = d_5 = d_{17} = d_{19} = 0$ ):

$$c_j = \sum_{i=0}^{62} a_i \beta^{-ij}, \quad d_j = \sum_{i=0}^{62} a_i \beta^j.$$

In order to perform the computations, we set up the correspondence of Part *e* between the 6-tuples of 0's and 1's and  $GF(2^6)$  in the following fashion: To 1  $\in GF(2^6)$

is associated the 6-tuple (0, 0, 0, 0, 0, 1) corresponding to the values of the polynomial

$$x + x^2 + x^4 + x^8 + x^{16} + x^{32}$$

at the values

$$1, \beta, \beta^2, \dots, \beta^5.$$

To  $\beta$  is associated the 6-tuple (0, 0, 0, 0, 1, 0), which is simply a shift along the code word length 63, generated by the recursion polynomial  $x^6 + x + 1$ , using the 6-tuple associated to 1  $\in GF(2^6)$  as the first 6 bits.

As in Part *e*, multiplication by the  $j$ th power of  $\beta$  is simply the 6-tuple obtained by shifting  $j$  positions along this code word from the 6-tuple to be multiplied. Thus,  $d_i$  say is given by the 6-tuple  $(a \cdot B, a \cdot TB, a \cdot T^2 B, \dots, a \cdot T^5 B)$ .

The first coordinate of  $d_3$  is obtained by starting with  $B$  and shifting 3 times as fast as usual and dotting with the received vector  $a$ . The second coordinate of  $d_3$  is obtained by shifting  $B$  one position and again dotting this speeded-up vector with  $a$ , and so on. In general, to obtain the first coordinate of  $d_e$ , we speed up the vector  $B$  a factor of  $e$  times and dot it with  $a$ . For the second coordinate of  $d_e$ ,  $B$  is shifted one position and then speeded up  $e$  times and dotted with  $a$ , and so on.

**h. Correction of  $t$  errors in the (63, 18) code.** Let us say that  $t$  errors have been made in the received vector  $a$ . Then, using the computed values for  $d_1, d_2, \dots, d_t$  and plugging them into the Newton formulae, the symmetric functions

$$\sigma_1, \sigma_2, \dots, \sigma_t$$

are obtained. Using the results of Part *d*, the symmetric functions  $\tau$  of the reciprocal roots are obtained. Plugging these values back into the Newton formulae, the power sums of the reciprocals of the error positions are obtained. In particular, the first, fifth, ninth, and twenty-first are obtained. All the others are computable and known from the  $c$ 's. For example, if  $\beta_i = 1, 2, \dots, t$  are the error positions, and if

$$T_j = \sum_i \beta_i^j,$$

then

$$T_3 = d_{15}^4, \quad T_7 = d_7^8, \quad T_{11} = d_{13}^4,$$

$$T_{13} = d_{11}^{16}, \quad T_{15} = d_3^{16}, \quad T_{17} = T_5^{16}, \quad T_{19} = d_{11}^4,$$

and the  $T_i$  are determined as above for  $i = 1, 5, 9, 21$ .

To the vector  $a$  is added the vector  $((h_a(\beta^i)))$ , where

$$h_a(t) = x + T_1x + T_1^2x^2 + T_1^{32}x^{32} + T_5x^5 + T_5^2x^{10} + \dots \\ + T_5^{32}x^{34} + T_3x^9 + T_3^2x^{18} + T_3^4x^{36} + T_{21}x^{21}T_{21}^2x^{42}.$$

Since we are only interested in the information bits we need only compute the first 18 coordinates of this vector and add these to the first 18 received coordinates. The result is the correct 18 information bits.

*i. Example for triple errors,  $t = 3$ .* Let us suppose three errors have been made on unknown positions  $\beta_1, \beta_2, \beta_3$ . Then we have

$$d_1 = \beta_1 + \beta_2 + \beta_3,$$

$$d_3 = \beta_1^3 + \beta_2^3 + \beta_3^3,$$

.

.

.

$$d_{10} = \beta_1^{10} + \beta_2^{10} + \beta_3^{10}.$$

Using Newton's formulae, and defining  $\sigma_1 = \beta_1 + \beta_2 + \beta_3$ ,  $\sigma_2 = \beta_1\beta_2 + \beta_1\beta_3 + \beta_2\beta_3$ ,  $\sigma_3 = \beta_1\beta_2\beta_3$ , we obtain

$$\sigma_1 = d_1;$$

$$\sigma_2 = \frac{d_1^2d_3 + d_5}{d_1^3 + d_3};$$

$$\sigma_3 = \frac{d_1d_5 + d_3^2 + d_1^3d_3 + d_1^6}{d_1^4 + d_3};$$

also,

$$\sigma_4 = \sigma_5 = \sigma_6 = \dots = 0.$$

Defining now

$$\tau_1 = \beta_1^{-1} + \beta_2^{-1} + \beta_3^{-1},$$

$$\tau_2 = (\beta_1\beta_2)^{-1} + (\beta_1\beta_3)^{-1} + (\beta_2\beta_3)^{-1},$$

$$\tau_3 = (\beta_1\beta_2\beta_3)^{-1},$$

we have, very easily,

$$\tau_1 = \frac{\sigma_2}{\sigma_3}, \quad \tau_2 = \frac{\sigma_1}{\sigma_3}, \quad \tau_3 = \sigma_3^{-1}.$$

Letting  $T_i = \beta_1^{-i} + \beta_2^{-i} + \beta_3^{-i}$ ; we have, again using the Newton formulae,

$$\tau_4 = \tau_5 = \tau_l = 0, \quad l > 5.$$

$$T_1 = \tau_1, \quad T_1^2 = T_2;$$

$$T_3 = T_2\tau_1 + T_1\tau_2 + \tau_3 = 0;$$

$$T_4 = T_2^2$$

$$T_5 + T_4\tau_1 + T_4\tau_3 + T_1\tau_4 + \tau_5 = 0;$$

$$T_7 + T_6\tau_1 + T_6\tau_2 + T_4\tau_3 + T_3\tau_4 + T_5\tau_5 + T_6\tau_6 + \tau_7 = 0.$$

We compute  $T_1, T_5, T_9, T_{21}$  very simply since everything else is known:

$$T_1 = \tau_1 = \frac{\sigma_2}{\sigma_3}; \quad T_3 = d_{15}^4;$$

$$T_5 = T_1^5 + T_3\tau_2 + \tau_3T_1^2;$$

$$T_9 = \frac{\sigma_2^5}{\sigma_3^5} + \frac{\sigma_1}{\sigma_3}T_3 + \frac{1}{\sigma_3} \frac{\sigma_2^2}{\sigma_3^2};$$

$$T_{21} = T_{20}\tau_1 + T_{19}\tau_2 + T_{18}\tau_3.$$

The  $h_a(\beta^i)$  are then computed and added to the received word to give the correct word.

### 3. Coding for the Binary Asymmetric Channel as Applied to Ranger Block III Command Detector,

E. C. Posner

*a. Summary.* The *Ranger* Command Detector studied in an article on extreme value theory (Sect. XIX-C-1 of this report) has in its present design the property that a 1 can be mistaken for a 0, but not a 0 for a 1. The channel that results is called the binary asymmetric channel. This short note introduces the mathematical study of such channels, and shows how a simple code can reduce the error probability in the *Ranger* Command Detector from high to low values, and with a minimum of equipment.

*b. Preliminaries.* In the theory of error-correcting codes, the binary symmetric channel has usually been considered. In this channel, there are two symbols, "0" and "1," and an error probability  $p$ . This error probability is

the probability that a transmitted 0 is received as a 1, and also the probability that a 1 is received as a 0. Thus, the name "symmetric."

A (systematic)  $(n, k)$  code with  $k$  information bits and  $n - k$  check bits is often used on the binary symmetric channel. These codes consist of words of length  $n$ . The first  $k$  are the  $k$  information bits to be sent, and the last  $n - k$  are the so-called check bits. These check symbols are Boolean functions of the  $k$  information bits.

It is well known that if one is using an  $(n, k)$  code on the binary symmetric channel, and one wishes to correct all single errors in transmission, then the code must have minimum distance at least 3. That is, every two code words in the code must differ in at least three positions. It will now be shown that minimum distance 2 often suffices in the binary asymmetric channel.

For example, consider the  $(2, 1)$  code consisting of the two code words 00 and 11. That is, 0 is coded as 00 and 1 as 11—the information symbol is merely repeated twice. This code does not correct all single errors in the binary symmetric channel. For, if one receives the codeword 10, one cannot tell whether the word 00 was sent, with an error in the first position, or whether 11 was sent, with an error in the second position.

However, the same  $(2, 1)$  code does in fact correct all single errors in the binary asymmetric channel, even though the minimum distance of the code is 2 instead of 3. To see this, note that the only possible error changes a 1 into a 0, not a 0 into a 1. Thus, if 00 is sent, 00 is received with certainty. If on the other hand 11 is sent, then 11 is received if no errors occur in transmission. But if a single error is made, the received word is 01 or 10, depending upon whether the single error occurs in the first or second position.

Thus, the  $(2, 1)$  code on the binary asymmetric channel is decoded as follows: if 00 is received, then the information bit is decoded as 0. On the other hand, if the received word contains a 1, the information bit is decoded as 1. Now let us see how the use of the  $(2, 1)$  code on the binary asymmetric channel reduces error probability.

**c. Error probabilities.** We shall examine the improvement obtainable with the  $(2, 1)$  code. Let  $r$  be the probability that 1 changes to 0. Then an information bit is decoded incorrectly if and only if the bit is a 1 and a double error occurs in transmission. It can be assumed that 0 and 1 are just as likely for the information bits, so

that the probability of a 1 information bit is  $\frac{1}{2}$ . And the probability of a double error is  $r^2$ , when the 11 word is sent. Thus, the output information bit error probability is  $r^2/2$ .

By the same reasoning, the uncoded bit error probability is  $r/2 = p$  say. So an input bit error probability of  $p$  yields after coding an output bit error probability of  $2(r/2)^2 = 2p^2$ . The price paid is that the information rate is cut in half.

An improvement is effected if  $p < \frac{1}{2}$ , since then  $2p^2 < p$ . (But note that if  $p > \frac{1}{2}$ , the usual option of interchanging 0 and 1 to get  $p < \frac{1}{2}$  does not work in the binary asymmetric channel. The case  $p > \frac{1}{2}$  would have to be considered separately.)

Of course, one is usually interested in  $r$  much less than 1. For example, in the *Ranger* Command Detector, the specifications call for a bit error probability of less than  $10^{-5}$ . If the  $(2, 1)$  code is used, this specification demands that  $2p^2 < 10^{-5}$ , or,  $p < 2.236 \times 10^{-3}$ . Then the probability that a 1 changes to a 0, that is,  $r$ , which is  $2p$ , must be less than  $4.47 \times 10^{-3}$ .

Now  $4.47 \times 10^{-3}$  is considered a rather high error probability. So this value could be easy to obtain. But to obtain  $r < 2 \times 10^{-5}$ , which is the value for  $r$  necessary if the specifications are to be met with no coding, can be much more difficult.

That is, by using the  $(2, 1)$  code, a "bad" detector can be made into a "good" one with a loss of only 50% in rate. For a command system in which rate is not as vital as low error probability, one would choose the coded system with lower rate. That is, in the present form of command detector in which the transition  $0 \rightarrow 1$  can be assumed not to occur, the  $(2, 1)$  code can be used to obtain detectors that meet the error probability specifications.

**d. Implementation of  $(2, 1)$  code.** In applications of error-correcting codes to the binary symmetric channel, one big problem present is that to obtain a dramatic decrease in output bit error probability, the decoding process quickly becomes complicated, and requires a big decoding machine. Such big machines are not necessary for the  $(2, 1)$  code used on the binary asymmetric channel.

For, one takes advantage of the fact that the transition  $0 \rightarrow 1$  does not occur. Then one decodes the information bit as a 1 if and only if one of the two symbols in the

received word is a 1. Otherwise, the information bit is called a 0.

The decoder for the (2,1) code, then, consists of a buffer of length two (Fig. 1). A *nor* gate with two inputs is connected to the two flip-flops of the buffer. Every other symbol period, the output of the gate is read. If the buffer contained two 0's, the output of the gate is a 0; if the buffer contained one or two 1's, the output of the gate is a 1. These indeed are the outputs required by the decoding scheme.

It will be seen that the decoder, as shown in Fig. 1, requires a minimum of extra logic over the method using no coding. In fact, the only extra logic required is two flip-flops for the buffer, one *nor* gate, and one flip-flop to read the output of the *nor* gate every other symbol time as required by the system clock. The (2,1) code is therefore entirely feasible and can be used in the *Ranger* Command Detector in its present form. The same code can be used for error-detection allowing repeat-requests.

*e. Areas of research.* The binary asymmetric channel offers many interesting areas of coding research:

- (1) Relate combinatorial properties of a given binary code to its binary asymmetric error-correction and -detection ability.

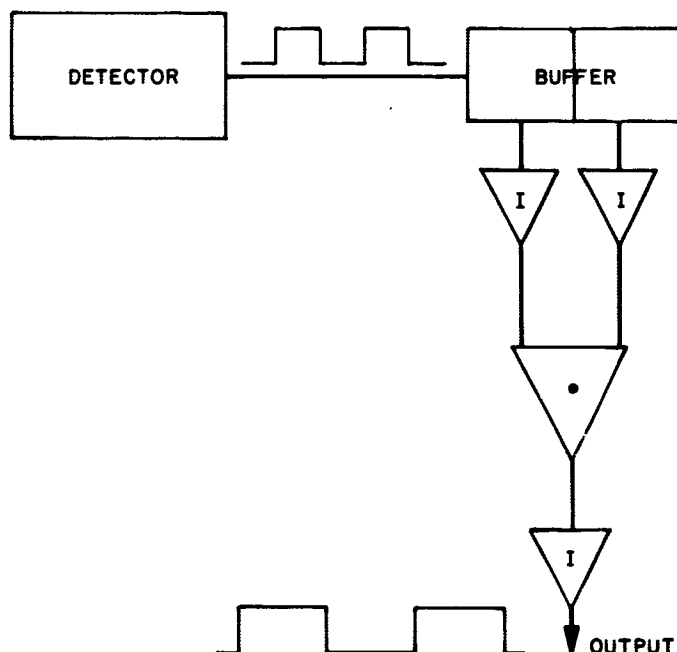


Fig. 1. Implementing the (2,1) code

- (2) Define "close-packed" for the binary asymmetric channel, and study the combinatorial structures that this definition implies.
- (3) Find codes and decoding procedures that take advantage of the peculiar properties of the binary asymmetric channel.
- (4) Consider the use of linear binary cyclic codes on the binary asymmetric channel. Find the improvement in decoder simplicity attainable with the binary asymmetric channel.
- (5) Estimate the way error probability drops with code length.

It is conjectured that the binary asymmetric channel yields huge decreases in decoder complexity.

## B. Detection and Filter Theory

W. B. Kendall, W. C. Lindsey, J. J. Stiffler,  
and R. C. Titsworth

### 1. A Comparison of Some Exact Results for the Biased Square-Law Sequential Detector to the Corresponding Approximate Classical Results, W. B. Kendall

*a. Introduction.* Sequential detectors are devices which detect the presence (or absence) of a signal in a noise background by sequentially observing an input waveform. At each instant, the sequential detector's output consists of one of three possible outputs (or decisions):

- (1) No signal is present (dismissal).
- (2) Signal is present (alarm).
- (3) Continue making observations.

The biased square-law sequential detector, which we will consider here, uses samples  $y_i$  of the envelope of the received waveform to calculate the sums

$$x_i = \sum_{j=1}^i [ky_j^2 - b]$$

where the bias  $b$  and the multiplier  $k$  are arbitrary non-negative constants which can be adjusted to suit the application. The sums  $x_i$ ,  $i = 1, 2, \dots$  are sequentially

compared to two thresholds until an  $x_i$  is found which is less than the lower threshold  $B$  or greater than the upper threshold  $A$ . Then, if  $x_i \leq B$  it is decided that no signal is present (dismissal), and if  $A \leq x_i$  it is decided that signal is present (alarm). For the  $y_i$  having the Rayleigh probability distribution  $f(y) = y \exp -y^2/2$ , i.e., for the received waveform consisting of Gaussian noise alone, the author (Ref. 13) has obtained exact expressions for the probability of (false) alarm and for the average number of observations required to end this test.

It is well known (Ref. 14) that the (optimum) sequential *probability-ratio* test for the incoherent detection of a sine wave in Gaussian noise requires that the sums

$$x_i = \sum_{j=1}^i [-\frac{1}{2}(a^2) + \ln I_0(ay_j)]$$

be compared to two thresholds  $A$  and  $B$  where  $B < A$ . Here  $a^2/2$  is the signal-to-noise power ratio when the signal is present, i.e., it is the "design" signal-to-noise ratio. When  $a$  is sufficiently small it is usually satisfactory to approximate  $\ln I_0(ay)$  by a constant plus the first term of its power series.

$$\ln I_0(ay) \approx \frac{1}{4}(a^2 y^2) - \beta.$$

The constant  $\beta \geq 0$  is usually taken to be  $a^4/8$ , which is the expected value of the next term in the power series when no signal is present (Ref. 14). Use of this approximation yields

$$x_i \approx \sum_{j=1}^i [(a^2/4)y_j^2 - (\beta + a^2/2)]$$

which suggests that for small  $a$ , and with  $k$  and  $b$  given by

$$k = a^2/4$$

and

$$b = \beta + a^2/2$$

the author's exact results should approach the classical results for sequential probability ratio tests. In the following parts of this article, we will show that this is indeed the case.

**b. Probability of false alarm.** In Ref. 13 it is shown that, for the biased square-law detector, the probability of

(false) alarm when no signal is present, which we shall denote by  $P_2(0)$ , is given by

$$P_2(0) = \frac{e^{-aDb} G[-DB; Db]}{e^{aD(A-B+b)} G[D(A-B+b); Db]} \quad B < 0 < A$$

where  $A$  and  $B$  are the thresholds,  $b$  is the bias, and

$$k = a^2/4$$

$$D = (2/a^2) \exp \{-2b/a^2\}$$

$$= 2/(a^2 \alpha)$$

$$\alpha = \exp \{2b/a^2\}$$

and

$$G(x; c) = 1 + \sum_{j=1}^{\infty} \frac{(jc - x)^j}{j!} \quad c \leq nc \leq x \leq (n+1)c$$

It is also shown in Ref. 13 that for  $x/c \geq 2$ , the function  $e^{ax} G(x; c)$  is given to three or more decimal places by

$$e^{ax} G(x; c) \approx \left( \frac{1}{1 - \gamma c} \right) e^{(\alpha - \gamma)x} - \left( \frac{1}{\alpha c - 1} \right) \quad \begin{cases} c < 1/e \\ \gamma < c < \alpha \end{cases}$$

where  $\alpha$  and  $\gamma$  are the two real roots of  $s = e^{sc}$ . (Thus,  $\ln \alpha = \alpha c$  and  $\ln \gamma = \gamma c$ .) This approximate expression for  $e^{ax} G(x; c)$  can be used in the evaluation of  $P_2(0)$  as long as  $-B/b \geq 2$  and the inequality

$$Db = (2b/a^2) \exp \{-2b/a^2\} < 1/e$$

is satisfied. This inequality is satisfied as long as  $2b/a^2 \neq 1$ , or, for  $b = \beta + a^2/2$ , as long as  $\beta \neq 0$ . Use of this approximation yields

$$P_2(0) \approx \frac{1 - \left[ \frac{1 - \ln \gamma}{\ln \alpha - 1} \right] e^{[(\ln \alpha - \ln \gamma)/b]B}}{e^{[(\ln \alpha - \ln \gamma)/b](A+b)} - e^{[(\ln \alpha - \ln \gamma)/b]B}} \quad \begin{cases} \beta \neq 0 \\ -\beta \geq a^2 + 2\beta \end{cases}$$

The two factors in square brackets in this equation can be written as

$$\left[ \frac{1 - \ln \gamma}{\ln \alpha - 1} \right] = 1 - \frac{a^2 \epsilon}{2\beta}$$

and

$$[(\ln \alpha - \ln \gamma)/b] = \frac{(2/a^2) [(4\beta/a^2) - \epsilon]}{1 + 2\beta/a^2}$$

where  $\epsilon$  is defined by

$$\begin{aligned}\epsilon &\triangleq \ln \gamma - 1 + 2\beta/a^2 \\ &< (2\beta/a^2) - [1 - e^{Db}]^{1/2} \\ &= (2\beta/a^2) - [1 - (1 + 2\beta/a^2) e^{-2\beta/a^2}]^{1/2} \\ &< (2\beta/a^2)^2/6.\end{aligned}$$

Thus, we have  $\epsilon \ll 2\beta/a^2$  when  $2\beta/a^2 \ll 1$ , so in this case  $P_2(0)$  can be further approximated by

$$P_2(0) \approx \frac{1 - e^{(2\beta/a^2)B}}{e^{(2\beta/a^2)(A+B+a^2/2)} - e^{(2\beta/a^2)B}}$$

$$\begin{cases} \beta \neq 0 \\ -B \geq a^2 + 2\beta \\ 2\beta/a^2 \ll 1 \end{cases}$$

It has been pointed out by Bussgang and Middleton (Ref. 14) and by Bussgang and Mudgett (Ref. 15) that the best biased square-law approximation to the optimum incoherent detector for small design signal-to-noise ratio  $a^2/2$  is obtained by letting  $\beta = a^4/8$ . For this case we get

$$P_2(0) \approx \frac{1 - e^B}{e^A - e^B} \quad \begin{cases} \beta = a^4/8 \\ -B \geq a^2 \\ a^2/2 \ll 1 \end{cases}$$

which agrees with Wald's (Ref. 16) classical result for sequential probability ratio tests. In Fig. 2 this small-signal approximation is shown along with the exact results for design signal-to-noise ratios of 5, 1, and 0.01.

**c. Average test duration.** For the biased square-law sequential detector in the absence of signal, the average test duration  $M_1(0)$  has been shown by the author (Ref. 13) to be given exactly by

$$M_1(0) = e^{-\alpha DB} H[-DB; Db] + P_2(0) \{1 - e^{\alpha D(A-B+b)} H[D(A-B+b); Db]\}$$

where the function  $H(x; c)$  is given by

$$H(x; c) = (n+1)e^{-\alpha x} - \sum_{j=1}^n \sum_{i=0}^{j-1} \frac{(jc-x)^i}{i! \alpha^{j-1}} \quad c \leq nc \leq x \leq (n+1)c$$

and the other notation is the same as that used above. Also given in Ref. 13 is an approximate expression for

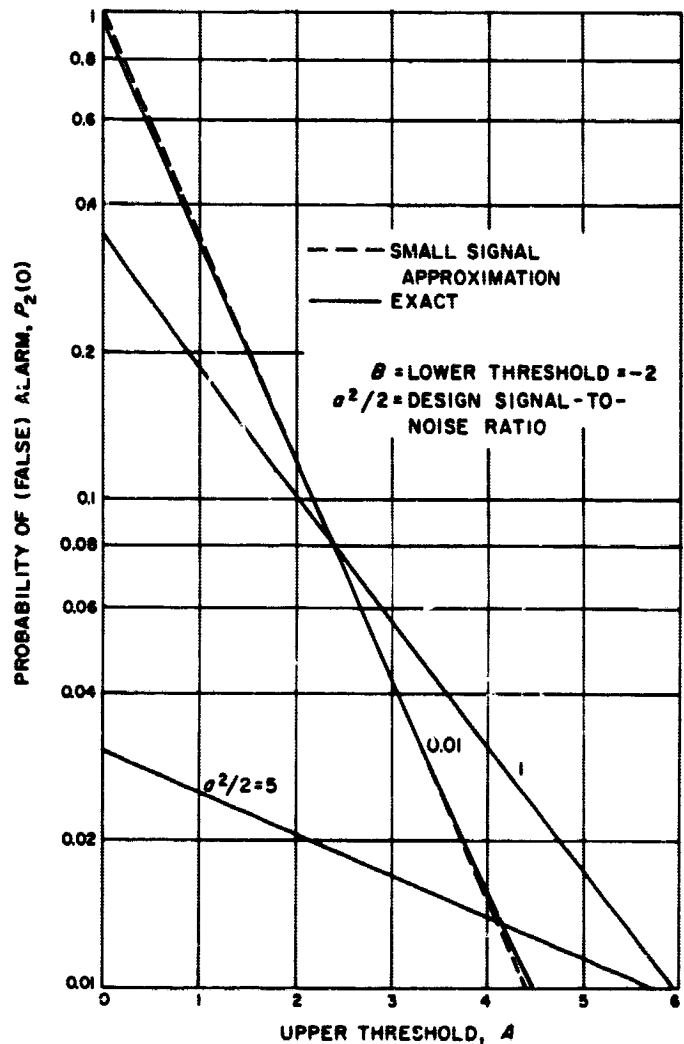


Fig. 2. Probability of alarm for the biased square-law sequential detector when the signal is not present (exact and small signal approximation)

$e^{\alpha x} H(x; c)$  which has the same accuracy and validity as that given above for  $G(x; c)$ . It is

$$\begin{aligned}e^{\alpha x} H(x; c) &\approx -\left(\frac{1}{1-\gamma c}\right)\left(\frac{\gamma}{\alpha-\gamma}\right)e^{(\alpha-\gamma)x} \\ &+ \left(\frac{1}{\alpha c-1}\right)\left[\alpha x + \frac{(\alpha c)^2}{2(\alpha c-1)} - 1\right] \\ &\quad \begin{cases} c < 1/e \\ \gamma < e < \alpha \end{cases}\end{aligned}$$

Use of this approximation yields

$$M_1(0) \approx [-B - P_2(0)(A-B) + E]/\beta$$

$$\begin{cases} \beta \neq 0 \\ -B \geq a^2 + 2\beta \end{cases}$$



where

$$E = \frac{a^4}{8\beta} \left\{ 1 + \left( \frac{2\beta}{a^2} \right)^2 - P_2(0) \left( 1 + \frac{2\beta}{a^2} \right)^2 + \frac{1 - \frac{2\beta}{a^2} + \epsilon}{1 - \frac{3a^2\epsilon}{4\beta} + \frac{a^4\epsilon^2}{8\beta}} \right. \\ \left. \times [P_2(0) e^{[(\ln a - \ln \gamma)/b](A-B+b)} - e^{-[(\ln a - \ln \gamma)/b]B}] \right\}$$

and where the above definitions and relations for  $\epsilon$  have been used. The term  $E$  is due to the "excess over the bounds," i.e., it is due to the fact that, at the end of the test, we have either  $x_i \leq B$  or  $A \leq x_i$ , rather than simply  $x_i = B$  or  $x_i = A$ . However, if the probability is high that  $x_i$  changes only a very small amount at each observation, then this excess is negligible, and we have simply

$$M_1(0) \approx \frac{-B - P_2(0)(A - B)}{\beta} \quad \begin{cases} \beta \neq 0 \\ -B \geq a^2 + 2\beta \\ E \ll -B \end{cases}$$

The condition for this last approximation to be applicable is that both the mean and the standard deviation of the random variables

$$x_i - x_{i-1} = (a^2/4)y_i^2 - b$$

are negligible compared to the values of the thresholds  $A$  and  $B$ . These quantities are given by

$$\langle x_i - x_{i-1} \rangle = -\beta$$

and

$$\langle (x_i - x_{i-1} + \beta)^2 \rangle^{1/2} = a^2/2$$

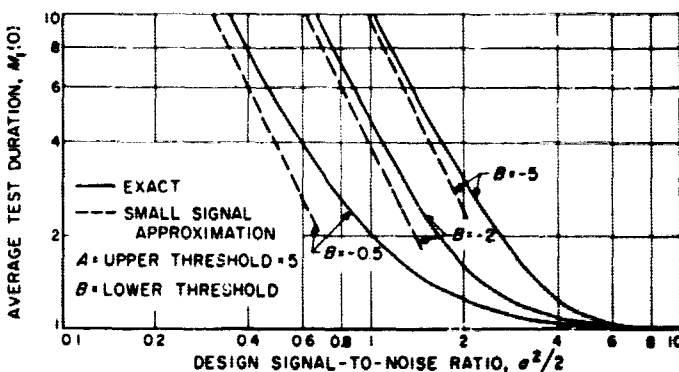


Fig. 3. Average test duration for the biased square-law sequential detector when the signal is not present (exact and small signal approximation)

where the pointed brackets denote expected values. Careful examination of the above expressions for  $E$  and  $P_2(0)$  shows that indeed  $E \rightarrow 0$  when both  $\beta$  and  $a^2$  individually become very small, yielding the classical result (Ref. 16)

$$M_1(0) \approx \frac{[P_2(0)]A + [1 - P_2(0)]B}{\langle x_i - x_{i-1} \rangle} \quad \begin{cases} 0 \neq \beta \ll 1 \\ a^2 \ll 1 \end{cases}$$

This small signal approximation is compared to the exact expression for  $M_1(0)$  in Fig. 3.

## 2. Frequency Demodulation, W. C. Lindsey

a. Introduction. In Ref. 17 we considered the following problem: A unit-variance random time-series  $m(t)$  was used to frequency modulate a transmitter, resulting in the emitted waveform

$$\zeta(t) = (2P)^{1/2} \sin[\omega t + \theta_1(t)] \quad (1)$$

where  $m(t)$  possessed the spectral density

$$S_m(s) = \frac{2a}{a^2 - s^2} \quad (2)$$

and

$$\theta_1(t) = K_f \int^t m(\tau) d\tau.$$

The waveform  $\zeta(t)$  was assumed to have been corrupted in passing through the channel by additive, white Gaussian noise with a single-sided spectral density of  $N_0$  w/cps. Demodulation of the observed data  $\zeta(t)$  at the receiver was accomplished by means of a phase-locked loop (PLL) followed by an appropriate output filter. The loop filter  $F(s)$  of the linearized PLL model was chosen such that the mean-squared phase error  $\phi$  was a minimum, while the output filter  $F_o(s)$  was selected such that the mean-squared frequency error was a minimum. In particular, it was found that the pole-zero plot of the optimum filters depended upon the signal power  $P$ , the modulation index  $m_f = K_f/a$ , and the noise spectral density  $N_0$ . Since these filters are, in general, difficult to implement, two alternate approaches for accomplishing the demodulation process were considered:

- (1) The PLL was preceded by a bandpass limiter. The pole-zero configuration of the loop and output filters was "fixed" such that the system would operate optimally under the worst channel conditions expected.

- (2) The same "fixed" PLL was preceded by an automatic gain control (AGC) amplifier. The performance characteristics were derived and graphically compared with those of the optimum demodulator.

Here we consider the same problem; however, the demodulator model will be modified so as to take into consideration the nonlinear element existing in the forward path of the "exact" PLL model. The sinusoidal nonlinearity will be taken into consideration by means of Booton's (Ref. 18) so-called *describing function* method. We shall compute the performance characteristics for all three demodulators and compare these results with those obtained in Ref. 17 for the linear model. Finally, we present the performance characteristics of a single-sideband or double-sideband amplitude-modulated system which operates in the presence of the same channel disturbance; i.e., demodulation is accomplished at the receiver by coherent-frequency translation of the observed data, and the result smoothed with a realizable Wiener filter.

**b. Quasi-linear receiver model.** The quasi-linear model, introduced by Develet (Ref. 19), is based on Booton's (Ref. 18) idea that any nonlinear device subjected to a Gaussian process may be replaced by an equivalent gain  $K$ . The concept is simple and may be briefly described as follows. Denote the input and response of an amplitude-sensitive element by  $x_i$  and  $x_r$ , respectively. Let the output  $x_r$  be written in the form

$$x_r = f(x_i)$$

where  $f$  is the *describing function* of the element. Then for a particular input process, we select that value of  $K$  for which the mean-squared error

$$\overline{(x_r - Kx_i)^2} = \overline{[f(x_i) - Kx_i]^2} \quad (3)$$

is a minimum. (The bar signifies an average taken in the statistical sense.) For the problem at hand,  $f = \sin[\ ]$ . Assuming the input  $\phi(t)$  to the nonlinear element is a zero-mean Gaussian process, Develet (Ref. 19) shows that

$$K_0 = \exp[-(\sigma_\phi^2/2)]$$

where  $\sigma_\phi^2$  is the mean-squared value of the phase error  $\phi(t)$ .

In passing, two comments are important. In order to use this gain, one further assumption is necessary, i.e., the process  $\phi(t)$  must be first-order stationary. Otherwise the gain  $K_0$ , hence the over-all receiver, of the element

is time-varying. Second is the fact that the nonlinear element is a part of the feedback system and, even though the input to the PLL may be Gaussian, the operation of a nonlinear element on a Gaussian process is, in general, non-Gaussian. These remarks should be kept in mind when employing the quasi-linear model; however, if we are going to take into consideration in the design the nonlinearity at all, this approximation seems to be as simple and mathematically tractable as any. As a matter of fact Viterbi (Ref. 20) shows that, for the first order loop and no modulation present, Develet's approximation is quite good so long as  $\sigma_\phi^2 < 1/2$ .

The equivalent quasi-linear model is depicted in Fig. 4. For reasons which will become obvious later, we define the following transfer functions, i.e.,

$$H(s) = \frac{y(s)}{\theta_1(s)}; \quad H_0(s) = \frac{K_r F(s)}{s\theta_1(s)} = \frac{\theta_2(s)}{\theta_1(s)} \quad (5)$$

Referring to Fig. 4, it is obvious that

$$F_0(s) = \frac{K_r H(s)}{s H_0(s)} \quad (6)$$

Further, the closed-loop transfer function of the quasi-linear PLL may be written by inspection from Fig. 4, namely,

$$H_0(s) = \frac{K_0 K_r (P)^N F(s)}{[s + K_0 K_r (P)^N F(s)]} \quad (7)$$

where

$$K_0 = \exp(-\sigma_\phi^2/2)$$

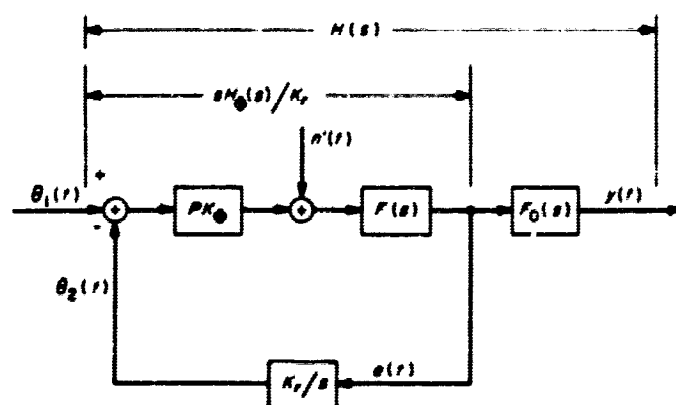


Fig. 4. Quasi-linear phase-lock receiver

and the loop filter  $F(s)$ , obtained from Eq. (7), is

$$F(s) = \frac{[K_r K_\phi (P)^{-1/2}]^{-1} s H_\phi(s)}{1 - H_\phi(s)} \quad (8)$$

From Eq. (5) we see that the output  $\theta_2(t)$  of the PLL is proportional to the amplitude of the input  $\theta_1(t)$ . Thus, the closed-loop response of the PLL is a function of both the amplitude of the transmitted signal and the phase error [Eqs. (7) and (8)].

Before leaving this discussion pertaining to the quasi-linear model we point out that the noise component  $n'(t)$  of Fig. 4 may be referred to the input and designated as  $n'_1(t)$ , a stationary, white Gaussian noise process with a single-sided spectral density

$$N'_0 = \frac{N_0}{P} \exp(\sigma_\phi^2) = \frac{N_0 K_\phi^{-2}}{P} \quad (9)$$

Thus, the noise which the equivalent receiver sees is now a function of the mean-squared phase error and the transmitter signal power.

**c. The optimum filters.** It is clear from Fig. 4 that there are two types of errors which are made at the receiver. One is the phase error  $\sigma_\phi^2$ , due to the receiver VCO's inability to follow exactly the incoming phase  $\theta_1(t)$ . To ensure proper operation of the PLL, i.e., make the loop operate as linearly as possible, it is necessary to minimize  $\sigma_\phi^2$  by selecting the loop filter  $F(s)$  properly. The other error is the frequency error  $\sigma_f^2$ , i.e., the total modulation error. For the receiver to reproduce the modulation waveform  $m(t)$  as faithfully as possible, and for the PLL to follow the observed data as closely as possible, it is necessary that we simultaneously minimize  $\sigma_f^2$  and  $\sigma_\phi^2$ .

It may seem, at first, that it is impossible to minimize the phase and frequency errors simultaneously; however, since the system is linear, the filter  $H_\phi(s)$  [hence the loop filter  $F(s)$ ] may be determined so as to minimize  $\sigma_\phi^2$ , and the over-all receiver filter  $H(s)$  may be adjusted to minimize  $\sigma_f^2$ . Having determined that loop filter  $F(s)$  which minimizes  $\sigma_\phi^2$  when  $H(s)$  is fixed, the ability to carry out the optimization procedure in this manner is a result of the quasi-linear receiver model.

The optimum filters derived in Ref. 17 are applicable here in that one needs only to replace  $N_0$  (Ref. 19) appear-

ing in the individual filter functions by  $N_0 \exp(\sigma_\phi^2)$ , i.e., (Fig. 4)

$$H(s) = \frac{a^2 (\gamma - 1)^2 s}{2K_t (s^2 + a\gamma s + a^2 \delta)} \quad (10)$$

where

$$\begin{aligned} \gamma &= (1 + 2\delta)^{1/2}; \quad \delta = m_f K_\phi (R)^{1/2} \\ m_f &= K_t/a; \quad R = P/aN_0 \end{aligned}$$

The individual loop and output filters become

$$\begin{aligned} F(s) &= [(P)^{1/2} K_\phi K_r]^{-1} \left[ \frac{a(\gamma - 1)s + a(\gamma + 1)}{s + a} \right] \\ F_\phi(s) &= \frac{a(\gamma - 1)^2 K_r}{2K_t [(\gamma - 1)s + a\delta]} \end{aligned} \quad (11)$$

**d. The output signal-to-noise and receiver threshold characteristic when the loop is preceded by an AGC amplifier.** The output "signal-to-noise" ratio  $\rho_0$ , i.e., the reciprocal of the mean-squared frequency error, is composed of two parts: the signal distortion  $S_d$  and the error due to the additive noise. Using the results given in Ref. 17,  $S_d$  is

$$S_d = \frac{4\delta\beta(\delta + \alpha) + 4\delta\beta(\alpha^2 - K) + K^2(\alpha + 1)}{4\delta\beta[(\alpha + 1)(\delta + \alpha) - \delta\beta]} \quad (12)$$

where

$$\begin{aligned} \alpha &= 1 + \beta(\gamma - 1) \\ K &= \beta(\gamma^2 - 1) - (\gamma - 1)^2 \\ \beta &= \exp\left(\frac{\sigma_\phi^2 - \sigma_1^2}{2}\right) \end{aligned}$$

and  $\sigma_1$  is the total phase error resulting in the fixed loop;  $\sigma_\phi$  is the phase error at the design values.

Similarly, that portion of the total phase error due to the additive noise is

$$\sigma_{n_1}^2 = \frac{N_1 (\gamma - 1)^2}{N_0 \alpha (\gamma - 1)^2}$$

where  $N_1$  is the actual spectral density of the additive noise process and  $N_0$  is the spectral density of the noise for which the loop was designed. Thus,  $\rho_0$  becomes

$$\rho_{\text{opt}} = \frac{1}{S_d + \frac{1}{\mu} \left[ \frac{(\gamma - 1)^2}{\alpha(\gamma + 1)^2} \right]} \quad (13)$$

$$\mu = \left( \frac{P}{\alpha N_0} \right) \left( \frac{\alpha N_1}{P} \right) = \frac{R_D}{R}$$

$R_D$  is the input design signal-to-noise ratio in a bandwidth of  $a/2\pi$  cps at the greatest communication range expected, and  $R$  is the actual input signal-to-noise ratio (due to changing channel conditions) referred to a bandwidth of  $a/2\pi$  cps. As a novel by-product of the analysis, the signal-to-noise ratio for the optimum receiver is obtained when  $R = R_D$ ; i.e.,

$$\rho_{\text{opt}} = \frac{(\gamma + 1)^2}{4\gamma} \quad (14)$$

This agrees with a result of Viterbi's (Ref. 17).

Further, the threshold locus of the PLL is given by

$$\sigma_1^2 = \frac{2m\dot{\gamma}}{\alpha\beta(\gamma^2 - 1)} \left[ 1 + \mu \frac{[2\beta(\gamma - 1)^2 + \gamma^2 - 1]}{(\gamma^2 - 1)} \right] \quad (15)$$

for the suboptimum or fixed receiver structure. Letting  $\mu = 1$ , we obtain the desired threshold locus

$$\sigma_1^2 = \frac{8m\dot{\gamma}}{(\gamma - 1)(\gamma + 1)^2} = C \quad (16)$$

In Part *h* we will have more to say about the constant  $C$ .

**e. Receiver performance for reception of a signal whose power is different from the design level.** From Eq. (10) it is clear that the closed-loop transfer function of the PLL depends on the received signal power  $P$ . In this part of this article, we wish to determine the effects on receiver performance when the transmitter power, say  $P_1$ , differs from the design level  $P_D$ . We shall assume further that the noise spectral density is  $N_0$  w/cps, and remains fixed.

Since the closed-loop transfer function of the PLL varies with the received signal power, it is necessary to determine the new transfer function when subjected to the true signal power  $P_1$ . Contrary to the linear case (where only the error due to noise affected the performance), we find that a change in signal power affects both the noise and signal error. This is, of course, due to the changing closed-loop transfer function which occurs in both of the error expressions.

From Eq. (7) we may rewrite, for an arbitrary signal power  $P_1$ , the quasi-linear PLL receiver transfer function as

$$H_{\bullet}(s) = \frac{K_r K_1 (P_1)^{1/2} F(s)}{s + K_r K_1 (P_1)^{1/2} F(s)} \quad (17)$$

where

$$K_1 = \exp(-\sigma_1^2/2)$$

and  $\sigma_1^2$  is the mean-squared phase error for this power level. Substitution of the fixed loop-filter  $F(s)$  [Eq. (11)] evaluated at design levels into Eq. (17) and rearranging gives

$$H'_{\bullet}(s) = \frac{\beta [a(\gamma - 1)s + a(\gamma + 1)]}{s^2 + as[1 + \beta(\gamma - 1) + a[(\gamma + 1)\beta]]} \quad (18)$$

where

$$\beta = \left( \frac{P_1}{P_D} \right)^{1/2} \exp\left(\frac{\sigma_1^2 - \sigma_1^2}{2}\right) \quad (19)$$

At  $\beta = 1$ ,  $H'_{\bullet}(s)$  reduces to  $H_{\bullet}(s)$  as it should. The overall fixed FM receiver filter is obtained from Eqs. (5) and (11), namely,

$$H'(s) = \frac{\beta a^2 (\gamma - 1) s}{2K_r [s^2 + a\{1 + \beta(\gamma - 1)\}s + a^2\delta\beta]} \quad (20)$$

The actual signal distortion is

$$S'_d = \frac{4\delta\beta(\delta + \alpha) + 4\delta\beta(\alpha^2 - \kappa) + \kappa^2(\alpha + 1)}{4\delta\beta[(\alpha + 1)(\delta + \alpha) - \delta\beta]} \quad (21)$$

where

$$\alpha = 1 + \beta(\gamma - 1)$$

$$\kappa = \beta(\gamma^2 - 1) - (\gamma - 1)^2$$

and  $\beta$  is defined by Eq. (19).

On the other hand, the variance of the noise error is

$$\sigma_{\epsilon_1}^2 = \frac{(\gamma - 1)^2}{\alpha(\gamma + 1)^2} \quad (22)$$

Hence, the actual signal-to-noise ratio  $\rho'_0$  is

$$\rho'_0 = \frac{1}{S'_d + \frac{(\gamma - 1)^2}{\alpha(\gamma + 1)^2}} \quad (23)$$

which is the required result.

The actual phase error  $\sigma_i^2$  which results from a change in received signal power may be written as

$$\sigma_i^2 = \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} S_{s_1}(s) |1 - H'_*(s)|^2 ds + \frac{1}{2\pi j} \int_{-j\infty}^{j\infty} S_{n'}(s) |H'_*(s)|^2 ds \quad (24)$$

and becomes, upon carrying out the necessary integration,

$$\sigma_i^2 = \frac{2m_i^2}{\alpha\beta(\gamma^2 - 1)} \left[ 1 + \frac{2\beta(\gamma - 1)^2 + \gamma^2 - 1}{(\gamma^2 - 1)} \right] \quad (25)$$

**f. Receiver performance when the PLL is preceded by a limiter and bandpass limiter.** In Part *d*, we investigated the demodulator performance when the PLL is preceded by an AGC amplifier which holds the signal level constant and the loop filter is designed in accordance with the signal power  $P$  and noise power  $N_0$  expected under worst channel conditioning. In this part, we again fix the loop filter as in Parts *d* and *e* and insert a bandpass limiter ahead of the PLL. Since the optimum demodulator requires the use of complex auxiliary servo loops which must be capable of continually adjusting the pole-zero configuration of the demodulator, it behooves the design engineer to seek means and ways of mechanizing the receiver so as to have it perform near optimally without the use of the exact equipment indicated by the deceptively simple adaptive filter equations. It appears, from an experimental standpoint, that the bandpass limiter may be an excellent engineering approximation to the complex servo system which we require for optimum performance.

The appropriate filter functions required in the derivation are given by Eqs. (18) and (20). As a matter of fact, we may make use of the results obtained in Part *e* for the signal distortion term. That portion of the error due to the new noise  $N_1$  is

$$\sigma_{n_1}^2 = \frac{N_1}{N_0} \left[ \frac{(\gamma - 1)^2}{\alpha(\gamma + 1)^2} \right] \quad (26)$$

The actual phase error which results from a change in received signal and noise power is found from Eq. (24). Appropriately substituting and integrating gives

$$\sigma_i^2 = \frac{2m_i^2}{\alpha\beta(\gamma^2 - 1)} \left[ 1 + \frac{N_1}{N_0} \left( \frac{2\beta(\gamma - 1)^2 + \gamma^2 - 1}{\gamma^2 - 1} \right) \right] \quad (27)$$

when Eq. (18) is substituted into Eq. (27) and the integration performed.

Thus, in order to obtain the resulting signal-to-noise ratio, we must relate the ratios  $N_1/N_0$  and  $P/P_D$  to the input-output signal-to-noise ratio relationship of the limiter. This, however, requires analyzing the limiter and we shall present this next.

**g. Signal-to-noise ratios in bandpass limiters.** In Ref. 21, a general analysis has been made relating output signal and noise powers for bandpass limiters having odd symmetry in their limiting characteristics. Assuming that the gain of the bandpass limiter is adjusted such that the output signal-plus-noise powers are those values for which the optimum loop was designed, one may show that (Ref. 17)

$$\frac{N_1}{N_0} = \frac{1 + R_D}{1 + R} \quad (28)$$

$$\frac{P}{P_D} = \frac{R}{R_D} \left( \frac{1 + R_D}{1 + R} \right) \quad (29)$$

where

$R$  = input signal-to-noise ratio of the bandpass limiter

$R_D$  = input signal-to-noise ratio for which the loop was designed

$P, N_1$  = bandpass limiter output noise and signal levels, respectively

$P_D, N_0$  = signal and noise levels for which the loop was designed

Thus, Eqs. (28) and (29) may be substituted into Eqs. (21) and (26) to yield the output "signal-to-noise" ratio, i.e.,

$$\rho'_0 = \frac{1}{S'_d + \left( \frac{1 + R_D}{1 + R} \right) \left[ \frac{(\gamma - 1)^2}{\alpha(\gamma + 1)^2} \right]} \quad (30)$$

where  $S_d'$  is given by Eq. (21) and

$$\alpha = 1 + \beta(\gamma - 1); \quad K = \beta(\gamma^2 - 1) - (\gamma - 1)^2$$

$$\beta = \left[ \frac{R}{R_d} \left( \frac{1 + R_d}{1 + R} \right) \exp(\sigma_1^2 - \sigma_2^2) \right]^{1/2} \quad (31)$$

This is the required result.

**h. Graphical results for the linear and quasi-linear demodulators.** In Fig. 5 we have illustrated graphically the performance of the linear FM demodulator for three different receiver configurations: (1) the optimum phase-locked loop demodulator, (2) the phase-locked demodulator (designed to be optimum for the most deleterious channel conditions) preceded by a bandpass limiter, and (3) an automatic gain control amplifier. For all three demodulators the threshold constant  $C$  has been set equal to unity. Experimental evidence supports this assumption. From Fig. 5 it is quite evident that the optimum demodulator outperforms either of the other two realiza-

tions; however, the "fixed" phase-locked loop, when preceded by the automatic-gain control-amplifier, is superior to the loop preceded by a bandpass limiter. The amount of superiority, however, becomes smaller as the modulation index  $m_f$  increases. In either of the two sub-optimum systems, the signal-to-noise ratio  $\rho$  becomes asymptotic (for large  $R$ ) to the reciprocal of the signal distortion  $S_d$  initially designed into the system. For comparison purposes, we indicate the performance of an amplitude-modulated double-sideband or single-sideband suppressed carrier system which demodulates the noisy received data by coherent frequency translation and smoothing the resulting waveform with a Wiener filter. The improvement obtained by using FM is clear.

Figs. 6 and 7 illustrate the performance of the same three systems when the quasi-linear receiver models are used. Fig. 7 has been prepared under the assumption that receiver threshold occurs when the total phase error equals  $\frac{1}{2}$  rad<sup>2</sup>, i.e.,  $C = 16$ , and Fig. 6 illustrates the performance of the three models for  $C = 8$ . Notice the higher threshold characteristics for the situation where  $C = 16$ .

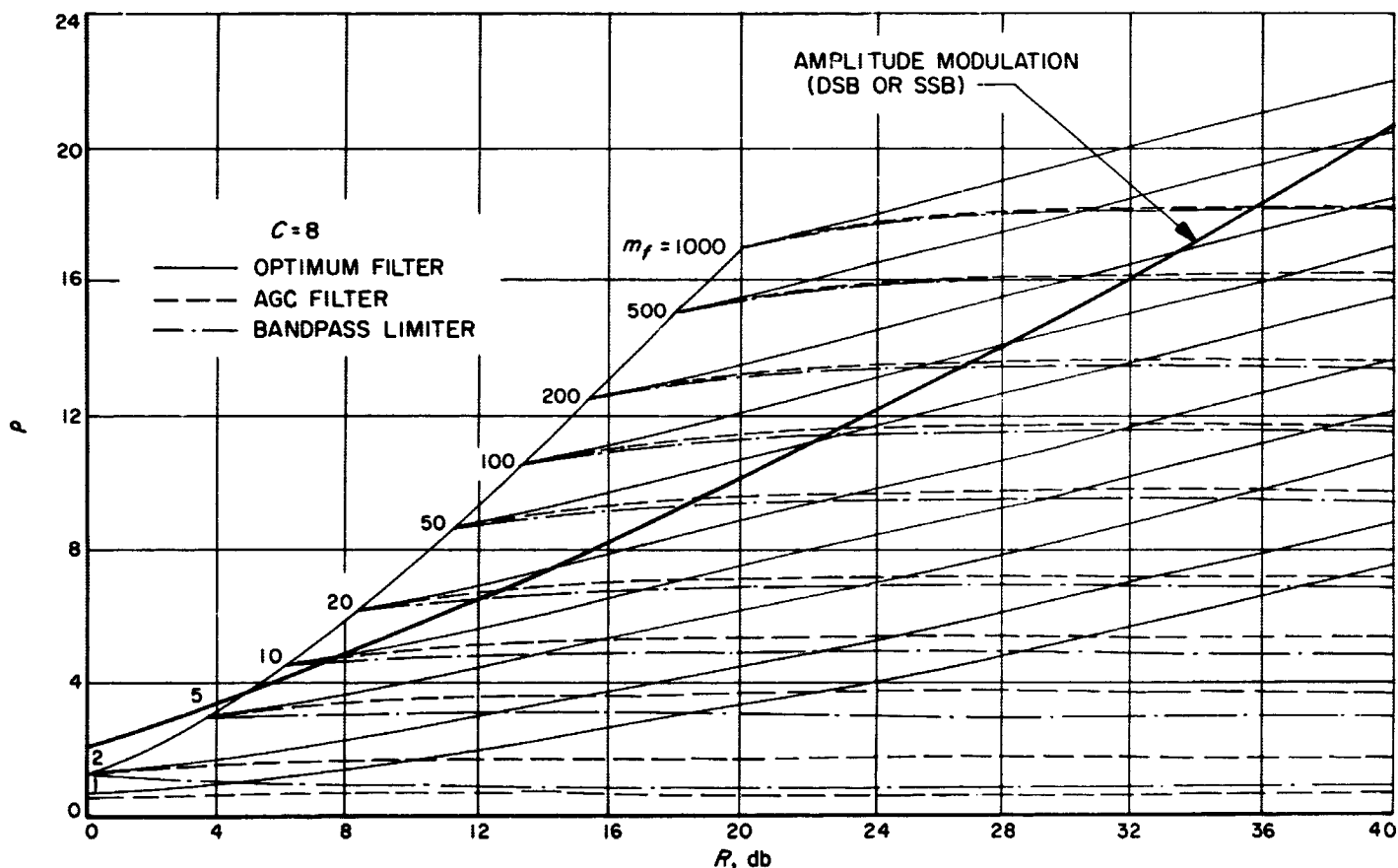


Fig. 5. Performance characteristics for the linear demodulator

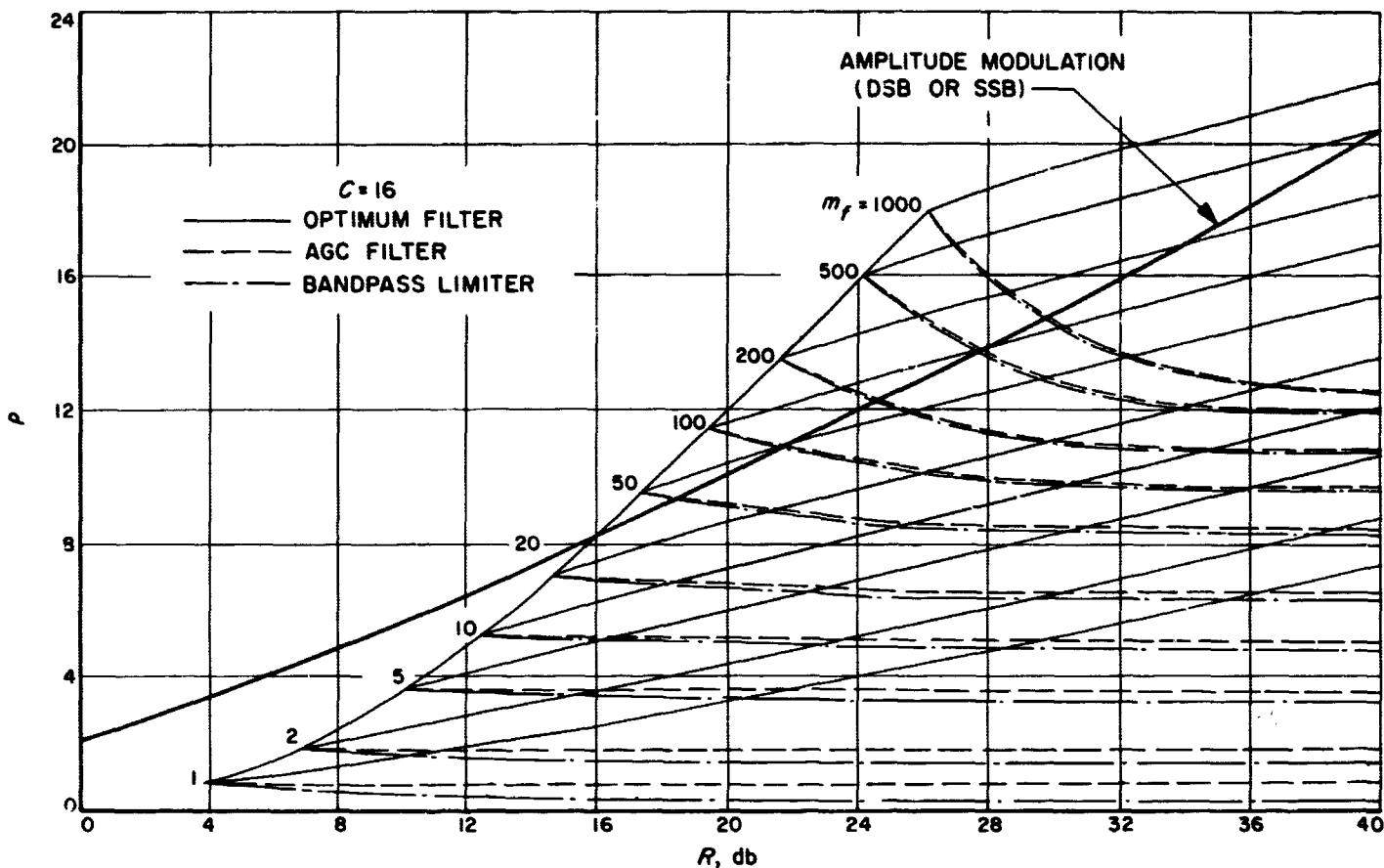


Fig. 6. Performance characteristics for the quasi-linear demodulator

The "dipping" behavior for either of the suboptimum quasi-linear systems may be explained as follows. As before, the asymptotic behavior (for large  $R$ ) is ultimately determined by the signal distortion  $S_d$  designed into the system. For the linear receiver models, this value of  $S_d$  was constant for all  $R \geq R_0$ . In the quasi-linear receivers, however, the signal distortion  $S_d$  is a function of the initial design parameters and the phase error, and is a minimum at threshold (worst channel conditions expected). As  $R$  increases beyond  $R_0$  the signal distortion increases until the phase error is so small that  $S_d$  is again determined by some function of the signal-to-noise ratio existing in the channel. These curves show the importance of keeping the system operating optimally. In fact, for  $C < 16$  it has been shown that the quasi-linear model specifies fairly accurately the performance of a phase-locked loop. As before, the "fixed" loop, when preceded by an AGC amplifier, outperforms the "fixed" loop preceded by a bandpass limiter.

Comparing Fig. 5 with Fig. 7, the linear system threshold characteristic is slightly lower than that of the nonlinear receiver. These results should prove beneficial to

engineers faced with the problem of designing second-order, phase-locked loop frequency demodulators.

### 3. Performance of a Class of Coherent Detectors, W. C. Lindsey

**a. Introduction.** In Ref. 22, we derived (in considerable generality) the output statistics of a coherent receiver which combines nonlinear functions of  $M$  samples taken from the output of both a set of matched filters and follow-up envelope detectors (Fig. 8). This receiver coalesces the two types of reception (commonly referred to as coherent and noncoherent) as extreme cases. For scatter-like channels the receiver operates on the envelope of the observed data, and the reception is usually termed noncoherent reception. For channels which do not scatter the transmitted waveform, the receiver operates coherently, while for the class of channels which lie somewhere between (the so-called "scatter-specular" channels) the receiver operates as a partially coherent detector. It is interesting to note that the structure for the partially coherent receiver is also optimum (with minor changes in the weighting of the output samples) for a problem

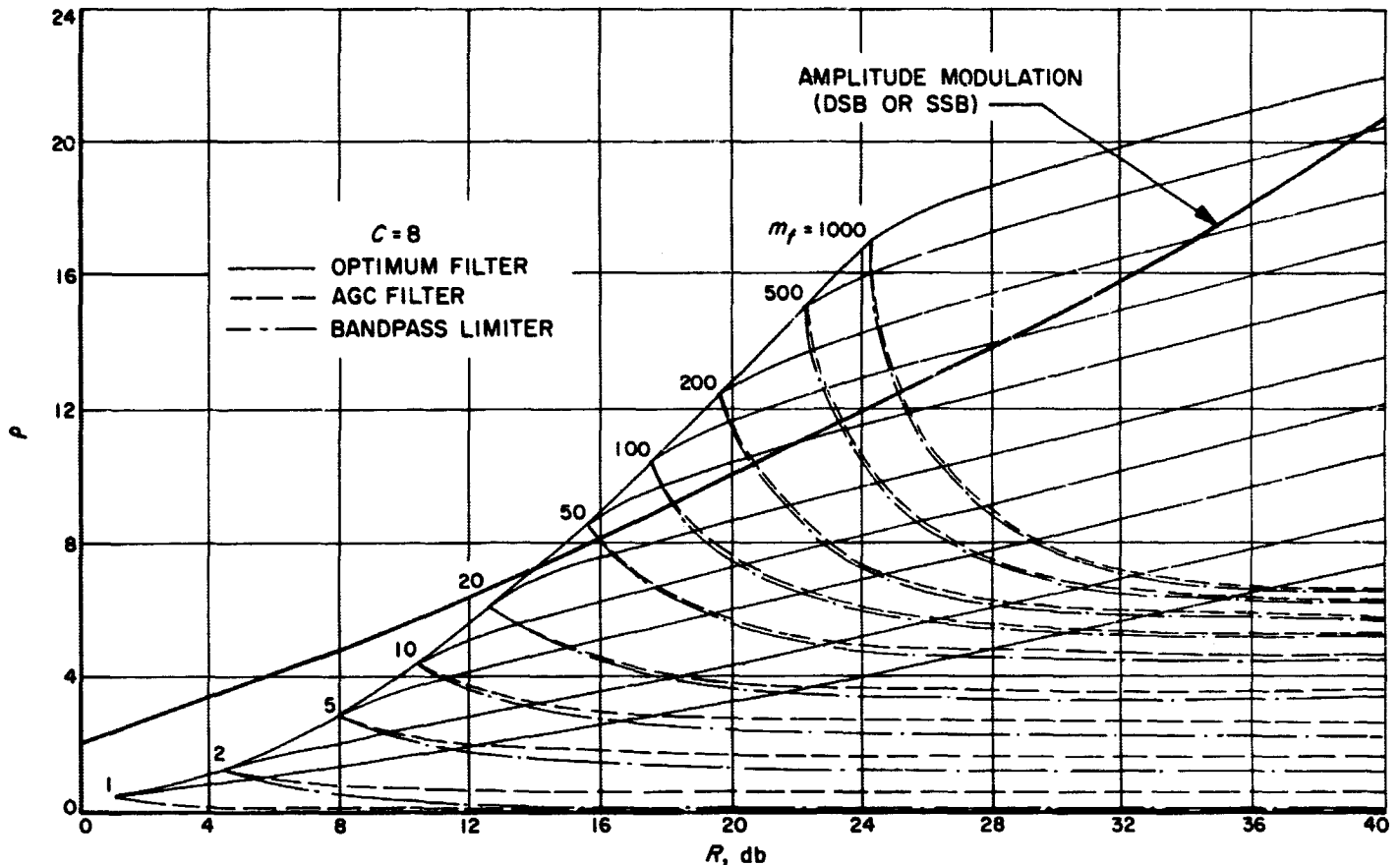


Fig. 7. Performance characteristics for the quasi-linear demodulator

considerably different from that considered here. Neglecting the detailed resemblances, it is sufficient to say that this optimum receiver structure arises when a coherent tracking device (known as a first-order, phase-locked loop) is used at the receiver for providing the receiver with a replica of the transmitted carrier.

Here we present the performance characteristics for the receiver of Fig. 8 and compare these with those of an easier implemented suboptimum square-law combiner. This receiver is in operation when switches S2 and S3 are open (Fig. 8). The results may be used as a guide when it comes to the engineering design of receivers and serve to point out the increase in system reliability obtained by inserting redundancy into the system.

Finally, the results may be applied to the following special problem. Suppose one desires the probability with which a command detector located in a spacecraft receives the wrong command upon approaching a certain planet. The waveform which the command detector

processes is composed of a direct-wave from the Earth-based transmitter and multiple reflections from the planet. The strength of these reflections depends on the reflectivity of the planetary surface, while the phase shift depends on the round-trip distance from the spacecraft to the oncoming planet, and the spacecraft antenna design. The model presumed depicts this situation in that the direct-wave from the Earth-based transmitter is of fixed strength while the planet serves as a Rayleigh-scatterer of the direct wave. We treat this problem, as a special case, i.e.,  $M = 1$ .

**b. Results and conclusions.** We briefly restate the problem. One of a finite set of known equiprobable signals  $\{\xi_k(t), k = 1, 2, \dots, N\}$  of limited time duration is transmitted through a random multichannel  $C$  resulting in a waveform of  $\psi(t)$  that is further corrupted by additive white Gaussian noise  $v(t)$  of single-sided spectral density  $N_0$  w/cps before being available as a waveform  $\zeta(t)$  to the receiver. The optimum receiver (in the Woodward sense) computes the set of a posteriori probabilities  $\{p(\xi_k|\zeta)\}$ . We specify the probability that the receiver errs in making its decision, and compare the



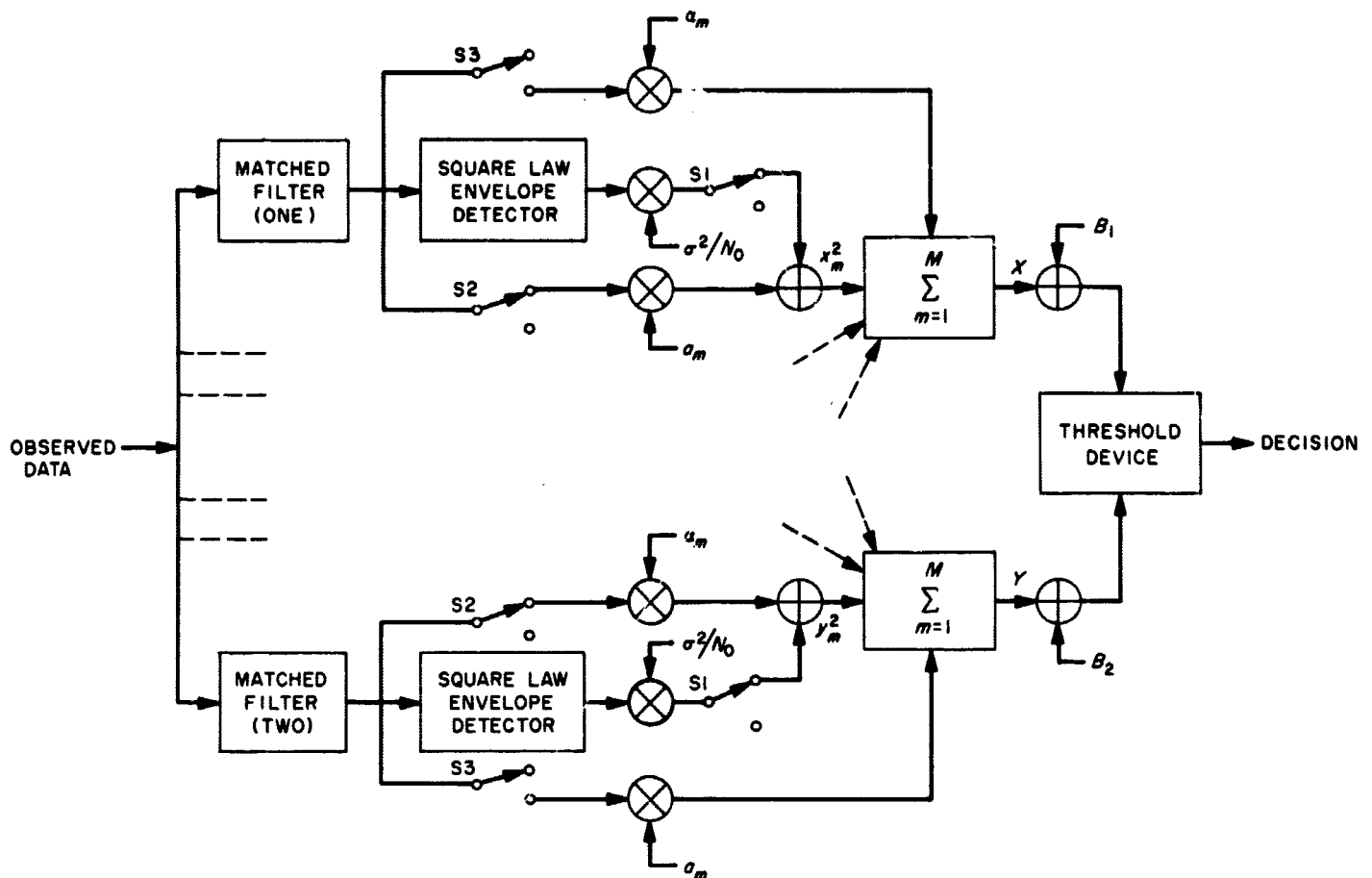


Fig. 8. Detection system

results with a more easily implemented, suboptimum receiver.

The strength and phase shift of each channel are considered random and are postulated to obey the "Rice" probability density. Physical justification for the use of this distribution is given elsewhere (Ref. 23); briefly, it is sufficient to describe the fading phenomenon as being composed of a specular component of fixed strength  $\alpha$ , phase shift  $\delta$ , and a scatter-component with Rayleigh-distributed amplitude of mean-squared value  $2\sigma^2$ , completely random phase. A parameter of importance in the final results is  $2^{1/2}\gamma = \alpha/\sigma$ . For  $\gamma = 0$ , the channel fading is Rayleigh, for  $\gamma > 0$  the fading becomes approximately Gaussian; whereas for  $\gamma = \infty$  there exists no fading. Hence, the fading model is sufficiently general to include several types of propagation media.

Assuming that the receiver knows a priori or as a result of measurement  $\alpha$ ,  $\delta$ ,  $N_0$ ,  $\sigma$ , and on the basis of the preceding hypotheses, it may be shown (Ref. 24)

that the ideal receiver conjectures that the transmitted signal was the one for which

$$R_k = \sum_{m=1}^M \frac{\sigma^2 |\rho_{mk}|^2 / N_0 + 2\alpha_m \rho_{mk}}{2N_0 (1 + \beta)} \quad (1)$$

is the largest. In Eq. (1),  $\beta = 2\sigma^2 E / N_0$  and  $E$  is the common signal energy;  $\beta$  is thus a measure of the received signal-to-noise ratio. The operations which the receiver must perform on the received data  $\zeta$  are expressed by the complex  $\rho$ 's in Eq. (2), namely,

$$\rho_{mk} = \int_0^T \zeta^*(t) \xi_k(t) dt \quad (2)$$

Physically, the  $\rho$ 's may be computed by cross-correlation techniques and  $\rho$  is twice the envelope of the physical cross-correlation function.

Under the foregoing assumption it is possible to show that the probability of correct reception is given by the rather general result (Ref. 22)

$$P_c = \int_0^\infty p(W) dW \left[ \int_0^{(1+\beta)^{1/2} W} p(Y) dY \right]^{N-1} \quad (3)$$

where the variables  $W$  and  $Y$  are non-central chi-squared variates of  $2M - 2$  degrees of freedom (Ref. 25) and possess non-centrality parameters  $\lambda_Y$  and  $\lambda_W$

$$\lambda_W = (1 + \beta)^{1/2} \lambda_Y; \quad \lambda_Y = \left( \frac{2}{\beta} \sum_{m=1}^M \gamma_m^2 \right)^{1/2} \quad (4)$$

(See Ref. 22.) Several interesting system performance characteristics may be deduced from Eq. (4), e.g., systems which transmit orthogonal codes through random and specular channels, etc. For the binary case (i.e.,  $N = 2$ ), it may be shown, using a result of Price (Ref. 25), that the error probability is given by

$$P_E(M) = Q(\mu, \epsilon) - \exp \left[ -\frac{\mu^2 + \epsilon^2}{2} \right] \times \left\{ \frac{r}{1 + r^2} I_0(\mu, \epsilon) + (r - r^{-1}) \sum_{m=1}^M C_m I_m(\mu, \epsilon) \right\} \quad (5)$$

where  $Q(\mu, \epsilon)$  is the well-known, tabulated Marcum  $Q$ -function,  $I_m(z)$  is the modified Bessel function,

$$C_m = \sum_{j=m+\delta_{m0}}^{M-1} (r + r^{-1})^{-2j-1} \quad (6)$$

and  $r = (1 + \beta)^{1/2}$ ,  $\mu = \lambda_Y [(1 + r^2)^{1/2}]^{-1}$ ,  $\epsilon = r \lambda_W [1 + r^2]^{-1}$ . Eq. (5) is valid for  $M > 2$ ,  $\mu \neq 0$ ,  $\epsilon \neq 0$ , and  $\delta$  is the Kronecker delta function. For  $M = 1$ ,  $C_m$  is zero for all  $m$ , and  $w$  obtain a well-known Turin result.

Eq. (5) has been evaluated for equal-strength fixed components and plotted in Fig. 9 versus total average received signal-to-noise ratio for  $\gamma^2 = 2$  (typical for spread  $F$ -echoes) and various values of  $M$ . For comparison purposes we present results for a more easily implemented noncoherent receiver which utilizes only the squared envelope detector outputs (Ref. 23). This receiver has the distinct advantage of obviating the need for measuring the channel parameters  $\alpha$ ,  $\delta$ ,  $\sigma$ , and  $N_0$ . Consequently, the system implementation problem is not as complex. In fact, if all specular components are zero and the expected power received from each channel is the same, the optimum combination rule for the  $M$  samples is square-law envelope detection (Ref. 24).

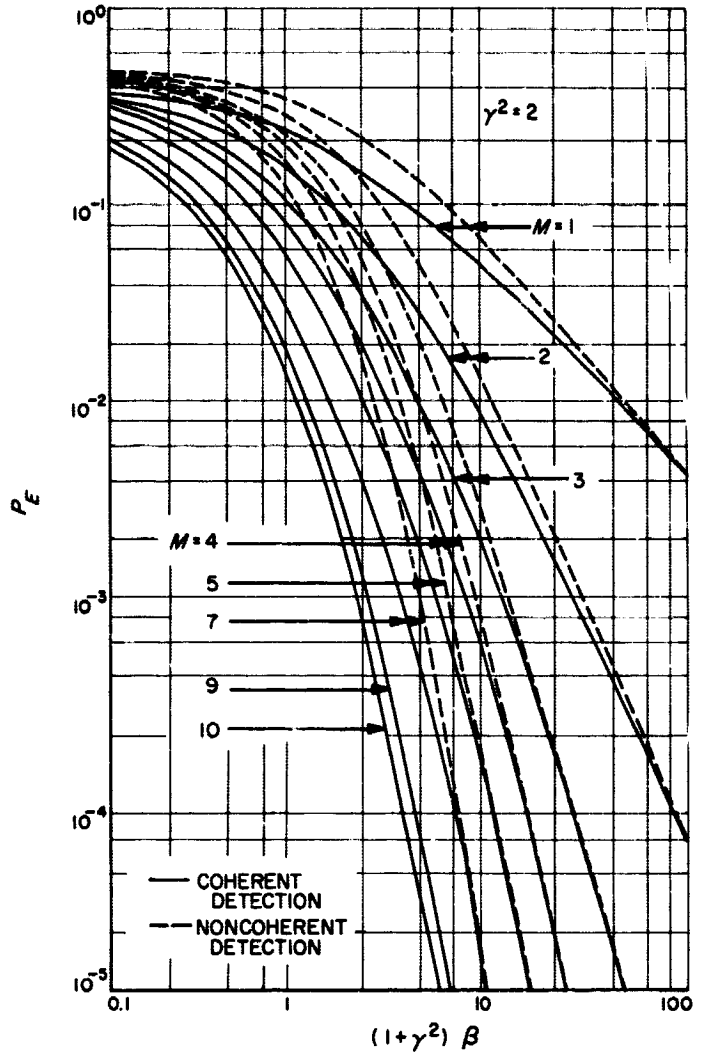


Fig. 9. Error probability versus average received signal-to-noise ratio

From the foregoing analysis and graphical results, the following conclusions are evidenced. As expected, multi-channel reception increases the reliability of communicating through a noisy-fading medium. Improvement is quite noticeable for values of  $M$  up to about 3; thereafter, improvement is not as drastic. As the channel noise  $N_0 \rightarrow \infty$  or  $\beta \rightarrow 0$ , the receiver operates as if there were no scatter components and almost all of the information is conveyed to the receiver via the specular components. For small  $\beta$ , the error probability approaches that of coherent reception in the face of white Gaussian noise. For large  $\beta$ , i.e., scatter channels, most of the information is conveyed to the receiver via the specular components, and the error probability approaches that of noncoherent square-law envelope detection. Thus, the error rate changes from that obtained when coherent reception is used and the channel is purely specular to that obtained

when the channel is purely random and noncoherent square-law reception is used.

If, in designing a particular system, where  $\gamma^2 = 2$ , requirements were such that the bit-error probability must remain less than  $10^{-3}$  (as is the case for most telemetry and command systems), the curves of Fig. 8 lead one to the practical conclusion that a system, which is to operate through a channel containing specular and random components, will be essentially optimum if it is designed to use square-law detection exclusively.

#### 4. Phase-Locked Synchronization With Sinusoidal Signals, J. J. Stiffler

**a. Introduction.** The synchronization problem in a communication system, in essence, involves the establishment of coherence between the transmitter and receiver clocks. In general, not only may the clocks not simultaneously record the same time instant, but they may actually be running at different rates. It is necessary to determine at the receiver both the rate and the time instant of the transmitter clock. Since these two quantities translate themselves into the frequency and phase, respectively, of a periodic time function, one synchronization scheme is to transmit such a function and to track it at the receiver with a phase-locked loop. Because of the possible frequency discrepancy between the two clocks, a second-order loop immediately suggests itself (Ref. 26). This summary considers some of the problems encountered when the periodic function is a sinusoid and a second-order loop is used at the receiver.

In Part *b* of this article, it is shown that the timing error in seconds is inversely proportional to the two-thirds power of the frequency of the sinusoid and that, consequently, the largest possible frequency should be used. It is observed, however, that the time ambiguity increases as the frequency increases, a factor which immediately places an upper bound on the frequency of the sinusoid. Because of this trade-off between ambiguity and accuracy, it is of interest to consider a multiple-loop scheme in which low-frequency sinusoids are transmitted to eliminate any ambiguities, and higher frequencies simultaneously transmitted to increase the accuracy. This method is investigated in Part *c*.

It might be supposed that an arbitrarily high accuracy could be obtained by using more and more sinusoids of higher and higher frequencies, each frequency enough higher than the previous so as to increase the tracking accuracy without introducing any new ambiguities. It is shown, however, that if the total available power is con-

strained, the number of sinusoids which can profitably be used in this manner is sharply limited. This is because the total power must be divided among more and more signals causing an increase in the tracking error of each, and thereby counteracting the above-mentioned tendency for the error to decrease through the use of higher frequencies.

**b. Tracking a single sinusoid.** In this part of the article, a second-order loop is to be used to track a sinusoid. Suppose the transmitter frequency is  $f_0$  cps, but that the transmitter and receiver clocks are running at different rates. If during  $\tau$  sec transmitter time, during which  $f_0\tau$  cycles have been transmitted, the apparent elapsed time at the receiver is  $r\tau$  sec, the received frequency is  $f_0\tau/r\tau = f_0/r$  cps. Thus, the difference in the two frequencies is

$$\Delta f = f_0 \left( \frac{r-1}{r} \right) \text{ cps} \quad (1)$$

and, for a given discrepancy in the clock rates, is directly proportional to the transmitter frequency  $f_0$ . (One of the causes of such a phenomenon might be a doppler shift due to the relative motion of the transmitter and the receiver.) If the maximum value of  $r$ , and hence the maximum frequency transient, is specified together with the signal-to-noise ratio, only two parameters remain to be chosen: the frequency  $f_0$  of the transmitter sinusoid and the loop bandwidth. Some observations relevant to this selection follow.

It will be supposed that it is important that the loop does not temporarily lose lock. In certain rather special situations, this may not be so important but, in general, loss of lock even temporarily, will mean the introduction of possible synchronization errors which are not easily eliminated. Thus, it is required that neither the rms error due to noise nor the transient error approach the value  $\pi/2$  rad. The standard deviation of the error due to the noise is well known to be

$$\sigma_n = \left( \frac{N_0}{P_c} \right)^{1/2} B_L^{1/2} \text{ rad} \quad (2)$$

where  $P_c$  is the carrier power,  $N_0$  the noise spectral density, and  $B_L$  the loop bandwidth. The transient error  $e(t)$  at any given instant due to a frequency step  $\Delta\omega$  at time  $t = 0$  can also be readily determined:

$$e(t) = \frac{1}{2\pi j} \int_{-j\infty}^{+j\infty} [1 - Y(s)] \frac{\Delta\omega}{s^2} e^{st} ds = \frac{\Delta\omega}{B_0/2^{1/2}} \times e^{-B_0 t/2^{1/2}} \sin \frac{B_0}{2^{1/2}} t, \quad t \geq 0 \quad (3)$$

where

$$Y(s) = \frac{2^{1/2} B_0 s + B_0^2}{s^2 + 2^{1/2} B_0 s + B_0^2}$$

is the loop transfer function and

$$B_0 = \frac{4(2)^{1/2}}{3} B_L \\ = \left( \frac{2^{1/2} \lambda \Delta \omega A_0}{(N_0)^{1/2}} \right)^{1/2}$$

is proportional to the noise bandwidth  $B_L$  of the loop (Ref. 26). The maximum transient error evidently occurs at time  $t = \pi(2)^{1/2}/4 B_0$  and is  $e_{max} = (\Delta \omega/B_0) e^{-\pi/4}$  rad. Consequently, one condition which must be satisfied is that

$$e_{max} + k\sigma_n < \frac{\pi}{2} \quad (4)$$

where  $k$  is chosen so that the probability that the instantaneous error exceed  $\pi/2$  rad at any given time is satisfactorily small. If Gaussian statistics and complete linearity are assumed, the probability of loss of lock at the instant of maximum transient error is just  $Pr(\text{loss of lock}) = \frac{1}{2} (1 - \text{erf } k)$ . The optimum value of  $B_L$  can be determined by differentiating the expression

$$e_{max} + k\sigma_n = \frac{3\Delta \omega}{4(2)^{1/2} B_L} e^{-\pi/4} + k \left( \frac{N_0}{P_c} \right)^{1/2} B_L^{1/2} \quad (5)$$

with respect to  $B_L$  and equating the result to zero, yielding the condition that:

$$B_L = \left[ \frac{3e^{-\pi/4}}{2(2)^{1/2} k (N_0/P_c)^{1/2}} \right]^{2/3} (\Delta \omega)^{2/3} = k_1 (\Delta \omega)^{2/3} \quad (6)$$

The frequency step is, of course, proportional to the signal frequency  $\omega$ .

So far as synchronization is concerned, it is not the error in radians that is significant, but the error in seconds. Since

$$\sigma_n = \left( \frac{N_0}{P_c} \right)^{1/2} \frac{B_L^{1/2}}{\omega} \text{ sec}$$

and

$$e_{max} = \frac{3\Delta \omega e^{-\pi/4}}{4(2)^{1/2} B_L \omega} \text{ sec} \quad (7)$$

both errors are proportional to

$$\frac{\Delta \omega^{1/2}}{\omega} \sim \frac{1}{\omega^{1/2}} \quad (8)$$

Consequently, it is advantageous to use the maximum practical frequency subject to the condition that the error, in radians, is bounded by  $\pi/2$ :

$$e_{max} + k\sigma_n = \frac{3}{2} \left[ \frac{3e^{-\pi/4}}{2(2)^{1/2}} k^2 \frac{N_0}{P_c} \left( \frac{r-1}{r} \right) \omega \right]^{1/2} \\ = \frac{3k}{2} \left( \frac{N_0 B_L}{P_c} \right)^{1/2} < \frac{\pi}{2} \quad (9)$$

where  $r$  designates the ratio of the two clock rates.

The disadvantage of this approach is readily apparent. As the frequency is increased beyond the reciprocal of the period of ambiguity  $T$ ,  $\omega > 2\pi/T$ , there is more than one position at which the loop may lock. In other words, referring to Fig. 10(a), it is seen that there is only one *stable* lock-in position each period. (The other zero crossing is an unstable position since the operation of the loop is to shift the phase of the VCO in the direction indicated by the arrows.) But in Fig. 10(b), there are a number of lock-in positions in each period, resulting in an ambiguity which must somehow be resolved.

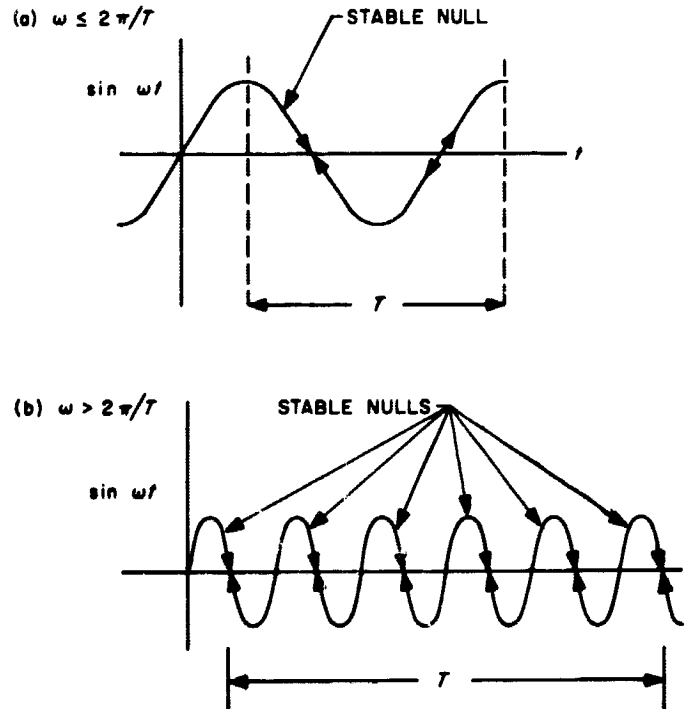


Fig. 10. Lock-in points in time  $T$ : (a)  $\omega \leq 2\pi/T$  and (b)  $\omega > 2\pi/T$

**c. Multiple loop systems.** This trade-off between accuracy and ambiguity suggests another approach to the synchronization problem. Instead of a single sinusoid, the waveform

$$f(t) = \sum_{i=1}^m a_i \cos \omega_i t \quad (10)$$

is used where  $2\pi/\omega_1 = T$ , the period of uncertainty. The receiver consists of  $m$  phase-locked loops, one centered at each of the frequencies  $\omega_i$ . The first loop centered at  $\omega_1$  then provides an unambiguous, though somewhat crude, reference. The following loops successively refine this synchronization estimate. It would be supposed that the error of the estimate would be less than that using a single loop alone, since the higher frequency loops provide greater accuracy at the same signal-to-noise ratio. The accuracy attainable with multiple loop systems will be the subject of this part of this article.

It was shown in Part *b* that, when the bandwidth was optimally chosen, the parameter

$$e_{max} + k\sigma_n = \frac{3k}{2} \left( \frac{N_0 B_L}{P_c} \right)^{1/2} \quad (11)$$

This parameter is related to the probability of loss of lock. (Note that the linear approximation to the phase-locked loop, on which this analysis has been based, ceases to be acceptable as the loop approaches loss of lock. Nevertheless, if the system parameters are such that  $e_{max} + k\sigma_n \ll \pi/2$ , it is clear that the linear approximation is valid, and that the probability of loss-of-lock is directly related to the value of  $k$ .)

Suppose that the  $i$ th loop is operating so that

$$(e_{max} + k\sigma_n)_i = \frac{\pi}{k_i} \quad (12)$$

for some satisfactorily large value of  $k$ . Then, clearly, the center frequency  $\omega_{i+1}$  of the  $(i+1)$ st loop can be as great as  $k_i \omega_i$  without introducing any ambiguity. Further, as long as  $k$  and  $k_i$  are large enough, the probability of loss of lock in the  $i$ th loop is negligibly small and the tracking error is essentially that of the  $(i+1)$ st loop. Let

$$K = \frac{2\pi}{3} \left[ \frac{3e^{-\pi/4}}{2(2)^{1/2}} k^2 N_0 \left( \frac{r-1}{r} \right) \right]^{1/2}$$

where the terms are as defined in Part *b*. Then, from Eqs. (9) and (12),

$$\frac{\pi}{K} \left( \frac{\omega_i}{P_i} \right)^{1/2} = \frac{\pi}{k_i},$$

where  $P_i = a_i^2/2$  is the signal power at the  $i$ th loop, and

$$\omega_{i+1} = k_i \omega_i = K P_i^{1/2} \omega_i^{1/2} \quad (13)$$

If  $m$  such loops are used, each related to the previous in this way, then the over-all error is given by

$$\begin{aligned} e_m &= (e_{max} + k\sigma_n)_m = \frac{\pi}{K} \left( \frac{\omega_m}{P_m} \right)^{1/2} \text{ rad} \\ &= \frac{\pi}{K} \frac{1}{P_m^{1/2} \omega_m^{1/2}} \text{ sec} \\ &= \frac{\pi}{K P_m^{1/2}} \frac{1}{[K P_{m-1}]^{1/2} [K P_{m-2}]^{1/2} \dots} \\ &\quad \times \frac{1}{[K P_{m-3}]^{1/2} \dots [K P_1]^{1/2} \omega_1^{1/2}} \end{aligned} \quad (14)$$

Since  $K$  is determined by the physical parameters of the problem, and since  $\omega_1 = 2\pi/T$  is fixed by the period of the ambiguity, the only parameters which can be varied are the terms  $P_i$ . It is desired to allocate the power such that the error [Eq. (14)], is minimized subject, of course, to the restriction that the total power is limited:

$$\sum_{i=1}^m P_i = P_c \quad (15)$$

This optimum power allocation can therefore be determined by equating to zero the derivatives with respect to  $P_i$  of the expression

$$\prod_{i=0}^{m-1} [P_{m-i}^{1/2}]^{(1/2)^i} + \lambda \sum_{i=0}^{m-1} P_{m-i} \quad (16)$$

This establishes the condition for an extremum (which is clearly a maximum) that

$$P_{m-v} = \left\{ - (1/\lambda) 1/3 (2/3)^v \prod_{i=0}^{m-1} [P_{m-i}^{1/2}]^{(1/2)^i} \right\}^{1/2} \quad (17)$$

But

$$P_c = \sum_{v=0}^{m-1} P_{m-v} = \left\{ \frac{- \prod_{i=0}^{m-1} [P_{m-i}^{1/2}]^{(1/2)^i}}{3\lambda} \right\}^{1/2} \sum_{v=0}^{m-1} (2/3)^{v/2} \quad (18)$$

and hence

$$\left[ \frac{-\prod_{i=0}^{m-1} [P_c^{1/2}]^{(1/2)^i}}{3\lambda} \right]^{1/2} = \left[ \frac{1 - (2/3)^{1/2}}{1 - (2/3)^{m/2}} \right] P_c$$

Consequently, the optimum power division is that in which

$$P_v = \left[ \frac{1 - (2/3)^{1/2}}{1 - (2/3)^{m/2}} \right] (2/3)^{(m-v)/2} P_c \quad (19)$$

and the ratio of the center frequency of the  $(v+1)^{\text{st}}$  loop to that of the  $v^{\text{th}}$  loop may be written, after some manipulation,

$$k_v = \frac{2\pi}{3} \left[ \frac{1 - (2/3)^{1/2}}{1 - (2/3)^{m/2}} \right]^{1/2} (2/3)^{(m-v)/2} \frac{1}{k} \left( \frac{P_c}{N_0 B_v} \right)^{1/2} \quad (20)$$

where

$$B_v = \left[ \frac{3e^{-\pi/4} \Delta\omega_v}{2(2)^{1/2} k \left( \frac{N_0}{P_c} \right)^{1/2}} \right]^{1/2}$$

would be the optimum bandwidth of  $v^{\text{th}}$  loop if all the power were used at that frequency only. The  $m^{\text{th}}$  loop error, in seconds, becomes

$$e_m = \frac{\pi}{k_m \omega_m} = \frac{\pi}{\omega_1} [3/2]^{4 - [(m+5)/2]} \left[ \frac{1 - (2/3)^{m/2}}{1 - (2/3)^{1/2}} \right]^{(1 - (1/2)^m)} \times \left[ \frac{k/\pi}{(P_c/N_0 B_1)^{1/2}} \right]^{3[1 - (1/2)^m]} \quad (21)$$

and decreases as  $m$  increases so long as

$$1 > \frac{e_m}{e_{m-1}}$$

or

$$\frac{1}{k} \left( \frac{P_c}{N_0 B_1} \right)^{1/2} \geq f(m) = \frac{(3/2)^{(m+5)/2}}{\pi [1 - (2/3)^{1/2}]^{1/2}} \times \left\{ \frac{[1 - (2/3)^{m/2}][(3/2)^{m-1} - (2/3)]}{[1 - (2/3)^{(m-1)/2}][(3/2)^{m-1} - 1]} \right\} \quad (22)$$

The expression on the right  $f(m)$  is plotted as a function of  $m$  in Fig. 11. It is noted that  $f(m)$  increases very rapidly with  $m$ . Note that

$$\left( \frac{P_c}{N_0 B_L} \right)^{1/2} = \frac{1}{\sigma_1} \quad (23)$$

where  $\sigma_1^2$  is the phase variance when only one loop is used at the same signal-to-noise ratio. Typically, this

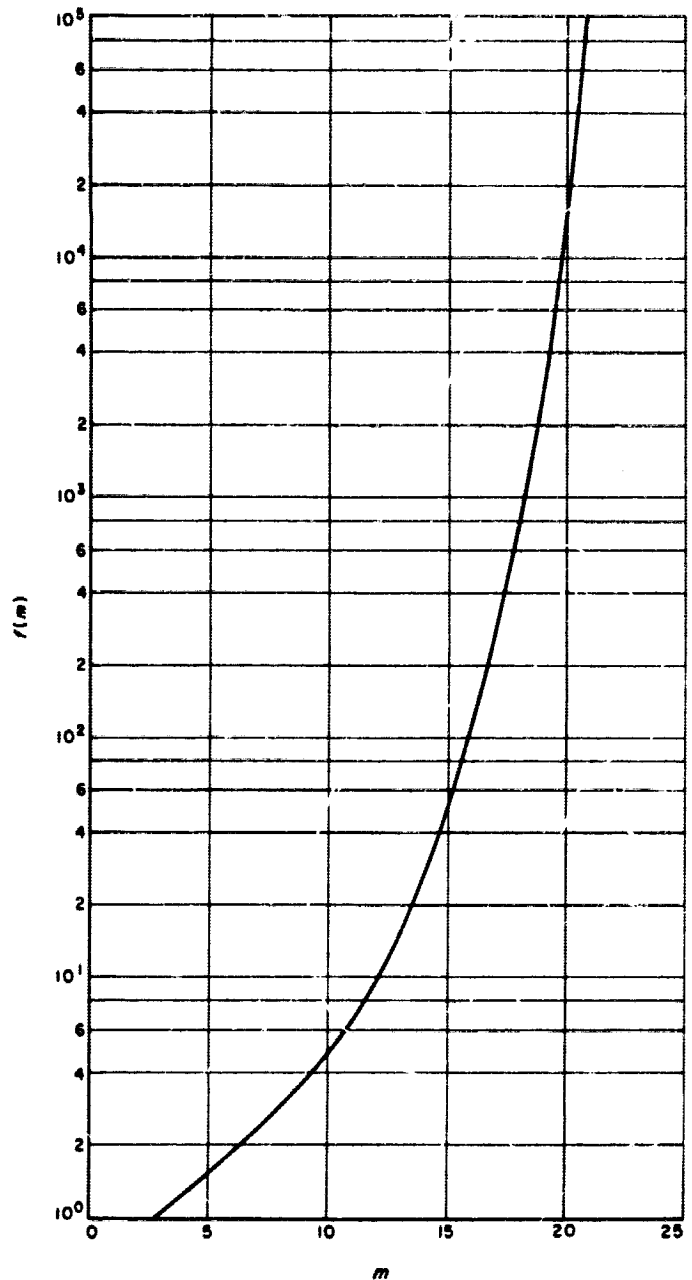


Fig. 11.  $f(m)$  versus  $m$  [see Eq. (22)]

might be on the order of  $\pi/40$  rad; and with  $k = 5$ , for example, this requires  $f(m) \leq 8/\pi$  or  $m \leq 7$  and the error [Eq. (21)] is reduced from  $T/80$  to  $T/776$ , or by a factor of nearly 10.

The results of this part of this article appear to be limited to the assumptions concerning the type of loop compensation required. Specifically, it was assumed that second-order loops were used, optimized so as to keep the transient error within bounds while minimizing the phase variance due to the noise. But, in general, it may

be said that the contribution to the phase error due to the additive noise will be proportional to  $B_L^{1/2}/\omega$  sec, while that due to the phase jitter at the source will be proportional to  $1/B_L$  sec. (Consider, for example, the respective errors when the loop is optimized to compensate for a random frequency jitter of the source.) Consequently, a trade-off between the two types of error does occur and a condition not unlike that obtained here will limit the number of loops which should be used. The parameter  $k$ , in fact, provides a useful weighting between the two types of error.

It is dubious that a multiple loop synchronization scheme would be practical in most communication situations. The slight advantage would generally be counteracted by the additional complexity of equipment necessary to lock onto several sinusoids as well as to combine the information from the different loops. In many systems it is necessary to use several loops, one for carrier synchronization, perhaps another for subcarrier synchronization, and even a third for symbol or bit synchronization. Nevertheless, the difference between frequencies will generally be so great that the knowledge of one type of synchronization cannot effectively be used to estimate or to improve the estimate of the others.

It is interesting to note that what has been said concerning synchronization is also applicable to ranging, since the two problems can both be solved by the accurate determination of the phase of a received signal. Ranging systems, in fact, have been designed which do use this technique for improving the range accuracy.

## 5. Optimization of Radar-Radiometric Cycle,

R. C. Tittsworth

**a. Introduction.** This article investigates the radar-radiometric cycle used in lunar and planetary radar systems, with a view of optimizing the cycle parameters so

that the maximum possible signal-to-noise ratio occurs at the output. In particular, it is found that the "receive  $S + N$ " portion of the cycle should be shorter than the "receive  $N$ " portion.

**b. The radar-radiometric cycle.** Consider the general cycle shown in Fig. 12. The transmitter is on for a time  $a$ , the receiver recovers for a time  $r$ , noise is received for a time  $b$ , signal plus noise for a time  $a$ , and noise again for a time  $c$ . We normalize time so that the cycle is based on a round-trip time of one unit.

We should point out that the cycle need not necessarily look exactly like that in Fig. 12. For example, transmission may be distributed in any fashion desired on  $(0, 1)$ , but the total transmission time is  $a$ , and the total recovery time is  $r$ .

The detector output we shall assume is

$$D(t) = \begin{cases} n(t) & 1-b < t < 1 \\ \mu + n(t) & 1 < t < 1+a \\ n(t) & 1+a < t < 1+a+c \end{cases}$$

where  $\mu$  is some mean value we wish to estimate, and  $n(t)$  is some noise term. The detector output is averaged over three intervals:

$$A_1 = \frac{1}{b} \int_{1-b}^1 D(t) dt = n_1$$

$$A_2 = \frac{1}{a} \int_1^{1+a} D(t) dt = \mu + n_2$$

$$A_3 = \frac{1}{c} \int_{1+a}^{1+a+c} D(t) dt = n_3$$

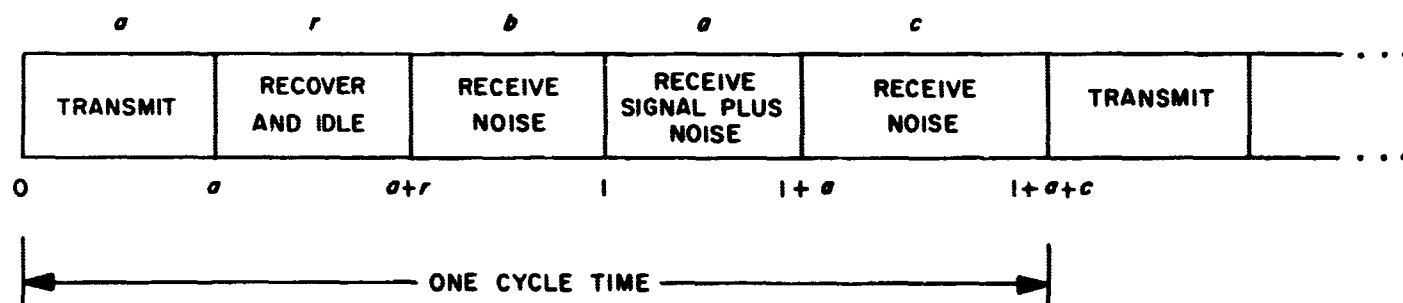


Fig. 12. Radar-radiometric cycle

The noise variances are, for any detector, approximated by (Ref. 27)

$$\text{var } n_1 = \frac{K}{b}$$

$$\text{var } n_2 = \frac{K'}{a}$$

$$\text{var } n_3 = \frac{K}{c}$$

$K$  and  $K'$  are slightly different due to the presence of signal in the detector, for that case. The estimate of  $\mu$  is

$$\hat{\mu} = k_1 A_1 + k_2 A_2 + k_3 A_3$$

We choose  $k_i$  to make  $\mu$  unbiased and minimal variance.

In order to simplify things, we shall assume weak signal conditions so that  $K \approx K'$  and  $E(n_i) = E(\gamma_i)$ . Then for  $\hat{\mu}$  to be unbiased

$$k_2 = 1$$

$$k = k_1 = -(1 + k_3)$$

and for  $\hat{\mu}$  to be minimum variance by adjustment of  $k$ ,

$$k = -\frac{b}{b+c}$$

so that the minimum variance of  $\hat{\mu}$  is

$$\sigma^2 = \min \text{var } \hat{\mu} = K \left[ \frac{1}{a} + \frac{1}{b+c} \right]$$

If several cycles are considered, say  $N$ , the variance is  $\sigma^2/N$ . Thus, in a time  $T = N \cdot (\text{cycle length})$ , the squared-mean-to-variance ratio is

$$\lambda = \frac{\mu^2 T}{K(1+a+c)} \cdot \frac{1}{\left( \frac{1}{a} + \frac{1}{b+c} \right)}$$

This is clearly monotone increasing in  $b$ , so  $b$  should be made as large as possible,  $b = 1 - a - r$ , where  $r$  is the minimum receiver recovery time. Then

$$\lambda = \frac{\mu^2 T a (1 - a - r + c)}{K(1+a+c)(1-r+c)}$$

We wish to choose  $a$  and  $c$  to maximize this ratio. For convenience, set

$$\rho = \frac{\lambda K}{\mu^2 T} = \frac{a(1+c-r-a)}{(1+c+a)(1+c-r)}$$

If  $\partial \rho / \partial a = 0$ , we have the condition

$$(1+c)^2 - r(1+c) - 2a(1+c) - a^2 = 0$$

On the other hand, if  $\partial \rho / \partial c = 0$ , we have

$$-(1+c)^2 + 2r(1+c) + 2a(1+c) + a^2 - ar - r^2 = 0$$

Hence, the point at which both derivatives vanish is

$$r(1+c) - ar - r^2 = 0$$

The cases of interest involve  $r > 0$ , so the zero slope point satisfies

$$r + a - c = 1$$

But  $c$  is positive, and  $a + r < 1$ , so this last equation does not have a solution as far as we are concerned. The conclusion here is that the maximum value of  $\rho$  must appear for values of  $a$  and  $c$  along the boundary.

The bounds on  $a$  are  $0 \leq a \leq 1 - r$ , and clearly the maximum is at the value  $a = 1 - r$ , rather than  $a = 0$ . With this value of  $a$ , there is a value of  $c$  which gives the maximum  $\rho$ , namely that positive value which satisfies

$$(1+c)^2 - 2(1+c) - 1 + 3r - r^2 = 0$$

This value is

$$c = [(1-r)(2-r)]^{1/2}$$

and, for  $a = 1 - r$ , the value of  $\rho$  is

$$\rho = \frac{(1-r)c}{(2-r+c)(1-r+c)}$$

*c. The standard cycle.* The common radar-radiometer cycle now in use uses the parameter values

$$a = b = c = r = \frac{1}{2}$$



This gives the value  $\rho_0$  for the "standard" cycle

$$\rho_0 = \frac{1}{8}$$

**d. Comparison of the optimum and standard cycles.**

Let us define  $\psi = \rho/\rho_0$  as the cycle enhancement ratio, or

$$\psi = 8\rho$$

The maximum enhancement occurs for  $r = 0$ , and this is

$$\begin{aligned} \psi_{max} &= 8 \cdot \frac{2^{1/2}}{[2 + (2)^{1/2}][1 + (2)^{1/2}]} \\ &= 1.4 \text{ db} \end{aligned}$$

The value of  $c$  for which  $\psi = 1$  is

$$c = \frac{1}{2} (5 - 6r)$$

or a value of  $r$ , since  $c$  is a function of  $r$ , of

$$\begin{aligned} r &= \frac{3}{4} - \left( \frac{9}{16} - \frac{17}{32} \right)^{1/2} \\ &= \frac{1}{4} \left( 3 - \frac{2^{1/2}}{2} \right) \\ &= 0.574 \end{aligned}$$

**e. Conclusion.** The optimum cycle parameters  $a$ ,  $b$ ,  $c$ , and  $r$  are related as follows:

$$a = 1 - r$$

$$b = 0$$

$$c = [(1 - r)(2 - r)]^{1/2}$$

The cycle is shown in Fig. 13. For any value of  $r$  set by the receiver, the enhancement of the optimum over the standard cycle is

$$\psi =$$

$$\frac{8(1 - r)[(1 - r)(2 - r)]^{1/2}}{\{2 - r + [(1 - r)(2 - r)]^{1/2}\} \{ (1 - 4) + [(1 - r)(2 - r)]^{1/2} \}}$$

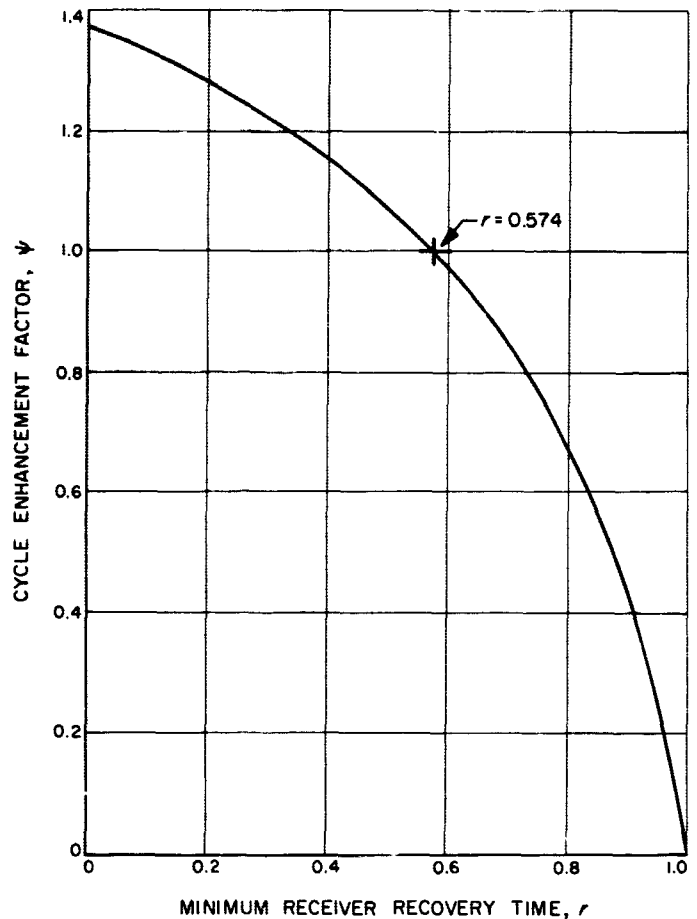


Fig. 14. Enhancement factor of optimized cycle, as a function of receiver recovery time

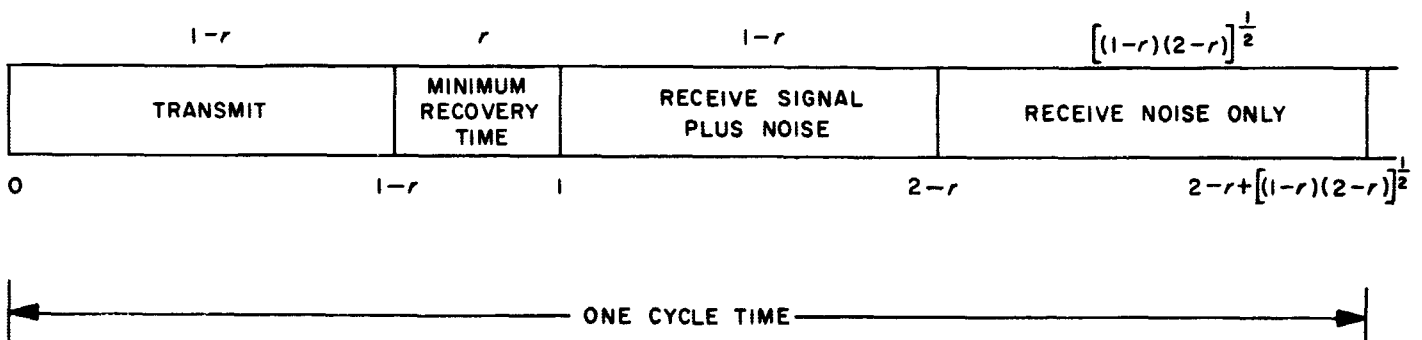


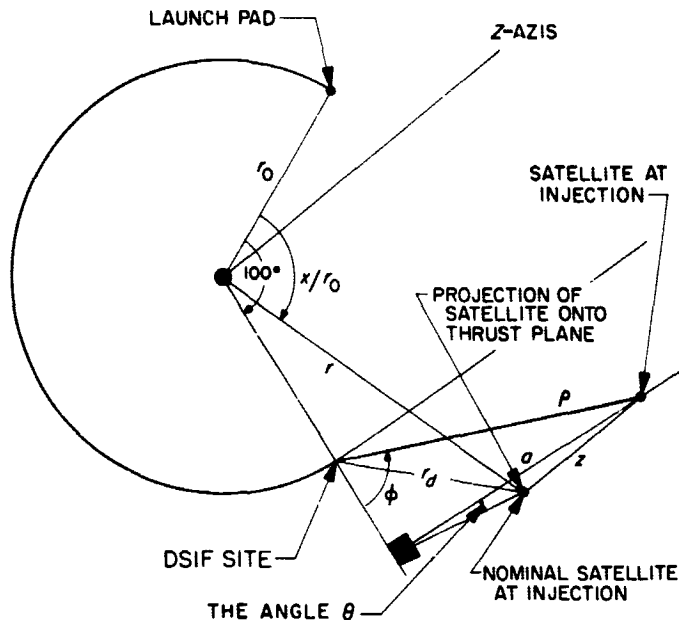
Fig. 13. Optimum radar-radiometric cycle

This function, shown in Fig. 14, is the ratio of the variance of the standard-cycle-estimator to the variance of the optimum-cycle-estimator. The figure shows 1.4-db improvement in the estimate at  $r = 0$  and a break-even at  $r = 0.574$ .

## 6. Conditional Frequency Uncertainty in Spacecraft Acquisition, E. A. Yerman

**a. Summary.** Assuming that the transmission frequency of a signal from a spacecraft is known, uncertainties in the spacecraft's velocity and position make it necessary to search in both frequency and angle in order to acquire the signal. This frequency uncertainty is proportional to the uncertainty in the radial component of spacecraft velocity due to the doppler effect. The purpose of this study has been to determine the effect on the uncertainty in this velocity and on its nominal value when the antenna look angles are used to condition the radial component of velocity. The results indicate that the length of the frequency uncertainty interval, at injection, can be reduced by a factor of about two for a representative trajectory. Furthermore, the conditional mean of the radial velocity can be expressed as a linear combination of the values of the look angles. The conditional means of the remaining three unconditional variables are also linear combinations of the values of the look angles.

***b. Introduction.*** Consider the model for obtaining the standard error in the radial component of spacecraft velocity at injection by transforming the assumed multinormal error distribution in the geocentric coordinates. These coordinates are  $\delta x$ ,  $\delta r$ ,  $\delta v$ ,  $\delta \tau$ ,  $\delta z$ ,  $\delta \dot{z}$ , where  $x$  is the great circle distance traversed in the thrust plane,  $r$  is the radial distance in the thrust plane,  $v$  is the magnitude of the thrust plane velocity,  $\tau$  is the angle between the thrust plane velocity vector and the local horizon,  $z$  is crossrange distance measured in the appropriate direction, and  $\dot{z}$  is crossrange velocity (Ref. 28). The error distribution is transformed to the corresponding multinormal error distribution in station-centered spherical coordinates. This method is an alternate to the method of Ref. 28, and gives additional information. The advantage of using spherical coordinates is that the resultant shift in the expected value of  $\hat{\rho}$ , due to conditioning, can be determined as well as the decrease in standard deviation of  $\hat{\rho}$ . In this new coordinate system,  $\rho$  is the radial distance of the spacecraft while  $\theta, \phi$  are the look angles (Fig. 15). The remaining three coordinates are the components of spacecraft velocity along the unit orthogonal vectors  $e_\rho$ ,  $e_\theta$ ,  $e_\phi$ , respectively. Other parameters of interest in Fig. 15 are the distance from the



**Fig. 15. Station centered spherical coordinates ( $\rho, \theta, \phi$ )**

tracking station to the projection of the spacecraft onto the nominal thrust plane,  $\tau_d$  (the thrust plane of an orbit is the plane in which the orbital motion takes place); the radius of the Earth,  $r_0$ ; the distance from the center of the Earth to the projection of the spacecraft onto the thrust plane,  $r$ ; the perpendicular distance from the spacecraft's position to the thrust plane,  $z$ ; the great circle distance traversed in the thrust plane,  $x$ ; and, finally, the distance from the spacecraft to the receiving antenna site zenith in the plane orthogonal to the zenith that passes through the spacecraft's position and its projection,  $a$ . By virtue of this transformation, it will then be possible to obtain the resultant *conditional* uncertainty in frequency (the conditional standard deviation of  $\delta\dot{\rho}$ ), as well as the *conditional* mean frequency, directly from the conditional multinormal error distribution in the coordinates  $\delta\rho$ ,  $\delta\dot{\rho}$ ,  $\delta(\rho\dot{\theta}\sin\phi)$ ,  $\delta(\rho\dot{\phi})$ , when knowledge of  $\theta$ ,  $\phi$  is used at an arbitrary time. (Once the multinormal error distribution in station-centered spherical coordinates is determined at injection, then the multinormal error distributions corresponding to arbitrary post-injection times can be obtained by mapping ahead in spherical coordinates by means of linear transformations.)

**c. Determination of transformation.** We now proceed to the determination of the transformation matrix,  $D$ , by which an error vector  $\delta Z^T = (\delta x, \delta r, \delta v, \delta \tau, \delta z, \delta \dot{z})$  can be transformed into an error vector  $\delta S^T = (\delta \rho, \delta \theta, \delta \phi, \delta \dot{\rho}, \delta (\rho \dot{\theta} \sin \phi), \delta (\rho \dot{\phi}))$  at injection, i.e.,  $\delta S = D \delta Z$ .

In Ref. 28, the following transformations were obtained:

$$\rho = \left[ r^2 + r_0^2 - 2rr_0 \cos \left( 100^\circ - \frac{x}{r_0} \right) + z^2 \right]^{1/2};$$

$$\theta = \sin^{-1} \frac{z}{\left[ z^2 + r^2 \sin^2 \left( 100^\circ - \frac{x}{r_0} \right) \right]^{1/2}};$$

$$\phi = \sin^{-1} \frac{\left[ z^2 + r^2 \sin^2 \left( 100^\circ - \frac{x}{r_0} \right) \right]^{1/2}}{\rho};$$

$$\dot{\rho} = \frac{\left[ r\dot{r} - r_0\dot{r} \cos \left( 100^\circ - \frac{x}{r_0} \right) - \dot{r}x \sin \left( 100^\circ - \frac{x}{r_0} \right) + z\dot{z} \right]}{\rho};$$

$$\dot{\theta} = \frac{1}{\left[ r \sin \left( 100^\circ - \frac{x}{r_0} \right) \right]} \left\{ \dot{z} - \frac{z \left[ z\dot{z} + r\dot{r} \sin^2 \left( 100^\circ - \frac{x}{r_0} \right) - \frac{r^2\dot{x}}{2r_0} \sin 2 \left( 100^\circ - \frac{x}{r_0} \right) \right]}{z^2 + r^2 \sin^2 \left( 100^\circ - \frac{x}{r_0} \right)} \right\};$$

$$\dot{\phi} = \frac{1}{\left[ r \cos \left( 100^\circ - \frac{x}{r_0} \right) - r_0 \right]} \left\{ \frac{z\dot{z} + r\dot{r} \sin^2 \left( 100^\circ - \frac{x}{r_0} \right) - \frac{r^2\dot{x}}{2r_0} \sin 2 \left( 100^\circ - \frac{x}{r_0} \right)}{\left[ z^2 + r^2 \sin^2 \left( 100^\circ - \frac{x}{r_0} \right) \right]^{1/2}} - \frac{\left[ z^2 + r^2 \sin^2 \left( 100^\circ - \frac{x}{r_0} \right) \right]^{1/2} \left[ z\dot{z} + r\dot{r} - r_0\dot{r} \cos \left( 100^\circ - \frac{x}{r_0} \right) - r\dot{x} \sin \left( 100^\circ - \frac{x}{r_0} \right) \right]}{\rho^2} \right\}$$

Since the time derivations  $\dot{\rho}, \dot{\theta}, \dot{\phi}$  are functions of  $x, r, \dot{x}, \dot{r}, z, \dot{z}$ , and the Z coordinate system uses the variables  $x, r, v, \tau, z, \dot{z}$ , it is necessary to find  $\dot{x}$  and  $\dot{r}$  as functions of  $v$  and  $\tau$ . By Fig. 16, it is seen that

$$\begin{aligned} \dot{r} &= v \sin \tau, \\ \dot{x} &= \frac{r_0 v \cos \tau}{r}. \end{aligned}$$

Hence,

$$\dot{\rho} = \frac{(v \sin \tau) \left[ r - r_0 \cos \left( 100^\circ - \frac{x}{r_0} \right) \right] - r_0 v \cos \tau \sin \left( 100^\circ - \frac{x}{r_0} \right) + z\dot{z}}{\rho};$$

$$\dot{\theta} = \frac{1}{\left[ r \sin \left( 100^\circ - \frac{x}{r_0} \right) \right]} \left\{ \dot{z} - \frac{z \left[ z\dot{z} + r v \sin \tau \sin^2 \left( 100^\circ - \frac{x}{r_0} \right) - \frac{1}{2} r v \cos \tau \sin 2 \left( 100^\circ - \frac{x}{r_0} \right) \right]}{z^2 + r^2 \sin^2 \left( 100^\circ - \frac{x}{r_0} \right)} \right\};$$

$$\dot{\phi} = \frac{1}{\left[ r \cos \left( 100^\circ - \frac{x}{r_0} \right) - r_0 \right]} \left\{ \frac{z\dot{z} + rv \sin \tau \sin^2 \left( 100^\circ - \frac{x}{r_0} \right) - \frac{1}{2} rv \cos \tau \sin 2 \left( 100^\circ - \frac{x}{r_0} \right)}{\left[ z^2 + r^2 \sin^2 \left( 100^\circ - \frac{x}{r_0} \right) \right]^{\frac{1}{2}}} - \frac{\left[ z\dot{z} + rv \sin \tau - r_0 v \sin \tau \cos \left( 100^\circ - \frac{x}{r_0} \right) - r_0 v \cos \tau \sin \left( 100^\circ - \frac{x}{r_0} \right) \right] \left[ z^2 + r^2 \sin^2 \left( 100^\circ - \frac{x}{r_0} \right) \right]^{\frac{1}{2}}}{\rho^2} \right\}.$$

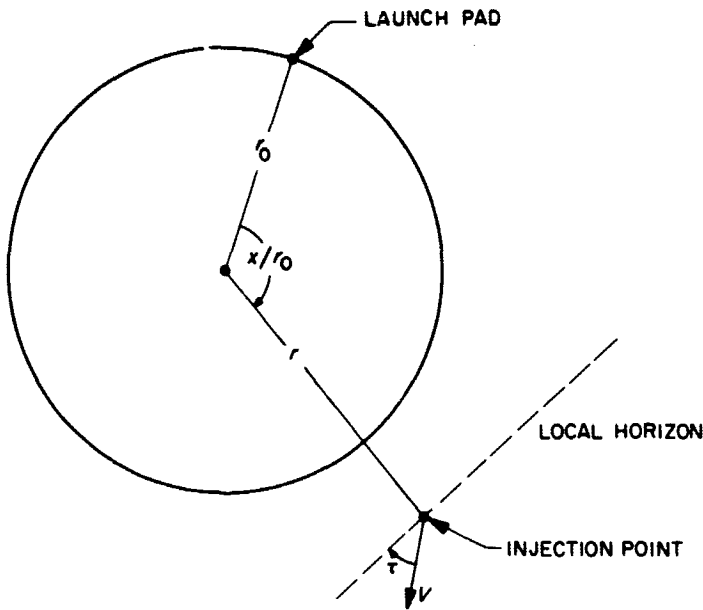


Fig. 16. The quantities  $V$  and  $\tau$  in the thrust plane

*d. The new covariance matrix.* The transformation matrix,  $D$ , which is a linear approximation to the exact transformation for small displacements, can now be computed and is given in Fig. 17 for the example trajectory considered in Ref. 28. If the covariance matrix of the  $Z$ -error coordinates at injection is  $V$ , then the covariance matrix of  $S$ -error coordinates at that time can be readily shown to be

$$\bar{V} = DVD^T.$$

This matrix is given in Fig. 18 for the representative case considered. The variance of  $\delta \dot{\rho}$  is seen to be  $1.23 \times 10^{-3} \text{ km}^2/\text{sec}^2$ .

*e. The covariance matrix and the regression line.* The  $4 \times 4$  covariance matrix  $B$  for the random variables  $\delta \rho$ ,  $\delta \dot{\rho}$ ,  $\delta (\rho \theta \sin \phi)$ ,  $\delta (\rho \dot{\phi})$  is given in Fig. 19 for the pre-injection trajectory in question. It is obtained from  $\bar{V}$  by assuming that  $\theta = \theta_0$  and  $\phi = \phi_0$  at injection. This new covariance matrix defines a conditional multinormal distribution in the variables  $\rho$ ,  $\dot{\rho}$ ,  $\rho \theta \sin \phi$ ,  $\rho \dot{\phi}$ . The variance

of  $\delta \dot{\rho}$  is seen to be  $3.80 \times 10^{-4} \text{ km}^2/\text{sec}^2$ . This means that the reduction, due to conditioning, in the square root of the variance of  $\delta \dot{\rho}$  is by a factor of

$$\frac{35.14 \text{ m/sec}}{19.51 \text{ m/sec}} = 1.8.$$

This ratio is in agreement with the results obtained by a somewhat different method considered in Ref. 28. In addition, the numerical values of velocity are very near those previously obtained (a decimal point was misplaced in the above-cited reference).

An explanation will now be given of how the conditional covariance matrix  $B$  and the regression lines are obtained. The multivariate normal distribution has a density given by

$$f(X_1, X_2, \dots, X_k) = \left( \frac{1}{2\pi} \right)^{k/2} (|\sigma^{ij}|)^{-1/2} \exp \left[ -\frac{1}{2} \sum_{i,j=1}^k \sigma^{ij} (X_i - \xi_i) (X_j - \xi_j) \right],$$

where the matrix  $(\sigma^{ij})$  is positive definite, the covariance matrix of the distribution is  $(\sigma_{ij}) = (\sigma^{ij})^{-1}$ , and the  $\xi_i$  represent the means of the random variables  $X_i$ . The conditional density function

$$f(X_1, X_2, \dots, X_r \mid X_{r+1}, \dots, X_k)$$

defines a multivariate normal distribution in the variables  $X_a$  ( $a = 1, 2, \dots, r$ ) with mean values given by  $\xi_a + m_a$  where

$$m_a = - \sum_{b=1}^r \sum_{p=r+1}^k \bar{\sigma}_{ab} \sigma^{bp} (X_p - \xi_p),$$

and the covariance matrix of the conditional distribution is equal to  $(\bar{\sigma}_{ab}) = (\sigma^{ab})^{-1}$  (Ref. 29, Chapt. 9). Hence, the mean value of each unconditional variate,  $X_a$ , is a linear

$$\begin{pmatrix}
 \frac{\partial \rho}{\partial x} = -0.9827 & \frac{\partial \rho}{\partial r} = 0.3123 & \frac{\partial \rho}{\partial v} = 0 & \frac{\partial \rho}{\partial \tau} = 0 & \frac{\partial \rho}{\partial z} = 0 & \frac{\partial \rho}{\partial \dot{z}} = 0 \\
 \frac{\partial \theta}{\partial x} = 0 & \frac{\partial \theta}{\partial r} = 0 & \frac{\partial \theta}{\partial v} = 0 & \frac{\partial \theta}{\partial \tau} = 0 & \frac{\partial \theta}{\partial z} = 0.0003 \text{ km}^{-1} & \frac{\partial \theta}{\partial \dot{z}} = 0 \\
 \frac{\partial \phi}{\partial x} = -0.0001 \text{ km}^{-1} & \frac{\partial \phi}{\partial r} = 0.0003 \text{ km}^{-1} & \frac{\partial \phi}{\partial v} = 0 & \frac{\partial \phi}{\partial \tau} = 0 & \frac{\partial \phi}{\partial z} = 0 & \frac{\partial \phi}{\partial \dot{z}} = 0 \\
 \frac{\partial \dot{\rho}}{\partial x} = -0.0002 \text{ sec}^{-1} & \frac{\partial \dot{\rho}}{\partial r} = 0.0011 \text{ sec}^{-1} & \frac{\partial \dot{\rho}}{\partial v} = -0.9441 & \frac{\partial \dot{\rho}}{\partial \tau} = 3.729 \text{ km/sec} & \frac{\partial \dot{\rho}}{\partial z} = 0 & \frac{\partial \dot{\rho}}{\partial \dot{z}} = 0 \\
 \frac{\partial (\rho \dot{\theta} \sin \phi)}{\partial x} = 0 & \frac{\partial (\rho \dot{\theta} \sin \phi)}{\partial r} = 0 & \frac{\partial (\rho \dot{\theta} \sin \phi)}{\partial v} = 0 & \frac{\partial (\rho \dot{\theta} \sin \phi)}{\partial \tau} = 0 & \frac{\partial (\rho \dot{\theta} \sin \phi)}{\partial z} = 0.0029 \text{ sec}^{-1} & \frac{\partial (\rho \dot{\theta} \sin \phi)}{\partial \dot{z}} = 0.9999 \text{ sec}^{-1} \\
 \frac{\partial (\rho \dot{\phi})}{\partial x} = 0.4947 \text{ sec}^{-1} & \frac{\partial (\rho \dot{\phi})}{\partial r} = 0.0009 \text{ sec}^{-1} & \frac{\partial (\rho \dot{\phi})}{\partial v} = -0.3503 & \frac{\partial (\rho \dot{\phi})}{\partial \tau} = -10.285 \text{ km/sec} & \frac{\partial (\rho \dot{\phi})}{\partial z} = 0 & \frac{\partial (\rho \dot{\phi})}{\partial \dot{z}} = 0
 \end{pmatrix}$$

Fig. 17. The matrix  $D$ 

$$\begin{pmatrix}
 352 \text{ km}^2 & 0 & -2.31 \times 10^{-2} \text{ km} & 0.633 \text{ km}^2/\text{sec} & 0 & -148 \text{ km}^2/\text{sec} \\
 & 9.46 \times 10^{-6} & 0 & 0 & 1.14 \times 10^{-4} \text{ km/sec} & 0 \\
 & & 3.86 \times 10^{-6} & -5.60 \times 10^{-3} \text{ km/sec} & 0 & 8.60 \times 10^{-3} \text{ km/sec} \\
 & & & 1.23 \times 10^{-3} \text{ km}^2/\text{sec}^2 & 0 & -0.259 \text{ km}^2/\text{sec}^2 \\
 \text{Symmetric} & & & & 1.66 \times 10^{-3} \text{ km}^2/\text{sec}^2 & 0 \\
 & & & & & 62.7 \text{ km}^2/\text{sec}^2
 \end{pmatrix}$$

Fig. 18. The covariance matrix  $\bar{V}$ 

$$\begin{pmatrix}
 192 \text{ km}^2 & 0.268 \text{ km}^2/\text{sec} & 0 & -86.7 \text{ km}^2/\text{sec} \\
 & 3.80 \times 10^{-4} \text{ km}^2/\text{sec}^2 & 0 & -0.121 \text{ km}^2/\text{sec}^2 \\
 & & 2.79 \times 10^{-4} \text{ km}^2/\text{sec}^2 & 0 \\
 \text{Symmetric} & & & 39.1 \text{ km}^2/\text{sec}^2
 \end{pmatrix}$$

Fig. 19. The conditional covariance matrix  $B$ 

function of the fixed variates,  $X_p$ , and defines the regression lines. Using this regression, the receiver changes the center frequency of the frequency search. Further details are given in Sect. XIX-B-7 of this volume.

*f. Example calculation.* We consider the general expression for the conditional expected value of  $\dot{\rho}$  where the unconditional expected values of  $\theta$ ,  $\phi$ , and  $\dot{\rho}$  are denoted by  $\theta_0$ ,  $\phi_0$ , and  $\dot{\rho}_0$ . From the results of Part e of this article, the conditional expected value of  $\dot{\rho}$  is  $\dot{\rho}_0 + m_4$  where

$$m_4 = - \sum_{b=1,4,5,6} \sum_{p=2,3} \bar{\sigma}_{4b} \sigma^{bp} (X_p - \xi_p),$$

$$m_4 = - \sum_{b=1,4,5,6} \bar{\sigma}_{4b} [\sigma^{b2} (\theta - \theta_0) + \sigma^{b3} (\phi - \phi_0)],$$

$$m_4 = - (\theta - \theta_0) \sum_{b=1,4,5,6} \bar{\sigma}_{4b} \sigma^{b2} - (\phi - \phi_0) \sum_{b=1,4,5,6} \bar{\sigma}_{4b} \sigma^{b3}.$$

Substituting in the required numerical values  $-(\sigma^{ij}) = \bar{V}^{-1}$  and  $(\bar{\sigma}_{ab}) = B$ ,

$$\begin{aligned}
 m_4 = & - (\theta - \theta_0) \{ (2.6839807 \times 10^{-1} \text{ km}^2/\text{sec}) (0 \text{ km}^{-1}) \\
 & + (3.8047093 \times 10^{-4} \text{ km}^2/\text{sec}^2) (0 \text{ sec/km}) \\
 & + (0 \text{ km}^2/\text{sec}^2) (4.3215158 \times 10^4 \text{ sec/km}) \\
 & - (1.2098310 \times 10^{-1} \text{ km}^2/\text{sec}^2) (0 \text{ sec/km}) \}
 \end{aligned}$$

$$\begin{aligned}
 & - (\phi - \phi_0) \{ (2.6839807 \times 10^{-1} \text{ km}^2/\text{sec}) \\
 & \times (1.8265267 \times 10^7 \text{ km}^{-1}) \\
 & + (3.8047093 \times 10^{-4} \text{ km}^2/\text{sec}^2) \\
 & \times (6.4017175 \times 10^7 \text{ sec/km}) \\
 & + (0 \text{ km}^2/\text{sec}^2) (0 \text{ sec/km}) \\
 & - (1.2098310 \times 10^{-1} \text{ km}^2/\text{sec}^2) \\
 & \times (4.0722258 \times 10^7 \text{ sec/km}) \}
 \end{aligned}$$

$$m_4 = - (\theta - \theta_0) \{0\} - (\phi - \phi_0) \{0.01407 \times 10^3 \text{ km/sec}\}$$

$$m_4 = - (\phi - \phi_0) \{14.07 \text{ km/sec}\}.$$

This result indicates that the conditional expected value of  $\dot{\rho}$ , at injection, is independent of the conditioning value of  $\theta$ . This fact is predictable, since variations in  $\theta_0 = 0$  have second-order effect on the value of  $\dot{\rho}$  and the analysis has been, rightly, to first order. Hence, it is unnecessary to determine the value of the angle  $\theta$  when calculating the new expected value of  $\dot{\rho}$  at injection times because the resultant shift due to conditioning the look angles  $\theta$ ,  $\phi$  is a function of the deviation  $(\phi - \phi_0)$  alone.

If it is assumed that the tracking equipment is looking at the nominal injection point at the time of injection, then  $\phi = \phi_0$  so that  $m_4 = 0$  and the conditional expected value of  $\dot{\rho}$  is  $\dot{\rho}_0$ . In those cases when other positions are to be considered at the time of injection, the value of  $\phi$  can be obtained from the antenna coordinate converter and the new expected value of  $\dot{\rho}$  computed. Once this has been done, the receiver frequency will be offset from the conditioned mean frequency by a suitable amount, and a frequency search initiated with fixed look angles. (See Sect. XIX-B-7 for further analysis.)

Finally, let us consider the frequency search interval for the unconditioned and the conditioned case. Assume a two-way transponder S-band system with carrier at 2388 Mc. Let the  $\phi$ -offset be one beamwidth of an 85-ft disk antenna, or 0.1 deg. This offset is 0.0017 in radians. The change in  $\dot{\rho}$  due to conditioning is then  $(0.0017) \times 14$  km/sec, or about 24 m/sec. Since 1 m/sec at two-way S-band corresponds to about 16 cps, the offset is about 382 cycles.

More significant is the *width* of the frequency interval. The unconditioned standard deviation of  $\dot{\rho}$  is 35 m/sec. Assuming a  $5\sigma$  search, as in Section XIX-B-7, the width of the frequency interval is 2786 cps. Using conditioning, the width is only  $1/1.8 \times 2786$  cps = 1548 cps. Thus, not only does conditioning reduce search time, but it also can allow for construction of narrower band receivers, and, hence, improved signal-to-noise ratio. Thus, incorporated into receiver design, the conditioning will result in still further savings in search time.

## 7. Frequency Uncertainty as a Function of Offset from Nominal, E. A. Yerman

**a. Introduction.** The standard procedure now used to acquire a signal from a spacecraft requires search in angle and then a search in frequency. The frequency search proceeds by sweeping the receiver VCO from the low end across a predetermined frequency uncertainty band. The sweep rate is determined by receiver parameters as the maximum constant rate at which lock can be achieved with certainty when the receiver frequency reaches the actual signal frequency. Thus, the search time is a proportionality constant times the frequency difference between the actual received frequency and the lower end of the frequency uncertainty band. In Sect. XIX-B-6, p. 215, and Ref. 30, it was shown that the uncertainty in signal frequency is reduced and the nominal value of signal frequency shifted when knowledge of the look angles at a given time is used. There was

an unconditional normal distribution whose mean may be considered zero without loss of generality, and an unconditional standard deviation  $\sigma_u$ . Also present was the conditional normal distribution of signal frequency, conditioned by the look angles, with mean  $\mu$  and standard deviation  $\sigma_c$ , with  $\sigma_c < \sigma_u$ . Given a required a priori probability  $\alpha$  of acquiring the signal frequency, one determines the parameter  $k > 0$  such that a symmetric search interval of the form  $(-k\sigma_u, +k\sigma_u)$  is required in the unconditioned case. A symmetric search interval of the form  $(\mu - k\sigma_c, \mu + k\sigma_c)$  is required for a priori probability  $\alpha$  in the conditioned case. The purpose of this article is to compare the expected values and variances of the time to find the signal frequency, when using the a priori unconditioned distribution, as a function of the conditional mean  $\mu$  which is not being used. The analysis also gives the probability of picking up the signal frequency as a function of the offset in the nominal value,  $\mu$ , when using the a priori distribution instead of the conditional distribution. The analysis makes use of certain properties of the truncated normal distribution.

**b. Moments of the truncated normal distribution.** We generalize the results given on pp. 247-248 of Ref. 31. Consider a random variable  $Y$  that possesses a normal distribution with mean  $m$  and standard deviation  $\sigma$ . This implies that the random variable  $Z = Y - m/\sigma$  has the distribution function

$$\Phi(z) = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^z e^{-t^2/2} dt, \quad (-\infty < z < +\infty).$$

Now, let  $W$  denote a random variable with truncated normal distribution function

$$F(u) = \frac{\Phi\left(\frac{w-m}{\sigma}\right) - \Phi\left(\frac{w_0-m}{\sigma}\right)}{\Phi\left(\frac{w_1-m}{\sigma}\right) - \Phi\left(\frac{w_0-m}{\sigma}\right)}, \quad (w_0 < w < w_1)$$

and let

$$\alpha_1 = E(W^1) = \frac{\int_{w_0}^{w_1} w^1 d\Phi\left(\frac{w-m}{\sigma}\right)}{\Phi\left(\frac{w_1-m}{\sigma}\right) - \Phi\left(\frac{w_0-m}{\sigma}\right)}$$

It follows that

$$\alpha_1 = m + \sigma(\lambda_1 - \lambda_2)$$

and

$$\alpha_z = m^2 + \sigma\lambda_1(w_0 + m) - \sigma\lambda_2(w_1 + m) + \sigma^2$$

where

$$\lambda_1 = \frac{\Phi'\left(\frac{w_0 - m}{\sigma}\right)}{\Phi\left(\frac{w_1 - m}{\sigma}\right) - \Phi\left(\frac{w_0 - m}{\sigma}\right)}$$

and

$$\lambda_2 = \frac{\Phi'\left(\frac{w_1 - m}{\sigma}\right)}{\Phi\left(\frac{w_1 - m}{\sigma}\right) - \Phi\left(\frac{w_0 - m}{\sigma}\right)}$$

*c. Expected values and variances of search times.* The results of Part *b* of this article are used throughout this part. Let us assume the conditioned frequency random variable  $X$  to be normally distributed with mean  $\mu$  and standard deviation  $\sigma_c$ . We now proceed to the determination of the expected values and variances of the time required to find the signal frequency when searching at the constant sweep rate the intervals  $(-k\sigma_u, +k\sigma_u)$  and  $(\mu - k\sigma_c, \mu + k\sigma_c)$  assuming that the spacecraft is actually present at the given look angles (Fig. 20). However, this problem is equivalent to finding the expected value

and variance of the random variable  $X + k\sigma_u$  with respect to the conditional distribution function

$$G(x) = \frac{\Phi\left(\frac{x - \mu}{\sigma_c}\right) - \Phi\left(\frac{-k\sigma_u - \mu}{\sigma_c}\right)}{\Phi\left(\frac{+k\sigma_u - \mu}{\sigma_c}\right) - \Phi\left(\frac{-k\sigma_u - \mu}{\sigma_c}\right)}, \quad (-k\sigma_u < x < +k\sigma_u)$$

and computing the expected value and variance of the random variable  $X - \mu + k\sigma_c$  whose distribution function is given by

$$H(x) = \frac{\Phi\left(\frac{x - \mu}{\sigma_c}\right) - \Phi(-k)}{\Phi(k) - \Phi(-k)}, \quad (\mu - k\sigma_c < x < \mu + k\sigma_c).$$

Define

$$\lambda_1 = \frac{\Phi'\left(\frac{-k\sigma_u - \mu}{\sigma_c}\right)}{\Phi\left(\frac{k\sigma_u - \mu}{\sigma_c}\right) - \Phi\left(\frac{-k\sigma_u - \mu}{\sigma_c}\right)},$$

$$\lambda_2 = \frac{\Phi'\left(\frac{+k\sigma_u - \mu}{\sigma_c}\right)}{\Phi\left(\frac{k\sigma_u - \mu}{\sigma_c}\right) - \Phi\left(\frac{-k\sigma_u - \mu}{\sigma_c}\right)},$$

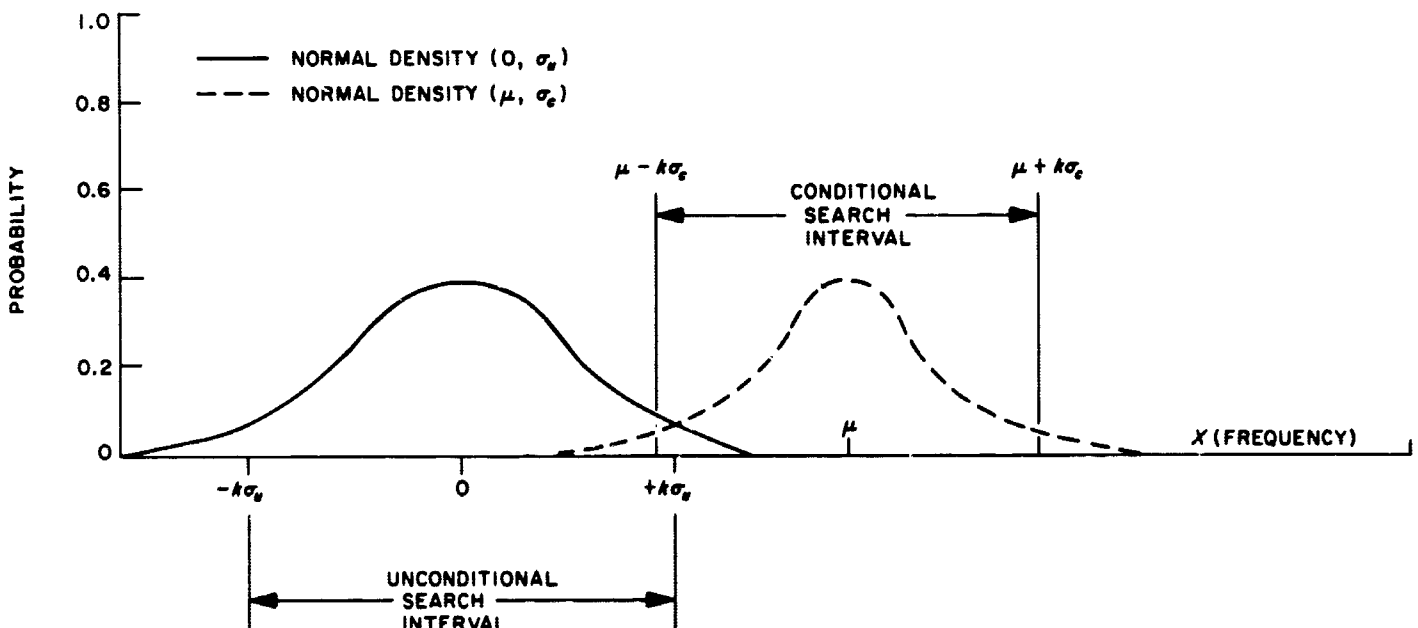


Fig. 20. The intervals to be searched

$$\gamma_1 = \frac{\Phi'(-k)}{\Phi(k) - \Phi(-k)},$$

and

$$\gamma_2 = \frac{\Phi'(k)}{\Phi(k) - \Phi(-k)}.$$

The following results are obtained:  $\xi$  denotes the event that the satellite is present at the given look angles, and  $P_\xi(\mu)$  is the probability, given  $\xi$ , of finding the signal when the offset is  $\mu$ , and the a priori distribution is used. Note that the probability of finding the signal when the conditional distribution of mean  $\mu$  is used is always exactly  $\alpha$  for any  $\mu$ . Also,  $P_\xi(\mu)$  averaged with respect to  $\mu$  is exactly  $\alpha$  by definition.

$$\begin{aligned} E(X + k\sigma_u | \xi) &= \int_{-k\sigma_u}^{+k\sigma_u} (x + k\sigma_u) dG(x) \\ &= \int_{-k\sigma_u}^{+k\sigma_u} x dG(x) + k\sigma_u, \end{aligned}$$

$$E(X + k\sigma_u | \xi) = \mu + k\sigma_u + \sigma_c(\lambda_1 - \lambda_2),$$

$$\begin{aligned} V(X + k\sigma_u | \xi) &= V(X | \xi) = \int_{-k\sigma_u}^{+k\sigma_u} x^2 dG(x) \\ &\quad - [\mu + \sigma_c(\lambda_1 - \lambda_2)]^2, \end{aligned}$$

$$\begin{aligned} V(X + k\sigma_u | \xi) &= \sigma_c^2 [1 - (\lambda_1 - \lambda_2)^2] \\ &\quad + \sigma_c [-\mu(\lambda_1 - \lambda_2) - k\sigma_u(\lambda_1 - \lambda_2)], \end{aligned}$$

$$\begin{aligned} E(X - \mu + k\sigma_c | \xi) &= \int_{\mu - k\sigma_c}^{\mu + k\sigma_c} (x - \mu + k\sigma_c) dH(x) \\ &= \int_{\mu - k\sigma_c}^{\mu + k\sigma_c} x dH(x) - \mu + k\sigma_c, \end{aligned}$$

$$E(X - \mu + k\sigma_c | \xi) = \mu + \sigma_c(\gamma_1 - \gamma_2) - \mu + k\sigma_c = k\sigma_c,$$

$$V(X - \mu + k\sigma_c | \xi) = V(X) = \int_{\mu - k\sigma_c}^{\mu + k\sigma_c} x^2 dH(x) - \mu^2,$$

$$V(X - \mu + k\sigma_c | \xi) = \sigma_c^2 [1 - k(\gamma_1 + \gamma_2)].$$

In order to compare these expected search times and the related variances, we define the ratios

$$E(\mu) = \frac{E(X + k\sigma_u | \xi)}{E(X - \mu + k\sigma_c | \xi)} = \frac{\mu + k\sigma_u + \sigma_c(\lambda_1 - \lambda_2)}{k\sigma_c}$$

and

$$\begin{aligned} V(\mu) &= \frac{V(X + k\sigma_u | \xi)}{V(X - \mu + k\sigma_c | \xi)} = \\ &= \frac{\sigma_c [1 - (\lambda_1 - \lambda_2)^2] + [-\mu(\lambda_1 - \lambda_2) - k\sigma_u(\lambda_1 - \lambda_2)]}{\sigma_c [1 - k(\gamma_1 + \gamma_2)]}. \end{aligned}$$

$E(\mu)$  is a measure of the loss (or gain) in the expected time to acquire the signal frequency as a function of  $\mu$  if the interval  $(-k\sigma_u, +k\sigma_u)$  is searched rather than the interval  $(\mu - k\sigma_c, \mu + k\sigma_c)$ , when in fact the signal frequency is normally distributed  $(\mu, \sigma_c)$ . Similarly,  $V(\mu)$  is a measure of the relative size of the variances in these search times. In Part d, sample calculations of  $E(\mu)$ ,  $V(\mu)$ , and  $P_\xi(\mu)$  are made for a representative case (Ref. 30 and Sect. XIX-B-6).

*d. Comparison for a representative case.* Assuming that the prescribed probability  $\alpha$  is 0.9999994, the symmetric search intervals of interest are  $(-5\sigma_u, +5\sigma_u)$  and  $(\mu - 5\sigma_c, \mu + 5\sigma_c)$ , i.e.,  $k = 5$ . Let  $\sigma_u/\sigma_c = 2$ , as is typical in Ref. 30 and Sect. XIX-B-6 of this volume. Substituting these values we have

$$\lambda_1 = \frac{\Phi'(-10 - \frac{\mu}{\sigma_c})}{\Phi(10 - \frac{\mu}{\sigma_c}) - \Phi(-10 - \frac{\mu}{\sigma_c})},$$

$$\lambda_2 = \frac{\Phi'(10 - \frac{\mu}{\sigma_c})}{\Phi(10 - \frac{\mu}{\sigma_c}) - \Phi(-10 - \frac{\mu}{\sigma_c})},$$

$$\gamma_1 = \gamma_2 = \frac{\Phi'(5)}{\Phi(5) - \Phi(-5)},$$

$$E(\mu) = \frac{\mu}{5\sigma_c} + 2 + \frac{1}{5}(\lambda_1 - \lambda_2),$$

and

$$\begin{aligned} V(\mu) &= \\ &= (1.000015) \left[ 1 - (\lambda_1 - \lambda_2)^2 - 10(\lambda_1 + \lambda_2) - \frac{\mu}{\sigma_c}(\lambda_1 - \lambda_2) \right]. \end{aligned}$$

$E(\mu)$  and  $V(\mu)$  are plotted in Figs. 21 and 22 for the case under consideration where  $0 \leq \mu < \infty$  since similar results would be obtained if the offset  $\mu$  were negative (pp. 164-168 of Ref. 32; Refs. 33 and 34).  $E(\mu)$  is strictly monotone increasing as  $\mu$  varies from zero to infinity, due to the fact that the average loss in time due



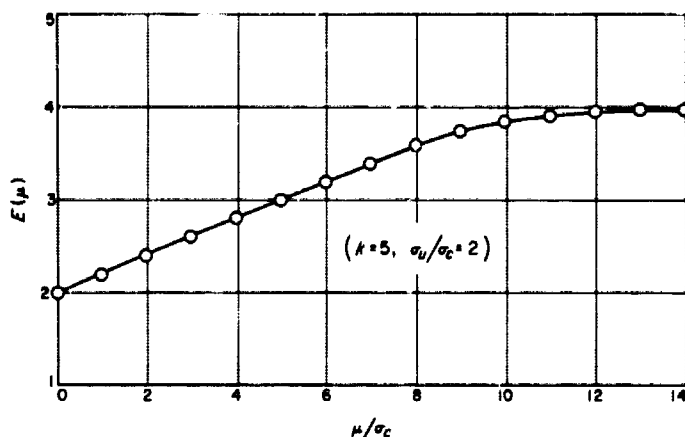


Fig. 21.  $E(\mu)$ , the ratio of expected search times, unconditioned to conditioned

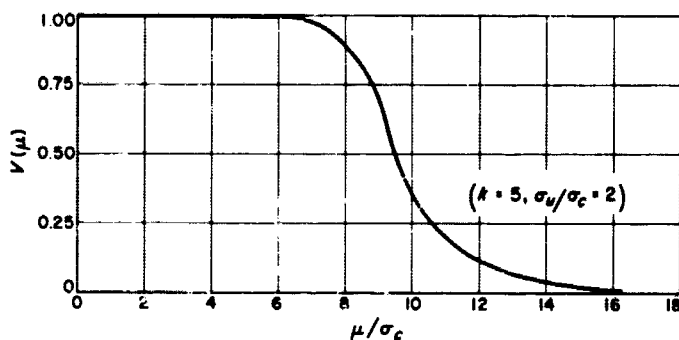


Fig. 22.  $V(\mu)$ , the ratio of the variances in search times

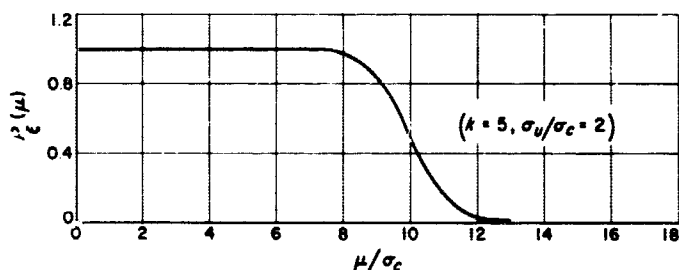


Fig. 23.  $P_\xi(\mu)$ , the probability of finding the signal using the unconditional distribution

to searching  $(-5\sigma_u, +5\sigma_u)$  instead of  $(\mu - 5\sigma_c, \mu + 5\sigma_c)$  under the assumption that the frequency has the conditional normal distribution becomes greater with increasing  $\mu$ . In fact,  $E(\mu) \rightarrow +2\sigma_u/\sigma_c$  as  $\mu \rightarrow +\infty$ . Although  $V(\mu)$  is strictly monotone decreasing with increasing  $\mu$ , it is clear that  $V(\mu)$  is nearly one for  $0 \leq \mu \leq 7\sigma_c$ . But as  $\mu \rightarrow +\infty$   $V(\mu) \rightarrow 0$ . The reason is that as  $\mu$  recedes from 0, the probability of finding the satellite in frequency given  $\xi$ , approaches zero, with all probability concentrated toward the left end.

The probability  $P_\xi(\mu)$  of picking up the true signal frequency when searching the interval  $(-k\sigma_u, +k\sigma_u)$  if the frequency is assumed to have the conditional normal distribution is given, in general, by

$$P_\xi(\mu) = \Phi\left(\frac{k\sigma_u - \mu}{\sigma_c}\right) - \Phi\left(\frac{-k\sigma_u - \mu}{\sigma_c}\right)$$

and, in the example,

$$P_\xi(\mu) = \Phi\left(10 - \frac{\mu}{\sigma_c}\right) - \Phi\left(-10 - \frac{\mu}{\sigma_c}\right)$$

$P_\xi(\mu)$  is very close to one for  $0 \leq \mu \leq 7\sigma_c$  and then drops off very quickly so that  $P_\xi(\mu)$  is effectively zero for large offsets in the conditional nominal value  $\mu$  (Fig. 23).

## C. Information Processing

E. C. Posner and I. Eisenberger

### 1. Extreme-Value Statistics for Estimation of Error Probabilities in Ranger Block III Command Detector, E. C. Posner

**a. Summary.** This article applies the statistical theory of extreme values (Refs. 35 and 36) to the problem of estimation of probability of error in the *Ranger* Command Detector<sup>1</sup> and in other communication systems. The method applies whenever errors are caused by large (one-sided) fluctuations of a random quantity, such as a voltage, for example. The basic idea is to use a short stretch of fluctuation data to estimate how long it would take for a random fluctuation to be so large as to cause an error. The technique is to be used whenever one must identify very low error probabilities and, in addition, the time or expense of taking sufficient data to estimate these probabilities by the law of large numbers is prohibitive. A block diagram of a proposed testing system is included.

**b. Review of Gumbel's theory.** E. J. Gumbel enunciated his theory of extreme values in Refs. 35 and 36. Since the applications to date have been in fields such as civil engineering and actuarial science rather than

<sup>1</sup>John C. Ashlock, "An Introduction to the *Ranger* Command Subsystem," private communication.

in communication theory, a brief review of his theory is in order.

Let  $z_j, 1 \leq j \leq n$ , be  $n$  independent samples from a continuous distribution of exponential type on the left. That is, the density function  $f(z)$  is less than a constant times  $e^{kz}$ , for some  $k > 0$ , as  $z \rightarrow -\infty$ . For  $n$  large, consider the minimum of the  $z_j$ , call it  $z_{\min}$ . Then the distribution of  $x = -z_{\min}$  for  $n$  large is the so-called first asymptotic (cumulative) distribution function

$$F(x) = \exp(-\exp(-\alpha(x-u))),$$

where  $\alpha, u$  are positive parameters (which depend on  $n$ , but  $n$  is fixed; Ref. 36, Chapt. 5).

Now let  $x_i, 1 \leq i \leq N$ , be  $N$  independent samples from this  $F$ . Let  $x_{(m)}, 1 \leq m \leq N$ , denote the  $x_i$  ordered by  $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(N)}$ . Graph  $x_{(m)}$  versus  $y = \log[(N+1)/m]$ . Then the  $N$  observations fall on the "Gumbel" line,  $y = \alpha(x-u)$ . That is,  $x = u + (y/\alpha)$ , if  $x$  is large. For then  $e^{-\alpha(x-u)}$  is small, and  $\exp(-\exp(-\alpha(x-u)))$  is approximately by  $1 - e^{-\alpha(x-u)}$ . Here  $m/(N+1)$  corresponds to  $1 - F(x)$ . And we have  $(1 - F(x)) \approx e^{-\alpha(x-u)}$ , which explains the equation.

In this theory,  $\alpha$  and  $u$  depend on the original unknown  $f(z)$ , but these two parameters need not be determined in advance. The parameters depend on  $f$  at large values of  $z$ , rather than on any central values of  $f$ , as is reasonable for a theory of extreme values.

**c. Need for Gumbel line method.** The Ranger Block III Command Detector uses a frequency shift keyed (FSK) input signal, but is essentially a "signal present" detection system that uses a tuning fork filter for its frequency-sensitive device. Any given bit is called a "1" bit if the envelope of the voltage out of the filter goes above a certain quantizing voltage at the time it is sampled; otherwise the bit is called a "0" bit. The quantizing level used at present in this detector is such that the probability of bit errors when transmitting a "1" and when transmitting a "0" are drastically different. In fact, in practice, never has a "0" been transmitted and observed to be received as a "1"; any error that is made is always a "1" transmitted and received as a "0." Thus, the channel is a so-called binary asymmetric channel. A companion article (Posner, Sect. XIX-A-3, p. 193 of this volume) discusses this channel with relevance to the Ranger Command Detector.

Specifications call for an error probability of less than  $10^{-5}$  at various temperatures and signal-to-noise condi-

tions. It is desired to test the detectors to see if they meet this specification. In the first place, the fork filter should be tested only by transmitting a "1"; the error probabilities obtained will be multiplied by  $1/2$  to take into account the fact that 1's and 0's are transmitted with approximately equal likelihood. Also, although each bit in actual operation lasts 1 sec, the mechanical properties of the fork are such that samples of voltage 0.1 sec apart are independent. The tests will use 10 samples per sec to take advantage of this fact by obtaining ten times as much data.

Now, an event with probability less than  $10^{-5}$  is at issue. The number of test conditions required is such that to test each detector for, say, 250,000 symbols would involve prohibitively much time. Each test would then take about 7 hr, so that weeks would have to be used to get a good estimate of the error probability at all the various test conditions.

The Gumbel line method takes advantage of the physical nature of the low probability event whose probability is to be estimated. Namely, the event is the event that a certain numerical random variable exceed a certain value. Extreme value theory is then applied to give an estimate of the error probability.

**d. Plan of use.** At present, there are available on magnetic tape 5-min recordings of the output of the tuning fork filter in the "signal present" condition, for a number of conditions of temperature and signal-to-noise. This corresponds to  $10 \times 5 \times 60 = 3000$  independent samples. The plan is to take the 3000 samples, the  $z_j$ , and break them up into 30 blocks of 100  $z_j$  each. Thus,  $n = 100$ , which should be large enough to have the asymptotic theory work. The value of  $n$  can be increased if necessary, but this reduces  $N = 3000/n$ . However, even for  $N = 30$  or even less, the Gumbel line gives good agreement with theory in a wide variety of applications, as Ref. 40 indicates.

The  $x_j, 1 \leq j \leq 3000/100 = 30$ , are defined as the  $-z_{\min}$  for the 30 blocks of 100  $z_j$  each. These  $x_j$  are then ordered into  $x_{(m)}, 1 \leq m \leq 30$ , where  $x_{(m+1)} \leq x_{(m)}$ . Then  $x_{(m)}$  is plotted on the vertical scale against  $\log[(N+1)/m]$  on the horizontal scale. A straight line is then fitted to these 30 points.

To estimate the error probability, the value  $z_{crit}$  obtained from the system parameters is the value of  $z$  below which the detector outputs a 0, that is, at which an error occurs, is indicated on the vertical axis, and a line

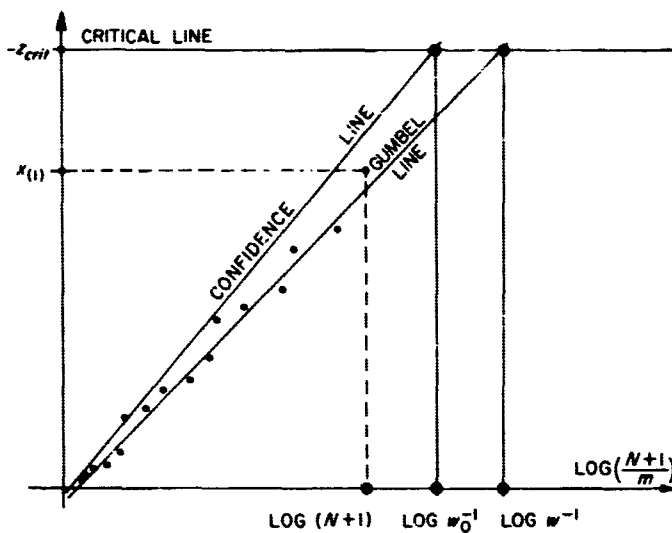


Fig. 24. The Gumbel line

drawn through  $-z_{crit}$  parallel to the horizontal axis (Fig. 24). This  $z_{crit}$  can be less than all the  $z_j$ ,  $1 \leq j \leq 3000$ . That is,  $-z_{crit}$  can exceed all  $x_i$ ,  $1 \leq i \leq 30$ . (In fact, the advantage of the Gumbel method shows up best in such situations.) The Gumbel line is extended, as usually will be necessary, until it intersects the "critical line." The value of the horizontal coordinate for that intersection is determined.

This value yields the estimate  $\hat{e}$  of the error probability as follows. Let the intersection coordinate be  $\log 1/w$ . Then  $w$  is the estimate of the probability that the random variable  $x$  exceed  $-z_{crit}$ .

However, since  $n = 100$  of the  $-z_j$ 's have the  $x_i$  as their maximum,  $w$  is the probability that at least one out of  $n$  independent  $z_j$  is less than  $z_{crit}$ . If  $e$  denotes the error probability in the system, so that  $e/2$  denotes the probability that a 1 is mistaken for a 0, then  $1 - w = (1 - (e/2))^n$ .

For  $e$  small, then, we obtain  $\hat{e} = 2w/n$ , the required estimator.

A functional block diagram of a test system is shown in Fig. 25. The system samples the incoming voltage from the command decoder at 10 times per second and converts this signal to a digital value. These  $z_j$  are read sequentially into a comparator which picks the smallest  $z_j$  of the 30 by comparing each  $z_j$  with the previous minimum  $z_j$ . If the now  $z_j$  is less than the previous minimum, the previous minimum is replaced with a new minimum. Otherwise, the previous minimum stays. After 30 samples, the  $-z_{min} = x_i$  is read into a buffer.

After all thirty of the  $x_i$  are stored in a buffer, the  $x_{(m)}$  are computed by arranging the  $x_i$  in decreasing order in a sorter. An arithmetic unit then finds the line of best fit, and computes the intersection of this line with the critical line. In this way,  $w$ , and ultimately the estimator  $\hat{e}$  of  $e$  is found. The actual Gumbel line data is also outputted for visual checking purposes.

**e. Confidence intervals.** The estimate of the probability obtained above is of little use without an estimate of the reliability to be attached to such estimates. One source of unreliability is the failure of the asymptotic theory to hold exactly for  $n$  samples  $z_j$ . This effect can only be found experimentally. Even assuming the asymptotic theory to hold for 100 samples, as has been found true in other disciplines, we still face the problem of assigning a confidence to the estimator  $\hat{e}$  of  $e$ .

This confidence interval shall be found by assigning a confidence to the estimate of  $w$ , and translating the confidence interval to  $\hat{e}$ . What is wanted is a one-sided confidence interval. That is, given a desired confidence coefficient  $\gamma$  (close to 1), find a  $w_0 > w$  such that the true value of  $w$  is less than  $w_0$  with probability  $\gamma$ . Then

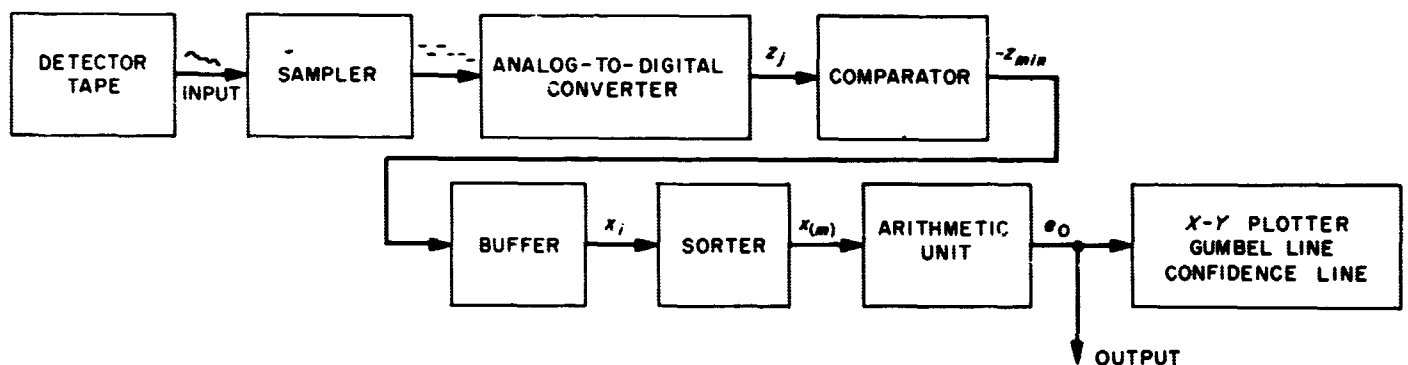


Fig. 25. Gumbel line error test system

the true  $e$  will be less than the corresponding  $e_0$  also with the probability  $\gamma$ ;  $e_0$  is naturally larger than the estimate  $\hat{e}$ . Thus,  $\hat{e}$  might be  $10^{-4}$ ,  $e_0$  might be  $10^{-3}$ , for a given  $\gamma$  such as 0.95. We then say that with probability 0.95, the true  $e$  is less than  $10^{-3}$ .

The way this one-sided confidence region for  $w$  is obtained is as follows. In Fig. 24, a  $\gamma$  confidence line above the Gumbel line is shown. This confidence line is constructed by one of several available techniques so that the true Gumbel line lies below the  $\gamma$  confidence line with probability  $\gamma$ . The intersection of the  $\gamma$  confidence line with the critical line gives the  $w_0$ , and thus the  $e_0$ .

The system of Fig. 25 should then also compute  $e_0$ , or, in fact, should compute  $e_0$  rather than  $\hat{e}$ . The test on the detector would then work as follows. Choose a  $\gamma$ , say 0.99; run the 5-min test and then compute  $e_0$ . If  $e_0$  is greater than  $10^{-5}$ , reject the command detector and return it to the shop for rework. Otherwise, accept the command detector subsystem.

If the possibility of more testing on the detector exists, the above procedure can be turned into a sequential procedure. Thus, if  $e_0$  is in a region of doubt, more tests would be called for. Some detectors would be rejected after just one 5-min run, and some accepted after just one run, but others would require a random number of runs before being accepted or rejected. However, in the initial use of the test system, the possibility of using a sequential test will not be present, since the voltage output records on magnetic tape will have been previously obtained before any tests. Therefore, 5 min of such data are all there will be to work with under each of the various test conditions.

**f. Conclusion.** A preliminary test system has been described above for estimating low error probabilities in the Ranger Block III Command Detector when only short samples of output are present. The system can be used to estimate error probabilities on all sorts of other systems where a 0 or 1 is detected by a voltage crossing a threshold. The system will be incorporated into the existing test system by J. Ashlock of Section 334.

We close with the remark that the previous uses of extreme value theory have not been to estimate error probability. Rather, the method has been used to determine for example how large a dam to build to cope with maximum floods. In the previous applications, the "threshold" was to be set in effect *after* determining the maximum voltage. In those previous applications, "threshold" corresponds to size of dam to be built, and "voltage"

to amount of water in the flood. In the present application, we are given a "dam" and must examine a short record of floods before accepting the dam. Due to this different nature of the use of the theory, there are therefore new areas of mathematical and statistical research to be investigated before the maximum benefit of this theory for error-probability testing can be achieved.

## 2. Statistics of Data Sources, E. C. Posner

**a. Summary.** In a recent issue of Volume III of this *Summary* (Ref. 37), a random pulse generator was described, to be used as a data and noise source in advanced engineering projects. Several output options are available, which yield data sources with different statistics. This note discusses the statistics of some of the options available with the random pulse generator. Such information is needed in order to interpret the results of experiments involving data compression and coding.

**b. Introduction.** In Ref. 38, a random pulse generator was described. The basic element of this machine is a 36-stage maximum-length shift register generator corresponding to the known primitive polynomial  $x^{36} + x^{11} + 1$ . The pseudorandom events that are to be called "pulses" are available in four options:

- (I) A given length  $l$ ,  $1 \leq l \leq 7$ , is selected. Also, a sequence of 0's and 1's of length  $l$ , called a *window*, is chosen. The window is allowed to have "don't-cares." A pulse is defined whenever the chosen window of length  $l$  occurs in the output of the pulse generator. After occurrence of the special sequence, the search begins afresh with *no memory*. Thus, if 11 is the special sequence, then the input 011100 records a pulse after the second 1, but not after the third 1.
- (II) A given number  $l$ ,  $1 \leq l \leq 7$ , of cumulated 1's constitutes a pulse. After a pulse, the counter is reset. Thus, if three 1's are sought, the input sequence 00111\*01000101\*1001 results in pulses before the stars.
- (III) A length  $l$ ,  $1 \leq l \leq 7$  is chosen, and consecutive *disjoint* windows of length  $l$  are considered. A window of length  $l$  is chosen, which is allowed to include "don't-care" positions. Pulses are defined whenever the given window occurs.
- (IV) The same as (III), including "don't-cares," but, instead of looking at non-overlapping windows, the window shifts by only one position each time. Thus, if the window 10 constitutes an event, then

the sequence 0110\*110\* yields pulses at the stars. The difference between (I) and (IV) is that (IV) does allow for memory.

As stated in Ref. 37, the statistics of option (I) are handled by the theory of recurrent events, and Ref. 37 describes them in such fashion. The same comment applies to option (II). Option (III) is even easier to describe, as the binomial distribution applies. It is in option (IV) that new theory must be developed, for the theory of recurrent events does not apply. For example, if a pattern of four consecutive 1's is sought, then, once a run of four 1's has occurred, the probability of a pulse on the very next bit is  $\frac{1}{2}$ . That is, the system does not start over at each pulse, as it did start in option (I). The purpose of this note is therefore to derive the formulas for the mean and variance of the number of pulses  $E_n$  in  $n$  bits under option (IV) with an arbitrary window of length  $l$ , including "don't-care" positions.

**c. Construction of the random variable.** Following Ref. 37, we represent the random variable  $E_n$ , the number of pulses in  $n$  bits, in the following fashion. Define a set of  $n$  random variables  $Y_i$  such that  $Y_i = 1$  or 0 according as a pulse occurs at bit  $i$  or not. Then, since  $E_n$  is the number of pulses in  $n$  bits, the random variable  $E_n$  can be expressed as:

$$E_n = \sum_{i=1}^n Y_i \quad (1)$$

Note that  $Y_1 = Y_2 = \dots = Y_{l-1} = 0$ , where  $l$  is the window length, since no pulse can occur until the  $l$ th bit (more or less by definition and construction of the machine). Since the case  $n$  large is primarily of interest, this end effect could be ignored.

Let us say that the window of length  $l$  has  $l - l^*$  "don't-care" positions, and  $l^*$  conditioned positions. Thus, if  $l = 5$ , the window 11 × 0 × has conditioned positions 1, 2, 4, and "don't-care" positions 3, 5. For  $l \leq i \leq n$ ,

$$\text{pr}(Y_i = 1) = \frac{1}{2^{l^*}},$$

since there are  $l^*$  positions forced by the window when a pulse occurs. Hence,  $E(Y_i) = 1/2^{l^*}$ ,  $l \leq i \leq n$ ;  $E(Y_i) = 0$ ,  $1 \leq i \leq l - 1$ . Since

$$E(E_n) = \sum_{i=1}^n E(Y_i),$$

Eq. (2) below is established as in Ref. 37:

$$E(E_n) = \frac{n - l + 1}{2^{l^*}}. \quad (2)$$

More difficult is finding the variance of  $E_n$ ,  $\text{Var } E_n$ , and it is to this direction that we now turn.

**d. The variance of  $E_n$ .** To find the variance of the number of pulses in  $n$  bits (where  $n \geq l$ ), we shall find  $E(E_n^2)$ , and subtract  $(E(E_n))^2$ . Now, since

$$E_n = \sum_{i=1}^n Y_i,$$

it follows that

$$\text{Var } E_n = \sum_{i,j=1}^n \text{Cov}(Y_i, Y_j). \quad (3)$$

An unfortunate property of the set  $\{Y_i\}$  is that the  $Y_i$  are not independent. Thus, as in Part *b* of this article, we observed that when a pulse occurs, the probability of a pulse on the very next bit is affected. Therefore, the covariance of  $Y_i$  and  $Y_j$  is not necessarily 0 for  $i \neq j$ .

Instead, we proceed as follows. For  $i$  or  $j < l$ ,  $\text{Cov}(Y_i, Y_j)$  is of course 0. For  $i = j \geq l$ ,

$$\begin{aligned} \text{Cov}(Y_i, Y_j) &= \text{Var } Y_i = E(Y_i^2) - (E(Y_i))^2 \\ &= \text{pr}(Y_i = 1) - (\text{pr}(Y_i = 1))^2 \\ &= \frac{1}{2^{l^*}} - \frac{1}{2^{2l^*}} = \frac{2^{l^*} - 1}{2^{2l^*}}. \end{aligned}$$

The difficult case is  $\text{Cov}(Y_i, Y_j)$  with  $l \leq i < j$ .

Let  $j - i = k \geq 1$ . First observe that  $\text{Cov}(Y_i, Y_{i+k})$  is a function of  $k$  only, since the joint distribution of  $Y_i$  and  $Y_{i+k}$ , for  $i \geq l$ , is independent of  $i$ . This follows from the fact that the occurrence of a pulse at a given position depends only on the present position and the previous  $l - 1$  positions. Thus, define  $f(k)$ , for  $k \geq 0$ , as  $\text{Cov}(Y_i, Y_{i+k})$ . Eq. (3) becomes

$$\text{Var } E_n = \sum_{i=l}^n \sum_{k=0}^{n-i} f(k), \quad (4)$$

or, rearranging the above sum,

$$\text{Var } E_n = \sum_{k=0}^{n-l} (n - l - k + 1) f(k) \quad (5)$$

Since  $f(0) = \text{Var } Y_i$  was calculated above as  $(2^{l*} - 1)/2^{2l*}$ , we need  $f(k)$ ,  $1 \leq k \leq n - l$ .

First let  $k \geq l$ . We claim that  $Y_i$  and  $Y_{i+k}$  are then independent. For the occurrence of a pulse more than  $l$  positions in the past does not affect the occurrence of a pulse in the present, since the length of the pulse-determining window is  $l$ . Then  $\text{Cov}(Y_i, Y_{i+k}) = 0$ ,  $k \geq l$ , so that  $f(k) = 0$ ,  $k \geq l$ . Formula (5) can then be written

$$\text{Var } E_n = \sum_{k=0}^{\min(l-1, n-1)} (n-l-k+1) f(k). \quad (6)$$

We thus turn our attention to evaluating  $f(k)$  for  $1 \leq k \leq l-1$ .

*e. Calculation of  $f(k)$ .* We use the formula

$$\text{Cov}(Y_i, Y_{i+k}) = E(Y_i Y_{i+k}) - E(Y_i) E(Y_{i+k}),$$

where  $i \geq l$ . Since  $E(Y_i) = E(Y_{i+k}) = 1/2^{l*}$ , we need only find  $E(Y_i Y_{i+k})$ . Now

$$\begin{aligned} E(Y_i Y_{i+k}) &= \text{pr}(Y_i = 1, Y_{i+k} = 1) \\ &= \text{pr}(Y_i = 1) \text{pr}(Y_{i+k} = 1 | Y_i = 1), \end{aligned}$$

where  $\text{pr}(A|B)$  denotes the conditional probability of  $A$  given  $B$ . Thus, the crucial computation is  $\text{pr}(Y_{i+k} = 1 | Y_i = 1)$ .

This conditional probability is computed as follows. If  $Y_i = 1$ , then the window of length  $l$  ending at position  $i$  was the window  $\mathfrak{A}$ . The hypothesis is thus that the window  $\mathfrak{A}$  ends at position  $i$ . How does this effect the occurrence of a pulse at position  $i+k$ ? The effect is one of two possible ones. Consider the overlap of  $\mathfrak{A}$  and  $\mathfrak{A}_k$ ;  $l-k$  positions are involved.

Namely, let  $W$  be the window of length  $l$  that actually occurred at  $i$ ;  $W$  is not  $\mathfrak{A}$  if  $\mathfrak{A}$  has "don't-care" positions. Consider  $W_{l-k}$ , the window of last  $l-k$  positions of  $W$ . If one of the positions of  $W_{l-k}$  contradicts  $\mathfrak{A}$  in a conditioned position, a pulse cannot occur at  $i+k$ . Let  $W_{l-k}$  not so contradict  $\mathfrak{A}$ . Then the probability of a pulse at  $i+k$  is  $1/2^{l*}$ , where  $l_k^*$  denotes the number of conditioned positions in the last  $k$  positions of  $\mathfrak{A}$ .

Thus, given that a pulse occurred at  $i$ , and that  $W$  was the window causing the pulse (if  $\mathfrak{A}$  has no "don't-cares," this extra condition is vacuous, as then  $W$  is always  $\mathfrak{A}$ ), the probability of a pulse at  $i+k$  is  $1/2^{l*}$  times 0 or 1

according as there is or is not a contradiction in  $W_{l-k}$  with the first  $l-k$  positions of  $\mathfrak{A}$ . Write this as

$$\text{pr}(Y_{i+k} = 1 | W) = \frac{1}{2^{l_k^*}} D(W_{l-k}). \quad (7)$$

Now we have

$$\text{pr}(Y_{i+k} = 1 | Y_i = 1) = \sum_W \text{pr}(W | Y_i = 1) \text{pr}(Y_{i+k} = 1 | W). \quad (8)$$

and  $\text{pr}(W | Y_i = 1) = 1/2^{l-l*}$ , for all  $2^{l-l*}$  possible  $W$ 's. Hence

$$\text{pr}(Y_{i+k} = 1 | Y_i = 1) = \frac{1}{2^{l-l*}} \frac{1}{2^{l_k^*}} \sum_W D(W_{l-k}). \quad (9)$$

But the sum

$$\sum_W D(W_{l-k})$$

is the number of  $W$  not contradicting  $\mathfrak{A}$  in its  $W_{l-k}$ . Let  $g(k) = 1$  or 0 according as there is or is not at least one such  $W$ , that is, according as there is or is not a contradiction between  $\mathfrak{A}_{n-k}$  and  $\mathfrak{A}$ . Define  $b_k$  as the number of unconditioned positions among the last  $l-k$  positions of  $\mathfrak{A}$  that are also unconditioned in the first  $l-k$  positions of  $\mathfrak{A}_k$ . Define  $t_k$  as the number of unconditioned positions in the first  $k$  positions of  $\mathfrak{A}$ . Then the number of  $W$  not contradicting  $\mathfrak{A}$  in  $W_{l-k}$  is  $g(k) 2^{t_k+b_k}$ , so that

$$\sum_W D(W_{l-k}) = 2^{t_k+b_k}. \quad (10)$$

Formula (9) can now be written

$$\text{pr}(Y_{i+k} = 1 | Y_i = 1) = g(k) 2^{-(l+l_k^*-l*-t_k-b_k)}. \quad (11)$$

Consequently,

$$\begin{aligned} E(Y_i Y_{i+k}) &= \text{pr}(Y_i = 1) g(k) 2^{-(l+l_k^*-l*-t_k-b_k)} \\ &= g(k) 2^{-(l+l_k^*-t_k-b_k)}, \end{aligned}$$

and

$$\text{Cov}(Y_i, Y_{i+k}) = g(k) 2^{-(l+l_k^*-t_k-b_k)} - 2^{-2l*}.$$

We have finally

$$f(k) = g(k) 2^{-(l+l_k^*-t_k-b_k)} - 2^{-2l*}, \quad 0 \leq k \leq l, \quad (12)$$

the required expression. Here  $g(k) = 1$  or  $0$  according as there is no contradiction between  $\mathfrak{A}$  and its translate by  $k$ , and

- (a)  $l_k^*$  is the number of *conditioned* positions in the last  $k$  position of  $\mathfrak{A}$ ;  
 (b)  $t_k$  is the number of *unconditioned* positions in the first  $k$  positions of  $\mathfrak{A}$ ;  
 (c)  $b_k$  is the number of jointly *unconditioned* positions in the last  $l - k$  positions of  $\mathfrak{A}$  and the first  $l - k$  positions of  $\mathfrak{A}$  shifted to the right by  $k$ .

(13)

*f. Simplifications in special cases.* This part of the article considers the simplifications available in special cases. If "don't-cares" are not allowed, then the following simplifications occur in Eq. (13):  $l_k^*$  becomes  $k$ ,  $t_k$  becomes  $0$ . Also,  $l^*$  is the same as  $l$ . Then Eq. (12) becomes

$$f(k) = 2^{-2l} \{g(k) 2^{l-k} - 1\}. \quad (14)$$

Eq. (6) now becomes

$$\text{Var } E_n = 2^{-2l} \sum_{k=0}^{\min(l-1, n-l)} (n-l-k+1) \{g(k) 2^{l-k} - 1\}. \quad (15)$$

Now consider the simplification in which "don't-cares" can be allowed, but no tail of the basic window of length  $k$  is a head of length  $k$ ,  $1 \leq k \leq n-1$ . That is, let  $g(k) = 0$ ,  $1 \leq k \leq n-1$ . Then Eq. (6) becomes (if  $n-l \geq l-1$  to avoid trivial complications):

$$\text{Var } E_n = 2^{-l*} \{(n-l+1) - 2^{-l*} (l-2)(n-\frac{1}{2}(l-1))\}. \quad (16)$$

We remark that in the above case, the theory of recurrent events applies.

Now consider the opposite case, in which  $g(k) = 1$ ,  $1 \leq k \leq n-1$ . That is, the window contains all 0's and "don't-cares," or all 1's and "don't-cares." For simplicity, no "don't-cares" are allowed in the below formula, and  $n-l$  is assumed  $\geq l-1$ :

$$\text{Var } E_n = 2^{-2l} \left\{ (2^{l+1} - l - 2)(n-l) + \frac{l(l+1)}{2} \right\}. \quad (17)$$

(As a check, let  $l=1$ ;  $\text{Var } E_n$  then is  $\frac{1}{4}$ , which is the known answer for the binomial distribution.) The mean  $E(E_n)$  for Eq. (17) is  $(n-l+1)/2^l$ .

Note that the variance is higher in the case of Eq. (17) than in the case in which the theory of recurrent events applies, that is, when each pulse starts the process over [option (I)]; this assertion can be checked using formulas from Ref. 38, Chapt. XIII. Of course, the mean number of pulses is also higher,  $(n-l-1)/2^l$  versus  $n/(2^{l+1}-2)$ . These facts are reasonable in advance of any calculation. Also for  $l$  large, asymptotically in  $n$ , we have  $n^{1/2} 2^{(l+1)/2}$  for option (I),  $n^{1/2} 2^{-(l+1)/2}$  for option IV, as the ratio of mean number of pulses to standard deviation of the number of pulses. These formulas are a precise measure of the greater dispersion attained in option IV.

Finally, let us consider the asymptotic formula for  $\text{Var } E_n$ . That is, let  $n \rightarrow \infty$  in Eq. (6). We obtain

$$\text{Var } E_n \sim \frac{n}{2^{2l*}} \left\{ \sum_{k=0}^{l-2} (g(k) 2^{-(l+1-k-t_k-b_k-2l*)} - 1) \right\}, \quad \text{in the general case;} \quad (18)$$

$$\text{Var } E_n \sim \frac{n}{2^{2l}} \left\{ \sum_{k=0}^{l-2} (g(k) 2^{l-k} - 1) \right\}, \quad \text{with no "don't-cares";} \quad (19)$$

$$\text{Var } E_n \sim \frac{n}{2^{2l}} (2^{l+1} - l - 2), \quad \text{for the all-1's word length } l. \quad (20)$$

Note that in Eqs. (18) through (20), the variance of  $E_n$  does increase linearly with  $n$ , as can be predicted in advance.

*g. Asymptotic distribution of  $E_n$ .* We have found the mean and variance of  $E_n$ ; now let us consider its asymptotic distribution. In the recurrent event case, observe that Ref. 38, p. 297, has the following theorem: *the number  $E_n$  of pulses in  $n$  bits is asymptotically normally distributed*. The same theorem holds for all cases considered in this paper. The proof is complicated by the fact that the random variables representing waiting times between successive pulses are not independent. Thus, the technique of Ref. 38 for proving the asymptotic normality theorem cannot be directly applied. However, it can be seen that the waiting time between the  $r^{\text{th}}$  and  $(r+1)^{\text{st}}$  pulse is almost independent of the waiting time between the  $s^{\text{th}}$  and  $(s+1)^{\text{st}}$  pulse, if  $s$  is sufficiently greater than  $r$ . This fact can be put in sufficiently quantitative form to make the proof of the above theorem hold. Details are omitted. It is to be noted that the asymptotic

distribution of  $E_n$  is not as important for many purposes as is the knowledge of how the pulses bunch. Thus, a window of all 1's tends to produce bursts of pulses, although the asymptotic distribution of  $E_n$  does not imply this property.

**h. Sample calculation.** We shall calculate the relevant values for the window  $1 \times 0$  of length 3. Here  $l = 3$ ,  $l^* = 2$ , so that  $E(E_n) = (n-2)/4$ . We next need  $f(k)$ ,  $0 \leq k \leq 2$ . Note that  $g(0) = 1$  as always,  $g(1) = 1$ ,  $g(2) = 0$ . To see that  $g(1) = 1$ , note that the sequence 110 yields a pulse, as well as the sequence 100. Hence, 1100 yields two successive pulses, so that  $g(1) = 1$ . And  $g(2) = 0$ , since, if  $1 \times 0$  occurs, no pulse can begin with the 0.

Also,  $l_0^* = 0$ ,  $t_0 = 0$ ,  $b_0 = 1$ ;  $l_1^* = 0$ ,  $t_1 = 0$ ,  $b_1 = 0$ ;  $l_2^* = 1$ ,  $t_2 = 1$ ,  $b_2 = 0$  (but these last values are not needed, since  $g(2) = 0$ ). Then

$$f(0) = 2^{-(3+0-0-1)} - 2^{-4} = \frac{3}{16};$$

$$f(1) = 2^{-(3+0-0-0)} - 2^{-4} = \frac{1}{16};$$

$$f(2) = 0 - 2^{-4} = -\frac{1}{16}.$$

We conclude that, for  $n-3 \geq 1$ , i.e., for  $n \geq 4$ ,

$$\begin{aligned} \text{Var } E_n &= (n-2)f(0) + (n-3)f(1) + (n-4)f(2) \\ &= \frac{3}{16}(n-2) + \frac{1}{16}(n-3) - \frac{1}{16}(n-4), \\ \text{Var } E_n &= \frac{3n-5}{16}. \end{aligned}$$

Thus,  $E(E_n)/(\text{Var } E_n)^{1/2} \sim n^{1/2}/3$ . As  $n \rightarrow \infty$ , the distribution of  $[E_n - (n/4)]/[(3n)^{1/2}/4]$  tends to the unit normal. For  $n=3000$ , we conclude that  $E_n$  is approximately normal with mean 750, standard deviation 25. For comparison purposes, a binomial distribution with  $n=3000$  and mean 750 would have probability of success  $p = 1/4$ ; hence, standard deviation  $[np(1-p)]^{1/2} = 3000 \cdot 1/4 \cdot 3/4^{1/2} = 25$  also.

### 3. An Almost Uniformly Most Powerful Test Using Quantiles, I. Eisenberger

**a. Summary.** The use of quantiles for data compression of space telemetry has been considered in Ref. 39; a mechanization of a quantile system has been considered in Ref. 40. Further work is necessary in order to study the uses that can be made of telemetry in the form of

quantiles. This article discusses one such use--the test of hypotheses, or, more accurately, the discrimination of one hypothesis from another.

Ref. 39 gave quantile tests for discriminating between means of normal distributions with the same known variance. In this report, quantile tests are given for discriminating between variances of normal distributions with the same known mean. More precisely, the problem of testing the simple null hypothesis

$$H_0: g(x) = g_1(x) = N(\mu, \sigma_1)$$

against the simple alternative hypothesis

$$H_1: g(x) = g_2(x) = N(\mu, \sigma_2)$$

is considered, where  $\sigma_2 > \sigma_1$  ( $\sigma_2 < \sigma_1$ ),  $\mu$  is known and the test is based on one, two, or four sample quantiles. Here,  $N(\mu, \sigma)$  denotes the normal distribution with mean  $\mu$  and variance  $\sigma^2$ . In each case the power function is derived and the efficiency of the test is determined when the sample size,  $n$ , is assumed to be large. The power is surprisingly large.

**b. Review of quantiles.** To define a quantile, consider a sample of  $n$  independent values,  $x_1, x_2, \dots, x_n$  taken from a distribution of a continuous type with distribution function  $G(x)$  and density function  $g(x)$ . The  $p$ th quantile, or the quantile of order  $p$  of the distribution, denoted by  $\xi_p$ , is defined as the root of the equation  $G(\xi) = p$ , that is,

$$p = \int_{-\infty}^{\xi_p} dG(x) = \int_{-\infty}^{\xi_p} g(x) dx$$

The corresponding sample quantile,  $Z_p$ , is defined as follows:

If the sample values are arranged in ascending order of magnitude

$$x_{(1)} < x_{(2)} < \dots < x_{(n)}$$

then  $x_{(i)}$  is called the  $i$ th order statistic and

$$Z_p = x_{[np] + 1}$$

where  $[np]$  is the greatest integer  $\leq np$ .

If  $g(x)$  is differentiable in some neighborhood of each quantile value considered, it has been shown that the



joint distribution of any number of quantiles is asymptotically normal as  $n \rightarrow \infty$  and that, asymptotically,

$$\begin{aligned} E(Z_p) &= \xi_p, \\ \text{Var}(Z_p) &= \frac{p(1-p)}{ng^2(\xi_p)}, \\ \rho_{12} &= \left[ \frac{p_1(1-p_2)}{p_2(1-p_1)} \right]^{1/2} \end{aligned}$$

where  $\rho_{12}$  is the correlation between  $Z_{p_1}$  and  $Z_{p_2}$ ,  $p_1 < p_2$ .

**c. The test for one quantile.** Now, assuming the limiting distribution and moments of the quantiles, denote by  $F(x)$  and  $f(x) = F'(x)$  the distribution function and density function, respectively, of the standard normal distribution, that is,

$$F(x) = \int_{-\infty}^x f(t) dt,$$

where  $f(x) = 1/(2\pi)^{1/2} e^{-1/2 x^2}$ , and let  $Z$  be the sample quantile of order  $p$ ,  $\xi$  the corresponding population quantile under  $H_0$  and  $\xi'$  the corresponding population under  $H_1$ . Then one has

$$p = \int_{-\infty}^{(\xi - \mu)/\sigma_1} f(x) dx = \int_{-\infty}^{(\xi' - \mu)/\sigma_2} f(x) dx \quad (1)$$

so that

$$\frac{\xi - \mu}{\sigma_1} = \frac{\xi' - \mu}{\sigma_2}$$

or

$$\xi' = \frac{\sigma_2}{\sigma_1} (\xi - \mu) + \mu$$

and

$$\begin{aligned} g_1(\xi) &= \frac{1}{\sigma_1} f\left(\frac{\xi - \mu}{\sigma_1}\right) \\ g_2(\xi') &= \frac{1}{\sigma_2} f\left(\frac{\xi' - \mu}{\sigma_2}\right) = \frac{1}{\sigma_2} f\left(\frac{\xi - \mu}{\sigma_1}\right) \end{aligned}$$

Hence, under  $H_0$ :

$$E(Z) = \xi_1 \quad \text{Var}(Z) = \sigma_1^2 a^2,$$

where

$$a^2 = \frac{F\left(\frac{\xi - \mu}{\sigma_1}\right) \left[1 - F\left(\frac{\xi - \mu}{\sigma_1}\right)\right]}{nf^2\left(\frac{\xi - \mu}{\sigma_1}\right)}$$

and under  $H_1$ :

$$E(Z) = \frac{\sigma_2}{\sigma_1} (\xi - \mu) + \mu, \quad \text{Var}(Z) = \sigma_2^2 a^2.$$

If the test is now based on the value of the statistic  $Z$ , the best critical (or rejection region) is determined by the *likelihood ratio inequality*

$$\frac{L(Z|H_0)}{L(Z|H_1)} = \frac{\exp\left[-\frac{1}{2} \left(\frac{Z - \xi}{\sigma_1 a}\right)^2\right]}{\exp\left[-\frac{1}{2} \left(\frac{Z - \xi'}{\sigma_2 a}\right)^2\right]} < k_1$$

which, by taking logarithms and completing the square in  $Z$  reduces to

$$\left(Z - \frac{\sigma_1 \mu + \sigma_2 \xi}{\sigma_1 + \sigma_2}\right)^2 > k^2, \quad \text{for } \sigma_2 > \sigma_1 \quad (2)$$

as the region providing the maximum power. Here,  $k$  is determined such that, under  $H_0$ , the probability of the inequality (2) occurring is equal to  $\epsilon$ , the significance level of the test. The power of the test,  $P_0$ , is determined as follows.

Under  $H_0$ :

$$\begin{aligned} P\left(-k + \frac{\sigma_1 \mu + \sigma_2 \xi}{\sigma_1 + \sigma_2} < Z < k + \frac{\sigma_1 \mu + \sigma_2 \xi}{\sigma_1 + \sigma_2}\right) \\ &= F\left(\frac{k + \frac{\sigma_1 \mu + \sigma_2 \xi}{\sigma_1 + \sigma_2} - \xi}{\sigma_1 a}\right) - F\left(\frac{-k + \frac{\sigma_1 \mu + \sigma_2 \xi}{\sigma_1 + \sigma_2} - \xi}{\sigma_1 a}\right) \\ &= F\left[\frac{k}{\sigma_1 a} - \frac{\xi - \mu}{\sigma_1 a} \left(\frac{1}{1 + \frac{\sigma_2}{\sigma_1}}\right)\right] \\ &\quad - F\left[\frac{-k}{\sigma_1 a} - \frac{\xi - \mu}{\sigma_1 a} \left(\frac{1}{1 + \frac{\sigma_2}{\sigma_1}}\right)\right] \\ &= F(b) - F(c) = 1 - \epsilon. \end{aligned}$$

Under  $H_1$ :

$$\begin{aligned}
 & P\left(-k + \frac{\sigma_1\mu + \sigma_2\zeta}{\sigma_1 + \sigma_2} < Z < k + \frac{\sigma_1\mu + \sigma_2\zeta}{\sigma_1 + \sigma_2}\right) \\
 &= F\left[\frac{\sigma_1ab + \zeta - \frac{\sigma_2}{\sigma_1}(\zeta - \mu) - \mu}{\sigma_2a}\right] \\
 &\quad - F\left[\frac{\sigma_1ac + \zeta - \frac{\sigma_2}{\sigma_1}(\zeta - \mu) - \mu}{\sigma_2a}\right] \\
 &= F\left[\frac{\sigma_1}{\sigma_2}b - \frac{\zeta - \mu}{\sigma_1a}\left(1 - \frac{\sigma_1}{\sigma_2}\right)\right] \\
 &\quad - F\left[\frac{\sigma_1}{\sigma_2}c - \frac{\zeta - \mu}{\sigma_1a}\left(1 - \frac{\sigma_1}{\sigma_2}\right)\right] \\
 &= 1 - P_0.
 \end{aligned} \tag{3}$$

Theoretically, the values of  $b$  and  $c$  depend upon the parameters of  $g_1(x)$  and  $g_2(x)$  as well as upon  $\epsilon$ . However, a few simple calculations show that for all practical purposes one can determine  $b$  by the relation  $F(b) = 1 - \epsilon$  and neglect as negligible the second term of the left-hand side of Eq. (3). With this simplification,  $P_0$  will be maximized if the order of the quantile is chosen to maximize  $\zeta - \mu/\sigma_1a$ . Setting equal to zero the derivative of this quantity with respect to  $\zeta - \mu/\sigma_1$ , one finds that this maximum occurs at  $p = F(\zeta - \mu/\sigma_1) = 0.9424$ . For this value of  $p$ ,  $\zeta - \mu/\sigma_1 = 1.575$  and  $a = 2.019/n^{1/2}$ . Inserting these values, one obtains as the optimum single quantile power function

$$P_0 = 1 - F\left[\frac{\sigma_1}{\sigma_2}b - 0.78(n)^{1/2}\left(1 - \frac{\sigma_1}{\sigma_2}\right)\right]$$

**d. Efficiency for one quantile.** If one defines the efficiency of the test as the ratio of  $P_0$  to  $P'_0$ , the power of the best test using all the sample values, it is of interest to determine  $P'_0$  in order to see how "good" the quantile tests are compared to what can be achieved using the entire sample. Now the likelihood ratio inequality in the case of the entire sample gives as the best critical region

$$\sum_{i=1}^n (x_i - \mu)^2 > k \quad (\sigma_2 > \sigma_1),$$

but it is more convenient to express this as

$$y = \left[2 \sum_{i=1}^n (x_i - \mu)^2\right]^{1/2} > k',$$

since it is shown in Ref. 41 that, asymptotically,  $y$  is distributed  $N[\sigma_1(2n)^{1/2}, \sigma_1]$  under  $H_0$  and  $N[\sigma_2(2n)^{1/2}, \sigma_2]$  under  $H_1$ . Hence, assuming the limiting distribution, which is a very good approximation for large  $n$ , one has

Under  $H_0$ :

$$p(y < k) = F\left(\frac{k - \sigma_1(2n)^{1/2}}{\sigma_1}\right) = F(b) = 1 - \epsilon;$$

Under  $H_1$ :

$$p(y < k) = F\left[\frac{\sigma_1}{\sigma_2}b - (2n)^{1/2}\left(1 - \frac{\sigma_1}{\sigma_2}\right)\right] = 1 - P'_0.$$

Fig. 26, for  $n = 200$ ,  $\epsilon = 0.01$ , shows the power and efficiency of the test using one optimal quantile. The efficiency is never less than 35% for any  $\sigma_2/\sigma_1$ . Furthermore, as in Ref. 39, the efficiency approaches 1 as  $\sigma_2/\sigma_1$  approaches 0 or  $\infty$ . Under these conditions,

$$b = 2.326, \quad k = 0.332\sigma_1 + \frac{1.575\sigma_1^2}{\sigma_1 + \sigma_2}.$$

Thus, if

$$\mu - \frac{1.243\sigma_1(\sigma_2 - 1.534\sigma_1)}{\sigma_1 + \sigma_2} < Z < \mu + 1.907\sigma_1$$

accept  $H_0$ . Otherwise reject  $H_0$ . The test is to be made at a significance level of 0.01.

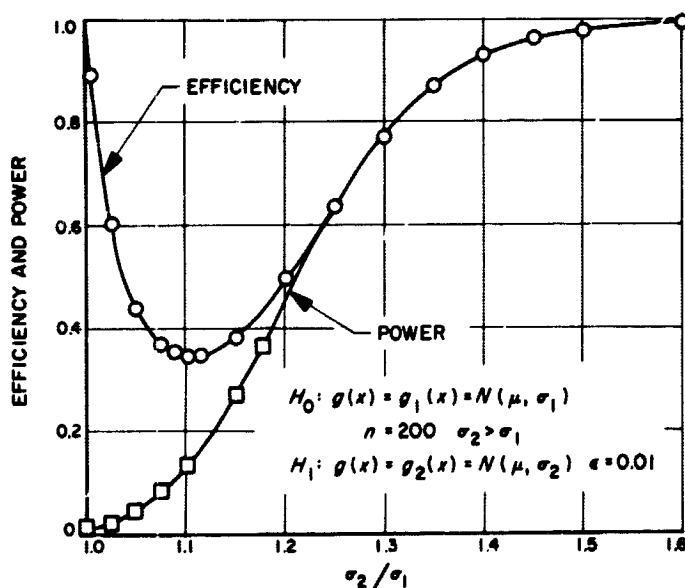


Fig. 26. Power and efficiency of test using one quantile

**e. Two quantiles.** Now let  $Z_1$  and  $Z_2$  be two sample quantiles of orders  $p_1$  and  $p_2 = 1 - p_1 > p_1$  and let  $x = Z_2 - Z_1$ . Then

Under  $H_0$ :

$$E(x) = 2(\mu - \xi_1), \quad \text{Var}(x) = 2\sigma_1^2 a^2,$$

where

$$a^2 = \frac{F\left(\frac{\xi_1 - \mu}{\sigma_1}\right) \left[1 - 2F\left(\frac{\xi_1 - \mu}{\sigma_1}\right)\right]}{nf^2\left(\frac{\xi_1 - \mu}{\sigma_1}\right)}$$

Under  $H_1$ :

$$E(x) = \frac{2\sigma_2}{\sigma_1}(\mu - \xi_1), \quad \text{Var}(x) = 2\sigma_2^2 a^2.$$

Simplifying the likelihood ratio inequality results in the best critical region given by

$$\left[x - \frac{2\sigma_2(\mu - \xi_1)}{\sigma_1 + \sigma_2}\right]^2 > k^2, \quad (\sigma_2 > \sigma_1).$$

Under  $H_0$ :

$$\begin{aligned} P\left(-k + \frac{2\sigma_2(\mu - \xi_1)}{\sigma_1 + \sigma_2} < x < k + \frac{2\sigma_2(\mu - \xi_1)}{\sigma_1 + \sigma_2}\right) \\ = F\left[\frac{k + \frac{2\sigma_2(\mu - \xi_1)}{\sigma_1 + \sigma_2} - 2(\mu - \xi_1)}{2^{1/2}\sigma_1 a}\right] \\ - F\left[\frac{-k + \frac{2\sigma_2(\mu - \xi_1)}{\sigma_1 + \sigma_2} - 2(\mu - \xi_1)}{2^{1/2}\sigma_1 a}\right] \\ = 1 - \epsilon; \end{aligned}$$

Under  $H_1$ :

$$\begin{aligned} P\left(-k + \frac{2\sigma_2(\mu - \xi)}{\sigma_1 + \sigma_2} < x < k + \frac{2\sigma_2(\mu - \xi)}{\sigma_1 + \sigma_2}\right) \\ = F\left[\frac{\frac{\sigma_1}{\sigma_2}b - 2^{1/2}\left(\frac{\mu - \xi_1}{\sigma_1 a}\right)\left(1 - \frac{\sigma_1}{\sigma_2}\right)}{2^{1/2}\sigma_1 a}\right] \\ - F\left[\frac{\frac{\sigma_1}{\sigma_2}c - 2^{1/2}\left(\frac{\mu - \xi_1}{\sigma_1 a}\right)\left(1 - \frac{\sigma_1}{\sigma_2}\right)}{2^{1/2}\sigma_1 a}\right] \\ = 1 - P_0. \end{aligned}$$

Again determining  $b$  such that  $F(b) = 1 - \epsilon$ , the orders of the two symmetric quantiles which maximize  $P_0$  are  $p_1 = 0.0659, p_2 = 0.9341$ . For these values,  $(\mu - \xi_1)/\sigma_1 = 1.508$ ,  $a = 1.869/n^{1/2}$  so that the optimum power function is given by

$$P_0 = 1 - \left[F\left(\frac{\sigma_1}{\sigma_2}b - 1.141(n)^{1/2}\left(1 - \frac{\sigma_1}{\sigma_2}\right)\right)\right]$$

We are considering only symmetric quantiles and we conjecture that this is the optimum procedure to adopt.

For  $n = 200$ ,  $\epsilon = 0.01$ , Fig. 27 shows the power and efficiency of the test using two quantiles. For this case if

$$-0.4341\sigma_1 + \frac{3.016\sigma_1(\sigma_2 - \sigma_1)}{\sigma_1 + \sigma_2} < x < 3.450\sigma_1$$

accept  $H_0$ . Otherwise reject  $H_0$ . The efficiency never drops below 67%.

**f. Four quantiles.** For the four quantile case, let  $Z_i$ ,  $i = 1, 2, 3, 4$  be four sample quantiles such that  $p_1 < p_2 < p_3 < p_4$  and  $p_1 + p_4 = p_2 + p_3 = 1$ . The test will be based on the linear combination

$$x = \alpha(Z_4 - Z_1) + \beta(Z_3 - Z_2)$$

determining  $\alpha$  and  $\beta$  as well as the orders of the quantiles so as to maximize  $P_0$ .

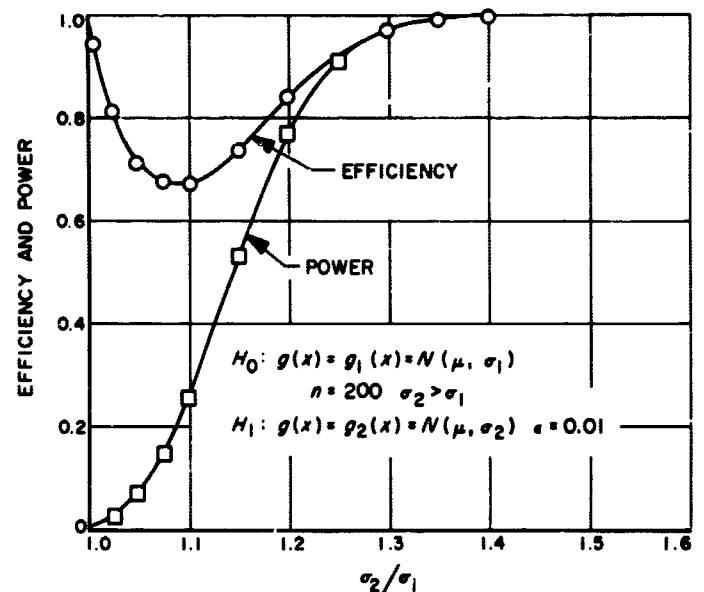


Fig. 27. Power and efficiency of test using two quantiles

Under  $H_0$ :

$$E(x) = 2\alpha(\mu - \zeta_1) + 2\beta(\mu - \zeta_2)$$

$$\text{Var}(x) = 2\sigma_1^2 [\alpha^2 a_3^2 + \beta^2 a_4^2 + 2\alpha\beta a_1 a_2 (\rho_{12} - \rho_{13})] = 2\sigma_1^2 \gamma^2$$

where

$$a_1^2 = \frac{F\left(\frac{\zeta_1 - \mu}{\sigma_1}\right) \left[1 - F\left(\frac{\zeta_1 - \mu}{\sigma_1}\right)\right]}{nf^2\left(\frac{\zeta_1 - \mu}{\sigma_1}\right)}$$

$$a_2^2 = \frac{F\left(\frac{\zeta_2 - \mu}{\sigma_1}\right) \left[1 - F\left(\frac{\zeta_2 - \mu}{\sigma_1}\right)\right]}{nf^2\left(\frac{\zeta_2 - \mu}{\sigma_1}\right)}$$

$$a_3^2 = \frac{F\left(\frac{\zeta_1 - \mu}{\sigma_1}\right) \left[1 - 2F\left(\frac{\zeta_1 - \mu}{\sigma_1}\right)\right]}{nf^2\left(\frac{\zeta_1 - \mu}{\sigma_1}\right)}$$

$$a_4^2 = \frac{F\left(\frac{\zeta_2 - \mu}{\sigma_1}\right) \left[1 - 2F\left(\frac{\zeta_2 - \mu}{\sigma_1}\right)\right]}{nf^2\left(\frac{\zeta_2 - \mu}{\sigma_1}\right)}$$

$\rho_{ij}$  = the correlation between  $Z_i$  and  $Z_j$ .

Under  $H_1$ :

$$E(x) = 2\alpha \frac{\sigma_2}{\sigma_1} (\mu - \zeta_1) + 2\beta \frac{\sigma_2}{\sigma_1} (\mu - \zeta_2)$$

$$\text{Var}(x) = 2\sigma_2^2 \gamma^2.$$

Omitting the details, one obtains finally

$$F\left\{\frac{\sigma_1}{\sigma_2} b - \frac{2\gamma\left(1 - \frac{\sigma_1}{\sigma_2}\right) \left[\alpha\left(\frac{\mu - \zeta_1}{\sigma_1}\right) + \beta\left(\frac{\mu - \zeta_2}{\sigma_1}\right)\right]}{\gamma}\right\} - F\left\{\frac{\sigma_1}{\sigma_2} c - \frac{2\gamma\left(1 - \frac{\sigma_1}{\sigma_2}\right) \left[\alpha\left(\frac{\mu - \zeta_1}{\sigma_1}\right) + \beta\left(\frac{\mu - \zeta_2}{\sigma_1}\right)\right]}{\gamma}\right\} = 1 - P_0.$$

As in the previous cases, we determine  $b$  such that  $F(b) = 1 - \epsilon$ . Moreover, from previous investigations in

the use of quantiles in estimating the parameters of normal distributions (Ref. 42), it is known that for

$$\alpha = 0.116 \quad \beta = 0.236$$

$$p_1 = 0.023 \quad p_2 = 0.127$$

$$p_4 = 0.977 \quad p_3 = 0.873$$

$P_0$  will be a maximum. Hence, the optimum power function is given by

$$P_0 = 1 - F\left[\frac{\sigma_1}{\sigma_2} b - 1.284(n)^{1/2} \left(1 - \frac{\sigma_1}{\sigma_2}\right)\right].$$

For  $n = 200$ ,  $\epsilon = 0.01$ , Fig. 28 shows the power and efficiency of the test using four quantiles. For this case if

$$\sigma_1 \left[1.001 \left(\frac{\sigma_2 - \sigma_1}{\sigma_1 + \sigma_2}\right)\right] - 0.1283\sigma_1 < x < 1.1295\sigma_1,$$

accept  $H_0$ ; otherwise, reject  $H_0$ . The efficiency in the four quantile case never drops below 82%.

**g. Conclusion.** In each case, since the lower limit of the acceptance interval depends on  $\sigma_2$  as well as on  $\sigma_1$ , strictly speaking the tests are not uniformly most powerful among quantile tests. However, in each case the

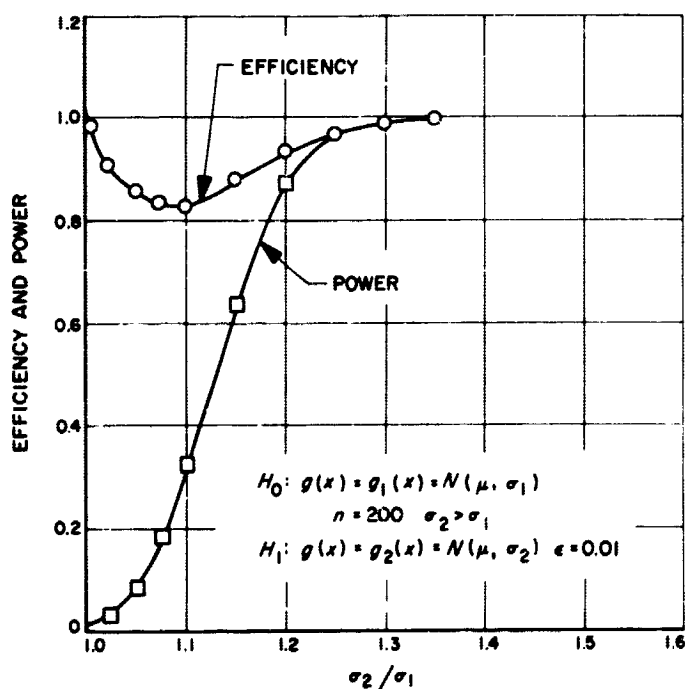


Fig. 28. Power and efficiency of test using four quantiles

probability that the test statistic will actually fall below the lower limit is practically zero even when  $H_0$  is true, so that the test is essentially one sided. Since the upper limit depends on the parameters of  $g_1(x)$  only, then, for all practical purposes, the tests are uniformly most powerful among quantile tests.

If  $\sigma_1 > \sigma_2$  is specified instead of  $\sigma_2 > \sigma_1$ , a similar analysis provides the following criteria.

For one quantile:

$$\text{If } \mu - \frac{1.243\sigma_1(\sigma_1 - 1.534\sigma_2)}{\sigma_1 + \sigma_2} < Z < \mu + 1.243\sigma_1$$

reject  $H_0$ ; otherwise, accept  $H_0$ .

For two quantiles:

$$\text{If } 0.4341\sigma_1 - \frac{3.016\sigma_1(\sigma_1 - \sigma_2)}{\sigma_1 + \sigma_2} < x < 2.582\sigma_1$$

reject  $H_0$ ; otherwise, accept  $H_0$ .

For four quantiles:

$$\text{If } \sigma_1 \left[ 0.1283 - 1.001 \left( \frac{\sigma_1 - \sigma_2}{\sigma_1 + \sigma_2} \right) \right] < x < 0.8729\sigma_1$$

reject  $H_0$ ; otherwise, accept  $H_0$ . In each case the optimum orders of the quantiles remain the same. If it is not known whether  $\sigma_1$  is greater or less than  $\sigma_2$ , the test can be modified slightly. The power is reduced of necessity. Details are omitted.

Figs. 26 through 28 show that the efficiency of this test drops at first as  $\sigma_2/\sigma_1$  increases, and then increases again to 1. The minimum efficiency for one quantile is about 35%, 67% for two quantiles, and 83% for four quantiles. Thus, the four quantile test is quite good uniformly in  $\sigma_2/\sigma_1$ , even when compared with the optimum test using all the samples.

## References

1. Golomb, S. W., *Sequences with Randomness Properties*, The Martin Co., Baltimore, 1955.
2. Zierler, N., "Linear Recurring Sequences," *Journal of the Society of Applied Mathematics*, Vol. 7, No. 1, March 1959.
3. Baumert, L., et al., "Coding Theory and its Applications to Communications Systems," Technical Report No. 32-167, Jet Propulsion Laboratory, Pasadena, California, 1961.
4. Titsworth, R. C., and Welch, L. R., "Modulation by Random and Pseudorandom Sequences," Technical Report No. 20-387, Jet Propulsion Laboratory, Pasadena, California, 1959.
5. Zygmund, A., *Trigonometric Series*, pp. 122-130, Dover Publishing Company, 1955.
6. Peterson, W. W., *Error Correcting Codes*, pp. 251-270, Wiley and Sons, 1961.
7. Watson, E. J., "Primitive Polynomials (mod 2)," *Mathematics of Computation*, No. 16, 1962.
8. Golomb, S. W., Welch, L. R., and Hales, A., "On the Factorization of Trinomials Over  $GF(2)$ ," Technical Report No. 20-189, Jet Propulsion Laboratory, Pasadena, California, 1959.

## References (Cont'd)

9. Peterson, W. W., "Error Correcting Codes," John Wiley & Sons, New York, 1961.
10. Mattson, H. F., and Solomon, G., "A New Treatment of Bose-Chaudhuri Codes," *Journal of the Society of Industrial and Applied Mathematics*, Vol. 4, No. 4, 1961.
11. Solomon, G., "A Note on A New Class of Codes," *Information and Control*, Vol. 4, No. 4, 1961.
12. Solomon, G., "Linear Recursive Sequences as Finite Difference Equations," Group Report 47.37, Lincoln Laboratories, Massachusetts Institute of Technology, 1960.
13. Kendall, W. B., *Performance of the Biased Square-Law Sequential Detector in the Absence of Signal*, The RAND Corporation, RM-3759-PR, July 1963.
14. Bussgang, J. J., and Middleton, D., "Optimum Sequential Detection of Signals in Noise," *IRE Transactions on Information Theory*, Vol. IT-2, pp. 5-18, December 1955.
15. Bussgang, J. J., and Mudgett, W. L., "A Note of Caution on the Square-Law Approximation to an Optimum Detector," *IRE Transactions on Information Theory*, Vol. IT-6, pp. 504-505, September 1960.
16. Wald, A., *Sequential Analysis*, John Wiley & Sons, New York, 1947.
17. Lindsey, W. C., "Optimum Frequency Demodulation," SPS 37-26, Vol. IV, pp. 227-234, Jet Propulsion Laboratory, Pasadena, California, April 30, 1964.
18. Booton, R. C., Jr., "The Analysis of Nonlinear Control Systems with Random Inputs," *Proceedings of the Nonlinear Circuit Analysis*, Vol. II, pp. 369-391, Polytechnic Institute of Brooklyn, New York, April 1953.
19. Develet, J. A., "A Threshold Criterion for Phase-Lock Demodulation," *Proceedings of the IRE*, pp. 349-356, February 1963.
20. Viterbi, A. J., "Phase-Locked Loop Dynamics in the Presence of Noise by Fokker-Planck Techniques," Technical Report No. 32-427, Jet Propulsion Laboratory, Pasadena, California, March 1963.
21. Davenport, W. B., Jr., "Signal-to-Noise Ratios in Band-Pass Limiters," *Journal of Applied Physics*, Vol. 24, pp. 720-727, June 1953.
22. Lindsey, W. C., "Performance Analysis for the Coherent Multireceiver in Random Multichannels," SPS 37-21, Vol. IV, pp. 154-156, Jet Propulsion Laboratory, Pasadena, California, June 30, 1963.
23. Lindsey, W. C., "Error Probabilities for Rician Fading Multichannel Reception of Binary and N-ary Signals," Technical Report No. 32-450, Jet Propulsion Laboratory, Pasadena, California, June 1963.
24. Turin, G. L., "Communication Through Noisy Random-Multipath Channel," *IRE Convention Record*, Part IV, 1956.
25. Price, R., "The Doubly Non-Central F-Distribution Expressed in Finite Terms," Lincoln Laboratories, Massachusetts Institute of Technology, TR 64G-5, September 1963.

## References (Cont'd)

26. Jaffe, L., and Rechtin, E., "Design and Performance of Phase-Lock Loops Capable of Near-Optimum Performance over a Wide Range of Input Signal and Noise Levels," *Transactions of the IRE, IT-1*, pp. 66-76, March 1955.
27. Titsworth, R. C., "Predetection Filters in Radiometric Receivers," SPS 37-26, Vol. IV, pp. 249-252, Jet Propulsion Laboratory, Pasadena, California, April 30, 1964.
28. Yerman, E. A., "Conditional Frequency Uncertainty in Spacecraft Acquisition," SPS 37-24, Vol. IV, pp. 192-195, Jet Propulsion Laboratory, Pasadena, California, December 31, 1963.
29. Mood, A. M., "Introduction to the Theory of Statistics," McGraw-Hill, New York, 1950.
30. Yerman, E. A., "Conditional Frequency Uncertainty in Spacecraft Acquisition," SPS 37-24, Vol. IV, pp. 192-195, Jet Propulsion Laboratory, Pasadena, California, December 31, 1963.
31. Cramer, H., "Mathematical Methods of Statistics," Princeton University Press, Princeton, 1946.
32. Feller, W., "An Introduction to Probability Theory and its Applications," Wiley and Sons, New York, 1959.
33. "Tables of Normal Probability Functions," U. S. Dept. of Commerce, National Bureau of Standards, *Applied Mathematics Series*, No. 23, 1953.
34. "Tables of the Exponential Function  $e^x$ ," U. S. Dept. of Commerce, National Bureau of Standards, *Applied Mathematics Series*, No. 14, 1961.
35. Gumbel, E. J., "Statistical Theory of Extreme Values and Some Practical Applications," National Bureau of Standards, *Applied Mathematics Series*, No. 33, 1954.
36. Gumbel, E. J., "Statistics of Extremes," Columbia University Press, New York, 1958.
37. "A Random Pulse Generator for use as a Data and Noise Source," SPS 37-27, Vol. III, pp. 112-115, Jet Propulsion Laboratory, Pasadena, California, May 31, 1964.
38. Feller, W., "An Introduction to Probability Theory and its Applications," Second Edition, Wiley and Sons, New York, 1957.
39. Eisenberger, I., "A Uniformly Most Powerful Test Using Quantiles," SPS 37-25, Vol. IV, pp. 194-198, Jet Propulsion Laboratory, Pasadena, California, February 29, 1964.
40. "Design of the Quantile System for Data Compression of Space Telemetry," SPS 37-27, Vol. III, pp. 103-112, Jet Propulsion Laboratory, Pasadena, California, May 31, 1964.
41. Cramer, H., *Mathematical Methods of Statistics*, p. 251, Princeton University Press, Princeton, New Jersey, 1946.
42. Eisenberger, I., and Posner, E., "Systematic Statistics Used for Data Compression in Space Telemetry," Technical Report No. 32-510, Jet Propulsion Laboratory, Pasadena, California, October 1, 1963.