

N71-21338 6

N71-21359

NASA CR-117440

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

*Space Programs Summary 37-66, Vol. III*

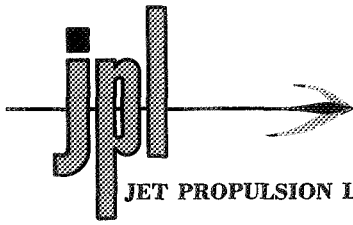
*Supporting Research and Advanced Development*

For the Period October 1 to November 30, 1970

**CASE FILE  
COPY**

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

December 31, 1970



**JET PROPULSION LABORATORY** *California Institute of Technology • 4800 Oak Grove Drive, Pasadena, California 91103*

## ANNOUNCEMENT

### CANCELLATION OF THE SPACE PROGRAMS SUMMARY (SPS)

With the completion of the SPS 37-66 series, the Space Programs Summary (SPS) is cancelled as a periodical publication of the Jet Propulsion Laboratory. The SPS is being replaced by a new publication, the JPL Quarterly Technical Review (QTR).

The QTR will be comprised of brief technical articles describing selected JPL work and a bibliography of recent JPL publications.

Current external recipients of the SPS will automatically receive the new QTR.



NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

*Space Programs Summary 37-66, Vol. III*

*Supporting Research and Advanced Development*

For the Period October 1 to November 30, 1970

JET PROPULSION LABORATORY  
CALIFORNIA INSTITUTE OF TECHNOLOGY  
PASADENA, CALIFORNIA

December 31, 1970

**SPACE PROGRAMS SUMMARY 37-66, VOL. III**

Copyright © 1971  
Jet Propulsion Laboratory  
California Institute of Technology

Prepared Under Contract No. NAS 7-100  
National Aeronautics and Space Administration

## Preface

The Space Programs Summary is a multivolume, bimonthly publication that presents a review of technical information resulting from current engineering and scientific work performed, or managed, by the Jet Propulsion Laboratory for the National Aeronautics and Space Administration. The Space Programs Summary is currently composed of four volumes:

- Vol. I. *Flight Projects* (Unclassified)
- Vol. II. *The Deep Space Network* (Unclassified)
- Vol. III. *Supporting Research and Advanced Development* (Unclassified)
- Vol. IV. *Flight Projects and Supporting Research and Advanced Development* (Confidential)



## Contents

### FUTURE PROJECTS

<b>I. Advanced Studies</b> . . . . .	1
A. Lunar Elevation Correction for Gravity Measurements <i>R. G. Brereton</i> . . . . .	1

### PROJECT ENGINEERING DIVISION

<b>II. Environmental Requirements</b> . . . . .	4
A. Multi-Phase Ammonia Water System (Rev. 1) <i>C. Haudenschild, NASA Code 124-12-07-04</i> . . . . .	4

### DATA SYSTEMS DIVISION

<b>III. Computation and Analysis</b> . . . . .	10
A. Unitary Similarity Transformation of a Hermitian Matrix to a Real Symmetric Tridiagonal Matrix <i>R. J. Hanson, NASA Code 129-04-20-03</i> . . . . .	10

### SPACE SCIENCES DIVISION

<b>IV. Physics</b> . . . . .	13
A. Response of Supersonic Laminar Boundary Layer to a Moving External Pressure Field <i>L. M. Mack, NASA Code 129-01-20-01</i> . . . . .	13
B. Solution of Partial Differential Systems <i>F. B. Estabrook, B. K. Harrison, and H. D. Wahlquist,         NASA Code 129-04-20-04</i> . . . . .	17

### TELECOMMUNICATIONS DIVISION

<b>V. Communications Systems Research</b> . . . . .	18
A. Digital Clean-Up Loop Transponder for Sequential Ranging System <i>W. J. Hurd, NASA Code 125-21-09-09</i> . . . . .	18
B. Performance of a Phase-Locked Loop During Loss of Signal <i>W. J. Hurd, NASA Code 125-21-09-09</i> . . . . .	28
C. Optimum Configurations for PSK/PM Systems <i>U. Timor, NASA Code 125-21-18-01</i> . . . . .	33
D. The Critical Problem and Coding Theory <i>N. White, NASA Code 125-21-09-01</i> . . . . .	36

## Contents (contd)

<b>VI. Communications Elements Research</b>	43
A. Spacecraft Antenna Research: Antenna Tolerances	
<i>D. Damlamayan, NASA Code 125-21-16-01</i>	43
<b>VII. Spacecraft Telecommunications Systems</b>	50
A. On the Stability of Second-Order Tracking Loops With Arbitrary Time Delay	
<i>M. K. Simon, NASA Code 125-21-09-06</i>	50
B. Signal Design for Single-Sideband Phase Modulation	
<i>H. D. Chadwick, NASA Code 164-06-50-01</i>	54
C. The Steady-State Performance of a Data-Transition Type of First-Order Digital Phase-Locked Loop	
<i>M. K. Simon, NASA Code 125-21-18-03</i>	59
D. Optimum Modulation Index for a Data-Aided, Phase-Coherent Communication System	
<i>M. K. Simon, NASA Code 125-21-18-03</i>	69

## GUIDANCE AND CONTROL DIVISION

<b>VIII. Spacecraft Power</b>	70
A. Review of Radioisotope Thermoelectric Generators for Outer Planet Missions	
<i>R. S. Caputo, NASA Code 120-27-41-06</i>	70
<b>IX. Guidance and Control Analysis and Integration</b>	76
A. Optimization and Reliability Calculations for Multi-Thermionic- Converter Systems	
<i>R. Szejn and K. Shimada, NASA Code 120-27-40-05</i>	76
B. A Proposed Laser Obstacle Detection Sensor for a Mars Rover	
<i>W. L. Kuriger, NASA Code 186-68-55-01</i>	80
C. Automated Test Techniques for Guidance and Control Subsystems	
<i>R. E. Williamson, NASA Code 186-68-54-04 and 120-60-04-05</i>	90
<b>X. Spacecraft Control</b>	92
A. The Attitude Control of a Flexible Solar Electric Spacecraft	
<i>E. L. Marsh, NASA Code 120-33-16-04</i>	92
B. Finite Element Modeling for Appendage Interaction With Spacecraft Control	
<i>P. W. Likins and E. L. Marsh, NASA Code 125-17-15-09</i>	100
C. Digital Gyro System (Phase I)	
<i>P. J. Hand, NASA Code 186-68-02-44</i>	105

## Contents (contd)

D. Stepper Motor Drive Electronics for the Solar Electric Thrust Vector Control Subsystem <i>W. E. Crawford, NASA Code 120-26-16-07</i> . . . . .	108
---	-----

### ENGINEERING MECHANICS DIVISION

<b>XI. Materials</b> . . . . .	111
A. Spacecraft Adhesives for Long Life and Extreme Environment <i>W. D. Roper, NASA Code 186-68-57-04</i> . . . . .	111
B. Cracking of Filter Layers in a High Performance Solar Cell Filter <i>W. Jaworski, NASA Code 124-09-26-05</i> . . . . .	115
C. Development of a Conical-Gregorian High-Gain Antenna <i>J. G. Fisher, NASA Code 124-08-26-01</i> . . . . .	118
<b>XII. Applied Mechanics</b> . . . . .	120
A. Creep Failure of Randomly Excited Structures <i>J.-N. Yang, NASA Code 124-08-26-02</i> . . . . .	120
B. Reliability of Randomly Excited Structures <i>J.-N. Yang and E. Heer, NASA Code 124-08-26-02</i> . . . . .	128
C. Boundary Layer Equations of a Heated Constrained Spherical Shell <i>H. E. Williams, NASA Code 124-08-26-07</i> . . . . .	136
D. Derivation of the Equations Governing Heated Shallow Shells of Revolution <i>H. E. Williams, NASA Code 124-08-26-07</i> . . . . .	141
<b>XIII. Electronic Parts Engineering</b> . . . . .	145
A. Radiation Effects on Electronic Parts: Literature Search and Data Evaluation <i>K. E. Martin, NASA Code 186-70-51-03</i> . . . . .	145
<b>XIV. Advanced Projects Development</b> . . . . .	148
A. TOPS Mechanical Devices <i>E. E. Sabelman, NASA Code 186-68-57-03</i> . . . . .	148
B. TOPS High-Gain Antenna <i>D. J. Starkey, NASA Code 186-68-57-03</i> . . . . .	157

### ASTRONICS DIVISION

<b>XV. Spacecraft Data Systems</b> . . . . .	160
A. Evaluation of Recording Tape and Heads for Spacecraft Magnetic Tape Recorder Applications <i>J. K. Hoffman, S. H. Kalfayan, and R. H. Silver, NASA Code 186-68-56-09</i> . . . . .	160

## Contents (contd)

B. The Implementation of $m$ -ary Linear Feedback Shift Registers With Binary Devices <i>M. Perlman, NASA Code 125-23-12-01</i> . . . . .	161
C. Extraction of Complete Subgraphs: Command Prefix Code for TOPS <i>N. Deo, NASA Code 186-68-56-01</i> . . . . .	164
D. Jupiter's Electron Dose Calculations on Metal Oxide Semiconductor Structures <i>S. P. Li and J. B. Barengoltz, NASA Code 186-68-56-10</i> . . . . .	166

## ENVIRONMENTAL SCIENCES

<b>XVI. Space Simulation</b> . . . . .	171
A. Holographic Study of Operating Compact-Arc Lamp <i>C. G. Miller and C. L. Youngberg, NASA Code 124-09-27-02</i> . . . . .	171

## PROPULSION DIVISION

<b>XVII. Solid Propellant Engineering</b> . . . . .	175
A. Nondestructive Testing of 1-W, 1-A Electro-Explosive Devices <i>V. J. Menichelli, NASA Code 180-24-51-01</i> . . . . .	175
B. Low-Pressure $L^*$ -Combustion Limits <i>L. D. Strand, NASA Code 128-32-90-01</i> . . . . .	179
C. Diffraction of a High-Order Gaussian Beam by an Aperture <i>A. C. R. Livanos, NASA Code 180-24-51-01</i> . . . . .	181
<b>XVIII. Polymer Research</b> . . . . .	187
A. Viscoelastic Behavior of Elastomers Undergoing Crosslinking Reactions <i>J. Moacanin and J. J. Aklonis, NASA Code 129-03-11-04</i> . . . . .	187
B. Energy Transfer in Bipyridilium (Paraquat) Salts <i>A. Rembaum, V. Hadek, and S. P. S. Yen, NASA Code 129-03-11-03</i> . . . . .	189
C. Electrical Properties of TCNQ Salts of Ionene Polymers and Their Model Compounds <i>V. Hadek, H. Noguchi, and A. Rembaum, NASA Code 129-03-11-03</i> . . . . .	192
D. Evaluation of Magnetic Recording Tapes: A Method for the Quantitative Determination of Stick-Slip <i>R. H. Silver, S. H. Kalfayan, and J. K. Hoffman, NASA Code 186-68-56-09</i> . . . . .	198
<b>XIX. Research and Advanced Concepts</b> . . . . .	201
A. Ion Thruster Electron Baffle Sizing <i>E. V. Pawlik, NASA Code 120-26-16-01</i> . . . . .	201



## Contents (contd)

B. An Ion Thruster Utilizing a Combination Keeper Electrode and Electron Baffle <i>E. V. Pawlik, NASA Code 120-26-16-01</i> . . . . .	204
C. Thrust Subsystem Design for Nuclear Electric Spacecraft <i>T. D. Masek, NASA Code 120-26-16-01</i> . . . . .	207
D. Ion Thruster Connectors <i>E. T. Hopper, NASA Code 120-26-16-01</i> . . . . .	211
<b>XX. Liquid Propulsion</b> . . . . .	213
A. High-Thrust Throttleable Monopropellant Hydrazine Reactors <i>T. W. Price, NASA Code 180-31-52-01</i> . . . . .	213
B. Injector Hydrodynamics Effects on Baffled-Engine Stability— A Correlation of Required Baffle Geometry With Injected Mass Flux <i>R. M. Clayton, NASA Code 128-31-90-04</i> . . . . .	222
 <b>MISSION ANALYSIS</b>	
<b>XXI. Tracking and Orbit Determination</b> . . . . .	233
A. Spectral Factorization in Discrete Systems <i>T. Nishimura, NASA Code 129-04-20-02</i> . . . . .	233
<b>Subject Index</b> . . . . .	236

# I. Advanced Studies

## FUTURE PROJECTS

### A. Lunar Elevation Correction for Gravity

Measurements, R. G. Brereton<sup>1</sup>

The gravitational constant and the mass of the moon are both known to several significant figures. The ratio of the mass of the earth to that of the moon has been recently determined from radio tracking of space probes to be 81.303 (Ref. 1). Thus,

$$\frac{E}{M} = 81.303 \quad (1)$$

Given  $M$ , the mass of the earth, as  $5.975 \times 10^{27}$  g then the mass of the moon  $E$  can be calculated to be  $7.349 \times 10^{25}$  g. Using this value for the mass of the moon, a radius  $r$  of 1738 km, and a value for the gravitational constant  $G$  of  $6.67 \times 10^{-8}$  cm<sup>3</sup>/g-s<sup>2</sup>, then the acceleration of gravity  $g$  on the lunar surface is

$$g = G \frac{M}{r^2} = 162.3 \text{ cm/s}^2 \text{ (162.3 gal)} \quad (2)$$

<sup>1</sup>Assigned to *Apollo* Lunar Exploration Office, NASA, Washington, DC.

Errors in our knowledge of the lunar radius may affect this value by as much as  $\pm 100$  mgal (1 gal = 1 cm/s<sup>2</sup> = 1000 mgal) and crustal anomalies by a like amount, but in general, if the shape of the lunar geoid is known and the value of the acceleration of gravity is known at one point, then the acceleration of lunar gravity can be computed for any other point. And conversely, if the value of the acceleration of gravity is known at all points on the lunar geoid, then the form of the level surface can be deduced.

Anomalies in the gravity field of the moon can be attributed to density contrasts in surface and subsurface rocks, to so-called terrain effects which lead to the free-air, Bouguer, and topographic corrections, and finally, shape effects. In order to use gravity data to interpret lunar structures and shape, it is necessary to make corrections for terrain effects on the data. In the following discussion, the various non-instrument corrections that will commonly be applied to lunar gravity data in order to relate all observations to a common datum are considered.

If the moon were a perfect fluid with no lateral variations in density, its surface would correspond to a

spheroid and the direction of gravity over this level surface would be everywhere perpendicular to it. Obviously, the lunar surface is not always parallel with a spheroid of reference and gravity measurements made on the actual surface will have to be corrected to the reference surface if gravity values are to be used for defining the shape of the moon or crustal structures. The largest corrections here will come from differences in elevation between observation points. Two corrections are required for this. They are the free-air and Bouguer corrections. The free-air correction takes into account the fact that gravity varies with elevation and measurements made at a higher elevation are farther from the center of mass and therefore have a smaller gravity value than measurements at a lower elevation. Since the moon's mass can be considered to be concentrated at its center, then if  $g_0$  represents the value of gravity at its reference surface and  $g_h$  the value at height  $h$ , the inverse-square law tells us that the gravity difference between the two points is

$$\frac{g_0}{g_h} = \frac{(r+h)^2}{r^2} = \frac{r^2 + 2rh + h^2}{r^2} \quad (3)$$

and

$$\frac{g_0}{g_h} = 1 + \frac{2h}{r} + \frac{h^2}{r^2}$$

Because  $h$  will be small compared to  $r$ , the term  $h^2/r^2$  can be neglected, and we can write

$$g_0 - g_h = \frac{2hg_h}{r} \quad (4)$$

This is also the same as the vertical gradient of gravity, where, from Eq. (2),

$$g = G \frac{M}{r^2} \text{ and } \frac{dg}{dr} = \frac{-2GM}{r^3} = \frac{-2g}{r} \quad (5)$$

Substituting,

$$\begin{aligned} \frac{dg}{dr} &= -\frac{2(162.3) \times 10^3}{1738 \times 10^3} \text{ mgal/m} \\ &= -0.187 \text{ mgal/m} \end{aligned} \quad (6)$$

The lunar free-air correction adjusts gravity observations for points made at different elevations to a common datum, but it does not account for the change in mass under the station in order to change its elevation. In

other words, to change the elevation of surface observation points requires that a specific volume of surface rock material is either added to or subtracted from the datum. This effect is called the Bouguer correction, which is always opposite in sign to the free-air correction. It is usual to assume either a slab or cylinder of infinite horizontal dimensions and finite vertical dimensions in order to compute an expression for the Bouguer correction. The value for the Bouguer correction is

$$\Delta B = 2\pi G \rho h = 0.042\rho \text{ mgal/m}$$

If we assume an average density  $\rho$  for lunar crustal rocks of say  $2.67 \text{ g/cm}^3$ , then the Bouguer correction is

$$0.112 \text{ mgal/m}$$

and the combined lunar elevation correction, that is, the free-air correction minus the Bouguer correction, is approximately  $0.075 \text{ mgal/m}$ . Interestingly, this value is less than one-half of the correction required on earth, and this has an important effect on gravimetric field operations on the moon since it reduces the needed accuracy for elevation control. It should be realized that the magnitude of the Bouguer correction will change for sedimentary areas in contrast to igneous areas, and also in response to regional investigations in contrast to local studies. The Bouguer correction is very sensitive to rock densities and in a local survey in hilly country it can serve as a precise tool for determining the average density of local rock units. Table 1 gives a range of lunar elevation corrections as a function of density.

**Table 1. Lunar elevation corrections**

Free-air correction, mgal/m	Density $\rho$ g/cm <sup>3</sup>	Bouguer correction, mgal/m	Elevation correction, mgal/m
0.187	1.5	0.063	0.124
0.187	2.0	0.084	0.103
0.187	2.5	0.105	0.082
0.187	3.0	0.126	0.061

A topographic correction will only be required in hilly areas. It may be thought of as a supplemental correction to the smooth infinite slab hypothesized in making the Bouguer correction, for it accounts for all material projecting above the Bouguer plain and all material needed to fill in hollows below the plain. It is always positive and, for most survey work, it will not be troublesome.

Tidal corrections for gravity measurements on the lunar surface, caused by the differential attraction of the moon by the earth as the former moves between perigee and apogee, may be larger than expected if the crust of the moon yields to the external force. The external force difference between perigee and apogee is very small, of the order of a milligal, but if the lunar crust yields to this force, then gravity readings on the surface will be affected. It is not likely that this effect will be significant for mobile surface exploration since the effect probably amounts to less than a milligal per lunation.

The scientific objectives and instruments for lunar gravity investigations were discussed in SPS 37-57, Vol. III, pp. 1-6, and SPS 37-62, Vol. III, pp. 1-5.

#### Reference

1. Melbourne, W. G., "The Determination of Planetary Masses From Radio Tracking of Space Probes and Planetary Radar," *Symposium C Space Probes, Part I*, presented at the 12th Plenary Meeting of COSPAR, Prague, Czechoslovakia, May 11-24, 1969.

## II. Environmental Requirements

### PROJECT ENGINEERING DIVISION

#### A. Multi-Phase Ammonia-Water System (Rev. 1)<sup>1</sup>, C. Haudenschild

##### 1. Introduction

In conjunction with the SR/AD Space Vehicle Design Criteria Monograph Program JPL develops monographs on various planetary environments. In the specific problem of modeling atmospheres of Jupiter and Saturn for the monographs one needs to consider the properties of the major condensible constituents—ammonia and water.

Pertinent information in references is primarily in a scattered tabular form.

In Ref. 1 Lewis makes a graphical representation of the  $\text{NH}_3\text{--H}_2\text{O}$  system from data in Refs. 2 and 3. For the production of model atmospheres it was felt that analytic expressions, in addition to a graphical display, would be very useful. So, an attempt was made to produce empirical equations relating quantities of interest displayed

<sup>1</sup>This article is a revision of "Multi-Phase Ammonia-Water System," C. Haudenschild, in SPS 37-64, Vol. III, pp. 4-9, and contains revised tabular data.

in Table 1. The equations give partial vapor pressure over solid  $\text{NH}_3$ , solid  $(\text{NH}_3)_2\text{H}_2\text{O}$  hydrate, solid  $\text{NH}_3\text{H}_2\text{O}$  hydrate, solid  $\text{H}_2\text{O}$ , and  $\text{NH}_3\text{--H}_2\text{O}$  solution phases. In addition, along boundaries of the regions, equations for freezing curve temperatures as a function of concentration are given. Finally a phase diagram similar to Lewis's is produced, displaying the equations graphically.

Table 1. Definition of terms

Symbol	Units	Definition
$T$	$^{\circ}\text{K}$	Temperature
$C$	$0.0 \leq C \leq 1.0$	Molar concentration of $\text{NH}_3$ in $\text{H}_2\text{O}$ solution
$P_2$	Torr	$\text{H}_2\text{O}$ vapor partial pressure
$P_3$	Torr	$\text{NH}_3$ vapor partial pressure

##### 2. Forms Relating the Quantities of Interest

To proceed, we must choose appropriate equation forms that can be fitted to the data available and incorporate our understanding of the physics of the situation. The following forms were chosen:

In the pure solid regions of  $\text{NH}_3$  or  $\text{H}_2\text{O}$ , equations of the form of the Clausius Clapeyron equation are chosen to relate  $P$  and  $T$  (Ref. 1):

$$\log P = \alpha - \beta/T \quad (1)$$

The constants  $\alpha$  and  $\beta$  can be determined numerically from any two data points at which  $P$  and  $T$  are specified.

Forms for the hydrate regions were taken from Ref. 1. For  $(\text{NH}_3)_2\text{H}_2\text{O}$  we get the equation

$$(2/3) \log P_3 + (1/3) \log P_2 = \alpha - \beta/T \quad (2)$$

Here the ratio of the coefficients of the logarithms of the pressures is the same as the ratio of  $\text{NH}_3$  and  $\text{H}_2\text{O}$  in the hydrate—that is, two to one—and the basic form of the Clausius Clapeyron equation is retained. For  $\text{NH}_3\text{H}_2\text{O}$  the appropriate equation is

$$(1/2) \log P_2 + (1/2) \log P_3 = \alpha - \beta/T \quad (3)$$

In each hydrate region one can obtain the two constants  $\alpha$  and  $\beta$  numerically from two data points, just as for the pure solid regions.

From the phase diagram displayed in Ref. 1 (p. 367) we infer that solution region equations obey:

$$\left( \frac{\partial \log P_3}{\partial \log P_2} \right)_T = \text{constant, independent of temperature} \quad (4)$$

That is, the isoconcentration lines are straight. Also, we infer that

$$\left( \frac{\partial \log P_3}{\partial \log P_2} \right)_T = - \frac{1 - C}{C}, \text{ also independent of temperature.} \quad (5)$$

This matches the slope of the solution region isotherm with the slopes in the hydrate regions, at those points where the molecular number ratios are identical in the two regions. Equations were chosen which retain a Clausius Clapeyron dependence and satisfy the above requirements; namely

$$\log P_2 = \log (1 - C) + A_0 + A_2 C^2 - \frac{a_0 + a_2 C^2}{T} \quad (6)$$

and

$$\log P_3 = \log C + B_0 + A_2 (C^2 - 2C) - \frac{b_0 + a_2 (C^2 - 2C)}{T} \quad (7)$$

The six constants  $A_0$ ,  $A_2$ ,  $B_0$ ,  $a_0$ ,  $a_2$ , and  $b_0$  can be obtained numerically from appropriately chosen combinations of tabular values of  $C$ ,  $T$ ,  $\log P_2$ , and  $\log P_3$ .

Finally, freezing curves can be obtained by equating the solution region partial pressures to solid region ones, and eliminating the pressures. There result equations which give  $T$  as a function of  $C$ .

### 3. Numerical Relations Among Quantities of Interest

Table 2 lists numerical equations for regions of solid and solution phases. Partial pressures of ammonia and water are given as function of temperature or of temperature

Table 2. Solid and solution equations

Region	Data source	Form	Equation
Solid $\text{H}_2\text{O}$	Table 3 (underlined entries)	Eq. (1)	$\log P_2 = 10.447 - 2673.01/T$ (8)
Solid $\text{NH}_3$	Table 4 (underlined entries)	Eq. (1)	$\log P_3 = 10.312 - 1691.77/T$ (9)
$(\text{NH}_3)_2\text{H}_2\text{O}$	Table 5 († entries)	Eq. (2)	$(2/3) \log P_3 + (1/3) \log P_2 = 9.777 - \frac{2051.185}{T}$ (10)
$\text{NH}_3\text{H}_2\text{O}$	Table 5 (§ entries)	Eq. (3)	$(1/2) \log P_2 + (1/2) \log P_3 = 9.790 - \frac{2219.28}{T}$ (11)
Solution	Table 5 (underlined entries)	Eq. (6)	$\log P_2 = \log (1 - C) + 9.488 + 1.743 C^2 - \left( \frac{2406.2 + 878.192 C^2}{T} \right)$ (12)
		Eq. (7)	$\log P_3 = \log C + 9.906 + 1.743 (C^2 - 2C) - \left( \frac{2149.65 + 878.192 (C^2 - 2C)}{T} \right)$ (13)

and concentration. The data from which the numerical coefficients in Table 2 were derived are found in Tables 3, 4, and 5.

Table 6 lists numerical freezing curves temperature equations as functions of concentration. These are formed by equating the source equation and eliminating  $\log P_2$ ,  $\log P_3$ , or both.

#### 4. Comparison With Data

Table 3 compares the values of  $\log P_2$  for solid  $H_2O$  from the tabular sources listed with  $\log P_2$  from Eq. (8). The largest discrepancy shown is near  $-99^\circ C$  and amounts to 0.02 in  $\log P_2$ .

Table 3. Vapor pressure over solid  $H_2O$  ice

Reference and page	$T$ , $^\circ C$	$T$ , $^\circ K$	$P_2$ , torr (From Ref.)	$\log P_2$ , torr (From Ref.)	$\log P_2$ , torr (From Eq. 8)
4, 1474	-99.0	174.16	$1.2 \times 10^{-5}$	-4.92	-4.901
4, 1474	-68.0	<u>204.46</u>	$2.36 \times 10^{-3}$	<u>-2.627</u>	-2.627
4, 1474	-46.0	227.16	0.0481	-1.318	-1.320
4, 1474	-20.1	253.06	0.0769	-1.114	-1.116
4, 1475	0.0	<u>273.16</u>	4.579	<u>0.661</u>	0.661

Underlined entries indicate data used to obtain constants in Eq. (8).

Table 4. Vapor pressure over solid  $NH_3$  ice

Reference and page	$T$ , $^\circ C$	$T$ , $^\circ K$	$P_3$ , torr (From Ref.)	$\log P_3$ , torr (From Ref.)	$\log P_3$ , torr (From Eq. 9)
2, 362	-77.6	<u>195.56</u>	45.82	<u>1.661</u>	1.661
3, D-140	-79.2	193.96	40.0	1.602	1.590
2, 362	-84.4	188.76	22.36	1.348	1.349
4, D-140	-91.9	181.26	10.0	1.000	0.979
3, D-140	-109.1	<u>164.06</u>	1.00	<u>0.000</u>	0.000

Underlined entries are data used to obtain constants in Eq. (9).

Table 4 compares values of  $\log P_3$  for solid  $NH_3$  from the tabular sources listed with  $\log P_3$  from Eq. (9). The largest discrepancy shown is near  $-91.9^\circ C$  and amounts also to 0.02 in  $\log P_3$ .

Table 5 was produced to display the accuracy of the solution and freezing curve equations. Column 1 gives the concentration considered on any particular line. Column 2 lists the tabular source from which  $T$  ( $^\circ K$ ),  $\log P_2$ , and  $\log P_3$  in Cols. 3, 4, and 5 were obtained. When an S in the "remarks" column appears, we are considering the solution region and the temperature in Col. 3 is transferred to Col. 7. F denotes freezing curve; hence,  $T$  in Col. 7 is computed from the appropriate freezing curve equation. Finally, based on the values of  $C$  in Col. 1 and  $T$  in Col. 7,  $\log P_2$  and  $\log P_3$  in Cols. 8 and 9 are obtained from Eqs. (12) and (13).

The largest discrepancy in  $\log P_3$  (Cols. 5 and 9 in Table 5) in the solution region occurs at  $C = 0.60$  and  $T = 189.31$  and is 0.15 (a factor of 1.4). Table values for  $\log P_2$  are only given in the interval  $0 \leq C \leq 0.2395$ . The largest discrepancy in  $\log P_2$  (Cols. 4 and 8 in Table 5) in the solution region occurs at  $C = 0.2181$  and  $T = 316.16$  and is 0.14 (a factor of 1.4).

For points along the freezing curves, discrepancies of  $\log P_2$  and  $\log P_3$  are magnified. That is, the freezing curve temperature is used to calculate the log pressure instead of the table temperature, as in the solution region. Hence, a double error is produced. The largest discrepancy in  $\log P_3$ , listed in Table 5, occurs at  $C = 0.25$  and is 0.387 (a factor of 2.4). Values of  $\log P_2$  are listed in the tables at only five points along the freezing curve, all of which have  $0 \leq C \leq 0.2$ . Among these the largest discrepancy occurs at  $C = 0.20$  with  $\log P_2$  0.33 below the tabular value (a factor of 2.1). One would expect relatively large discrepancies along the freezing curve in the region of  $0.2 \leq C \leq 0.35$ , because  $T$  is a very sensitive function of  $C$  there (Fig. 1). In fact, if one scans the freezing curve temperature, the largest discrepancy of  $9.8^\circ K$  occurs at

Table 5. Vapor pressure over  $NH_3-H_2O$  solution

C	Reference and page	$T$ , $^\circ K$ (from Ref.)	$\log P_2$ , torr (from Ref.)	$\log P_3$ , torr (From Ref.)	Remarks	$T$ , $^\circ K$ (See Subsection 4.)	$\log P_2$ , torr (Eq. 12)	$\log P_3$ , torr (Eq. 13)
1.00	2, 362	—	—	—	S	<u>273.16</u>	$-\infty$	<u>3.508</u>
1.00	↓	195.56	—	1.661	F	<u>195.59</u>	$-\infty$	<u>1.662</u>
0.95	↓	192.66	—	1.537	F	192.95	-6.818	1.544
0.95	↓	213.16	—	2.194	S	213.16	-5.246	2.170

Table 5 (contd)

C	Reference and page	T, °K (from Ref.)	log P <sub>2</sub> , torr (from Ref.)	log P <sub>3</sub> , torr (From Ref.)	Remarks	T, °K (See Subsection 4.)	log P <sub>2</sub> , torr (Eq. 12)	log P <sub>3</sub> , torr (Eq. 13)
0.90	2, 362	213.16	—	2.167	S	213.16	-4.726	2.129
0.90		188.76	—	1.367	F	189.01	-6.594	1.361
0.85		184.16	—	1.155	F	183.71	-6.628	1.103
0.85		213.16	—	2.135	S	213.16	-4.341	2.074
0.812		180.16	—	0.956	Q, F	178.74†	-6.790†	0.847†
0.80		181.96	—	1.022	F	181.35	-6.463	0.931
0.80		223.16	—	2.365	S	223.16	-3.396	2.281
0.75		189.31	—	1.248	F	189.45	-5.442	1.146
0.70		223.16	—	2.245	S	223.16	-2.891	2.113
0.70		193.46	—	1.320	F	193.61	-4.831	1.190
0.6667		194.36	—	1.306	N, F	194.36†	-4.603†	1.137†
0.65		194.16	—	1.240	F	194.19	-4.533	1.088
0.65		223.16	—	2.156	S	223.16	-2.677	2.010
0.60		223.16	—	2.043	S	223.16	-2.482	1.893
0.60		189.31	—	0.934	F	191.36	-4.509	0.841
0.5855		186.81	—	0.813	Q, F	189.91	-4.553	0.740
0.55		191.16	—	0.0820	F	191.77	-4.264	0.699
0.50		193.96	—	0.732	N, F	193.91‡	-3.918‡	0.609‡
0.50		243.16	—	2.278	S	243.16	-1.176	2.166
0.45		193.16	—	0.508	S	193.16	-3.796	0.386
0.45		191.31	—	0.432	F	191.76	-3.894	0.328
0.40		184.16	—	-0.174	F	185.24	-4.203	-0.178
0.35		263.16	—	2.248	S	263.16	-0.038	2.202
0.35		173.26	—	-1.046	S	174.03	-4.930	-0.995
0.346		172.86	—	-1.081	Q, F	<u>172.96‡</u>	<u>-5.007‡</u>	<u>-1.075‡</u>
0.30		273.16	—	2.292	S	273.16	0.392	2.264
0.30		204.46	—	0.204	F	196.20	-3.177	-0.179
0.25		227.16	—	0.806	F	217.36	-1.851	0.419
0.2395	3, 1467	273.16	0.448	2.067	S	273.16	0.476	2.036
0.2181		316.16	1.578	2.760	S	316.16	1.721	2.848
0.2054		273.16	0.477	1.916	S	273.16	0.517	1.892
0.20		241.91	-0.600	1.045	F	234.93	-0.931	0.775
0.1546		273.16	0.613	1.710	S	273.16	0.571	1.646
0.15	2, 362	253.00	—	0.125	F	249.46	-0.268	0.958
0.10	2, 362	261.21	0.215	1.076	F	<u>261.34</u>	<u>0.219</u>	0.988
0.0987	3, 1467	333.16	2.132	2.575	S	333.16	2.212	2.616
0.0964		273.16	0.724	1.394	S	273.16	0.622	1.291
0.05		267.81	0.466	0.868	F	270.87	0.578	0.815
0.0408		333.16	2.159	2.136	S	<u>333.16</u>	2.246	<u>2.136</u>
0.00		—	—	—	S	373.16	3.040	-∞
0.00	2, 363	273.16	0.661	—	F	278.22	0.839	-∞

†Data used to obtain constants in Eq. (10) for (NH<sub>3</sub>)<sub>2</sub>H<sub>2</sub>O hydrate.‡Data used to obtain constants in Eq. (11) for (NH<sub>3</sub>)<sub>2</sub>H<sub>2</sub>O hydrate.

Underlined entries indicate data used to obtain constants in Eqs. (12) and (13) for solution region.

S = point of consideration lies in solution region.

F = point of consideration lies on freezing curve.

Q = quadrupole point.

N = nose of freezing curve where the H<sub>2</sub>O and NH<sub>3</sub> number ratio are equal in the solution and adjacent hydrate regions.



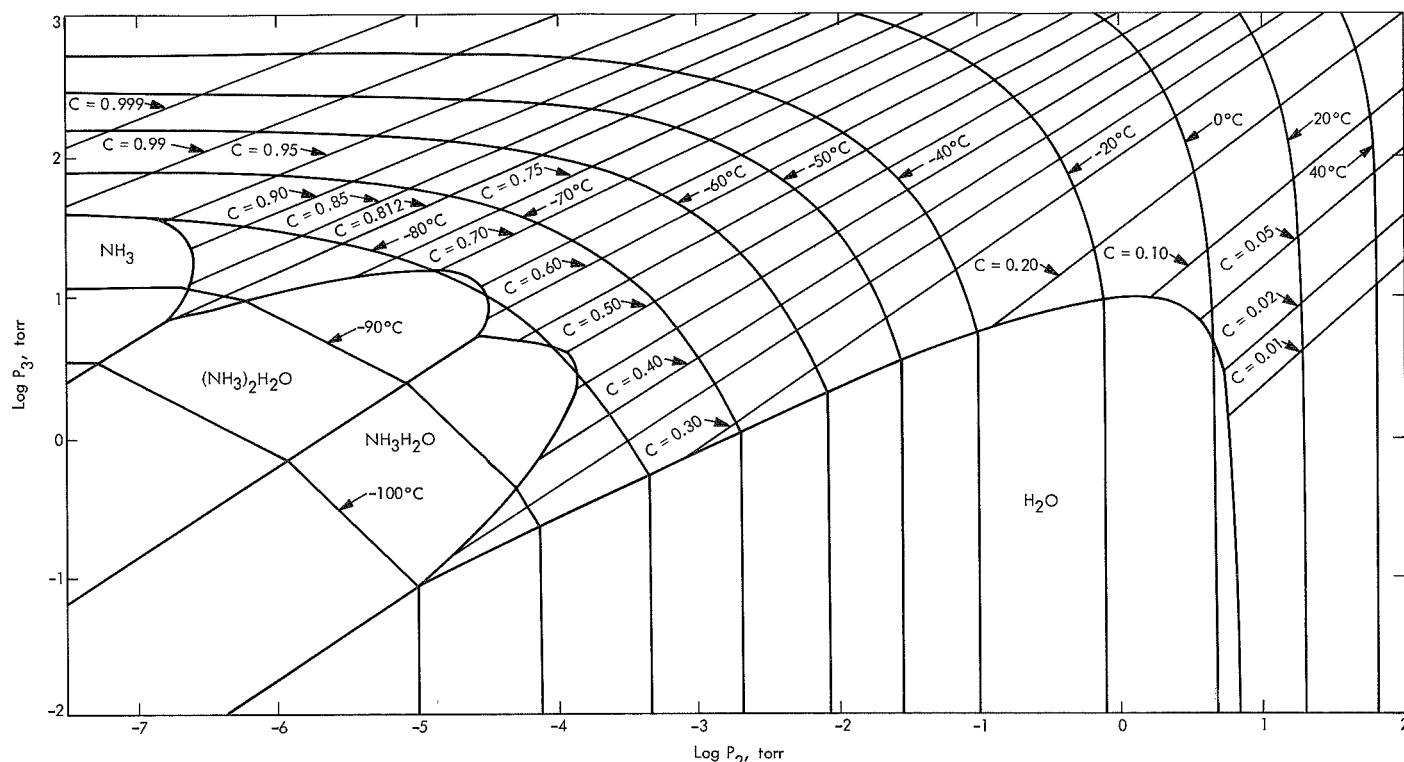
Table 6. Freezing curve equations

Region of concentration	Quantities eliminated	Source equations	Freezing curves
$0.812 \leq C \leq 1.00$ Solid $\text{NH}_3$	$\log P_3$	(9), (13)	$T = \frac{457.88 + 878.192(C^2 - 2C)}{\log C - 0.406 + 1.743(C^2 - 2C)} \quad (14)$
$0.5855 \leq C \leq 0.812$ $(\text{NH}_3)_2\text{H}_2\text{O}$	$\log P_2$ $\log P_3$	(10), (12), (13)	$T = \frac{183.982 - 1170.92C + 878.192C^2}{(1/3) \log [C^2(1 - C)] - 0.01033 - 2.324C + 1.743C^2} \quad (15)$
$0.346 \leq C \leq 0.5855$ $\text{NH}_3\text{H}_2\text{O}$	$\log P_2$ $\log P_3$	(11), (12), (13)	$T = \frac{117.29 + 1756.384(C^2 - C)}{\log [-(C^2 - C)] - 0.186 + 3.486(C^2 - C)} \quad (16)$
$0.0 \leq C \leq 0.346$ Solid $\text{H}_2\text{O}$	$\log P_2$	(8), (12)	$T = \frac{-266.81 + 878.192C^2}{\log (1 - C) - 0.959 + 1.743C^2} \quad (17)$

$C = 0.25$ . The next largest freezing curve discrepancies occur at the quadrupole points: at  $C = 0.5855$ ,  $T$  is  $3^\circ\text{K}$  below table values and at  $C = 0.812$ ,  $T$  is  $1.4^\circ\text{K}$  below table value. However, the  $\log P_3$  values remain within 0.11 of the tabular values there.

A phase diagram of the  $\text{NH}_3\text{--H}_2\text{O}$  system was constructed (Fig. 1) for comparison with the phase diagram

from Ref. 1. Lines of constant concentration are straight and range from  $0.01 \leq C \leq 0.999$ . Isothermal curves range from  $T = -100^\circ\text{C}$  to  $T = +40^\circ\text{C}$ . Equations (12) and (13) were used to plot the solution region. The freezing curves (Eqs. 14 to 17) were used to plot the solid-solution boundaries. The lines separating two solid phases are found by equating two adjacent solid pressure equations and finding the ratio of  $\log P_2$  to  $\log P_3$ .

Fig. 1. Phase diagram for the  $\text{NH}_3\text{--H}_2\text{O}$  system prepared from Eqs. (8) through (17)

## 5. Conclusions

Empirical equations produced for all regions agree remarkably well with the table values. Discrepancies over the entire region shown in Fig. 1 are within a factor of about 2.5, with the poorest fit occurring along the freezing curve in the vicinity of  $C \simeq 0.25$ . Qualitatively the results agree with those of Lewis (Ref. 1), who employed graphical representations of the  $\text{NH}_3\text{-H}_2\text{O}$  system rather than analytic ones such as those discussed above. It appears likely that much better fits cannot be obtained without more data and more complicated relations among the quantities of interest.

The relations derived enable one to predict cloud composition and temperature of formation in a model atmosphere containing ammonia and water required for space vehicle design criteria. In particular, the equations were

used in preparation of the Jupiter and Saturn environmental design criteria monograph atmospheric models.<sup>1</sup>

---

<sup>1</sup>Divine, T. N., *The Planet Jupiter (1970)*, Monograph to be published by National Aeronautics and Space Administration, Washington, D.C.

## References

1. Lewis, J. S., "The Clouds of Jupiter and the  $\text{NH}_3\text{-H}_2\text{O}$  and  $\text{NH}_3\text{-H}_2\text{S}$  System," *Icarus*, Vol. 10, No. 3, 1969, pp. 365-378.
2. *International Critical Tables*, Vol. 3. Edited by F. W. Washburn. McGraw-Hill Book Co., New York, 1928.
3. *Handbook of Chemistry and Physics*, (50th Ed.), Chemical Rubber Pub. Co., Cleveland, Ohio, 1969.
4. *Handbook of Chemistry*, (Rev. 10th Ed.). Edited by N. A. Lange. McGraw-Hill Book Co., New York, 1967.

### III. Computation and Analysis

#### DATA SYSTEMS DIVISION

##### A. Unitary Similarity Transformation of a Hermitian Matrix to a Real Symmetric Tridiagonal Matrix, R. J. Hanson

Two known theorems of finite dimensional linear algebra are the following:

**Theorem 1.** *Every  $n \times n$  real symmetric matrix  $\mathbf{A}$ , ( $\mathbf{A}^T = \mathbf{A}$ ), can be written as*

$$\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^T \quad (1)$$

*Here  $\mathbf{V}$  is an  $n \times n$  real orthonormal matrix, ( $\mathbf{V}^T\mathbf{V} = \mathbf{V}\mathbf{V}^T = \mathbf{I}_n = n \times n$  identity matrix), and  $\mathbf{D}$  is a diagonal matrix of real scalars.*

**Theorem 2.** *Every  $n \times n$  hermitian matrix  $\mathbf{A}$ , ( $\bar{\mathbf{A}}^T = \mathbf{A}$ ), can be written as*

$$\mathbf{A} = \mathbf{V}\mathbf{D}(\bar{\mathbf{V}})^T \quad (2)$$

*Here  $\mathbf{V}$  is an  $n \times n$  complex unitary matrix ( $(\bar{\mathbf{V}})^T\mathbf{V} = \mathbf{V}(\bar{\mathbf{V}})^T = \mathbf{I}_n$ ) and  $\mathbf{D}$  is a diagonal matrix of real scalars. In Eq. (2) the bar over  $\mathbf{V}$  denotes complex conjugation.*

These theorems have been well known for a long time (Ref. 1). Frequently there are applications where the

matrices  $\mathbf{V}$  and  $\mathbf{D}$  of Eqs. (1) and (2) must be computed numerically.

A great deal of research has been devoted to this numerical problem. This has resulted in one of the most astounding algorithms in all of numerical mathematics. It is usually called the "QR algorithm for symmetric tridiagonal matrices" (Ref. 2). *With this algorithm the construction of the matrices  $\mathbf{V}$  and  $\mathbf{D}$  of Eq. (1) can be done with remarkable speed and accuracy in the special case that  $\mathbf{A}$  is real and tridiagonal, or zero above and below the first super- and sub-diagonal (Refs. 3 and 4).*

The remarks which follow will show how an orthonormal or unitary transformation  $\mathbf{W}$  can be explicitly constructed so that

$$\mathbf{A} = \mathbf{W}\mathbf{B}\mathbf{W}^T \quad (3)$$

if  $\mathbf{A}$  is symmetric, or

$$\mathbf{A} = \mathbf{W}\mathbf{B}(\bar{\mathbf{W}})^T \quad (4)$$

if  $\mathbf{A}$  is hermitian.

The matrix  $\mathbf{B}$  of Eq. (3) or (4) is real, symmetric, and tridiagonal. Thus the powerful and elegant QR algorithm

can be used to compute  $\mathbf{X}$  and  $\mathbf{D}$  of the right side of the equation:

$$\mathbf{B} = \mathbf{XDX}^T \quad (5)$$

Here  $\mathbf{X}$  is an orthonormal  $n \times n$  matrix and  $\mathbf{D}$  is a real diagonal matrix of scalars. Finally,  $\mathbf{V} = \mathbf{WX}$ . The details regarding the reduction of real symmetric matrices to tridiagonal form have appeared in the literature in detailed algorithmic form (Ref. 5).

To this author's knowledge the details for the reduction of the general hermitian matrix to tridiagonal, real symmetric form have not appeared in the literature.

It should be remarked that there is a naive method for finding the expansion of Eq. (2) using a computational capacity for the case of real symmetric matrices. To see this, write the  $n \times n$  hermitian matrix

$$\mathbf{A} = \mathbf{A}_1 + i\mathbf{A}_2, \quad (i = \sqrt{-1})$$

where the  $\mathbf{A}_j$ , ( $j = 1, 2$ ), are real.

Then the  $2n \times 2n$  symmetric matrix

$$\hat{\mathbf{A}} = \begin{pmatrix} \mathbf{A}_1 & -\mathbf{A}_2 \\ \mathbf{A}_2 & \mathbf{A}_1 \end{pmatrix}$$

has each of the eigenvalues of  $\mathbf{A}$  repeated as pairs, and the corresponding  $2n$  real eigenvectors  $\hat{\mathbf{v}}_j = (\hat{\mathbf{v}}_{1j}^T, \hat{\mathbf{v}}_{2j}^T)^T$  of  $\hat{\mathbf{A}}$ , where each  $\hat{\mathbf{v}}_{lj}$ , ( $l = 1, 2$ ), is a real  $n$  vector, are such that  $\mathbf{v}_j = \hat{\mathbf{v}}_{1j} + i\hat{\mathbf{v}}_{2j}$  is an eigenvector of  $\mathbf{A}$ . The matrix  $\mathbf{V}$ , whose columns are the appropriate members of the  $\mathbf{v}_j$ , is unitary.

This reformulation of Eq. (2) is *unsatisfactory* for at least three reasons:

- (1) About twice as much storage and computer time is used for computing the expression of Eq. (1) for a  $2n \times 2n$  real matrix as for the expansion of Eq. (2) for an  $n \times n$  hermitian matrix.
- (2) Accuracy is unnecessarily lost with twice as many computer operations.
- (3) There is a possible programming difficulty in "re-shuffling" the eigenvalues and eigenvectors of  $\hat{\mathbf{A}}$  to obtain an expansion as in Eq. (2).

For these reasons a subroutine HERMQR has been implemented for the computation of the hermitian eigenproblem. (The subroutine is currently available in the

JPL Computation and Analysis Section Subroutine Library.)

To ease the notation which follows we define

$$\mathbf{C}^H = (\bar{\mathbf{C}})^T \quad (6)$$

Thus  $\mathbf{C}^H$  denotes the conjugate transpose of any general  $m \times n$  complex matrix  $\mathbf{C}$ .

We next state

**Lemma 1. (Householder Unitary Transformation).** Let  $\mathbf{x} = [x_1, \dots, x_n]^T$  be an arbitrary complex  $n$  vector. Let  $k \geq 0$  be an integer such that  $k + 1 \leq n$ .

Define

- (1) The integer  $\ell = k + 1$
- (2) The complex number  $(c, s) = \begin{cases} x_\ell / |x_\ell|, & x_\ell \neq 0, (c^2 + s^2 = 1)^\dagger \\ (1, 0), & x_\ell = 0 \end{cases}$
- (3) The scalar  $t = -(c, s) (|x_\ell|^2 + \dots + |x_n|^2)^{1/2}$
- (4) The scalars  $u_\ell = x_\ell - t, u_j = x_j, (j = \ell + 1, \dots, n)$
- (5) The vector  $\mathbf{u} = [0, \dots, 0, u_\ell, u_{\ell+1}, \dots, u_n]^T$
- (6) The scalar  $b = \bar{t}u_\ell$

Then the transformation

$$\mathbf{Q} = \mathbf{I}_n + \mathbf{u}\mathbf{u}^H/b \quad (7)$$

satisfies:

- (1)  $\mathbf{Q}$  is unitary, and hermitian  
( $\mathbf{Q}\mathbf{Q}^H = \mathbf{Q}^H\mathbf{Q} = \mathbf{Q}^2 = \mathbf{I}_n$ )
- (2)  $\mathbf{Q}\mathbf{x} = [x_1, \dots, x_n, t, 0, \dots, 0]^T$

(See Ref. 6.)

We also state

**Lemma 2.** Let  $\mathbf{A}$  be an  $n \times n$  hermitian matrix with  $n \geq 3$  and write

$$\mathbf{A} = \begin{pmatrix} a & \mathbf{x}^H \\ \mathbf{x} & \mathbf{B} \end{pmatrix}$$

where  $a$  is a real scalar,  $\mathbf{x}$  is an  $(n-1)$  complex vector and  $\mathbf{B}$  is  $(n-1) \times (n-1)$  and hermitian.

<sup>†</sup>SPS 37-55, Vol. III, pp. 28-30.

From Lemma 1, define  $Q$  so that

$$Q[a, \mathbf{x}^T]^T = [a, t, 0, \dots, 0]^T$$

where  $|t| = \|\mathbf{x}\| = \text{euclidean length of } \mathbf{x}$ .

Then the similarity transformation  $A_1 = QAQ^H = QAQ$  satisfies:

$$(1) \quad A_1 = \begin{pmatrix} a & \bar{t} & 0 & \dots & 0 \\ t & & & & \\ 0 & & B_1 & & \\ \vdots & & & \ddots & \\ 0 & & & & \end{pmatrix}$$

where  $B_1$  is hermitian.

(2)  $B_1$  is computed from the identity

$$A_1 = A + \mathbf{y}\mathbf{u}^H + \mathbf{u}\mathbf{y}^H$$

where  $\mathbf{y} = (\mathbf{I}_n + \mathbf{u}\mathbf{u}^H/2b)\mathbf{A}\mathbf{u}/b$ .

By successive application of Lemma 2,  $n-2$  transformations,  $Q_1, \dots, Q_{n-2}$ , can be constructed so that

$$Q_{n-2} \dots Q_1 A Q_1 \dots Q_{n-2} = T$$

where

$$(8) \quad T = \begin{pmatrix} a_1 & \bar{b}_1 & & & 0 \\ & \cdot & & & \\ b_1 & & & \cdot & \\ & \cdot & & \cdot & \\ & & & & \bar{b}_{n-1} \\ 0 & & & b_{n-1} & a_n \end{pmatrix}$$

is a tridiagonal hermitian matrix.

In conclusion, the following are summarized:

- (1) The matrix product  $Y = Q_1 \dots Q_{n-2}$  can overwrite the entirety of the working array which holds  $A$  in storage without additional computer storage requirements.
- (2) A diagonal unitary transformation,  $Z$ , can easily be constructed so that  $B = ZTZ^H$  is real, tridiagonal and symmetric. The eigenvalues and eigenvectors of  $B$  can be found by the QR algorithm.
- (3) The entire computation can be arranged so that (a) only the upper triangular part of  $A$  is used for computing, (b) the eigenvectors for  $A$  overwrite the entirety of the working array which holds  $A$  in storage, and (c) only  $n$  additional complex entries are required in the machine.

## References

1. Gantmacher, F. R., *Matrix Theory*, Vol. I, pp. 331-338. Chelsea Publishing Co., Inc., New York, 1959.
2. Bowdler, H., Martin, R. S., Reinsch, C. H., and Wilkinson, J. H., "The QR and QL Algorithms for Symmetric Matrices," *Numer. Math.*, Vol. II (Handbook Series Linear Algebra), pp. 293-306, 1968.
3. Martin, R. S., and Wilkinson, J. H., "The Implicit QL Algorithm," *Numer. Math.*, Vol. 12 (Handbook Series Linear Algebra), pp. 377-383, 1968.
4. Wilkinson, J. H., "Global Convergence of the QR Algorithm," *Proceedings of the International Federation for Information Processing Congress*, Vol. 1, pp. 130-133, 1968.
5. Martin, R. S., Reinsch, C. H., and Wilkinson, J. H., "Householder's Tridiagonalization of a Symmetric Matrix," *Numer. Math.*, Vol. 11 (Handbook Series Linear Algebra), pp. 181-195, 1968.
6. Wilkinson, J. H., *The Algebraic Eigenvalue Problem*, pp. 48-50. Clarendon Press, Oxford, London, 1965.

## IV. Physics

### SPACE SCIENCES DIVISION

#### A. Response of Supersonic Laminar Boundary Layer to a Moving External Pressure Field,

L. M. Mack

##### 1. Introduction

In a supersonic wind tunnel the laminar boundary layer on a model in the test section is irradiated by sound waves generated by the turbulent boundary layers on the tunnel side walls. It has long been thought that the effect of these sound waves is to cause the early transition from laminar to turbulent flow that is observed in wind tunnels. It is probably not feasible to calculate the response of a laminar boundary layer to a sound field as complicated as that existing in a wind tunnel, but a simpler, although closely related, problem can be solved by making use of the methods developed for linear stability theory. This problem was formulated many years ago by Lees and Lin (Ref. 1), but has never been solved. The boundary layer is replaced by a parallel flow having the velocity and temperature profiles of the true boundary layer at some particular location, and the radiating turbulent boundary layer on the tunnel ceiling is replaced by a wavy wall moving in the direction of the flow with velocity  $c_r^*$ , or dimensionless velocity  $c_r = c_r^*/U_1^*$ , where

$U$  is the mean velocity and the subscript 1 refers to the free stream. The asterisks refer to dimensional quantities. When the waviness is of small amplitude and the relative free-stream Mach number  $M_{R1} = M_1(1 - c_r)$  is supersonic ( $M_1$  = free-stream Mach number), a disturbance field is produced which propagates along the family of Mach waves which extend in the downstream direction. It is the interaction of this moving pressure field with the specified parallel flow which is to be studied.

##### 2. Method of Solution

The laminar stability theory has been extensively studied over the last few years<sup>1</sup> by means of two computer programs, one for the inviscid theory and the other for the viscous theory. Both of these programs make use of analytic solutions of the stability equations in the free stream to provide the initial conditions for the numerical integration, which extends from the edge of the boundary layer to the wall. In the inviscid theory there are two solutions, and when  $M_{R1} > 1$ , one solution represents a family of incoming Mach waves, and the second solution

---

<sup>1</sup>Mack, L. M., *Boundary Layer Stability Theory*, Rev. 1, 1969 (JPL internal document).

a family of outgoing Mach waves. When  $M_{R1} < 1$ , the outgoing, or reflected, wave family (for neutral and amplified disturbances) becomes a subsonic pressure field with exponential decay upward (the positive  $y$  direction), and the incoming-wave family becomes a subsonic pressure field increasing exponentially upward.

Stability theory uses only the first solution. However, the second solution is the flow field produced by a moving wavy wall located above the boundary layer, and, therefore, in the present problem it represents the given external pressure field. When two independent numerical solutions are formed by integrating the inviscid equations through the boundary layer, the boundary condition of zero normal velocity ( $v' = 0$ ) at the wall can be satisfied for any values of  $c_r$  and  $\alpha$  (the dimensionless wave number). The situation in the viscous theory is similar, except that in addition to the incoming and reflected waves there are three viscous solutions which are also needed in order to satisfy the wall boundary conditions of zero velocity and temperature fluctuations.

The two computer programs have been modified to calculate the boundary-layer response to an incoming wave. In the inviscid theory, the two first-order equations are integrated for specified  $\alpha$  and  $c_r$  for both an incoming wave of specified magnitude and a reflected wave of arbitrary magnitude. Both  $\alpha$  and  $c_r$  are real, so that the waves are neutral in the sense of stability theory, and neither amplify nor damp with respect to space or time. What is being calculated is the steady-state response to a forcing function (the external pressure field) considered to have existed for an indefinite period of time. The amplitude and phase of the reflected wave are adjusted to satisfy the boundary condition at the wall, and then the equations are integrated again from the free stream to the wall for the combined solution to provide the final output. Because  $c_r$  is real, there is a singularity on the real  $y$  axis at the critical point ( $U = c_r$ ). In the viscous theory, the eight first-order equations are integrated from the free stream to the wall for the three viscous solutions which go to zero as  $y \rightarrow \infty$  along with the incoming and reflected wave solutions.

A combination of the five solutions satisfies the four boundary conditions at  $y = 0$  for specified  $c_r$ ,  $\alpha$ ,  $R$  (Reynolds number), and a specified magnitude of the incoming wave. In both theories about 50 quantities are computed at each output point. Chief among these are the rms fluctuation amplitudes and phase angles of the

six flow variables ( $u'$ ,  $v'$ ,  $w'$ ,  $p'$ ,  $\rho'$ ,  $T'$  in conventional notation), plus the mass-flow and stagnation-temperature fluctuation amplitudes and phases, vorticity, Reynolds stress and energy-balance terms.

### 3. Numerical Results

A series of calculations have been carried out at  $M_1 = 4.5$  with the two programs. This Mach number was chosen because it is the Mach number at which Kendall has made extensive measurements (SPS 37-62, Vol. III, pp. 43-47). The phase velocity used was  $c_r = 0.65$ , which is indicated by the measurements to be the most probable velocity of the sound sources in the turbulent side-wall boundary layers. The relative free-stream Mach number is then 1.575 for a two-dimensional wave ( $\psi = 0$  deg). For oblique waves (in the plane formed by the  $u'$  and  $w'$  components), the relevant Mach number is that normal to the wave front. Hence for  $0 \text{ deg} < \psi < 51 \text{ deg}$  all disturbances are supersonic; for  $\psi \geq 51 \text{ deg}$  they are subsonic. The phase velocity of the sonic two-dimensional disturbance is 0.778. Only supersonic disturbances will be considered here.

The dimensionless frequency used in stability theory is

$$F = \frac{\omega^* \nu_1}{U_1^{*2}} = \frac{\alpha c_r}{R}$$

The quantity  $\omega^*$  is the circular frequency, and  $\nu_1^*$  the free-stream kinematic viscosity coefficient. The reference length is  $(x^* \nu_1^* / U_1^*)^{1/2}$ , so that  $R = R_x^{1/2}$ , where  $R_x$  is the usual free-stream x-Reynolds number. Since  $c_r$  is constant,  $\alpha/R$  is also constant for a fixed-frequency disturbance. The calculations were made for five frequencies in a range suggested by experiment. Since a hot-wire anemometer measures primarily local mass-flow fluctuations, attention will be directed to the ratio of the rms mass-flow fluctuation,  $(MF)'$ , to the mass-flow fluctuation of the incoming wave in the free stream at  $y = 0$ ,  $(MF)'_i$ .

One of the curves in Fig. 1 shows the distribution through the boundary layer of  $(MF)' / (MF)'_i$  for a typical case as computed from the viscous theory. There is a sharp peak near  $y/\delta = 0.84$ , and even though there is no amplification in the sense of stability theory, the peak  $(MF)'$  in the boundary layer is 16 times the  $(MF)'$  of the incoming wave alone. It can also be seen from Fig. 1 that in the free stream the  $(MF)'$  of the combined incoming and reflected waves is less than 2. Also of interest is the fact that the peak  $(MF)'$  occurs not near  $y_{cr}$ , the

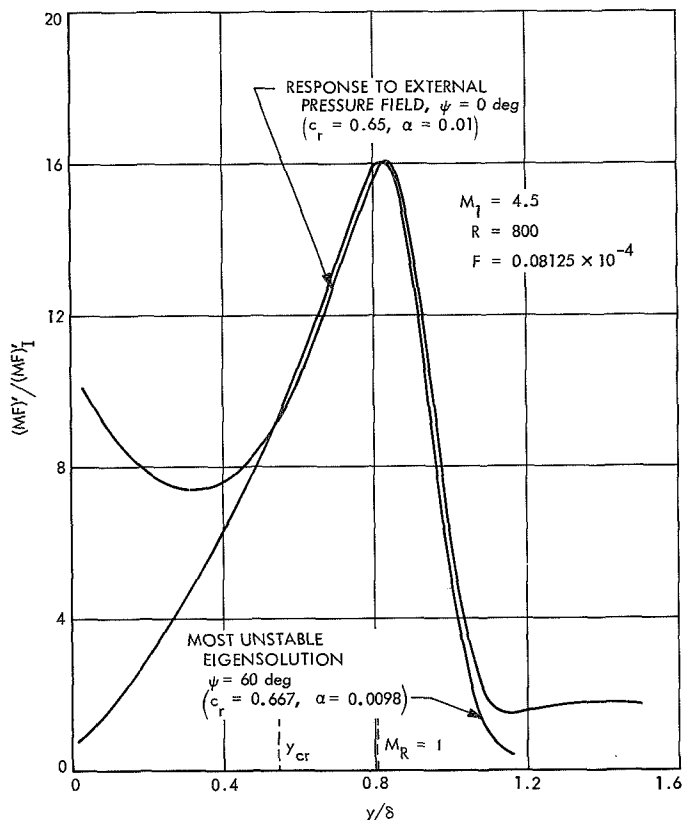


Fig. 1. Distribution through boundary layer of mass flow fluctuation amplitude

critical point, but near the point where the local relative Mach number is unity. It is not known whether this last feature is significant as the position of the peak decreases from  $y/\delta = 0.96$  at  $R = 50$  to  $0.79$  at  $R = 1200$  for the frequency of Fig. 1. However, in the inviscid theory there seems to always be a peak near  $M_R = 1$ , except when masked by the singularity. The second curve in Fig. 1 shows  $(MF)'$ , scaled to the same peak value as the first curve, of the  $\psi = 60$  deg eigensolution at the same Reynolds number and frequency. This eigensolution is slightly amplified with a temporal amplification rate about one fourth the amplification rate of the most unstable first-mode frequency at  $R = 800$  ( $F \times 10^4 \approx 0.6$ ). This particular comparison is given because it is desirable to know if the forced response of the boundary layer can be distinguished from instability waves, and since the  $\psi = 60$  deg disturbances are the most unstable, they are the ones expected to occur in the boundary layer.

Figure 2 gives the ratio  $(MF)'_{\max}/(MF)'_1$ , where  $(MF)'_{\max}$  is the peak value of  $(MF)'$  in the boundary layer, as a function of  $R$  for five different frequencies with  $\psi = 0$

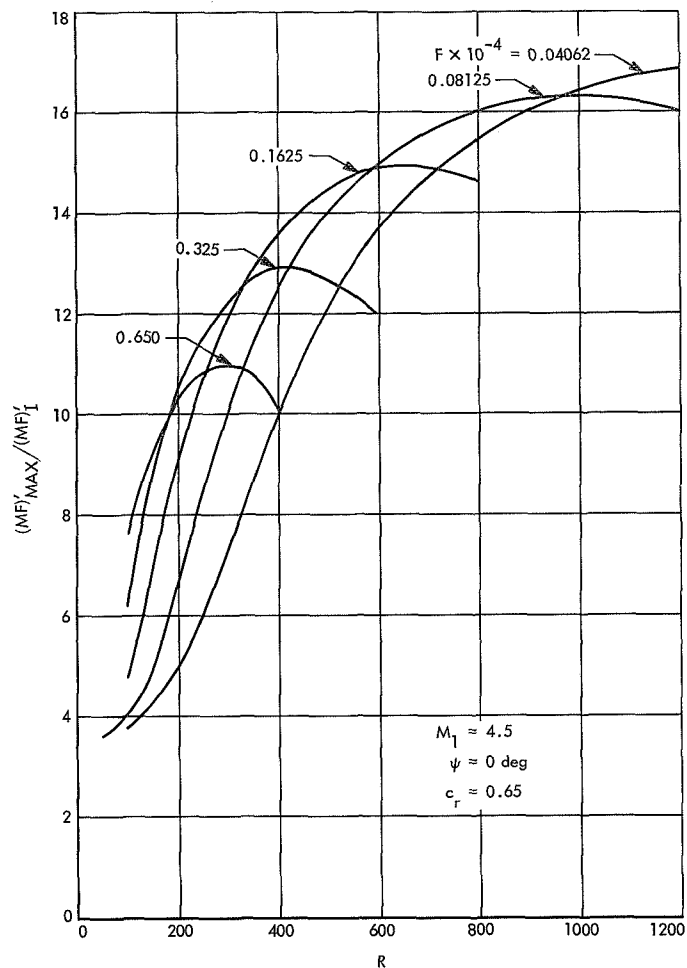


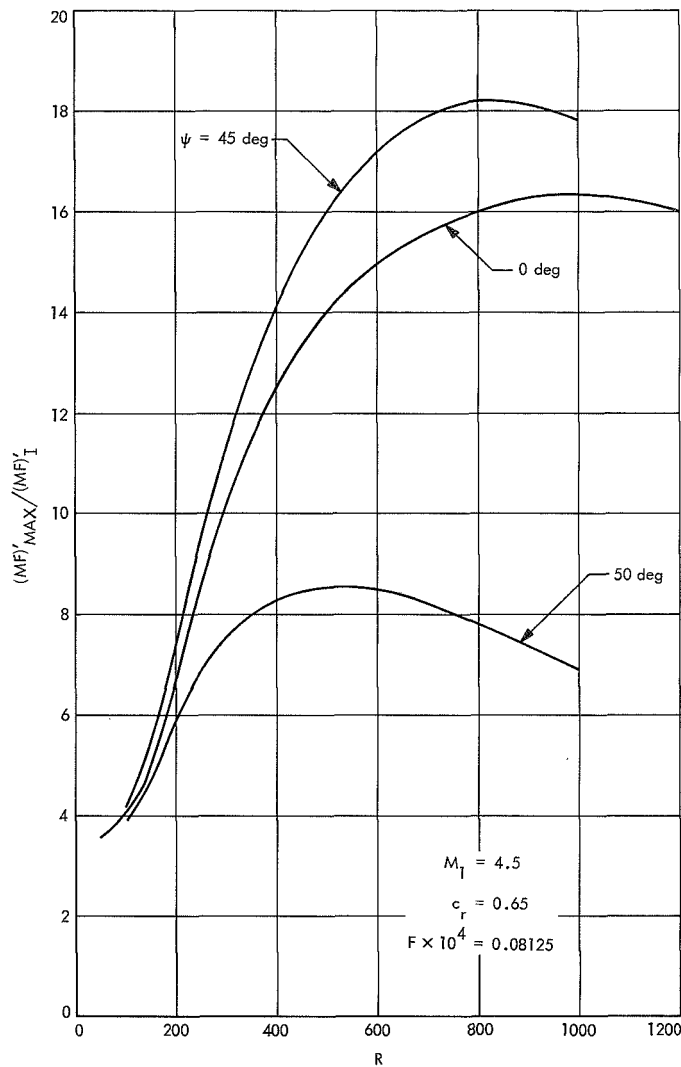
Fig. 2. Effect of frequency on boundary layer response to external pressure field

deg. It is seen that  $(MF)'_{\max}$  increases steadily from the leading edge up to a certain Reynolds number whose value decreases with increasing frequency. Again it is emphasized that the disturbances are all neutral, but nevertheless there is an effective amplification of 10 to 20 times over the free-stream  $(MF)'$  of the external pressure field. Once the disturbances reach the unstable region, they may be further amplified by the instability mechanism.

Figure 3 shows the effect of wave angle on the response for one of the frequencies of Fig. 2. The maximum response is found at  $\psi = 45$  deg, and then falls off rapidly for higher angles. At  $\psi = 50$  deg, the relative Mach number normal to the wave front is 1.012.

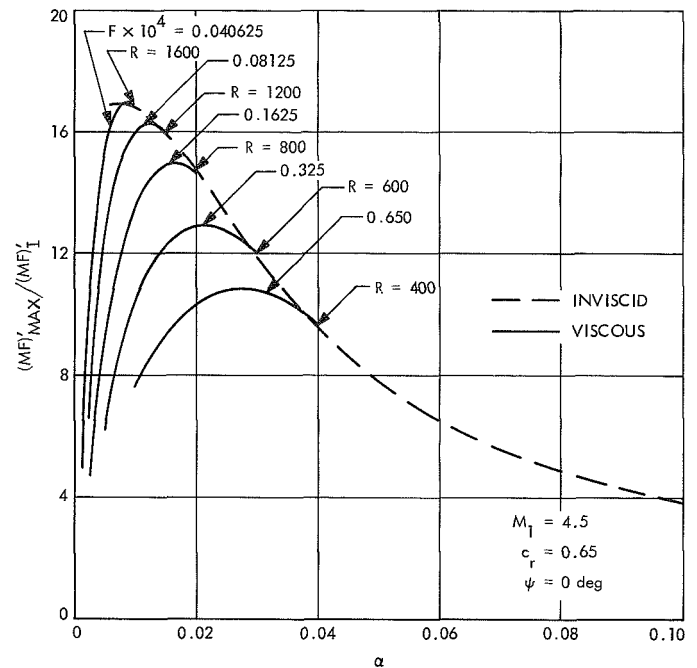
Finally, Fig. 4 shows the relation between the viscous and inviscid theories. The inviscid curve is not precisely located because the output points were spaced 0.07





**Fig. 3. Effect of wave angle on boundary-layer response to external pressure field**

apart in  $y/\delta$ , and the  $(MF)'_{\max}$  plotted is the largest numerical value occurring in the output. At sufficiently large  $\alpha$ , the viscous and inviscid theories give almost identical results, except for the singularity and different



**Fig. 4. Comparison of inviscid and viscous peak mass-flow fluctuations as functions of  $\alpha$**

boundary conditions of the inviscid theory. Since  $\alpha$  increases proportionally to  $R$  for a fixed frequency, Fig. 4 shows that in the inviscid theory there is no apparent amplification at all. The peak mass-flow fluctuation decreases monotonically starting at the leading edge. The amplification region is a strictly viscous phenomenon. The viscous curves join the inviscid curve at an  $\alpha R$  of about 18, so the longest viscous region occurs at the lowest frequencies.

#### Reference

1. Lees, L., and Lin, C. C., *Investigation of the Stability of the Laminar Boundary Layer in a Compressible Fluid*, NACA TN No. 1115. National Advisory Committee on Aeronautics, Washington, D.C., 1946.

## B. Solution of Partial Differential Systems,

F. B. Estabrook, B. K. Harrison,<sup>1</sup> and H. D. Wahlquist

Exact solutions of the sets of coupled nonlinear partial differential equations of physics (fluid physics, general relativity, etc.) are of great value for the physical insight they can give into extreme cases not susceptible to numerical or approximate treatment. The search for such solutions is, however, not a systematic process, depending in practice on intuitive skill, and physical or geometrical analogy for the divination of successful ansätze. In a first attempt at a more systematic methodology, we have considered, as a practical process of applied mathematics, the use of the generators of the invariance group of a given set of equations. Our method for finding these appears to be new; it is based on Cartan's geometric formulation of partial differential equations in the language of exterior differential forms, and introduces the fields of isovectors which generate the invariance transformations. With these we indeed find certain specializations of the set considered, which may be soluble exactly; in particular, we refer to classes of so-called similarity solutions and to algebraically special solutions based on invariant subspaces. Other methods for systematic discovery of exact solutions must also have close connections to this underlying group structure.

---

<sup>1</sup>Senior Resident Research Associate; Physics Dept., Brigham Young University, Provo, Utah.

A feature of this theory is that the operations for finding the isovector fields are: (1) algebraically routine and (2) involve many simultaneous first-order linear equations. As a consequence, formula manipulation by computer can have a very important application here. Having already achieved the capability of computer manipulation of exterior differential forms a FORMAC program, denoted ISOVECT, was developed which derives the equations for the generators of the invariance group of a given set of differential forms. It may ultimately prove feasible analytically to solve these equations by computer, also.

A detailed exposition of this new theory is to be published in the literature (Ref. 1). Application is there made to the one-dimensional heat equation (general similarity solutions are discussed), the vacuum Maxwell equations (algebraically special solutions), the Korteweg-de Vries equation of plasma physics (similarity solutions, conservation theorems), one-dimensional compressible fluid dynamics (hodograph solutions), the Lambropoulos equation (similarity solutions), and the cylindrically symmetric Einstein-Maxwell equations of general relativity (solution generation methods).

### Reference

1. Harrison, B. K., and Estabrook, F. B., "Geometric Approach to Invariance Groups and Solution of Partial Differential Systems," *J. Math. Phys.* Vol. 12, Feb. 1971 (in press).

## V. Communications Systems Research

### TELECOMMUNICATIONS DIVISION

#### A. Digital Clean-Up Loop Transponder for Sequential Ranging System, W. J. Hurd

##### 1. Introduction

The range from a ground tracking station to a spacecraft is determined by measuring the round-trip time required for a ranging code sequence signal to be transmitted from the station to the spacecraft and transponded back to the ground station. In typical existing spacecraft, the ranging transponders demodulate and limit the received ranging signal and then remodulate it onto a different frequency RF carrier for retransmission to the earth. A major drawback of such transponders is that both the received ranging signal and the receiver noise are transmitted on the downlink. This means that when the uplink ranging channel SNR is below unity, most of the downlink ranging power is wasted on transmitting receiver noise rather than ranging signal. The problem will be extremely severe on outer planet missions. For a *Mariner*-type spacecraft at a range of 31 AU, the approximate distance to Neptune, the uplink SNR will be as low as  $-46$  dB, even using a 210-ft transmitting antenna and a 400-kW transmitter.

To overcome the problem of retransmission of receiver noise, a digital signal processing system has been developed to regenerate the ranging signal in the spacecraft for transmission to the ground. The system, called a clean-up loop, phase locks on the received ranging signal and generates a replica of the received ranging code. The regenerated ranging signal has no amplitude noise, but a small amount of phase jitter, corresponding to the phase error in the phase-locked regeneration loop. This phase noise can be averaged out on the ground. Since little downlink transmitter power is consumed by transmission of the phase noise, the increase in downlink SNR due to the clean-up loop is approximately equal to the ratio of uplink receiver noise to signal, or up to 46 dB. Less power need be devoted to the ranging signal, on both the uplink and the downlink, and accurate ranging can be accomplished in a shorter time.

##### 2. Sequential Component Ranging System

The clean-up loop is designed to operate in conjunction with a binary coded sequential code-component ranging system, called the  $\mu$ -system (SPS 37-52, Vol. II, pp. 46-49;

SPS 37-57, Vol. III, pp. 72-81, and Ref. 1) rather than with the parallel component pseudonoise system. The sequential component system, without a clean-up loop, was used successfully in the post-encounter phases of *Mariners VI* and *VII*. In the  $\mu$ -system, square waves at successively lower frequencies are transmitted sequentially. First the range is measured to high resolution, but with 2  $\mu$ sec ambiguities, using a 500-kHz square wave. Then a 250-kHz square wave is transmitted, and half of the ambiguities are eliminated. Lower and lower frequencies are transmitted until all ambiguities are resolved. The clean-up loop transponder operates by phase locking on each code component frequency as it is received by the spacecraft. A square wave in phase with the received signal is generated for modulation onto the downlink carrier and transmission to earth.

The information bearing components of the ranging code,  $c_k(t)$ , are square waves in phase with sine waves, and are defined by

$$c_k(t) = \text{sgn} [\sin (2^{-k} 2\pi f_0 t)] \quad \text{for } k = 1, 2, \dots, K$$

where  $f_0$  is approximately 1 MHz. The number of code components,  $K$ , is chosen to be large enough so that the period of  $c_K$  is sufficiently long to resolve all *a priori* uncertainty in range. In the prototype clean-up loop,  $K$  was chosen to be 16.

When using the clean-up loop,  $c_1$  is not transmitted by itself, but is used as a suppressed subcarrier for the other components. Its frequency, 500 kHz, determines the basic resolution of the system. The actual sequence of transmitted waveforms is

$$\{w_k(t), k = 2, 3, \dots, 16\}$$

where

$$w_k = c_1 c_k \quad \text{for } k = 2, 3, \dots, 9, 11, 13, 15$$

and

$$w_k = c_1 c_k c_7 \quad \text{for } k = 10, 12, 14, 16$$

The purpose of introducing the third component,  $c_7$ , into  $w_k$  for  $k = 10, 12, 14$ , and  $16$  is so that adjacent components decorrelate in a shorter length of time. Components  $c_8$  and  $c_9$  are orthogonal over 256  $\mu$ sec, the period of  $c_8$ , and all other adjacent pairs are orthogonal over 128  $\mu$ sec or less.

Operationally, the first component ( $w_2$ ) is transmitted for a time long enough for the clean-up loop to acquire frequency and phase lock on this component. The trans-

mitter then switches to  $w_3$ , then  $w_4$ , etc. The clean-up loop must determine when the received signal changes from  $w_k$  to  $w_{k+1}$ , make a binary decision as to the phase of  $w_{k+1}$ , and change the phase-locked loop (PLL) reference signal to track the new component. Note that  $w_{k+1}$  can arrive in one of two possible phase relationships with respect to  $w_k$ , since consecutive transitions of  $c_k$  correspond alternately to positive and negative transitions of  $c_{k+1}$ . The determination of the phase relationships between the  $c_k$  corresponds to resolution of the ambiguities in range measurement.

Since the downlink SNR is normally lower than the uplink SNR, the time required to measure range to the desired accuracy is determined primarily by the downlink SNR, as discussed by Goldstein (SPS 37-52, Vol. II). However, because the decision time parameters of the clean-up loop are determined by the uplink design point SNR, each code component must be transmitted for some minimum length of time, even when the SNR is high. Conservatively, the minimum time is about 2 to 3 min for acquisition of the first component and about 12 seconds for each of the other components, for a total of 5-6 min. At the lowest SNRs, it may be desirable to allow more time for initial acquisition, especially if the frequency offset is not small.

### 3. Implementation of Clean-up Loop

In this section we describe the first implementation of the clean-up loop. Methods for improving the particular design and possible changes in design philosophy which would result in major changes are presented in *Subsection 6*.

A fundamental block diagram of the clean-up loop, which would probably not change with any modifications, is shown in Fig. 1. The loop has three major portions: the PLL, the lock detector, and the code generator

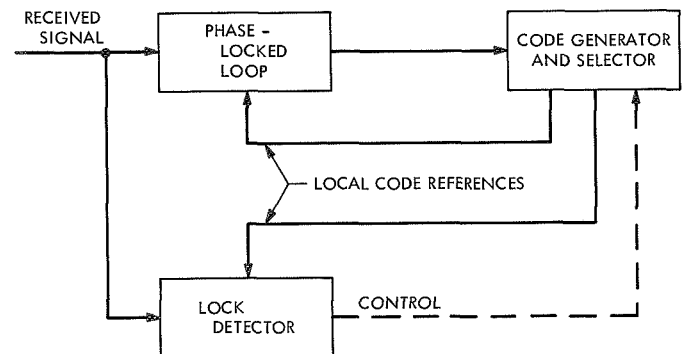


Fig. 1. Clean-up loop simplified block diagram

and selectors. The present implementation is entirely digital, with the exception of the voltage-controlled oscillator (VCO), which is controlled by a digital-to-analog converter (DAC). Because the ranging signal in *Mariner*-type transponders is most readily available at the output of a hard limiter, the input to the clean-up loop is assumed to be binary. This simplifies the implementation of the PLL and the lock detectors, since products can be realized with exclusive or gates and additions and accumulations can be realized with counters. The input signal  $S$  is sampled at a 4-MHz rate, synchronously with the local reference clock. Every fourth sample occurs at the transition of  $\hat{c}_1$ , the locally generated replica of  $c_1$ . These samples are used by the phase detector, and the three samples between each transition of  $\hat{c}_1$  are used by the lock detector.

**a. Phase-locked loop.** The PLL consists of the phase detector, a linear digital first order filter, a nonlinear digital low pass filter, a DAC and a VCO, as shown in Fig. 2. Since the nonlinear filter has no effect on the loop dynamics except as a gain control, the loop can be modeled as a perfect integrator second-order loop, with open-loop transfer function

$$F(s) = \frac{A}{s} \left( K_1 + \frac{K_2}{s} \right) \quad (1)$$

where  $A$ , the phase detector gain, depends on the input SNR. The bandwidth of the loop is varied during acquisition by changing the parameters of the digital filters upon command of the lock detector. To maintain a fixed damping ratio,  $K_1$  is varied in proportion to loop bandwidth and  $K_2$  is varied in proportion to the square of loop band-

width. This is accomplished by varying the gain of the nonlinear filter in direct proportion to loop bandwidth, which changes both  $K_1$  and  $K_2$ , and by varying the integrator gain in the linear filter directly with loop bandwidth, which affects only  $K_2$ .

**Phase detector.** The phase detector is of the transition tracking type (Ref. 2). An estimate of the sign of the phase error is made each time a transition occurs in the local replica of the code component being tracked. The code generator and selector informs the phase detector when such a transition occurs and its direction. The product of the transition direction and the sample of the received signal taken at the time of the transition forms an estimate of the sign of the phase error. For example, if the sample is positive (negative) and the transition is negative going, then the sampling time is most likely early (late) compared to the true transition time in the received signal. The phase estimate is implemented simply as the exclusive-or of the input signal sample and the value of  $\hat{w}_k$  at the instant immediately preceding the possible transition time. This value is gated into the nonlinear filter only when  $\hat{w}_k$  changes, i.e., whenever there truly is a transition.

**Nonlinear filter.** The nonlinear filter consists of two up/down counters and some gating logic (SPS 37-61, Vol. II, pp. 63-68). The inputs to the filter are all 1, 0, or -1, and the outputs are all 1 or -1; however, the filter low-frequency gain can easily be varied by changing the number of stages in one of the counters. For the wide, medium, and narrow loop bandwidths, respectively, the gain is either 32, 8, or 2. To achieve a gain of  $2^N$ , i.e., an average output equal to  $2^N$  times the average

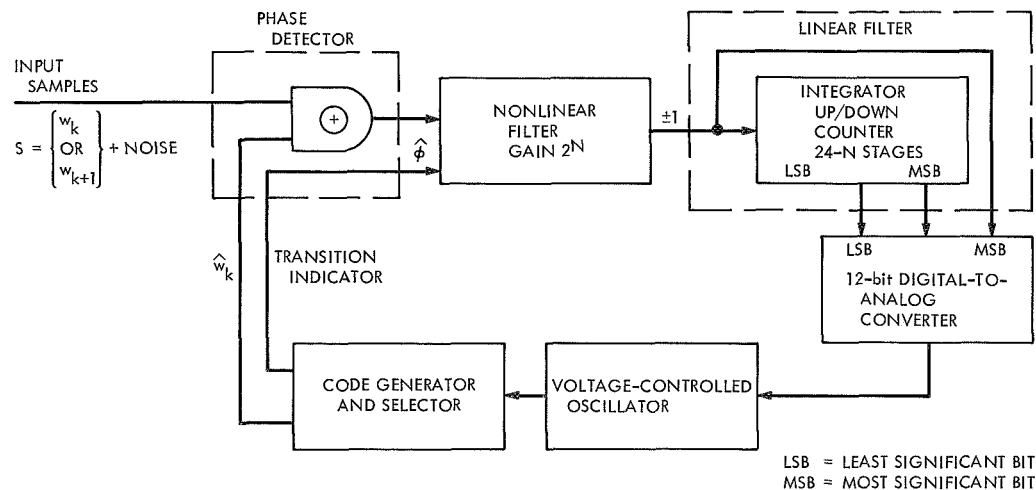


Fig. 2. Implementation of phase-locked loop

input, the number of +1 outputs less the number of -1 outputs must be  $2^N$  times the number of +1 inputs less the number of -1 inputs. To accomplish this, a count stored in the two counters is incremented or decremented by  $2^N$  for each +1 or -1 input, and decremented or incremented by one for each +1 or -1 output. The filter output is the sign of the value in the counters.

**Linear filter.** Since the output of the nonlinear filter is always either +1 or -1, the integrator portion of the loop filter can be simply implemented with an up/down counter. The number of stages in the counter is varied with loop bandwidth to control  $K_2$ . When the nonlinear filter has gain  $2^N$ , the number of stages in the integrator-counter is  $24 - N$ , so that  $K_2$  varies as  $2^{2N}$ . The proportional control component of the linear filter is fixed, so  $K_1$  varies only as the nonlinear filter gain,  $2^N$ .

**Digital-to-analog converter.** The proportional control and integrated components of the loop filter output are summed in one digital-to-analog converter. This can be accomplished because of the careful scaling of the digital filters. This scaling is such that the proportional control component controls the most significant bit of the DAC, and the integrator-counter controls the remaining 11 bits. The actual voltage output of the DAC is of course made compatible with the range and gain of the VCO.

**b. Code generator and selector.** The code generator and selector is shown in Fig. 3. The 4-MHz VCO clocks a binary divider chain called the C-counter, whose stage outputs are labeled  $C_{-1}, C_0, C_1, \dots, C_{24}$ . Outputs  $C_1$  to  $C_{16}$  ( $\hat{c}_1$  to  $\hat{c}_{16}$ ) are the local replicas of the received code sequences, and stages  $C_{17}$  to  $C_{24}$  are required for system timing. A 4-stage counter, called the K-counter, keeps account of which sequence is being tracked. This counter controls two data selectors whose outputs are  $\hat{c}_k$  and  $\hat{c}_{k+1}$  when the K-register contains the binary representation of  $k + 1$ . The reference signals  $\hat{w}_k$  and  $\hat{w}_{k+1}$  and the transition indication information are simply derived from  $\hat{c}_k$ ,  $\hat{c}_{k+1}$ ,  $C_1$  and  $C_7$ .

Whenever the lock detector decides that the received signal has changed from  $w_k$  to  $w_{k+1}$ , the K-counter is incremented, which has the effect of replacing  $\hat{c}_k$  by  $\hat{c}_{k+1}$  and  $\hat{c}_{k+1}$  by  $\hat{c}_{k+2}$ . At the same time, if the lock detector decides that the phase of the new received sequence is such that  $\hat{c}_{k+1}$  must be inverted, an invert pulse is gated through the demultiplexer to the direct set line of the appropriate flip-flop in the C-counter, under control of the K-counter. Since lock decisions are made when  $C_{21}$

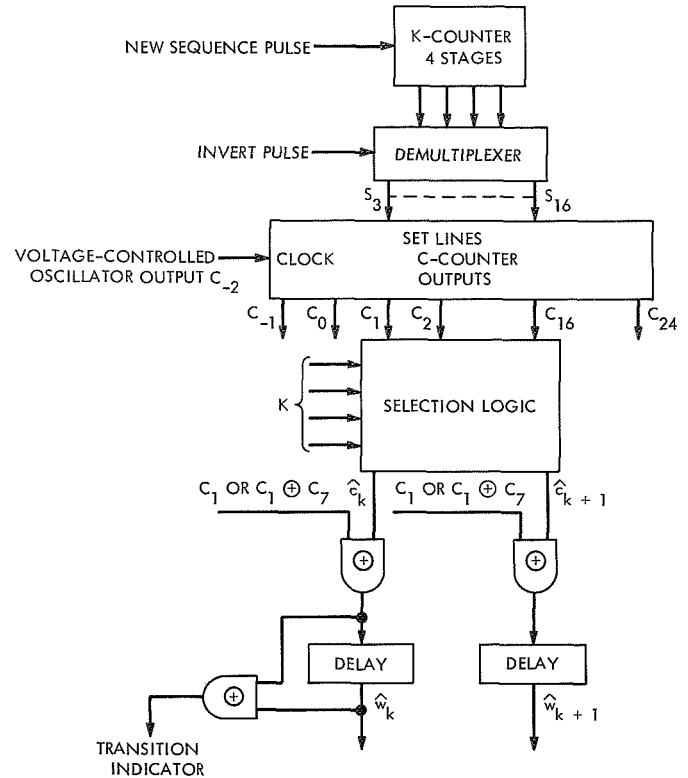


Fig. 3. Code generator and selector

switches high,  $C_{k+1}$  is zero when this pulse arrives, so that the effect is always to invert  $\hat{c}_{k+1}$ .

**c. Lock detector.** The lock detector is shown in Fig. 4. Its functions are: (1) during initial acquisition, to detect when phase lock on  $w_2$  is achieved and to control the loop bandwidth narrowing procedure, (2) during tracking of the various  $w_k$ , to determine when the received signal changes to  $w_{k+1}$  and the sign of  $w_{k+1}$  relative to the local replica, and (3) to determine when the loop is not locked on any component so that the reacquisition can be initiated.

**Initial acquisition mode.** At the beginning of the acquisition procedure, all registers and counters are set to zero. The two most significant bits of the control ( $M$ ) counter control whether the clean-up loop is in the acquisition mode or the tracking mode, and the PLL bandwidth. This is shown in Table 1. During acquisition, the A-counter correlates the received signal  $S$  with  $\hat{w}_2$ . Only the samples of  $S$  occurring halfway between the transitions of  $\hat{c}_1$  are used. The A-counter counts down when  $S$  and  $\hat{w}_2$  are equal, and up otherwise, and the result is sampled when  $C_{24}$  comes high, which is after  $2^{23}$  samples. At this time, A and  $C_{24}$  are reset, and the M-counter

Table 1. Control counter states

$M_3$	$M_4$	Mode	Bandwidth
0	0	Acquisition	Wide
1	0	Acquisition	Medium
0	1	Acquisition	Narrow
1	1	Tracking	Narrow

is incremented if the final count in A was less than  $-2^{15}$ , which is detected by a negative (one) sign bit and at least one zero in the most significant eight magnitude bits. When the threshold is exceeded four times, the state of  $M_3$  and  $M_4$  changes, so the bandwidth is narrowed or the mode is switched to tracking. If  $M_3$  does not change within eight trials, it is assumed that the loop is not in lock and the system is reset so that the acquisition procedure starts anew.

**Tracking mode.** The tracking mode begins when the M-counter reaches  $M_3 = M_4 = 1$ . The A- and B-counters, respectively, now correlate  $S$  with the difference and sum of  $\hat{w}_k$  and  $\hat{w}_{k+1}$  over intervals of length  $2^{20}$   $\mu$ sec using the three samples of  $S$  taken between each transition of  $\hat{c}_1$ . In particular, A counts only when  $\hat{w}_k \neq \hat{w}_{k+1}$ , B counts only when  $\hat{w}_k = \hat{w}_{k+1}$ , and when counting, both counters count down if  $S = \hat{w}_k$ . When  $C_{21}$  comes high, the signs

of the counters are sampled and reset and  $C_{21}$  is reset. The counter values at the sampling times are

$$S_A = -\frac{1}{2} \sum_{3 \times 2^{20} \text{ counts}} S(\hat{w}_k - \hat{w}_{k+1}) \quad (2)$$

and

$$S_B = -\frac{1}{2} \sum_{3 \times 2^{20} \text{ counts}} S(\hat{w}_k + \hat{w}_{k+1}) \quad (3)$$

where  $S$ ,  $\hat{w}_k$  and  $\hat{w}_{k+1}$  have the values  $\pm 1$ , and the arithmetic is real arithmetic. The index of summation is not shown. When the summations are completed, both  $S_A$  and  $S_B$  will most likely be negative if the signal input is  $w_k$ . If  $w_{k+1}$  is present,  $S_A$  will most likely be positive if  $w_{k+1} = \hat{w}_{k+1}$  and  $S_B$  will be positive if  $w_{k+1} = -\hat{w}_{k+1}$ . When either is positive, the new sequence pulse is generated and the K-counter is incremented (unless  $k = 16$ ), and  $C_{k+1}$  is inverted if  $S_B > 0$ . At the sampling times,  $S_A$  and  $S_B$  should never both be positive, although this will occur with probability 1/4 if neither  $w_k$  nor  $w_{k+1}$  is the input signal. For example, this would occur if the transmitter has returned to  $w_2$ . If both  $S_A$  and  $S_B$  are positive, it is assumed that the loop is out of lock and acquisition is reinitiated.

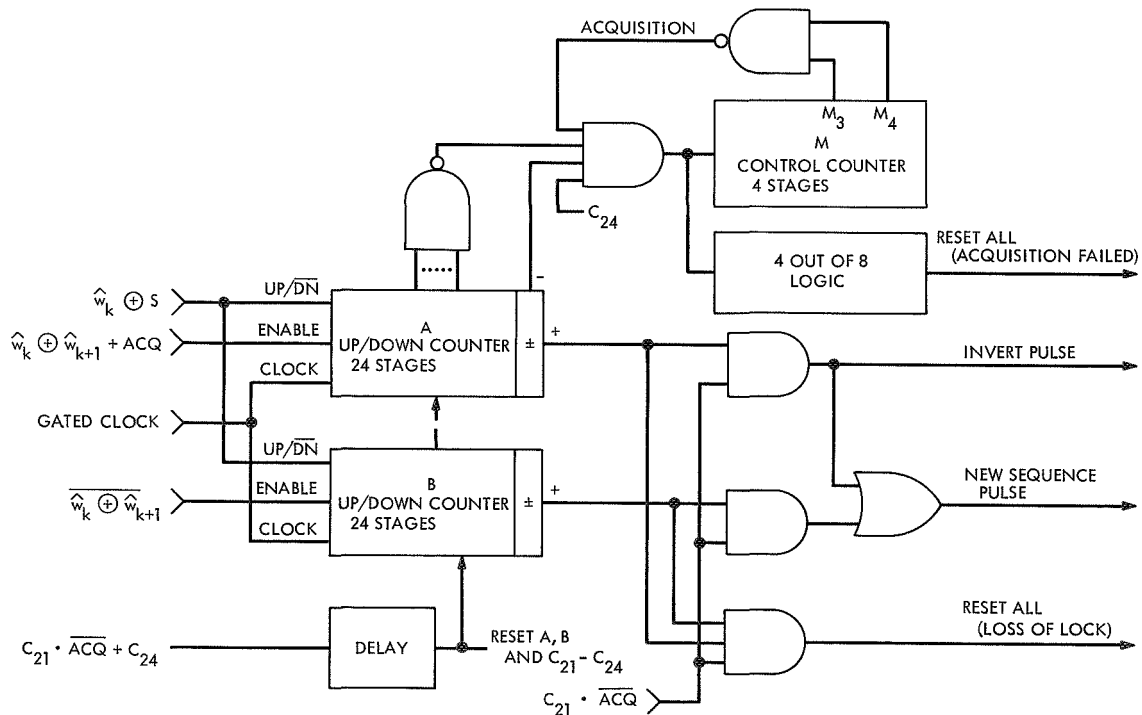


Fig. 4. Lock detector

#### 4. Performance Analysis

The design goal for the clean-up loop is to achieve a high ( $> 0.99$ ) probability of successful transponding when the ranging signal SNR at the limiter output is  $-46$  dB in a bandwidth of  $1.5$  MHz. The model assumed for analysis of the system performance was chosen for simplicity of analysis and because this model was easily duplicated for the experimental verification of performance, the results of which are given in Subsection 5. The block diagram is shown in Fig. 5. White gaussian noise is added to the signal  $w_k(t)$ , and the result is passed through an RC low-pass filter with time constant  $\alpha = 1/6$   $\mu$ sec, and then hard limited and sampled at a  $4$  MHz rate and converted to a logic signal. The low-pass filter has the same  $1.5$  MHz bandwidth as the *Mariner* transponder.

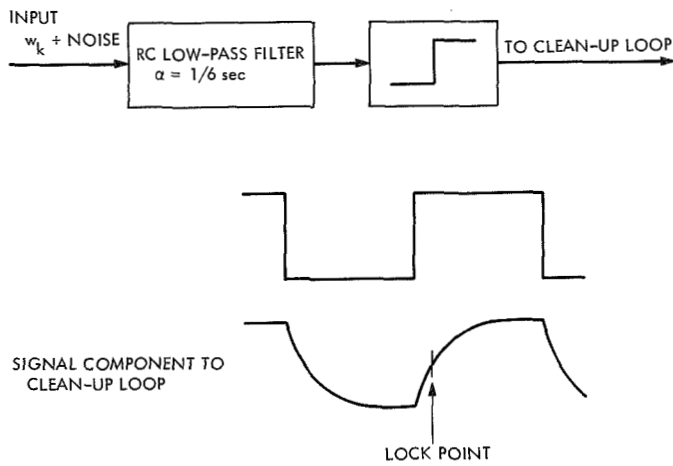


Fig. 5. Model for analysis and experimental tests

At low signal-to-noise ratios, it can be assumed that the signal component of the limiter output has the same shape as at the RC filter output, i.e., piecewise exponential, and that noise is independent of the signal. The noise has unity power and correlation function

$$\rho(\tau) = \frac{2}{\pi} \arcsin [\exp(-|\tau/\alpha|)] \quad (4)$$

The assumed signal waveform is shown with respect to  $w_k$  in Fig. 6. Since the RC time constant is short compared to the  $1$ - $\mu$ sec minimum time between transitions, it can be assumed that the exponential has reached its target voltage, called  $\pm V$ , by the time a transition occurs. The correct lock point or reference sampling point is assumed to be the point that the waveform crosses zero. Assuming

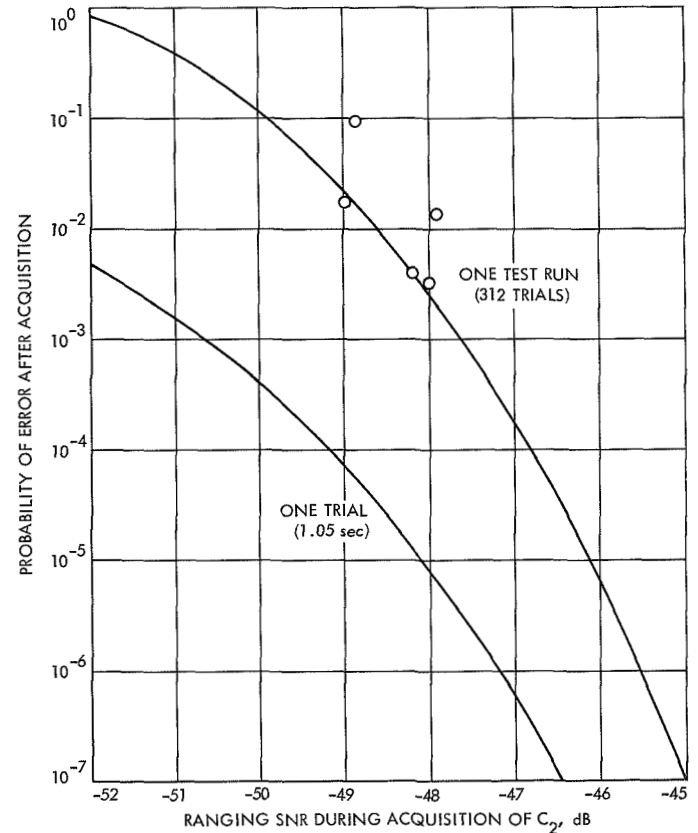


Fig. 6. Probability of error after acquisition

$t = 0$  at the positive zero crossing, the equation of the rising exponential is

$$v(t) = -V + 2V [1 - \exp(-t/\alpha)] \quad (5)$$

and the delay from the transition of  $w_k$  to the zero crossing is  $\alpha \ln(2) = 0.115$   $\mu$ sec.

The results are presented in terms of  $R$ , the SNR at the limiter output when the input is  $w_2$  plus noise, that is, during acquisition. This SNR is

$$R = \frac{5}{6} V^2 \quad (6)$$

For tracking the other components, the SNR is approximately the same as for a  $500$ -kHz square wave input, and is thus approximately

$$R_T = \frac{2}{3} V^2 = \frac{4}{5} R \quad (7)$$

or, in decibels, a  $0.96$ -dB drop from  $R$ .



*a. Phase detector gain and noise density.* The phase detector gain is defined as the derivative with respect to phase of the expected value of the phase detector output. Phase, denoted by  $\phi$ , will always be phase with respect to 500 kHz. The phase detector output is zero when there is no transition, and otherwise has average value equal to the absolute value of  $v(t)$  at the phase error in question, i.e., at  $t = (\phi/\pi) \mu\text{sec}$ . Letting  $P_T$  be the average number of transitions, the phase detector gain is

$$\begin{aligned} A &= \left. \frac{dv(t)}{dt} \frac{dt}{d\phi} \right|_{\phi=0} \\ &= \frac{P_T V}{\pi \alpha} = \frac{1}{\pi \alpha} \left( \frac{6}{5} R \right)^{1/2} \\ &= 2.09 R^{1/2} P_T \text{ units/rad} \end{aligned} \quad (8)$$

At  $R = -46$  dB,  $A = 5.24 \times 10^{-3}$  for acquisition and  $A = 1.05 \times 10^{-2}$  for tracking.

Because the sampling rate at the phase detector is much less than the Nyquist rate, the noise samples are essentially independent. Hence  $N_0$ , the one-sided noise spectral density at zero frequency, is  $2P_T \times 10^6 \text{ units}^2/\text{Hz}$ .

*b. PLL performance.* The parameters of the phase-locked loop were chosen to achieve a reasonable loop bandwidth and for convenience in design of the loop filter and DAC connections. For the narrow-loop bandwidth, the parameters are

$$K_1 = 88.0 \text{ rad/sec/unit}$$

and, since the integrator-counter has 23 stages and is updated every microsecond,

$$K_2 = 2^{-23} \cdot 10^6 K_1 = 10.5 \text{ rad/sec}^2/\text{unit}$$

The damping ratio is

$$\zeta = \frac{K_1}{2} (A/K_2)^{1/2} \quad (9)$$

At  $R = -46$  dB,  $\zeta = 1.0$  for acquisition, and  $\zeta = 1.414$  for tracking. Similarly, the one-sided loop bandwidths are given by

$$b_L = \frac{1 + 4\zeta^2}{4} \frac{K_2}{K_1} \quad (10)$$

At  $R = -46$  dB, the bandwidth is 0.26 Hz for tracking, and the three bandwidths for acquisition are 2.32, 0.58, and 0.145 Hz.

*Steady-state phase error.* The steady-state rms error for each case is

$$\sigma_\phi = \left[ \frac{N_0 b_L}{A^2} \right]^{1/2} \quad (11)$$

Evaluating this at  $R = -46$  dB,  $\sigma_\phi$  is 0.0725 rad for tracking, and 0.275, 0.138, and 0.0688 rad for acquisition. These correspond to loop SNRs of 22.8 dB for tracking and 11.2, 17.2, and 23.2 dB for acquisition.

*Acquisition.* The acquisition performance of a phase-locked loop depends on the loop parameters, the phase detector nonlinearity, signal-to-noise ratio, and frequency offset. No adequate theory exists to predict probability of lock as a function of time, or any similar statistic. Heuristically, for a second-order loop one can say that acquisition will occur with high probability in a time less than about  $10/b_L$  when the frequency offset does not exceed about  $b_L/2$ , and when the steady-state loop SNR is high enough so that cycle slipping is infrequent after acquisition. The pertinent parameters during acquisition are very similar for the clean-up loop and for the symbol synchronizer for low SNR coded systems (Ref. 2 and SPS 37-61, Vol. II, pp. 63-68). Both have perfect integrator second-order loops and transition tracking phase detectors (decision directed for the symbol synchronizer), and in both cases  $b_L$  is narrowed by a factor of 4 after acquisition. Acquisition time experiments were run for the symbol synchronizer at a symbol SNR of  $-3$  dB, which resulted in a 12-dB loop SNR in the acquisition bandwidth. This compares to a loop SNR of 11.2 dB for the clean-up loop. Two pertinent experimental results for the symbol synchronizer (SPS 37-61, Vol. II) are: (1) that for a frequency offset of  $b_L/2$  Hz, acquisition occurs with probability 0.998 within  $16/b_L$  sec, and (2) that if the bandwidth is narrowed to one-fourth of its original value  $8/b_L$  sec after start of acquisition, the loop will be in lock at that time and remain in lock upon bandwidth narrowing with probability 0.99. Applying these results to the clean-up loop, acquisition should occur with probability 0.99 to 0.998 in 3.5 to 7.0 sec, provided that the frequency offset is less than 1.16 Hz, reflected to the center frequency of 500 kHz.

*c. Acquisition lock detection.* During initial acquisition,  $S$  is sampled at  $1\text{-}\mu\text{sec}$  intervals and correlated against  $\hat{w}_2$  using sums over  $2^{23}$  samples. Since the sampling rate is well below the Nyquist rate, the variance of the sum is approximately  $\sigma^2 = 2^{23}$ . When the phase error is zero, samples are taken 0.5 and 1.5  $\mu\text{sec}$  after each transition. Since the time constant is  $1/6 \mu\text{sec}$ , the means of  $S\hat{w}_2$  at

these two times are  $V(1 - e^{-3})$  and  $V(1 - e^{-9})$ . The mean of the summation is thus

$$n = 2^{22} V (2 - e^{-3} - e^{-9}) = 1.95 \times 2^{22} V \quad (12)$$

The summation is compared to a threshold of  $T = 2^{15}$ , and the probability of not exceeding this threshold when the phase error is zero is

$$P_{\text{MISS}} = 0.5 \operatorname{erfc} \left( \frac{(m - T)}{\sqrt{2\sigma}} \right) \quad (13)$$

For  $R = -46$  dB,  $P_{\text{MISS}} \approx 2 \times 10^{-5}$ . The probability of failing to exceed the threshold 5 times in eight trials, which would be required for false recycling of the acquisition procedure, is about  $10^{-22}$ .

It is also possible for false alarms to occur, i.e., for  $T$  to be exceeded when the loop is not in lock. For this case, the mean of the summation is zero, assuming the loop is slipping cycles. The reason that a correlation time of  $2^{23} \mu\text{sec} = 8.4 \text{ sec}$  is used is that when the loop is slipping cycles, it will most certainly slip several cycles in this time. If the loop slipped only part of a cycle during the correlation, the mean of the integration could be quite large, depending on the exact phase relationships. However, if the phase error is changing at a rate of less than one cycle in 8.4 sec, the error is most likely approaching zero, i.e., the loop is achieving lock. For these reasons, it is justified to assume that the mean of the correlation is zero when the loop is not in lock. The probability of false alarm during initial acquisition is then

$$P_{\text{FA}} = 0.5 \operatorname{erfc} \left( \frac{T}{\sqrt{2\sigma}} \right) = 0.5 \operatorname{erfc}(8) \quad (14)$$

which is infinitesimal.

**d. New sequence detection.** After lock is acquired and the loop bandwidth is narrowed, the lock detector makes a decision every 1.05 sec as to whether the received signal  $w_k$  is equal to  $\hat{w}_k$ ,  $\hat{w}_{k+1}$ , or  $-\hat{w}_{k+1}$ , according to Eqs. (2) and (3). Each decision is called a trial. Because the PLL has no reference signal from the time  $w_k$  changes until this change is detected, the rms phase error increases, and it is important to detect changes before a cycle slip can occur. Also, the rms phase error is larger at the instant when the new reference signal is applied than in the steady state. Since this can degrade the lock detector performance, the first 8 trials after detection of a new sequence are inhibited, which places a constraint of about 10 trials or 10.5 sec on the minimum time for which each

sequence must be transmitted. After the 8 trial waiting periods, the phase error statistics will have essentially reached steady state.

The probability of a false decision on any trial is the probability that either  $S_A$  or  $S_B$  has the wrong sign. These events are independent, since half of the terms in  $S_A$  are zero and the other half of the terms in  $S_B$  are zero.  $S_A$  and  $S_B$  are also identically distributed, except for the signs of the means. The mean and variance of  $S_A$  and  $S_B$  were calculated taking into account the phase jitter, assuming the loop to be in the steady state. The mean was determined as a function of phase error, and then averaged over  $\phi$ , assuming  $\phi$  to be gaussian. The variance was calculated by adding the variance of the mean to the variance due to noise, and the probability  $P_s$  of having the wrong sign was calculated assuming  $S_A$  and  $S_B$  to be gaussian. Finally, the probability of an error in a trial is

$$P_E = 1 - (1 - P_s)^2 \approx 2P_s$$

This probability is plotted as a function of  $R$  in Fig. 6.

As shown in Fig. 6, the probability of making an error on a given trial is very small, for example,  $P_E = 0.18 \times 10^{-7}$  at  $R = -46$  dB. This means that the probability is correspondingly high and that a change in  $w_k$  will be detected on the first trial which starts after the change occurs. This requires, at most, 2.1 sec. (It can also be shown that the probability of an error on a trial during which a change in  $w_k$  occurs is less than  $P_E$ .)

It is shown elsewhere (Section B) that the increase in mean square phase error after loss of signal at  $t = 0$  is

$$\begin{aligned} \frac{\sigma_\phi^2(t)}{\sigma_\phi^2(t=0)} &= 1 + \frac{16\zeta^2}{4\zeta^2 + 1} (2b_L t) + \left( \frac{8\zeta^2 + 1}{4\zeta} \right) \left( \frac{4\zeta}{4\zeta^2 + 1} \right)^3 \\ &\times (2b_L t)^2 + \frac{1}{3} \left( \frac{4\zeta}{4\zeta^2 + 1} \right)^4 (2b_L t)^3 \end{aligned}$$

For the tracking loop parameters,  $\sigma_\phi(t = 2.1 \text{ sec}) = 2.4\sigma_\phi(0) = 0.174 \text{ rad}$  with  $R = -46$  dB. Since this rms error corresponds to a loop SNR of over 15 dB, it is unlikely that the loop will slip a cycle in this time. We conclude that the new sequence will be detected within 2.1 sec and without a cycle slip, with probability approximately  $1 - P_E$ .

## 5. Experimental Verification

An experimental test program has been initiated to verify the theoretical performance of the clean-up loop.

Results to date indicate that the performance after initial acquisition is as anticipated, within the tolerance of the test parameters.

The experimental test setup is shown in Fig. 5. For initial acquisition, component  $w_2$  was generated for about 8 min. Then components  $w_3$  through  $w_{16}$  were generated for  $32 \times 2^{20}$   $\mu$ sec each, or 32 trial intervals. Acquisition was deemed to have occurred if the change from  $w_2$  to  $w_3$  was detected properly. Since detection is inhibited for 8 detection intervals after detection of a new component, there were 312 trials during which an error could occur, 24 on each of 13 components. The theoretical probability of an error in 312 trials is shown in Fig. 6, with the dots denoting experimental results. The dots represent experiments over 428 to 1460 cases, which required approximately a week for each experiment. The results agree with theory to within 1 dB, with the discrepancies attributed to inaccuracy in measuring the RC filter bandwidth and the signal and noise powers, to variations in the noise statistics over the considerable time required for each test point, and to statistical fluctuations.

## 6. Problem Areas and Suggestions for Improvement

Up to this point we have discussed the clean-up loop as implemented in the prototype model. We now consider some of the problems encountered, and suggest methods for improvement.

The major problem area concerns initial acquisition and the long-term frequency stability of the VCO. Also of major importance is a reduction in the power required by the high-speed digital logic and simplification of the logic. Less important, but still worthy of consideration, is the approximate 2-dB improvement which is attainable by utilizing the received ranging signal before it is hard limited.

*a. Frequency stability and acquisition.* The uplink ranging frequency is locked to the RF carrier frequency, which is doppler compensated and adjusted to the best lock frequency of the spacecraft RF VCO. As the clean-up loop is implemented, rapid acquisition will occur only when the frequency offset is less than  $b_L/2$ , or 1.16 Hz out of 500 kHz at the design point SNR. The difference between the center frequencies of the RF and ranging channel VCOs must therefore be maintained to within about  $\pm 2$  ppm over the spacecraft temperature range, for the life of the mission, which might be about 10 yr. This is a difficult, if not impossible, requirement to satisfy with independent VCOs. If adequate VCOs cannot be obtained, two possibilities for solving the problem are:

- (1) to use a more elaborate acquisition procedure, and
- (2) to reference the ranging channel frequency to the spacecraft RF VCO.

*Acquisition procedures.* An obvious way to enable acquisition over a wider range of frequency offsets is to sweep the VCO. This could be achieved quite simply by inserting counts into the integrator up/down counter in the linear filter, thereby sweeping the VCO control voltage. After acquisition is detected, the sweeping would be disabled. The major disadvantage to this procedure is that the first ranging component would have to be transmitted for a long enough time for the frequency to be swept across the band of possible offsets.

A second procedure might be preferable if offsets are known to occur mainly because of aging, rather than because of shorter term effects, like temperature drift. If this is the case, the center frequency could be set at the start of acquisition to the last known lock frequency. This frequency is determined by the value in the integrator-counter the last time the clean-up loop was known to be in lock. The disadvantage to this procedure is that the VCO might be set to the wrong initial frequency if the loop was really out of lock the last time it was thought to be in lock. This disadvantage may not be too serious, however, because recovery could be accomplished by transmitting the first component for a long enough time for acquisition to occur even with a high-frequency offset.

*Referencing to RF VCO.* The major disadvantage to using the RF VCO frequency as a reference for the ranging channel is that it makes the operation of the clean-up loop dependent on the configuration and operation of the rest of the spacecraft. This is not desirable, for one would like to have a mission independent ranging transponder. A second disadvantage is that cycle slipping in the RF phase-locked loop would cause phase changes in the ranging PLL. This is a less serious disadvantage, because ranging is otherwise dependent on accurate carrier lock.

Two ways to utilize a reference from the RF VCO are shown in Fig. 7. In both cases,  $f_r$  is the required ranging clock frequency, and  $f_r \pm f_c$  is a frequency derived from the RF VCO, and therefore precisely related to the ranging frequency. The required ranging frequency  $f_r$  is obtained by mixing  $f_r \pm f_c$  with  $f_c$ . The first method (Fig. 7a) is a standard mixer-modulator VCO, with the higher frequency having a precise reference. Since the VCO frequency,  $f_c$ , is fairly low, it can have a center frequency tolerance which is fairly high percentage of its center frequency, and still be a small percentage of  $f_r$ .

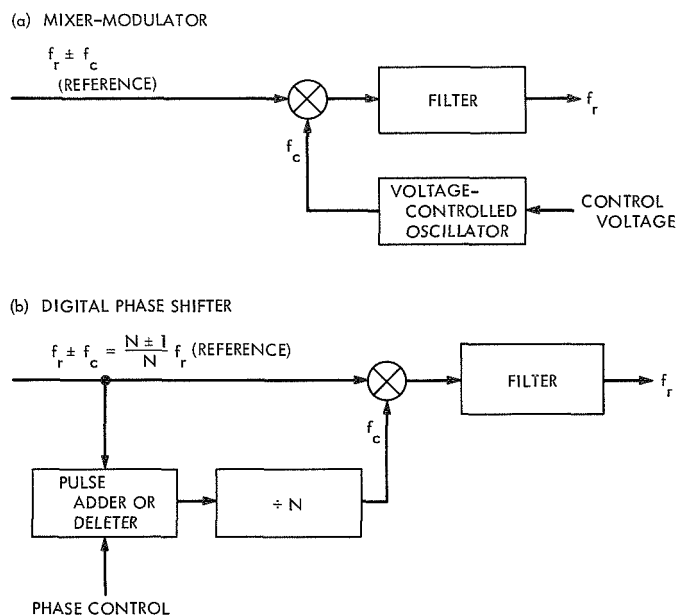


Fig. 7. Use of RF VCO reference

The remainder of the PLL could be unchanged. The second method, Fig. 7b, uses a digital phase shifter rather than a VCO. In this case,  $f_c$  is  $f_r/N$  for some convenient integer  $N$ . Since  $f_c$  is derived from  $f_r$ , its frequency is exact. Phase control of  $f_c$  is accomplished by digitally phase shifting  $f_c$  by adding or deleting pulses into the frequency divider. The phase is changed by  $2\pi/N$  each time a count is added or deleted. This method has the advantage that the frequencies are exact, so that a first-order PLL is sufficient. Furthermore, no VCO or DAC is required, since the system is entirely digital.

**b. Power reduction.** The prototype implementation of the clean-up loop was not designed to minimize power. All of the logic circuits are high-speed transistor-transistor logic even where the speed is not required, and the 80 odd integrated circuits consume about 12 W. It is obvious that some power can be saved by using slower circuits where feasible, but greater savings are achievable by reducing the need for high-speed circuits.

The circuits which consume the most power are the up/down counters. In the case of the lock detectors, this

problem can be overcome by using up counters only, after presetting to the proper value. This is possible because the total number of up and down counts during each detection interval is known, so that sufficient information is contained in the up counts only. In the case of the integrator-counter in the linear filter, the speed requirements could be reduced by updating the counter less often, say every 16  $\mu\text{sec}$  rather than every 1  $\mu\text{sec}$ . This would also eliminate one 4-bit counter chip, and would enable use of an integrated circuit DAC rather than the larger DAC now utilized because of its higher switching speed. The reduction in sampling rate can be accomplished by proper modifications to the nonlinear filter.

**c. Use of non-limited input signal.** The input signal to the clean-up loop is now assumed to be hard limited, but about a 2-dB gain can be realized by using a non-limited signal. This would enable the lock detection interval to be reduced and the PLL bandwidth to be increased, at the expense of some complication in the mechanization. First, an automatic gain control (AGC) system and an analog-to-digital converter would be required. For the PLL, it would be necessary to modify the filters to accept multi-level inputs. One way to do this would be to modify the nonlinear filter so as to have multilevel inputs and binary outputs. Modifications to the lock detectors would be more complicated. Adders with up/down counters for overflow would be required rather than up counters only, and selection of a threshold for initial lock detection would be difficult because of inaccuracies in any AGC system. Overall, it appears that the possible 2-dB improvement in performance would not justify the increased complexity of the clean-up loop.

## References

1. Martin, W. L., "A Binary-Coded Sequential Acquisition Ranging System," in *Proceedings of the Third International Conferences in System Sciences*, University of Hawaii, Honolulu, January 1970.
2. Hurd, W. J., and Anderson, T. O., "Digital Transition Tracking Symbol Synchronizer for Low SNR Coded Systems," *IEEE Trans. Commun. Technology*, Vol. COM-18, April 1970, pp. 141-147.

## B. Performance of a Phase-Locked Loop During Loss of Signal, W. J. Hurd

### 1. Introduction

The signal input to a phase-locked loop (PLL) is sometimes lost after the loop has reached steady-state conditions, as can occur in the article on ranging clean-up loops (Section A). This loss may be accidental, due to fading, for example, or it may be deliberate due to a change in the transmitted signal which causes the received signal to be orthogonal to all phase shifts of the loop reference signal. In either case, it may be desirable for the loop to remain in lock with the original input signal.

Timor (SPS 37-61, Vol. III, pp. 33-37) has considered the first passage problem for the phase error of a first-order PLL when the signal is lost, the major result being the probability that a cycle slip occurs up to time  $t$  after loss of signal. In most applications, however, the probability distribution of phase error is of more interest, for the closest lock point is the correct one whenever the absolute phase error is less than  $\pi$ , regardless of whether or not complementing cycle slips have occurred. Here we determine the joint probability distribution of phase error and all other state variables for any order PLL, conditioned on the state of the loop at the time of loss of signal. When the linearized model for the loop is valid, which is usually the case when few cycle slips are allowed, we determine the unconditional distribution of the state variables.

The increase in phase error variance as a function of time after loss of signal is shown graphically for the first-order loop with and without frequency offset and for the perfect integrator second-order loop with various damping ratios.

### 2. Model

The PLL model used is shown in Fig. 1. The noise  $n(t)$  is assumed to be white with one-sided spectral density  $N_0$ ;  $\phi(t)$  is the phase error, and  $\Omega_0$  is the offset between the frequency of the input signal and the voltage-controlled oscillator (VCO) center frequency. The sign of  $\Omega_0$  is chosen so that positive  $\Omega_0$  causes a positive change in  $\phi$ . The phase detector nonlinearity  $g(\phi)$  has slope  $A = g'(0)$  at  $\phi = 0$ , so the open-loop transfer function of the linearized model for the loop is  $AF(s)/s$ . Note that  $g(\phi)$  is dependent on the amplitude of the input signal, and is zero for all  $\phi$  when the input signal is zero or is orthogonal to all phase shifts of the VCO output signal.

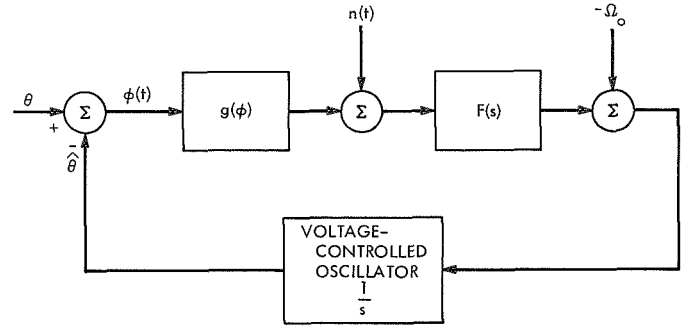


Fig. 1. Model for phase-locked loop

### 3. Problem and General Solution

Suppose that  $F(s)$  is the transfer function of an  $(N - 1)$ th order linear system, so that the loop is  $N$ th order. The state of the loop is completely described by  $\phi(t)$  and  $N - 1$  other state variables. These state variables should be judiciously chosen so as to have no white noise components so that the variances are finite. Suppose that the signal input to the loop is lost at time  $t = 0$ , i.e.,  $g(\phi) = 0$  for  $t > 0$ . We obtain the means and cross-correlation functions of all state variables for  $t \geq 0$ , conditioned on the initial conditions of the state variables at  $t = 0$ . If  $n(t)$  is gaussian, we thus obtain the conditional distribution of the state variables. Furthermore, if the loop has reached steady state by time  $t = 0$  and the loop signal-to-noise ratio is high (so that the linear approximation  $g(\phi) = A\phi$  is valid) then the state variables are jointly gaussian at  $t = 0$  and we can obtain the unconditional distribution of the state variables for  $t > 0$ .

Observe that setting  $A$  equal to zero at  $t = 0$  is equivalent to opening the loop at the output of the phase detector nonlinearity  $g(\phi)$ . The response of the system is then just the response of the open-loop system to  $n(t)$  plus the initial conditions. Let

$$X = [x_1(t), \dots, x_N(t)]$$

be the vector of the state variables, with  $\phi(t) = x_1(t)$ . Because the system is linear, we can write

$$X = Y + W \quad (1)$$

where  $Y = (y_1, \dots, y_N)$  is the portion of  $X$  due to the initial conditions and frequency offset and  $W = (w_1, \dots, w_N)$  is the portion due to the noise  $n(t)$ .

Since  $n(t)$  is zero mean, so are the  $w_i(t)$ . Hence, conditioned on  $Y(t = 0)$  the mean of  $x_i(t)$  is  $y_i(t)$ , and the covariance function of  $x_i$  and  $x_j$  is the covariance of  $w_i$

and  $w_j$ . Given  $Y(t=0)$ ,  $Y(t)$  can be evaluated for  $t \geq 0$  by several well-known techniques.

To evaluate the covariances, let  $h_i(t)$  be the weighting function relating to  $w_i(t)$  to  $n(t)$ . In other words, if  $H_i(s)$  is the transfer function from  $n(t)$  to  $w_i(t)$ , then  $h_i(t)$  is the inverse Laplace transform of  $H_i(s)$ . We can then write

$$w_i(t) = \int_0^t n(\tau) h_i(t - \tau) d\tau \quad (2)$$

The covariance of  $w_i$  and  $w_j$  is simply

$$\begin{aligned} R_{ij}(t_i, t_j) &= E \{w_i(t_i) w_j(t_j)\} \\ &= \int_0^{t_i} \int_0^{t_j} h_i(t_i - \alpha) h_j(t_j - \beta) \\ &\quad \times E \{n(\alpha) n(\beta)\} d\beta d\alpha \end{aligned} \quad (3)$$

Since

$$E \{n(\alpha) n(\beta)\} = \frac{N_0}{2} \delta(\alpha - \beta)$$

the integral with respect to  $\beta$  is equal to

$$\frac{N_0}{2} h_j(t_j - \alpha)$$

for  $0 \leq \alpha \leq t_j$  and zero elsewhere, so

$$R_{ij}(t_i, t_j) = \frac{N_0}{2} \int_0^{\min(t_i, t_j)} h_i(t_i - \alpha) h_j(t_j - \alpha) d\alpha \quad (4)$$

For  $t = t_i = t_j$ , this specializes to

$$\begin{aligned} R_{ij}(t, t) &= \frac{N_0}{2} \int_0^t h_i(t - \alpha) h_j(t - \alpha) d\alpha \\ &= \frac{N_0}{2} \int_0^t h_i(\alpha) h_j(\alpha) d\alpha \end{aligned} \quad (5)$$

#### 4. First-Order Loop

For the first-order PLL, let  $F(s) = K$  and denote the steady-state probability density of phase error by  $p_s(\phi)$ . If the linearized model is valid,  $p_s(\phi)$  is gaussian with mean equal to the static phase error  $\omega_0/AK$ , and variance

$$\sigma_s^2 = \frac{N_0 K}{4A} = \frac{N_0 b_L}{A^2} \quad (6)$$

where  $b_L$  is the one-sided loop bandwidth,  $AK/4$ . The density  $p_s(\phi)$  is also well known when the phase detector nonlinearity is sinusoidal, i.e.,  $g(\phi) = A \sin \phi$ , and is given by Viterbi (Ref. 1) and elsewhere.

To determine the statistics of  $\phi(t)$  after the signal is lost at  $t = 0$ , let  $\phi_0 = \phi(0)$  and let  $\phi_n(t) = w_1(t)$  be the component of  $\phi(t)$  due to  $n(t)$  for  $t > 0$ . Then

$$\phi(t) = \phi_0 + \Omega_0 t + \phi_n(t) \quad (7)$$

and

$$\phi_n(t) = -K \int_0^t n(t) dt \quad (8)$$

The noise term  $\phi_n(t)$  is independent of  $\phi_0$ , since the noise is white, and has zero mean and variance  $\sigma_n^2 = K^2 N_0 t / 2$ . It is, of course, gaussian if  $n(t)$  is, so the conditional distribution of  $\phi(t)$  is

$$p(\phi(t) | \phi_0) = \frac{1}{(2\pi)^{1/2} \sigma_n} \exp \left[ -\frac{(\phi - \phi_0 - \Omega_0 t)^2}{2\sigma_n^2} \right] \quad (9)$$

The unconditional distribution of  $\phi(t)$  for  $t > 0$  is

$$p_t(\phi) = \int_{-\infty}^{\infty} p(\phi(t) | \phi_0) p_s(\phi_0) d\phi_0 \quad (10)$$

and, when  $p_s(\phi_0)$  is gaussian,  $\phi(t)$  is gaussian with mean

$$\frac{\Omega_0}{AK} + \Omega_0 t \quad (11)$$

and variance

$$\sigma_t^2 = \sigma_s^2 + \sigma_n^2 = \frac{N_0 K}{4A} (1 + 2AKt) \quad (12)$$

It is interesting to rewrite this expression in terms of the two-sided loop bandwidth  $w_L$ ,

$$\sigma_t^2 = \frac{N_0 w_L}{2A^2} (1 + 4w_L t) \quad (13)$$

When the signal is removed, the variance of phase error doubles in one fourth of an inverse loop bandwidth, or one half of one loop time constant.

#### 5. Second-Order Loop

For a second-order loop we can find the unconditional probability density of  $\phi$  (and of the other state variable) only when the loop is linearized. This is because the joint

density of the two state variables in the steady state is not known, even though  $p_s(\phi)$  is known for some  $g(\phi)$ .

Suppose the loop filter is

$$F(s) = K \frac{1 + \tau_2 s}{1 + \tau_1 s} \quad (14)$$

If we let  $K_2 = K/\tau_1$  and  $K_1 = K_2\tau_2$ , then  $F(s)$  reduces to the filter for a perfect integrator second-order loop as  $\tau_1 \rightarrow \infty$  with  $K_1$  and  $K_2$  constant,

$$F(s) = K_1 + K_2/s, \quad \tau_1 = \infty \quad (15)$$

Appropriate state variables can be chosen by factoring  $F(s)/s$  into

$$\frac{F(s)}{s} = \frac{K_2}{s + 1/\tau_1} \left[ \frac{K_1}{K_2} + \frac{1}{s} \right] \quad (16)$$

and considering the factors of  $F(s)$  to be implemented in series, with the  $K_2/(s + 1/\tau_1)$  factor first, as shown in Fig. 2. We then chose  $\phi$  and the negative of the output of the first factor,  $\omega$ , as the state variables. This choice is appropriate both because  $\omega$  has no white noise component and because  $E\{\omega\} = -\Omega_0$  for the perfect integration case.

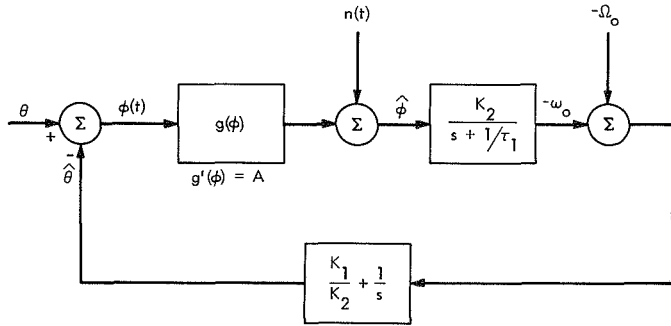


Fig. 2. Model for second-order loop

In the formulation of Eq. (1),

$$X(t) = (\phi(t), \omega(t)) = (x_1(t), x_2(t)) = Y + W$$

where  $Y$  is the portion of  $X$  due to initial conditions and  $\Omega_0$ , and  $W$  is the portion due to noise for  $t > 0$ .

**a. Conditional distribution.** The components of  $\phi$  and  $\omega$  due to  $\Omega_0$  and the initial conditions can be obtained by inspection. For the imperfect integrator,  $\tau_1 < \infty$ ,

$$y_2[t|Y(0)] = \omega_0 \exp(-t/\tau_1) \quad (17)$$

and

$$\begin{aligned} y_1[t|Y(0)] &= \frac{K_1}{K_2} [\Omega_0 + y_2(t)] + \int_0^t [\Omega_0 + y_2(\tau)] d\tau + C \\ &= \phi_0 + \Omega_0 t + \omega_0 \left( \tau_1 - \frac{K_1}{K_2} \right) [1 - \exp(-t/\tau_1)] \end{aligned} \quad (18)$$

For the perfect integrator,  $\tau_1 = \infty$ ,

$$y_2[t|Y(0)] = \omega_0 \quad (19)$$

and

$$y_1[t|Y(0)] = \phi_0 + (\Omega_0 + \omega_0)t \quad (20)$$

Note that  $y_1$  and  $y_2$  are the means of  $\phi$  and  $\omega$  for  $t > 0$ , conditioned on  $\phi_0$  and  $\omega_0$ .

To determine the conditional variances of  $\phi$  and  $\omega$ , we evaluate the variances of  $w_1(t)$  and  $w_2(t)$ , the components of  $\phi(t)$  and  $\omega(t)$ , respectively, due to white noise after  $t = 0$ . For the imperfect integrator case, the weighting functions from  $n(t)$  to  $w_2$  and  $w_1$  are

$$h_2(\tau) = u(\tau) K_2 \exp(-\tau/\tau_1) \quad (21)$$

where  $u(\tau)$  is the unit step function, and

$$h_1(\tau) = u(\tau) K_2 \left[ \frac{1}{\tau_1} + \left( 1 - \frac{1}{\tau_1} \right) \exp(-\tau/\tau_1) \right] \quad (22)$$

Similarly, for  $\tau_1 = \infty$ ,

$$h_2(\tau) = K_2 u(\tau) \quad (23)$$

and

$$h_1(\tau) = u(\tau) [K_1 + K_2 \tau] \quad (24)$$

Applying Eq. (5), the variance of  $\omega$  conditioned on  $\phi_0$  and  $\omega_0$  is

$$\sigma_2^2(t) = \frac{N_0 K_2^2 \tau_2}{4} [1 - \exp(-2t/\tau)], \quad \tau_1 < \infty \quad (25)$$

$$= \frac{N_0 K_2^2 t}{2}, \quad \tau_1 = \infty \quad (26)$$

and the variance of  $\phi$  conditioned on  $\phi_0$  and  $\omega_0$  is

$$\sigma_{\phi}^2(t) = \frac{N_0 K_2^2}{2} \left( \frac{t}{\tau_1^2} + \left( 1 - \frac{1}{\tau_1} \right) [1 - \exp(-t/\tau_1)] + \frac{(\tau_1 - 1)^2}{2\tau_1} [1 - \exp(-2t/\tau_1)] \right), \quad \tau_1 < \infty \quad (27)$$

$$= \frac{N_0}{2} \left[ K_1^2 t + K_1 K_2 t^2 + \frac{K_2^2}{3} t^3 \right], \quad \tau_1 = \infty \quad (28)$$

Similarly, the conditional covariance of  $\phi(t)$  and  $\omega(t)$  is

$$\sigma_{\phi\omega}^2(t) = \frac{N_0 K_2^2}{2} \left[ \exp(-t/\tau_1) + \left( \frac{\tau_1 - 1}{2} \right) \exp(-2t/\tau_1) \right], \quad \tau_1 < \infty \quad (29)$$

$$= \frac{N_0}{2} (K_1 K_2 t + \frac{1}{2} K_2^2 t^2), \quad \tau_1 = \infty \quad (30)$$

At time  $t$ , the state variables  $\phi$  and  $\omega$ , conditioned on  $\phi_0$  and  $\omega_0$ , are jointly gaussian with means  $y_1$  and  $y_2$ , variances  $\sigma_1^2$  and  $\sigma_2^2$ , and covariance  $\sigma_{12}^2$ .

**b. Unconditional distributions.** To obtain the unconditional distributions of  $\phi(t)$  and  $\omega(t)$ , we must utilize the distributions of  $\phi_0$  and  $\omega_0$ . The means are given by Eqs. (17) to (20), so it remains only to obtain the second moments when the linear PLL model is assumed and  $\phi_0$  and  $\omega_0$  are jointly gaussian.

In terms of the damping ratio  $\zeta$ , and the loop natural frequency  $\beta$ , the closed-loop transfer functions from  $n(t)/A$  to  $\phi$  and  $\omega$ , respectively

$$F_1(s) = \frac{\left( 2\zeta\beta - \frac{1}{\tau_1} \right) s + \beta^2}{s^2 + 2\zeta\beta s + \beta^2} \quad (31)$$

and

$$F_2(s) = \frac{\beta^2 s}{s^2 + 2\zeta\beta s + \beta^2} \quad (32)$$

where

$$2\zeta\beta = AK_1 + 1/\tau_1$$

and

$$\beta^2 = AK_2$$

Using the steady state linear theory, the second moments are obtained by integrating the (cross) spectral densities, e.g.,

$$E\{\phi_0\omega_0\} = \frac{N_0}{2A^2} \int_{-\infty}^{\infty} F_1(j\omega) F_2^*(j\omega) \frac{d\omega}{2\pi} \quad (33)$$

We present further details only for the perfect integrator case,  $\tau_1 = \infty$ . For this case,

$$E\{\phi_0^2\} = \frac{N_0}{2A^2} \frac{\beta}{2} \left( 2\zeta + \frac{1}{2\zeta} \right) \quad (34)$$

$$E\{\omega_0^2\} = \frac{N_0}{2A^2} \frac{\beta^3}{4\zeta} \quad (35)$$

and

$$E\{\phi_0\omega_0\} = \frac{N_0}{2A^2} \frac{\beta^2}{2} \quad (36)$$

The two-sided loop bandwidth is

$$w_L = \frac{\beta}{2} \left( 2\zeta + \frac{1}{2\zeta} \right) \text{Hz} \quad (37)$$

The moments of  $Y$  can now be obtained using Eqs. (19) and (20). The means are zero, since  $E\{\omega_0\} = -\Omega_0$ . The second moments are

$$\begin{aligned} E\{y_1^2\} &= E\{(\phi_0 + \omega_0 t)^2\} \\ &= \frac{N_0}{2A^2} \left[ w_L + \beta^2 t + \frac{\beta^3}{4\zeta} t^2 \right] \end{aligned} \quad (38)$$

$$E\{y_2^2\} = E\{\omega_0^2\} \quad (39)$$

and

$$E\{y_1 y_2\} = E\{\phi_0 \omega_0\} + E\{\omega_0^2\} t \quad (40)$$

To obtain the moments of

$$X = [\phi(t), \omega(t)]$$

we need merely add the moments of  $W$  and  $Y$ . In particular, the variance of  $\phi(t)$  is

$$\begin{aligned} \sigma_{\phi}^2(t) &= \frac{N_0}{2A^2} \left[ w_L + (4\zeta^2 + 1) \beta^2 t \right. \\ &\quad \left. + \left( 2\zeta + \frac{1}{4\zeta} \right) \beta^3 t^2 + \frac{\beta^4}{3} t^3 \right] \end{aligned} \quad (41)$$



Since  $\beta = 4\zeta w_L / (4\zeta^2 + 1)$ , we can express the increase in the variance of  $\phi$  with time as

$$\frac{\sigma_\phi^2(t)}{\sigma_\phi^2(0)} = 1 + \frac{16\zeta^2}{4\zeta^2 + 1} w_L t + \frac{8\zeta^2 + 1}{4\zeta} \left( \frac{4\zeta}{4\zeta^2 + 1} \right)^3 (w_L t)^2 + \frac{1}{3} \left( \frac{4\zeta}{4\zeta^2 + 1} \right)^4 (w_L t)^3 \quad (42)$$

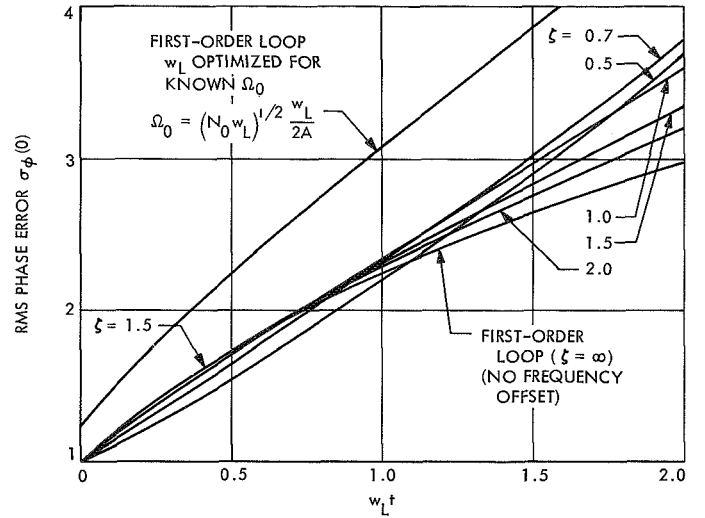
## 6. Comparison of Phase Error for First- and Second-Order Loops

The variance of phase error at time  $t$  sec after loss of signal is given by Eq. (12) for the first-order loop and by Eq. (42) for the perfect integrator second-order loop. To obtain the mean square phase error, we must add the square of the static phase error to the variance. This static phase error is zero for the second-order case, and is  $\Omega_0/AK$  for the first-order case. To present comparative results, it is convenient to normalize the rms phase error by  $\sigma_\phi(0)$ , which eliminates dependence on  $N_0$ ,  $A$ , and  $w_L$ . Since there is no static error for the second-order case, the rms phase error, normalized by  $\sigma_\phi(0)$ , depends only on  $t$  and  $\zeta$ . For the first-order case, it depends only on  $t$  and  $\Omega_0$ .

Figure 3 shows the rms phase error, normalized by  $\sigma_\phi(0)$ , for the second-order loop with several damping ratios and for the first-order loop with no frequency offset and with offset

$$\Omega_0 = (N_0 w_L)^{1/2} \frac{w_L}{2A}$$

This particular frequency offset was chosen because it corresponds to the case when  $w_L$  is optimized for given  $N_0$ ,  $A$ , and  $\Omega_0$ , in the sense that rms phase error is minimized when the signal is present. The rms phase error at



**Fig. 3. RMS phase error versus time after loss of signal for second-order phase-locked loop**

$t = 0$  is thus as small as is possible with a first-order loop. When there is no frequency offset, there is little difference in performance between the first- and second-order loops for small  $w_L t$ , but the first-order loop appears to be better for large  $w_L t$ . This is deceptive, however, for when there is any frequency offset, the optimum bandwidth first-order loop is worse than a second-order loop of the same bandwidth. It is interesting that the damping ratio of the second-order loop is not an important parameter, so that its selection would normally be determined by other considerations in the design of a system.

## Reference

1. Viterbi, A. J., "Principles of Coherent Communication," Chap. 4. McGraw-Hill Book Co., New York, 1966.

## C. Optimum Configurations for PSK/PM Systems, U. Timor

### 1. Introduction

Let  $\hat{a}_1(t), \dots, \hat{a}_N(t)$  be  $N$  binary data sequences to be transmitted over a single coherent RF channel by phase modulating the carrier with biphas-modulated squarewave subcarriers (phase-shift keyed/phase modulated). The transmitted signal is

$$s(t) = \sqrt{2P} \sin[\omega_c t + \theta(t)] \quad (1)$$

where  $P$  is the total power and  $\theta(t)$  is a  $2^N$ -level function of the data according to the type of modulation scheme used. Two methods previously investigated are:

**a. Conventional method (Ref. 1).** This method uses a linear combination of the data streams, i.e.,

$$\theta(t) = \sum_{k=1}^N \theta_k \hat{a}_k(t) \text{sq}(\omega_k t) = \sum_{k=1}^N \theta_k a_k(t) \quad (2)$$

where  $\theta_k$  are the modulation angles,  $\text{sq}(\omega_k t)$  are square-wave subcarriers of frequency  $\omega_k$ , and  $a_k(t) = \hat{a}_k(t) \text{sq}(\omega_k t)$ .

**b. Interplex method (SPS 37-62, Vol. III, pp. 57-60; and SPS 37-63, Vol. III, pp. 34-37).** This method performs some intermultiplications of data sequences to yield the modulating angle

$$\theta(t) = \theta_1 a_1(t) + a_1(t) \sum_{k=2}^N \theta_k a_k(t) \quad (3)$$

Comparing these two modulation schemes, it was shown (SPS 37-65, Vol. III, pp. 27-31) that the Interplex method out-performs the conventional one when  $N$  is small, while for large  $N$ , the power allocation to the data channels determines which configuration is superior.

The above methods are but two of many possible configurations of generating the  $2^N$ -level phase  $\theta(t)$ . The question arises whether, under some given circumstances, there is another scheme which out-performs both the conventional and Interplex methods. We therefore consider the general modulation angle

$$\theta(t) = \sum_{k=1}^N \theta_k x_k(t) \quad (4)$$

where  $x_k(t)$  are binary waveforms generated by the data sequences  $a_k(t)$ . Clearly the conventional scheme  $x_k(t) = a_k(t)$ , and the Interplex method  $x_1(t) = a_1(t)$ ,  $x_k(t) = a_1(t) a_k(t)$   $k \geq 2$ , are special cases of Eq. (4).

It will be shown that for all  $N$  and all possible power allocation to the data and RF channels, either the Interplex or the conventional scheme are optimum. This optimum might not be unique; that is, some other configuration could have the same efficiency. However, no other configuration could out-perform both the conventional and Interplex system. Therefore, to find the optimum performance, it is enough to consider those two schemes.

### 2. System Optimization

Let the transmitted signal be

$$S(t) = \sqrt{2P} \sin\left(\omega t + \sum_{k=1}^N \theta_k x_k(t)\right) \quad (5)$$

where  $x_k(t)$  are binary waveforms generated by the data.

Correlating the received signal with a coherent reference ( $\sqrt{2} \cos \omega t$ ) and its quadrature ( $\sqrt{2} \sin \omega t$ ), and performing subcarrier demodulation and filtering, we could separate any of the  $2^N - 1$  signals of the form

$$y_{\underline{\lambda}}(t) = \prod_{k=1}^N [x_k(t)]^{\lambda_k} \quad (6)$$

where

$$\underline{\lambda} = \{\lambda_1, \dots, \lambda_N\}$$

is a binary vector ( $\lambda_k = 0$  or  $1$ ) and can assume all  $2^N - 1$  non-zero combinations. The relative power of the signal  $y_{\underline{\lambda}}(t)$  is given by

$$P_{\underline{\lambda}}(t) = \prod_{k=1}^N (\sin^2 \theta_k)^{\lambda_k} (\cos^2 \theta_k)^{(1-\lambda_k)} \quad (7)$$

Let  $\alpha_k = P_k/P_1$ ,  $k = 1, \dots, N$  be the required ratio of powers in the  $k$ th channel to the first one, and let  $\tilde{P}_{\text{RF}}$  be the minimum RF power required to track the frequency and phase.

An allowable configuration  $C$  of the system is defined as any collection of  $N$  independent signals from the  $2^N - 1$  possible signals  $y_{\underline{\lambda}}(t)$  and corresponding modulation angles  $\theta_1, \dots, \theta_N$  which satisfy the required ratios  $\alpha_k$ ,  $k = 1, \dots, N$  and which yield an RF power greater than or equal to  $\tilde{P}_{\text{RF}}$ .

An optimum configuration  $\hat{C} = \{\hat{y}_k(t), \hat{\theta}_k, k = 1, \dots, N\}$ , not necessarily unique, is one which maximizes the total data power among all allowable configurations.

Without loss of generality we can assume that:

$$\left. \begin{aligned} \theta_k &\leq \frac{\pi}{2}, & k = 1, \dots, N \\ \theta_1 &= \max_k \theta_k \end{aligned} \right\} \quad (8)$$

The optimum configurations are given by the following theorem.

**THEOREM 1.** *Let*

$$\hat{C} = \{\hat{y}_k(t), \hat{\theta}_k, k = 1, \dots, N\}$$

*be an optimum configuration satisfying the required RF power and data power ratios.*

*Then  $\hat{C}$  is either identical or equivalent (with respect to power allocation) to*

$$\tilde{C} = \{\tilde{y}_k(t), \tilde{\theta}_k, k = 1, \dots, N\}$$

*where  $\tilde{C}$  is given by*

$$\begin{aligned} \tilde{y}_1(t) &= x_1(t) \\ \tilde{y}_k(t) &= x_k(t) & \text{if } \hat{\theta}_1 \leq \frac{\pi}{4} \\ &= x_1(t) x_k(t) & \text{if } \hat{\theta}_1 > \frac{\pi}{4} \end{aligned} \quad (9)$$

$k = 2, \dots, N$

The proof is given in *Paragraph 3*.

To obtain estimates of the data sequences at the receiver we assign

$$\tilde{y}_k(t) = a_k(t), \quad k = 1, \dots, N \quad (10)$$

From Theorem 1, we obtain the modulating binary signals:

$$\begin{aligned} x_1(t) &= \tilde{y}_1(t) = a_1(t) \\ x_k(t) &= a_k(t), & \hat{\theta}_1 \leq \frac{\pi}{4} \\ &= a_1(t) a_k(t), & \hat{\theta}_1 > \frac{\pi}{4} \end{aligned} \quad (11)$$

We thus have the following result:

**THEOREM 2.** *Given the data power ratios  $\alpha_k$  ( $k=1, \dots, N$ ) and the required RF power, either the conventional system or the Interplex system is optimal. Other configurations might achieve the same efficiency but could never out-perform both of them.*

### 3. Proof of Theorem 1

*a. Case I:  $\hat{\theta}_1 \leq \pi/4$*

From Eq. (8) we have

$$\hat{\theta}_k \leq \frac{\pi}{4} \quad k = 1, \dots, N, \quad (12)$$

and therefore

$$\cos^2 \hat{\theta}_k \geq \sin^2 \hat{\theta}_k \quad k = 1, \dots, N \quad (13)$$

Each term  $\sin^2 \theta_k$  in  $P_\lambda$  (Eq. 7) indicates the presence of  $x_k(t)$  in  $y_\lambda(t)$  (Eq. 6), and therefore must appear at least once in  $P_1, \dots, P_N$  (otherwise the  $N$  signals

$$\hat{y}_1(t), \dots, \hat{y}_N(t)$$

will be dependent). However, if for some  $k$ ,  $\sin^2 \theta_k$  appears more than once, we can substitute it with  $\cos^2 \theta_k$  and thus increase the total data power (Eq. 13) while preserving signal independence. It can be shown that due to the continuity of the  $\theta_k$ 's, the increase in data power can be redistributed to preserve the required power ratios. Thus the data power is maximized if each  $\sin^2 \theta_k$  appears once and only once to yield:

$$\left. \begin{aligned} \tilde{y}_k(t) &= x_k(t) & k = 1, \dots, N \\ P_k &= \sin^2 \theta_k \prod_{i=1}^N \cos^2 \theta_i \end{aligned} \right\} \quad (14)$$

Moreover if  $\theta_1 < \pi/4$  the optimum is unique

QED. Case I

*b. Case II:  $\hat{\theta}_1 > \pi/4$*

#### LEMMA 1

*If  $\hat{\theta}_1 > \pi/4$  then there exists an optimum configuration satisfying:*

$$\hat{y}_k(t) = x_1(t) z_k(t), \quad k = 1, \dots, N \quad (15)$$

*where  $z_k(t)$  is independent of  $x_1(t)$ .*

**Proof**

Suppose this is not the case. Then for every optimum configuration there exists some  $\tilde{y}_j(t)$  independent of  $x_1(t)$ . Hence

$$P_j = \cos^2 \theta_1 A_j(\theta_2, \dots, \theta_N) \quad (16)$$

since  $\sin^2 \theta_1 > \cos^2 \theta_1$ , the signal

$$u(t) = x_1(t) \tilde{y}_j(t) \quad (17)$$

has a power

$$P_u = \sin^2 \theta_1 A_j(\theta_2, \dots, \theta_N) > P_j \quad (18)$$

Therefore  $u(t)$  belongs to the optimum set, i.e.,  $u(t) = \tilde{y}_l(t)$ , otherwise the data power would be increased by substituting  $\tilde{y}_j(t)$  with  $u(t)$ .

From Eq. (16) it is evident that  $A_j(\theta_2, \dots, \theta_N)$  must include some  $\sin^2 \theta_i$  term which we can assume without loss of generality to be  $\sin^2 \theta_2$ . Thus

$$P_j = \cos^2 \theta_1 \cdot \sin^2 \theta_2 B(\theta_3, \dots, \theta_N) \quad (19)$$

$$P_l = \sin^2 \theta_1 \cdot \sin^2 \theta_2 B(\theta_3, \dots, \theta_N)$$

and

$$P_l > P_j \quad (20)$$

Consider the signal  $v(t)$  associated with the power

$$P_v = \sin^2 \theta_1 \cdot \cos^2 \theta_2 B(\theta_3, \dots, \theta_N) \quad (21)$$

Clearly  $\tilde{y}_j(t)$ ,  $\tilde{y}_l(t)$  and  $v(t)$  are linearly dependent, and thus, at most, two of them could belong to the optimum set. Also,  $\theta_1 \geq \theta_2$  and therefore

$$P_v \geq P_j \quad (22)$$

Thus  $\tilde{y}_j(t)$  has the smallest power among the three dependent signals and can be excluded from the optimum set, contrary to our assumption.

QED. Lemma 1

**LEMMA 2**

*There exists an optimum solution satisfying*

$$\hat{\theta}_k \leq \frac{\pi}{4} \quad k = 2, \dots, N \quad (23)$$

**Proof**

Let  $\hat{\theta}_i$ ,  $i = 1, \dots, N$  be an optimum set of modulation angles and suppose we have, for some  $k$ ,  $\hat{\theta}_k > \pi/4$ .

Define

$$\tilde{\theta}_k = \frac{\pi}{2} - \hat{\theta}_k < \frac{\pi}{4} \quad (24)$$

Since  $\sin^2 \tilde{\theta}_k = \cos^2 \hat{\theta}_k$  and  $\cos^2 \tilde{\theta}_k = \sin^2 \hat{\theta}_k$  we could modify our data set to satisfy

$$\tilde{P}_j = \sin^2 \tilde{\theta}_k A_j = \cos^2 \hat{\theta}_k A_j = \hat{P}_j \quad (25)$$

and

$$\tilde{P}_j = \cos^2 \tilde{\theta}_k A_j = \sin^2 \hat{\theta}_k A_j = \hat{P}_j$$

Repeating for all  $\hat{\theta}_j > \pi/4$  we obtain a new set of  $N$  signals, having the same data power (and thus optimum) and satisfying  $\tilde{\theta}_k \leq \pi/4$ ,  $k = 2, \dots, N$ .

QED. Lemma 2

Consider the optimum solution satisfying (by Lemma 2):

$$\left. \begin{array}{l} \hat{\theta}_1 > \pi/4 \\ \tilde{\theta}_k \leq \pi/4, \quad k = 2, \dots, N \end{array} \right\} \quad (26)$$

by Lemma 1,

$$\hat{y}_k(t) = x_1(t) z_k(t), \quad k = 1, \dots, N$$

For  $k \geq 2$  we have  $\cos^2 \hat{\theta}_k \geq \sin^2 \hat{\theta}_k$ , and arguing as in Case I, we can show that each  $x_k(t)$  appears once and only once, thus

$$\hat{y}_k(t) = x_1(t) x_k(t), \quad k = 2, \dots, N \quad (27)$$

The  $N$ th signal is  $\hat{y}_1(t) = x_1(t)$ , which has the highest power, since

$$\begin{aligned} P_1 &= \sin^2 \theta_1 \prod_{k=2}^N \cos^2 \theta_k \\ &> \sin^2 \theta_1 \cdot \sin^2 \theta_j \prod_{\substack{k=2 \\ k \neq j}}^N \cos^2 \theta_k \\ &= P_j \end{aligned} \quad (28)$$

QED.

#### 4. Conclusion

Given the design conditions, that is the power ratios in the data channels, and the minimum required RF power, it was shown that the optimum modulation scheme is either the Interplex or the conventional method. Which of the two is superior depends on the design conditions; also, other configurations might achieve the same performance. However, in no case can another configuration out-perform both Interplex and conventional system.

#### D. The Critical Problem and Coding Theory,

N. White

##### 1. Introduction

In this paper we shall investigate in detail the interesting fact that the "critical problem" of combinatorial geometry contains as a special case the problem of determining the largest dimension possible for a linear code over  $GF(q)$  of fixed length and minimum distance. In *Subsection 2* we introduce the reader to the critical problem and its relationship to coding theory, and present a sketch of Crapo and Rota's solution of the critical problem in terms of the characteristic polynomial of a certain geometric lattice. In the remainder of the paper we restrict attention to  $q = 2$ ; i.e., binary codes. In *Subsection 3* we show that the critical problem for distance three (single-error correcting) codes has a nice solution in terms of combinatorial geometry. And while this problem has been solved before, our proof is instructive and shows how the highly developed techniques of combinatorial geometry are beginning to cast light on coding theory. In *Subsection 4* we give techniques which are useful for computing the characteristic polynomials for distance- $d$  codes,  $d \geq 4$ , and show how the calculations for a code of block length  $n$  can in fact be performed in a space of  $n - 3$  dimensions. These results take on especially satisfying form for  $d = 5$ ; we show the equivalence of the  $d = 5$  problem to a certain generalized coloring problem. One hopes that the  $d = 5$  coding problem can be solved in this manner. At present it is unsolved, and the techniques of conventional combinatorial coding theory seem exhausted on this subject. Finally, in *Subsection 5* we obtain a recursion for the characteristic polynomial for  $d = 4$  codes. The

Therefore, to find the best performance achievable by an  $N$ -channel phase-shift-keyed/phase-modulated digital coherent system, the investigation can be narrowed to these two schemes.

#### Reference

1. Lindsey, W. C., "Design of Block-Coded Communication Systems," *IEEE Trans. Commun. Tech.*, Vol. COM-15, No. 4, August 1967.

paper concludes with a short table of the known characteristic polynomials, and should provide useful information from which future theorems may be induced.

##### 2. The Critical Problem

Let  $V_q^n$  be a vector space of dimension  $n$  over  $GF(q)$ , the finite field of order  $q$ . If  $x \in V_q^n$  we denote

$$x = (x^{(1)}, x^{(2)}, \dots, x^{(n)})$$

Let  $S$  be any set of nonzero vectors in  $V_q^n$ , thought of as *error patterns*. We define an *optimal linear code with respect to  $S$*  to be any subspace  $U \subseteq V_q^n$  of maximum dimension such that  $S \cap U = \phi$ . For any  $S$  we denote some such subspace by  $U(S)$ , even though  $U$  in general is not unique. If  $k = k(S) = \dim(U(S))$ , then  $U(S)$  is an  $(n, k)$  linear (block) code on  $q$  symbols such that no two of its code words differ by a vector in  $S$ . Thus, the code can be used to detect error patterns from  $S$ .

Several examples we will use are given below; other classes, such as those arising from burst errors, are also of interest.

$$\begin{aligned} B_w^n &= \{x \in V_q^n \mid x \neq 0, \quad \text{Hamming weight}(x) \leq w\} \\ D_w^n &= \{x \in V_q^n \mid \text{Hamming weight}(x) = w\} = B_w^n - B_{w-1}^n \\ E_w^n &= \{x \in V_q^n \mid x \neq 0, \quad \text{Hamming weight}(x) \text{ is even} \\ &\quad \text{and} \leq w\} \end{aligned}$$

Thus  $U(B_w^n)$  is a linear code of maximum rate with given block length  $n$  and minimum weight  $w + 1$ . We note that

if  $\zeta: V_2^n \rightarrow V_2^{n+1}$  is the operation of adding a parity check bit to a binary code word, then  $\zeta(V_2^n) \cong E_{n+1}^{n+1}$ , and  $\zeta(B_w^n) \cong E_w^{n+1}$  if  $w$  is even.

Any subspace of dimension  $\ell$  in  $V_q^n$  is the intersection of  $n - \ell$  hyperplanes. Thus the problem of finding  $k(S)$  is equivalent to finding the minimum number of hyperplanes  $H_1, H_2, \dots, H_c$  such that  $H_1 \cap H_2 \cap \dots \cap H_c \cap S = \phi$ , by taking  $U(S) = H_1 \cap H_2 \cap \dots \cap H_c$ . Crapo and Rota (Ref. 1) call this problem the *critical problem* of combinatorial geometry, with the solution  $c = c(S) = n - k(S)$  called the *critical exponent*. We now sketch Crapo and Rota's solution to the critical problem in terms of the characteristic polynomial of  $S$ .

We begin with the partially ordered set  $L(S)$  of all subspaces of  $V_q^n$  which are spanned by elements of  $S$ , ordered by inclusion. By intersecting these subspaces with  $S$ , we have equivalently  $L(S) = \{A \subseteq S \mid A = \bar{A}\}$ , again ordered by inclusion, where  $\bar{A} = (\text{span } A) \cap S$  is the *closure* of  $A \subseteq S$ .  $L(S)$  is a finite geometric lattice with *bottom element*  $0 = \phi$  and *top element*  $1 = S$ . However, we will restrict our notation to the case  $L(S)$ , so that the reader will require no previous knowledge of geometric lattices or the equivalent concept of combinatorial geometries (Ref. 1), although the concepts are more general.

The *rank function*  $r: L(S) \rightarrow \mathbb{Z}$ , (the integers) is given, for  $x \in L(S)$ , by  $r(x) = \text{dimension}(x)$ , considering  $x$  as a sub-set of  $V_q^n$ . It is elementary to observe that  $r(x) = k$  if and only if  $x$  has a *basis* of cardinality  $k$ , where a basis is an independent subset of maximum cardinality.

On  $L(S)$ , or more generally, on any locally finite partially ordered set, we have the *Möbius function* (Ref. 2)  $\mu_{L(S)}(x, y): L(S) \times L(S) \rightarrow \mathbb{Z}$ , satisfying

- (1)  $\mu_{L(S)}(x, y) = 0$  unless  $x \leq y$  in  $L(S)$
- (2)  $\mu_{L(S)}(x, x) = 1$  for all  $x \in L(S)$
- (3)  $\sum_{y: x \leq y \leq z} \mu_{L(S)}(x, y) = 0$  if  $x < z$  in  $L(S)$

The *characteristic polynomial* of  $L(S)$ ,  $p(\lambda; L(S)) = p(\lambda; S)$  is defined by

$$p(\lambda; L(S)) = \sum_{x \in L(S)} \mu_{L(S)}(0, x) \lambda^{r(1) - r(x)}$$

**THEOREM 1.** (Ref. 1, p. 16.4.) *The number of ordered sequences  $L_1, L_2, \dots, L_d$  of  $d$  linear functionals*

$$L_i: V_q^n \rightarrow GF(q)$$

*not necessarily distinct, such that*

$$\left( \bigcap_{i=1}^d \ker L_i \right) \cap S = \phi$$

*is the number  $q^{d(n-r(S))} p(q^d; S)$ . Thus  $p(q^d; S) = 0$  for  $d = 0, 1, 2, \dots, c(S) - 1$ , and  $p(q^d; S) > 0$  for  $d \geq c(S)$ .*

It is interesting to note that we get information about the largest subspace of  $V_q^n - S$  by studying the internal geometric structure of  $S$ , as expressed in  $L(S)$ .

We now examine the critical problem for the case  $q = 2$ ,  $S = B_w^n$ . We let  $a_{ij} \dots_k \in V_2^n$  be the binary vector

$$\begin{aligned} a_{ij}^{(l)} \dots_k &= 1 \text{ if } l = i, j, \dots, k \\ &= 0 \text{ otherwise} \end{aligned}$$

### 3. The Critical Problem for $B_2^n$ , $q = 2$

The following example is not new, but it motivates the study of the critical problem.

**LEMMA 2**  $L(B_2^n) \cong P_{n+1}$ , *the lattice of partitions of a set of  $n + 1$  elements, where  $P_{n+1}$  is ordered by  $\sigma_1 \leq \sigma_2$  if  $\sigma_1$  is a refinement of  $\sigma_2$ .*

**Proof.**  $B_2^n \cong E_2^{n+1} = D_2^{n+1}$ . Let  $X = \{x_1, x_2, \dots, x_{n+1}\}$  and let  $A$  be a closed subset of  $D_2^{n+1}$ . We define the blocks of a partition  $\sigma(A)$  on  $X$  to be the equivalence classes of the relation  $x_i \sim x_j$  if and only if  $a_{ij} \in A$  or  $i = j$ . It is clear that  $\sim$  is an equivalence relation; in particular, if  $x_i \sim x_j$  and  $x_j \sim x_k$ , then either  $i = j$  or  $j = k$ , hence  $x_i \sim x_k$ ; or else  $a_{ij}, a_{jk} \in A$ , hence  $a_{ik} = a_{ij} + a_{jk} \in A$ ; and again  $x_i \sim x_k$ . Thus  $\sigma: L(B_2^n) \rightarrow P_{n+1}$ .

On the other hand, given  $\pi \in P_{n+1}$ , let

$$\sigma^{-1}(\pi) = \{a_{ij} \in S_2^n \mid x_i \approx x_j, i \neq j\}$$

where  $\approx$  is the equivalence relation whose classes are the blocks of  $\pi$ . Let  $a_{ki} \in \text{span } (\sigma^{-1}(\pi))$ . Since we are working over  $GF(2)$ , a minimal subset of  $\sigma^{-1}(\pi)$  whose span contains  $a_{ki}$  must be a simple circuit with  $a_{ki}$  removed,  $a_{im}, a_{mn}, \dots, a_{hk}$ . Thus  $x_i \approx x_k$  and  $a_{ki} \in \sigma^{-1}(\pi)$ , and  $\sigma: L(B_2^n) \cong P_{n+1}$ .

**QED.**

But  $P_{n+1} \cong LC(C_{n+1})$ , the lattice of contractions of the complete graph  $C_{n+1}$  on  $n + 1$  vertices, by contracting all edges between vertices corresponding to elements of the same block of a partition. Here  $LC(G)$  is ordered so that  $G$  itself is the bottom element, for any graph  $G$ .

The reader may have noticed that the weight-two vectors  $D_2^{n+1}$  correspond in the obvious way to the edges of  $C_{n+1}$  under the above isomorphisms. Similarly, any graph  $G$  on  $n+1$  vertices corresponds to some subset  $S_G \subseteq D_2^{n+1}$ , with  $L(S_G) \cong LC(G)$ . Furthermore,  $p(\lambda; S_G) = 1/\lambda C(\lambda; G)$  where  $C(\lambda; G)$  is the chromatic polynomial of  $G$  (e.g., Ref. 3, pp. 234-235).

Thus for  $G = C_{n+1}$

$$p(\lambda; B_2^n) = (\lambda - 1)(\lambda - 2) \cdots (\lambda - n)$$

Hence by Theorem 1

$$k(B_2^n) = n - c(B_2^n) = n - [\log_2 n] - 1$$

It is interesting to note that the solution of the critical problem for all  $S_G$  such that  $G$  is a planar graph is equivalent to the famous four-color problem, since  $G$  is four-colorable if and only if  $C(S_G) \leq 2$ .

#### 4. The Critical Problem for $B_w^n$ , $w \geq 3$ , $q = 2$

By the following proposition, it suffices to consider only the case that  $w$  is even.

**PROPOSITION 3.**  $k(B_{2l}^n) = k(B_{2l+1}^{n+1})$  for  $l, n \geq 0$ , hence  $C(B_{2l}^n) = C(B_{2l+1}^{n+1}) - 1$ .

*Proof.* The parity check operator

$$\xi: V_2^n \xrightarrow{\cong} E_{n+1}^{n+1}$$

as defined in Subsection 1, induces  $B_{2l}^n = E_{2l}^{n+1}$  and  $U(B_{2l}^n) \cong U'$  for  $U'$  a subspace of  $E_{n+1}^{n+1}$  missing  $E_{2l}^{n+1}$ ,

$$r(U') = r[U(B_{2l}^n)]$$

Thus  $U'$  has minimum weight  $\geq 2l + 2$ , and  $k(B_{2l+1}^{n+1}) \geq k(B_{2l}^n)$ .

Since  $B_{2l+1}^{n+1} \neq \phi$ ,

$$k = r[U(B_{2l+1}^{n+1})] < n + 1$$

hence  $U(B_{2l+1}^{n+1})$  can be expressed as the row space of a  $k$  by  $(n+1)$  matrix  $M$  which has dependent columns. Hence we can delete one column to get a matrix  $M'$  with  $\text{rank } M' = \text{rank } M$ . Let  $U' =$  the row space of  $M'$ ,  $U' \subseteq V_2^n$ . Then  $U'$  has minimum weight  $\geq 2l + 1$ , since  $U(B_{2l+1}^{n+1})$  has minimum weight  $\geq 2l + 2$ , and

$$r(U') = r[U(B_{2l+1}^{n+1})]$$

Thus  $k(B_{2l}^n) \geq k(B_{2l+1}^{n+1})$ , establishing equality. QED.

**LEMMA 4.**  $c(B_w^n)$  and  $k(B_w^n)$  are monotone increasing in  $n$ , and

$$c(B_w^{n+1}) - c(B_w^n) = 1 - (k(B_w^{n+1}) - k(B_w^n)) = 0 \text{ or } 1$$

*Proof.* Since

$$c(B_w^n) + k(B_w^n) = n$$

it suffices to show

$$k(B_w^n) \leq k(B_w^{n+1}) \leq k(B_w^n) + 1$$

By the isomorphism

$$V_2^n \cong H_{n+1} \subseteq V_2^{n+1}$$

where

$$H_{n+1} = \{x \in V_2^{n+1} \mid x^{(n+1)} = 0\}$$

$k(B_w^{n+1}) \geq k(B_w^n)$  is trivial.

It suffices to show that

$$\text{rank}(U(B_w^{n+1}) \cap H_{n+1}) \geq k(B_w^{n+1}) - 1$$

since

$$k(B_w^n) \geq \text{rank}(U(B_w^{n+1}) \cap H_{n+1})$$

Let  $U(B_w^{n+1})$  have the basis

$$B = \{b_1, b_2, \dots, b_k\}, \quad k = k(B_w^{n+1})$$

Now we assume that some

$$b_i^{(n+1)} \neq 0, \quad 1 \leq i \leq k$$

for otherwise  $U(B_w^{n+1}) \subseteq H_{n+1}$  and we are done. Suppose

$$B' = \{b'_j \in U(B_w^{n+1}) \cap H_{n+1} \mid b'_j = b_j + b_j^{(n+1)} b_i, \\ j \neq i, \quad 1 \leq j \leq k\}$$

is dependent. Then for some

$$J \subseteq \{1, 2, \dots, i-1, i+1, \dots, k\}$$

we have

$$\sum_{j \in J} b'_j = 0$$

Then

$$\sum_{j \in J} b_j + \delta b_i = 0$$

for some  $\delta \in GF(2)$ , contradicting the independence of  $B$ . Thus  $B'$  is independent, and

$$\text{rank}(U(B_{w_0}^{n+1}) \cap H_{n+1}) \geq |B'| = k - 1. \quad \text{QED.}$$

If  $L$  is any lattice, and  $x, y \in L$ ,  $x \leq y$ , then we define the *interval lattice*

$$[x, y]_L = \{z \in L \mid x \leq z \leq y\}$$

**LEMMA 5.** If  $x \in L(S)$ ,  $S \subseteq V_q^n$ ,  $r(x) = k$ , then  $[x, 1]_{L(S)}$  is also a geometric lattice. Indeed,  $[x, 1]_{L(S)} \cong L(R)$  for some  $R \subseteq V_q^{n-k}$ .

**Proof.** Let  $x$  have a basis  $\{b_1, b_2, \dots, b_k\}$ , and let  $x_i = \{b_{k-j+1}, b_{k-j+2}, \dots, b_k\}$ , for  $j = 1, 2, \dots, k$ . Then

$$[x, 1]_{L(S)} = [x_k, 1] \\ ([x_{k-1}, 1] \\ \vdots \\ ([x_2, 1] \\ ([x_1, 1]_{L(S)})) \vdots \vdots)$$

because  $0, x_1, x_2, \dots, x_k = x$  is simply a totally ordered chain between 0 and  $x$  in  $L(S)$ , and subintervals of intervals of  $L$  are obviously intervals of  $L$ . Each interval in the expression above has height one less than its subscript-lattice, thus it suffices to prove the lemma for the case  $r(x) = 1$ . Then  $x = \{a\}$  for some  $a \in S$ ,  $a \neq 0$ .

For some  $i \leq n$ ,  $a^{(i)} \neq 0$ . Let  $\rho: V_q^n \rightarrow V_q^{n-1}$  by

$$\rho(x) = x - \frac{x^{(i)}}{a^{(i)}} a$$

where we identify  $V_q^{n-1}$  with

$$\{y \in V_q^n \mid y^{(i)} = 0\}$$

Let  $R = \rho(S) \subseteq V_q^{n-1}$ . Let  $Z \in [a, 1]_{L(S)}$  and let

$$y \in (\text{span } \rho(Z)) \cap R$$

Then

$$y = \sum_{x_j \in \rho(Z)} \alpha_j x_j = \rho(y')$$

for some  $y' \in S$ ,  $\alpha_j \in GF(q)$ . Thus

$$y' = \sum_j \alpha_j x'_j + \delta a$$

for

$$x'_j \in \rho^{-1}(x_j)$$

and  $\delta \in GF(q)$ . But  $a \in Z$ . Thus  $y' \in (\text{span } Z) \cap S$ , and since  $Z$  is closed,  $y' \in Z$ . Thus  $y \in \rho(Z)$ , proving that  $\rho(Z)$  is closed  $\subseteq R$ . Hence  $\rho$  induces a function also denoted  $\rho$ :  $\rho: [x, 1]_{L(S)} \rightarrow L(R)$ .

On the other hand, if

$$Y \in L(R), \rho^{-1}(Y) = (\text{span } (Y, a)) \cap S$$

hence  $\rho^{-1}(Y) \in [x, 1]_{L(S)}$  since  $\rho^{-1}(Y)$  is obviously closed. Thus  $\rho: [x, 1]_{L(S)} \cong L(R)$ . QED.

We now use a result of Crapo (Ref. 4, p. 606), specialized to geometric lattices of the type  $L(S)$ . We note that  $T \subseteq S \subseteq V_q^n$  implies  $L(T) \subseteq L(S)$  canonically.

**PROPOSITION 6.** If  $T \subseteq S \subseteq V_q^n$ , then

$$\lambda^{r(S)-r(T)} p(\lambda; T) = \sum_{x \in M(S, T)} p(\lambda; [x, 1]_{L(S)})$$

where

$$M(S, T) = \{x \in L(S) \mid y \in L(T), y \leq x \text{ in } L(S) \text{ implies } y = 0\}$$

We again consider exclusively characteristic 2. Let

$$F_w^n = E_w^n - D_2^n = \left\{ x \in V_2^n \mid \text{weight}(x) = 4, 6, 8, \dots, 2 \left\lceil \frac{w}{2} \right\rceil \right\}$$

**PROPOSITION 7.** For  $n \geq 5$ ,  $h \geq 2$ ,

$$p(\lambda; F_{2h}^n) = p(\lambda; B_{2h}^{n-1}) + \binom{n}{2} p(\lambda; B_{2h}^{n-2}) + \binom{n}{3} p(\lambda; B_{2h}^{n-3})$$

**Proof.** In Proposition 6, let  $T = F_{2h}^n \subseteq S = E_{2h}^n$ . No  $x \in M(S, T)$  may contain  $a_{ij}$  and  $a_{kl}$  for any distinct  $i, j, k, l$ , for then  $a_{ijkl} = a_{ij} + a_{kl} \in x$ . Therefore,  $M(S, T) = \{0\}$ ; the  $\binom{n}{2}$  points  $a_{ij} \in S - T = D_2^n$ , and the  $\binom{n}{3}$  lines  $\{a_{ij}, a_{jk}, a_{ik}\}$  for distinct  $i, j, k$ .

By Lemma 5,  $[a_{ij}, 1] \cong L(R)$ , where

$$R = \{x + x^{(i)} a_{ij} \mid x \in S\} = \rho(S)$$

Under the usual identification,

$$V_2^{n-1} = \{x \in V_2^n \mid x^{(i)} = 0\}$$



$\rho$  is the identity on  $V_2^{n-1}$ , hence  $\rho(F_{2h}^{n-1}) = E_{2h}^{n-1}$ . On the remainder

$$\{x_\varepsilon E_{2h}^n \mid x^{(i)} = 1\}$$

weight  $(\rho(X)) \leq \text{weight}(x)$ . Since  $\rho$  maintains parity, we see  $R = E_{2h}^{n-1}$ . Thus

$$[a_{ij}, 1]_{L(S)} \cong L(E_{2h}^{n-1}) \cong L(B_{2h}^{n-2})$$

The line  $\{a_{ij}, a_{ik}, a_{jk}\} = \ell_{ijk}$  has a basis  $\{a_{ij}, a_{jk}\}$ . Thus by Lemma 5 and the preceding case,

$$[\ell_{ijk}, 1]_{L(S)} \cong L(E_{2h}^{n-2}) \cong L(B_{2h}^{n-3})$$

For  $n \geq 5$ ,  $h \geq 2$ ,  $r(F_{2h}^n) = r(E_{2h}^n)$ , so we have

$$p(\lambda; F_{2h}^n) = p(\lambda; B_{2h}^{n-1}) + \binom{n}{2} p(\lambda; B_{2h}^{n-2}) + \binom{n}{3} p(\lambda; B_{2h}^{n-3})$$

QED.

**Corollary 8.**  $c(F_{2h}^n) = c(B_{2h}^{n-3})$ , hence  $k(F_{2h}^n) = 3 + k(B_{2h}^{n-3})$ .

**Proof.** Let  $\sigma: V_2^{n-3} \subseteq V_2^n$  by  $[\sigma(x)]^{(i)} = 0$  if  $i \geq n-2$ . Then, in  $V_2^n$ ,

$$U' = \text{span}(\sigma(U(B_{2h}^{n-3})), a_{n-2}, a_{n-1}, a_n)$$

contains only vectors of weights  $\geq 2h+1$  or  $\leq 3$  in  $V_2^n$ .

Thus

$$U' \cap F_{2h}^n = \phi, k(F_{2h}^n) \geq 3 + k(B_{2h}^{n-3})$$

and

$$c(F_{2h}^n) \leq c(B_{2h}^{n-3})$$

On the other hand, if  $c = c(B_{2h}^{n-3})$ , let

$$Q_c(\lambda) = (\lambda-1)(\lambda-2)(\lambda-4) \cdots (\lambda-2^{c-1})$$

By Theorem 1

$$Q_c(\lambda) \mid p(\lambda; B_{2h}^{n-3})$$

hence by Lemma 3

$$Q_c(\lambda) \mid p(\lambda; B_{2h}^{n-2}) \text{ and } p(\lambda; B_{2h}^{n-1})$$

Hence

$$Q_c(\lambda) \mid p(\lambda; F_{2h}^n)$$

and

$$c(F_{2h}^n) \geq c(B_{2h}^{n-3}) \quad \text{QED.}$$

It is difficult to use Proposition 6 to excise vectors of weight greater than two, since vectors of weight greater than those in  $S$  will usually be introduced into the geometries  $[x, 1]_{L(S)}$ .

Perhaps the most interesting case of the preceding proposition and corollary is  $2h = 4$ , since  $F_4^n = S_4^n$ . This set of vectors is transitive under permutation of the coordinates, and with such symmetry one might expect a solution similar to that of  $D_2^n \cong B_2^{n-1}$  in Subsection 3. Such a solution would immediately give the optimum rate of a 2-error correcting code. The following observation, which generalizes the equivalence between the critical problem for  $D_2^n$  and the coloring problem for  $C_n$ , may be a step in that direction.

Let  $L$  be a linear functional

$$\sum_{i=1}^n \alpha_i x^{(i)}, \quad \alpha_i \in GF(2)$$

The Boolean function corresponding to  $L$  is the function

$$\alpha_L = (\alpha_1, \alpha_2, \dots, \alpha_n): N = \{1, 2, \dots, n\} \rightarrow GF(2)$$

where  $\alpha_L(i) = \alpha_i$ . The linear functionals on  $V_2^n$  are in one-to-one correspondence with the Boolean functions  $\alpha: N \rightarrow GF(2)$ . Let

$$Y_L = \{i \in N: \alpha_i = 1\}$$

and for  $x \in S$ , let

$$E_x = \{i \in N: x^{(i)} = 1\}$$

Then

$2^{d(n-r(S))} p(2^d, S)$  = the number of ordered sequences of  $d$  linear functionals  $L_1, \dots, L_d$  such that

$$\left( \bigcap_{i=1}^d \ker L_i \right) \cap S = \phi$$

= the number of ordered sequences of  $d$  subsets (including  $\phi$ )  $Y_1, Y_2, \dots, Y_d$  of  $N$  such that for each  $x \in S$

$$|Y_j \cap E_x| \text{ is odd for some } j, 1 \leq j \leq d$$

We consider the binary relation  $R \subseteq A_1 \times A_2$  where  $A_1 = 2^N = \{Y \subseteq N\}$  and  $A_2 = \{E_x \subseteq N : x \in S\}$ , and  $YRE_x$ , for  $Y \in A_1$ ,  $E_x \in A_2$ , if and only if  $|Y \cap E_x|$  is odd. Then  $c(S) = \text{minimum } \{|A'| : A' \subseteq A_1, \text{ and every } E_x \in A_2 \text{ is related to some } Y \in A'\}$ . For  $S = D_3^n$ ,  $A_2 = \{X \subseteq N : |X| = 4\}$ , and the binary relation should be easier to handle than the corresponding relation for  $S = B_3^n$ .

### 5. Recursion for $p(\lambda; B_3^n)$

We know  $k(B_3^n)$  from Subsection 3 by applying Proposition 3, but that does not give us the characteristic polynomial. Using Proposition 6, we obtain a recursion for  $p(\lambda; B_3^n)$  involving  $p(\lambda; D_3^n)$ .

Let  $T = B_3^n - D_2^n \subseteq B_3^n = S$  in Proposition 6. The rank  $k$  elements of  $M(S, T)$  are the closures of independent sets of  $k$  weight-two vectors, which are partitions of an  $n$ -element set into  $n - k$  parts, as in Subsection 3.  $[a_{ij}, 1]_{L(B_3^n)} \cong L(B_3^{n-1})$  as in Proposition 7, since  $\rho(B_3^{n-1}) = B_3^{n-1}$  and  $\rho$  does not increase weights. Thus if  $r(x) = k$ ,  $x \in M(S, T)$ ,  $[x, 1]_{L(B_3^n)} \cong L(B_3^{n-k})$ . Thus

$$p(\lambda; T) = \sum_{k=0}^{n-1} S(n, n-k) p(\lambda; B_3^{n-k})$$

where

$S(n, m)$  = the Stirling number of the second kind  
= the number of partitions of an  $n$ -element set into  $m$  parts

We re-apply Proposition 6 to the case  $T = D_3^n$ ,

$$S = B_3^n - D_2^n = D_3^n \cup D_1^n$$

$M(S, T) = \{\text{the } 0\text{-, } 1\text{- and } 2\text{-element subsets of } D_1^n\}$ .  $[a_i, 1]_{L(S)} \cong L(B_3^{n-1})$  because in the notation of Lemma 5,

$$\rho(D_3^{n-1} \cup D_1^{n-1}) = D_3^{n-1} \cup D_1^{n-1}$$

since  $\rho$  fixes  $x$  if  $x^{(i)} = 0$ . But for  $a_{ijk} \in D_3^n \subset S$ ,  $\rho(a_{ijk}) = a_{jk}$ .

Therefore  $\rho(S) = B_3^{n-1}$ . Similarly

$$[\{a_i, a_j\}, 1]_{L(S)} \cong L(B_3^{n-2})$$

Thus we have

$$\begin{aligned} p(\lambda; D_3^n) &= p(\lambda; B_3^n - D_2^n) + np(\lambda; B_3^{n-1}) + \binom{n}{2} p(\lambda; B_3^{n-2}) \\ &= np(\lambda; B_3^{n-1}) + \binom{n}{2} p(\lambda; B_3^{n-2}) + \sum_{k=1}^n S(n, k) p(\lambda; B_3^k) \\ &= np(\lambda; B_3^{n-1}) + \binom{n}{2} p(\lambda; B_3^{n-2}) \\ &\quad + \text{the } n\text{th coefficient of} \end{aligned}$$

$$\sum_{k=1}^n \frac{p(\lambda; B_3^k) x^k}{(1-x)(1-2x) \cdots (1-kx)}$$

The term for  $k = n$  is just  $p(\lambda; B_3^n)$ , which may be transposed with  $p(\lambda; D_3^n)$  to give

$$\begin{aligned} p(\lambda; B_3^n) &= p(\lambda; D_3^n) - np(\lambda; B_3^{n-1}) - \binom{n}{2} p(\lambda; B_3^{n-2}) \\ &\quad - \text{the } n\text{th coefficient of} \end{aligned}$$

$$\sum_{k=1}^{n-1} \frac{p(\lambda; B_3^k) x^k}{(1-x)(1-2x) \cdots (1-kx)}$$

Table 1. Values of  $\tilde{p}$

$n$	$c(B_3^n)$	$k(B_3^n)$	$\tilde{p}(\lambda; B_3^n)$	$c(D_3^n)$	$k(D_3^n)$	$\tilde{p}(\lambda; D_3^n)$
4	4	0	1	1	3	$\lambda^2$
5	4	1	$(\lambda - 15)$	2	3	$(\lambda - 1 \pm i)$
6	5	1	$(\lambda - 25)$	3	3	$(\lambda - 4 \pm 2i)$
7	6	1	$(\lambda - 35)$	4	3	$(\lambda - 10 \pm 4.5i)$
8	6	2	$(\lambda - 49.5 \pm 8.3i)$	4	4	$(\lambda - 15.5)(\lambda - 19.75 \pm 9.9i)$
9	7	2	$(\lambda - 64 \pm 16.7i)$	5	4	$(\lambda - 30.3)(\lambda - 32.3 \pm 18.8i)$
10	7	3	$(\lambda - 95)(\lambda - 81.5 \pm 31.4i)$	6	4	$(\lambda - 51.6)(\lambda - 47.7 \pm 32.2i)$
11	7	4	$(\lambda - 83.3 \pm 44.6i)(\lambda - 133.7 \pm 44.98i)$	6	5	$(\lambda - 68.0 \pm 1.6i)(\lambda - 65.5 \pm 50.0i)$
12	8	4		7	5	$(\lambda - 101.75 \pm 34.1i)(\lambda - 82.25 \pm 67.7i)$
$n$	$c(B_3^n)$	$k(B_3^n)$	$\tilde{p}(\lambda; B_3^n)$	$c(D_3^n)$	$k(D_3^n)$	$\tilde{p}(\lambda; D_3^n)$
3	3	0	1	1	2	1
4	3	1	$(\lambda - 7)$	1	3	$(\lambda - 1)^2$
5	4	1	$(\lambda - 10)$	1	4	$(\lambda - 2.08 \pm .4i)(\lambda - 2.4 \pm 2.35i)$
6	4	2	$(\lambda - 13 \pm 2.45i)$	1	5	$(\lambda - 1.99)(\lambda - 387 \pm 1.05i)(\lambda - 4.6 \pm 5.3i)$
7	4	3	$(\lambda - 14.8)(\lambda - 16.6 \pm 5.0i)$	1	6	$(\lambda - 2.0002)(\lambda - 3.95)(\lambda - 6.8 \pm 2.5i)(\lambda - 7.2 \pm 9.2i)$

## 6. Values for $\tilde{p}$

Table 1 gives a set of values for

$$\tilde{p}(\lambda; S) = \frac{p(\lambda; S)}{(\lambda-1)(\lambda-2)(\lambda-4) \cdots (\lambda-2^{g(S)-1})}$$

for  $S = B_4^n, D_4^n, B_3^n, D_3^n$ . We denote  $(\lambda - a + bi)(\lambda - a - bi)$  by  $(\lambda - a \pm bi)$ .

## References

1. Crapo, H. H., and Rota, G.-C., *Combinatorial Geometries*, preliminary edition, M.I.T. Press, Cambridge, Mass., 1970.
2. Rota, G.-C., "On the Foundations of Combinatorial Theory" I, *Z. Wahrschein. und Verw. Gebiete*, Vol. 2, pp. 340-368, 1964.
3. Ore, O., "Theory of Graphs," *AMS Colloquium Publications*, Vol. 38, 1962.
4. Crapo, H. H., "Mobius Inversion in Lattices," *Archiv der Math.*, Vol. 19, pp. 595-607, 1968.

## VI. Communications Elements Research

### TELECOMMUNICATIONS DIVISION

#### A. Spacecraft Antenna Research: Antenna Tolerances, D. Damlamayan

##### 1. Introduction

The actual performance of an antenna always falls short of ideal performance. The reason for this is that the actual aperture distribution deviates from the ideal one due to various errors which may be either systematic or random.

Systematic errors are those which, once the cause is known, may be calculated accurately. Errors due to the displacement of the feed from the focus of a parabolic reflector or a lens are of this type.

Random errors are due to unpredictable deviations of the antenna parameters from their design value. Departure of the surface of a parabolic reflector or of a lens from its true shape gives rise to random errors.

In this article, the effect of various errors on the gain of antennas representable by a field distribution over an aperture is discussed, with particular reference to parabolic reflectors and lenses. The related subject of pattern degradation has not been considered. The subject of antenna tolerances has been studied quite extensively in recent years. Some of the more interesting works are cited in Refs. 1-6.

The axial gain of a highly directional aperture antenna, radiating close to the  $\theta = 0$  direction, whose aperture distribution is  $F(x, y)$ , can be written as

$$G_0 = \frac{4\pi}{\lambda^2} \frac{\left| \int F(\underline{x}) dS \right|^2}{\int \left| F(\underline{x}) \right|^2 dS} \quad (1)$$

where the integration is carried over the aperture ( $xy$ ) plane. This result is most simply and directly obtained by considering the plane wave spectrum representation of the aperture field.

As our interest lies mainly in reflector and lens antennas, it will be sufficient to assume only a pure phase error. If such a phase error  $\delta(\underline{x})$  is introduced, the gain becomes

$$G = \frac{4\pi}{\lambda^2} \frac{\left| \int F(\underline{x}) e^{i\delta(\underline{x})} dS \right|^2}{\int \left| F(\underline{x}) \right|^2 dS} \quad (2)$$

$\delta(\underline{x})$  will be a deterministic function when systematic errors are considered and a random function when we treat random errors.

## 2. Random Errors

Let us first consider the gain loss resulting from random phase errors on the aperture plane caused by the deviations of the antenna surface from the ideal. Following Ruze (Refs. 1 and 5), we assume the phase error  $\delta(\underline{x})$  at any point on the aperture to be a random sample from a single gaussian distribution with zero mean and variance  $\bar{\delta}^2$ . It follows that the surface deviations are assumed to be distributed uniformly over the aperture.

The numerator of Eq. (2), which represents the power radiated axially, can be written as a double surface integral:

$$P = \iint F(\underline{x}_1) F^*(\underline{x}_2) \exp \{i[\delta(\underline{x}_1) - \delta(\underline{x}_2)]\} dS_1 dS_2 \quad (3)$$

Letting  $\underline{x}_1 - \underline{x}_2 = \underline{\tau}$  as the vector difference between two integration points on the aperture, and  $y(\underline{\tau}) = \delta(\underline{x}_1) - \delta(\underline{x}_2)$  as the phase difference of two points on the aperture spaced a distance  $\tau$  apart, we get

$$P = \iint F(\underline{x} + \underline{\tau}) F^*(\underline{x}) e^{iy(\underline{\tau})} dS_\tau dS_x \quad (4)$$

Defining now the aperture illumination auto-correlation function as

$$\phi(\underline{\tau}) = \frac{\int F(\underline{x} + \underline{\tau}) F^*(\underline{x}) dS_x}{\int |F(\underline{x})|^2 dS_x} \quad (5)$$

we obtain for the gain the simple expression

$$G = \frac{4\pi}{\lambda^2} \int \phi(\underline{\tau}) e^{iy(\underline{\tau})} dS_\tau \quad (6)$$

Since  $\delta(\underline{x})$  and hence  $y(\underline{\tau})$  are random functions, the quantity of interest is the expected value of the gain  $\bar{G}$ . The correlation distance of the phase errors,  $c$ , is defined as that distance at which the phase errors become essentially independent. Then for  $\tau \gg c$ ,  $y(\underline{\tau})$  becomes the difference of two independent normal random variables, hence it is also normally distributed with variance

$$\overline{y^2(\underline{\tau})} \xrightarrow{\tau \gg c} 2\bar{\delta}^2 \quad (7)$$

On the other hand, for  $\tau = 0$  the mean square difference in phase error is obviously zero:

$$\overline{y^2(0)} = 0 \quad (8)$$

We now assume a convenient form for the variance function  $\overline{y^2(\tau)}$ , which fits these end conditions, such as

$$\overline{y^2(\tau)} = 2\bar{\delta}^2 (1 - e^{-\tau^2/c^2}) \quad (9)$$

Another assumed form for the variance function with a similar smooth behavior does not significantly alter the result. Now it can be shown that

$$\overline{e^{iy(\tau)}} = \exp[-\bar{\delta}^2 (1 - e^{-\tau^2/c^2})] \quad (10)$$

Hence

$$\begin{aligned} \bar{G} &= \frac{4\pi}{\lambda^2} e^{-\bar{\delta}^2} \int \phi(\tau) \exp(\bar{\delta}^2 e^{-\tau^2/c^2}) dS_\tau \\ &= G_0 e^{-\bar{\delta}^2} + \frac{4\pi}{\lambda^2} e^{-\bar{\delta}^2} \sum_{n=1}^{\infty} \int \phi(\tau) \frac{(\bar{\delta}^2)^n}{n!} e^{-n\tau^2/c^2} dS_\tau \end{aligned} \quad (11)$$

where  $G_0$  is the gain for no phase errors. Because of the exponential factor the integral essentially vanishes for  $\tau > c$ . Now  $\phi(0) = 1$ , and assuming that  $c$  is small compared to the aperture dimensions we may set  $\phi(\tau) = 1$  in evaluating the integral. Extending the integration limit to infinity, we finally obtain

$$\bar{G} = G_0 e^{-\bar{\delta}^2} + \left(\frac{2\pi c}{\lambda}\right)^2 e^{-\bar{\delta}^2} \sum_{n=1}^{\infty} \frac{(\bar{\delta}^2)^n}{n \cdot n!} \quad (12)$$

The reduction in axial gain is given by

$$\frac{G}{G_0} = e^{-\bar{\delta}^2} + \frac{1}{\eta} \left(\frac{2c}{D}\right)^2 e^{-\bar{\delta}^2} \sum_{n=1}^{\infty} \frac{(\bar{\delta}^2)^n}{n \cdot n!} \quad (13)$$

where we have considered a circular aperture of diameter  $D$  and  $\eta$  is the aperture efficiency defined by

$$G_0 = \eta \left(\frac{\pi D}{\lambda}\right)^2$$

For small tolerance losses and for correlation regions  $c$  small compared to the aperture diameter  $D$ , we may write

$$G = G_0 e^{-\bar{\delta}^2} = \eta \left( \frac{\pi D}{\lambda} \right)^2 e^{-(2\pi/\lambda)^2 \bar{\xi}^2} \quad (14)$$

where  $\bar{\xi}^2$  is the variance of optical path error. The maximum gain is achieved at the wavelength

$$\lambda_m = 2\pi \sqrt{\bar{\xi}^2} \quad (15)$$

Finally, the rms surface deviations  $\epsilon$  must be related to the variance of phase or optical path errors for various antennas. For shallow reflectors (large  $f/D$ )

$$\sqrt{\bar{\xi}^2} = 2\epsilon \quad (16)$$

while for lenses of large  $f/D$

$$\sqrt{\bar{\xi}^2} = \epsilon |n - 1| \quad (17)$$

where  $n$  is the refractive index of the lens. Practical considerations such as low surface reflections require that  $|n - 1| \sim 0.5$ . Then for the same small surface deviations, the fractional loss in gain in parabolic reflectors is about 4 times as great as in lenses. For the same surface deviation, the maximum gain for a lens will be achieved at a wavelength about 4 times smaller than for parabolic reflectors.

It is possible to take into account the curvature of the reflector or of the lens surface. Let the rms axial and normal surface deviations be denoted respectively as  $\Delta z$  and  $\Delta h$  for both reflectors and lenses. Then (Fig. 1)

$$PQ = \Delta z \quad \text{and} \quad PN = \Delta h \quad (18)$$

with

$$\frac{\Delta h}{\Delta z} = \cos \theta \quad (19)$$

In the case of a reflector, the rms optical path error is (Fig. 1a)

$$\begin{aligned} \sqrt{\bar{\xi}^2} &= PP' + PR = \Delta z + \Delta z \cos 2\theta \\ &= 2\Delta z \cos^2 \theta \end{aligned} \quad (20)$$

Also, for a parabolic reflector we have ( $\alpha = 2\theta$ ),

$$x = FP \sin 2\theta = \frac{2f}{1 + \cos 2\theta} \sin 2\theta = 2f \tan \theta \quad (21)$$

Hence

$$\cos \theta = \frac{1}{\sqrt{1 + \left(\frac{x}{2f}\right)^2}} \quad (22)$$

and

$$\begin{aligned} \sqrt{\bar{\xi}^2} &= \frac{2\Delta z}{1 + \left(\frac{x}{2f}\right)^2} \\ &= \frac{2\Delta h}{\sqrt{1 + \left(\frac{x}{2f}\right)^2}} \end{aligned} \quad (23)$$

The curvature must then have the effect of reducing the gain loss.

Considering the lens, the rms optical path error can be seen to be (Fig. 1b)

$$\begin{aligned} \sqrt{\bar{\xi}^2} &= |PP' - nP'R| \\ &= PP' |1 - n \cos \alpha| \\ &= \frac{\Delta h}{\cos(\alpha + \theta)} |1 - n \cos \alpha| \end{aligned} \quad (24)$$

On the other hand, for a lens,

$$x = FP \sin \alpha = \frac{(n-1)f}{n \cos \alpha - 1} \sin \alpha \quad (25)$$

The last two relations together with Snell's law,  $\sin(\alpha + \theta) = n \sin \theta$ , enable one to obtain an expression for  $\sqrt{\bar{\xi}^2}$  involving only  $x/f$ ,  $n$  and either  $\Delta h$  or  $\Delta z$ .

It is easy to see that these expressions for the rms optical path errors in reflectors and lenses reduce to the ones given above for large  $f/D$ . In fact, for large  $f/D$ ,  $\alpha \approx 0$ ,  $\theta \approx 0$  and we immediately obtain

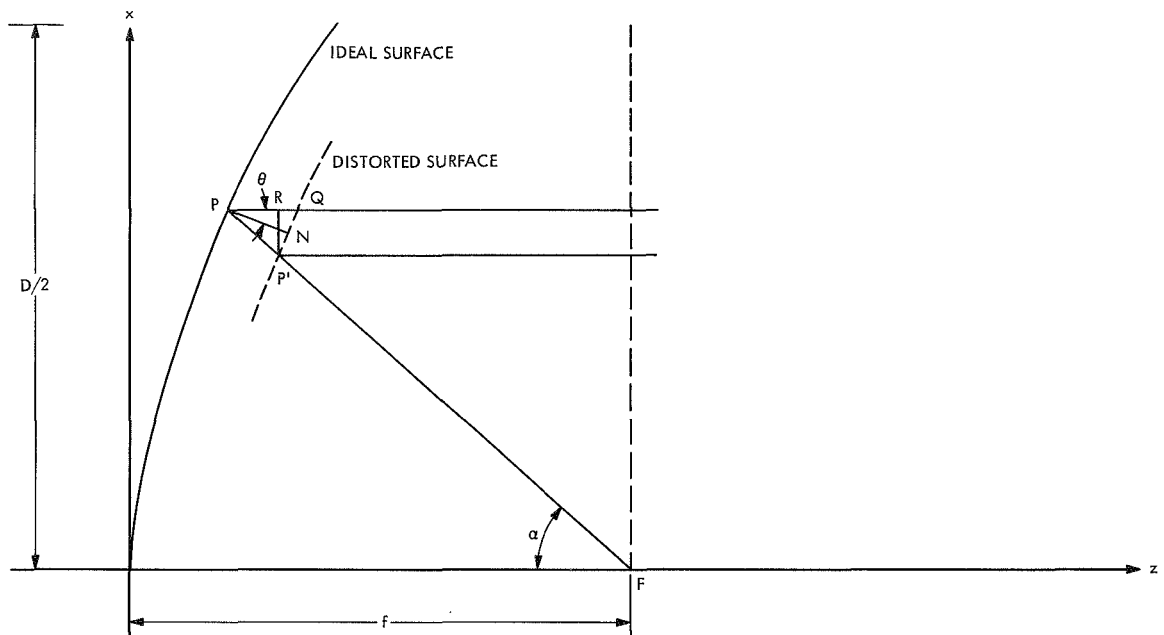
$$\sqrt{\bar{\xi}^2} = 2\Delta z = 2\Delta h \quad (26)$$

for reflectors, and

$$\sqrt{\bar{\xi}^2} = \Delta h |1 - n| = \Delta z |1 - n| \quad (27)$$

for lenses.

(a) PARABOLIC REFLECTOR



(b) LENS

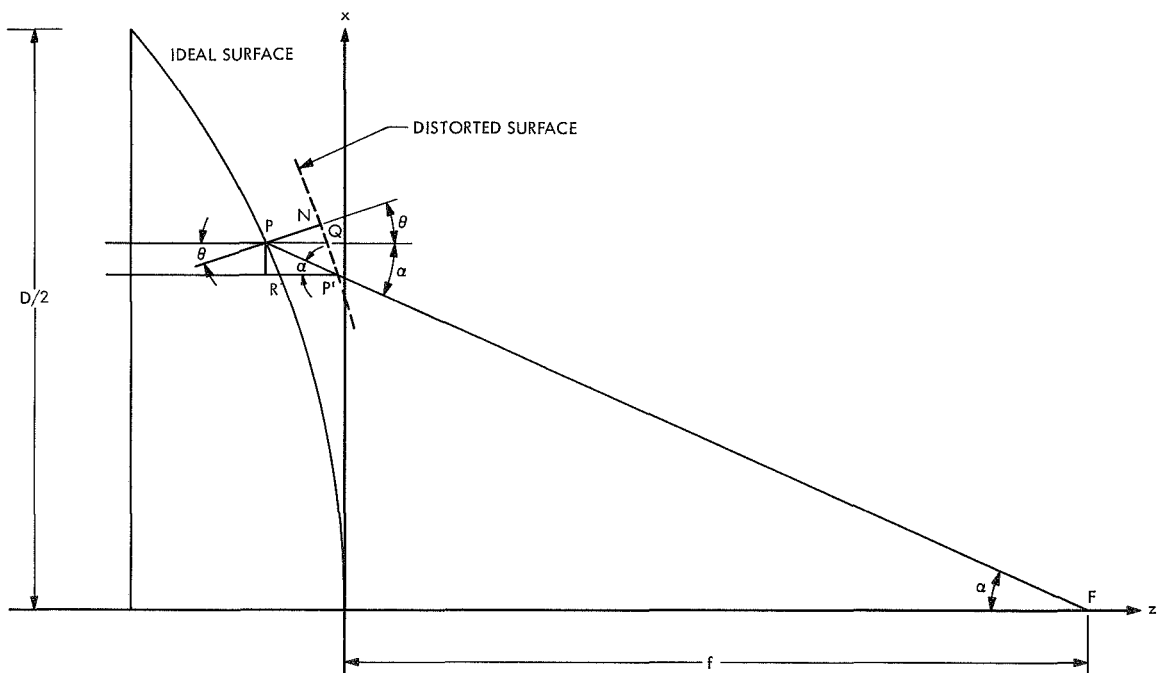


Fig. 1. Optical path length errors

In this connection, an extension of the above theory to surface deviations which are not distributed uniformly over the aperture must also be mentioned (Ref. 6). It is found that the assumption of uniform error distribution in general gives a lower value for the axial gain of an antenna whose surface deviations have regional variations over the aperture.

### 3. Systematic Errors

We next consider systematic errors due to defocussing of the source. The ratio of the gain to no error gain can be written as

$$\frac{G}{G_0} = \frac{\left| \int F(\underline{x}) e^{i\delta(\underline{x})} dS \right|^2}{\left| \int F(\underline{x}) dS \right|^2} \quad (28)$$

For convenience the phase error will be measured with respect to the illumination weighted mean phase error, which we call  $\delta_0(\underline{x})$ . Then

$$\delta_0(\underline{x}) = \delta(\underline{x}) - \frac{\int F(\underline{x}) \delta(\underline{x}) dS}{\int F(\underline{x}) dS} \quad (29)$$

Assuming in-phase illumination and small phase errors, expanding the exponential in a power series, we obtain

$$\frac{G}{G_0} = \frac{\left| \int F(\underline{x}) \exp[i\delta_0(\underline{x})] dS \right|^2}{\left| \int F(\underline{x}) dS \right|^2} = 1 - \bar{\delta}_0^2 \quad (30)$$

where

$$\bar{\delta}_0^2 = \frac{\int F(\underline{x}) \delta_0^2(\underline{x}) dS}{\int F(\underline{x}) dS} \quad (31)$$

This well-known simple result is very useful in calculating the gain loss in various circumstances. For small phase errors, Eq. (30) is similar to Eq. (13) or (14). There are a number of differences, however, which should not be ignored. First, Eq. (13) is not restricted to small phase errors, and an extension of the theory that led to Eq. (13) is able to give information on the radiation pattern as well. Then Eq. (30) depends on the illumination function

$F(\underline{x})$  whereas Eq. (13) does not. This is only so because in the statistical analysis the phase errors are assumed to be uniformly distributed. In this case  $\bar{\delta}_0^2$  as given by Eq. (31) would also be independent of the illumination function.

The loss in gain incurred as a result of the feed displacement axially by a distance  $a$  from the focus  $F$  to a point  $F'$  is considered next. The optical path difference between an axial ray and one reaching the aperture at the point  $x$  is the same for both lens and parabolic reflector antennas of focal length  $f$ . Referring to Fig. 2 and noting that  $n$  is the refracting index for the lens and equals unity for the parabolic reflector, we have for both cases

$$\begin{aligned} \xi &= F'N - (F'O + nOP) \\ &\approx (FN - FR) - (F'O - nOP) \\ &= (FO + nOP - FR) - (F'O + nOP) \\ &= a(1 - \cos \alpha) \end{aligned} \quad (32)$$

Now for shallow reflectors we write

$$\cos \alpha = \frac{1 - \left(\frac{x}{2f}\right)^2}{1 + \left(\frac{x}{2f}\right)^2} = 1 - \frac{1}{2} \left(\frac{x}{f}\right)^2 + \frac{1}{8} \left(\frac{x}{f}\right)^4 + \dots \quad (33)$$

with a similar expression for lenses. Hence, in either case, lumping the constants together, we write

$$\delta(x) = \frac{2\pi}{\lambda} \xi(x) = c_2 x^2 + c_4 x^4 \quad (34)$$

Considering only the quadratic term and a uniform illumination, we obtain

$$\delta_0(x) = c_2 x^2 - \frac{\Delta}{3} \quad (35)$$

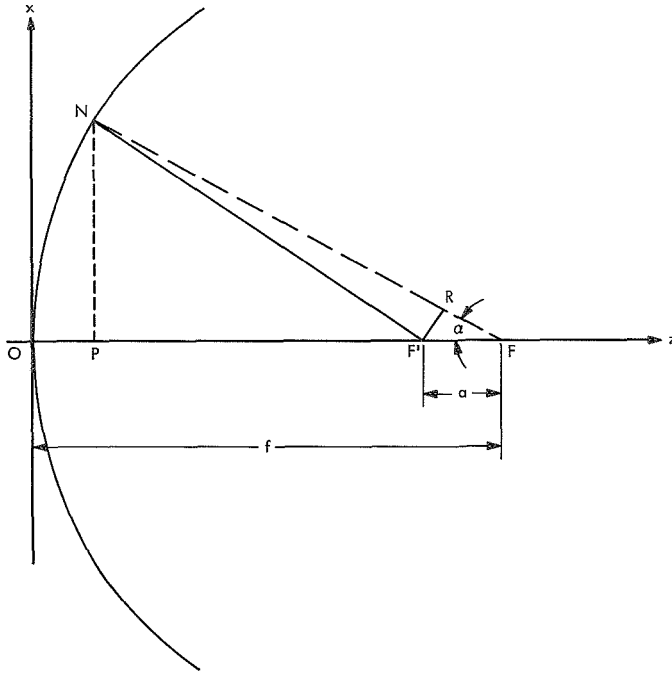
where  $\Delta$  is the phase error at the edge of the aperture. Then

$$\frac{G}{G_0} = 1 - \frac{1}{A} \int_A \left( c_2 x^2 - \frac{\Delta}{3} \right)^2 dA = 1 - \frac{4}{45} \Delta^2 \quad (36)$$

Thus in the case of uniform illumination an extreme phase error of about 1.1 radians can be tolerated if the loss in



(a) PARABOLIC REFLECTOR



(b) LENS

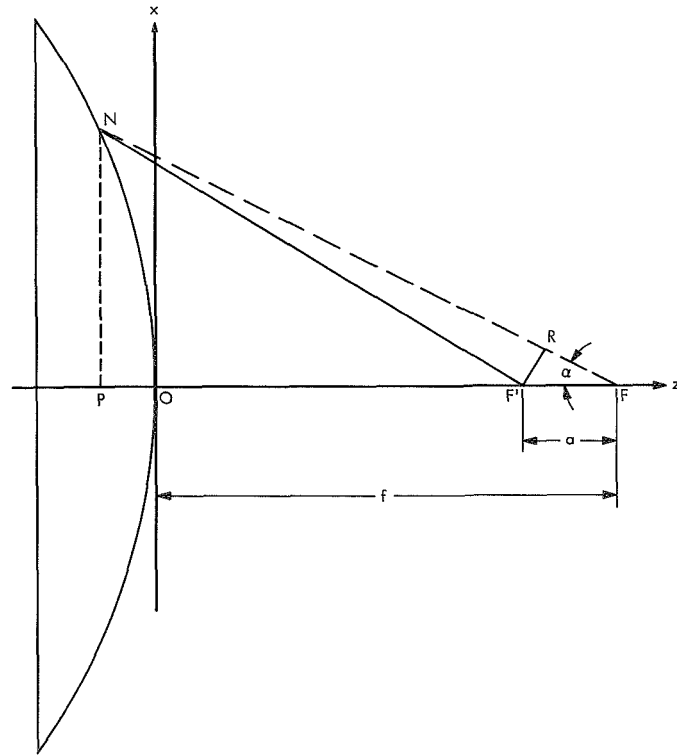


Fig. 2. Axial displacement of the feed from the focus

directivity is not to exceed 10%. This calculation for shallow reflectors is not a realistic one for communication antennas which may have small values of  $f/D$  ratios and is intended merely as an illustration.

The loss of gain resulting from the movement of the feed point away from the focus in a direction perpendicular to the axis can be calculated in a similar way. In this case, for small phase errors across the aperture, we can write

$$\delta(x) = c_1 x + c_3 x^3 \quad (37)$$

The constants may be calculated from the geometry for lenses and parabolic reflectors. The linear term causes the radiation pattern to tilt by an angle  $c_1 \lambda / 2\pi$  without change of shape and hence has no bearing on the loss of gain. Only the cubic term contributes to the reduction in gain which, however, is small since the cubic errors are strong near the edge of the aperture where usually the illumination is reduced.

It must be mentioned in this connection that the two surfaces of a lens can be chosen so as to minimize the loss of gain resulting from the feed movement, an option which does not exist for parabolic reflectors, and which makes lenses particularly suitable for wide-angle scanning.

Finally, the effect of systematically misplaced plates in metal-plate lenses will be considered. If the theoretical distance between the plates is  $s$  and the actual distance is  $s + \Delta s$ , where  $|\Delta s| \ll s$ , the index of refraction for the theoretical situation and the change in the index are

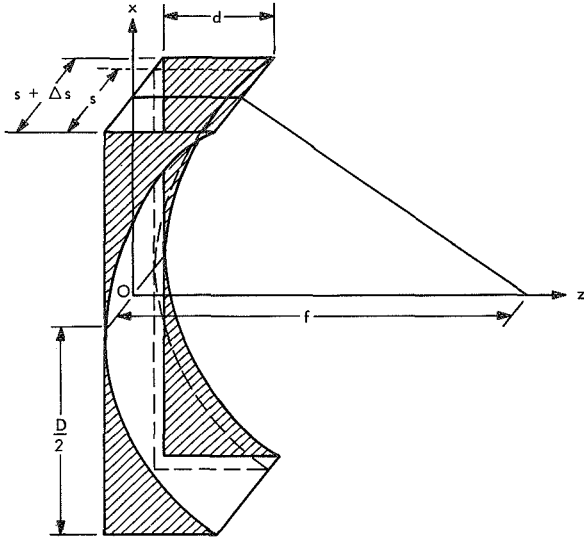
$$n = \sqrt{1 - \left(\frac{\lambda}{2s}\right)^2} \quad (38)$$

and

$$\Delta n = \frac{\partial n}{\partial s} \Delta s \quad (39)$$

Then the phase error becomes

$$\begin{aligned} \delta(x) &= \frac{2\pi}{\lambda} z(x) \Delta n \\ &= \frac{\pi\lambda}{n} z(x) \frac{\Delta s}{2s^3} \\ &= \frac{4\pi}{\lambda} \frac{z(x)}{\lambda} \frac{(1 - n^2)^{3/2}}{n} \Delta s \end{aligned} \quad (40)$$



**Fig. 3. Incorrect plate location in metal-plate lenses**

The equation of the surface  $z(x)$  can be found, referring to Fig. 3, from the relation

$$f = nz + \sqrt{x^2 + (f - z)^2}$$

which, upon squaring, becomes

$$(1 - n^2)z^2 - 2f(1 - n)z + x^2 = 0 \quad (41)$$

These results also apply for random spacings, where the spacings  $s$  form a random sample from a single gaussian distribution with zero mean and rms deviation  $\Delta s$ . The gain loss may now be calculated from the expression for  $\delta(x)$ . We see that increasing the index of refraction or decreasing the thickness of the lens lowers the phase error. Thus reducing the thickness of a lens by stepping it helps to ease this tolerance, although, in general, stepping a lens causes a loss in gain because of radiation scattered in "shadow" areas.

As an example consider a metal-plate lens of large  $f/D$  ratio. In this case,

$$z(x) \approx \frac{f}{2(1-n)} \left(\frac{x}{f}\right)^2 \quad (42)$$

and a quadratic phase error is introduced. For a uniform aperture distribution and a quadratic phase error  $\bar{\delta}_0^2$  has already been calculated. If the lens is stepped such that its thickness never exceeds  $\lambda/(1-n)$ , we obtain

$$\bar{\delta}_0^2 \leq \left(\frac{4\pi}{\lambda}\right)^2 \frac{1}{2} \Delta s^2 \quad (43)$$

for  $0.4 \leq n \leq 0.6$ .

#### 4. Conclusion

It has been shown quantitatively that the manufacturing tolerances for lenses can be less stringent than for reflectors. This is still so even if the possibility of errors arising from plate-spacing tolerances are considered. Easier tolerances and capability for better scanning are, in fact, the main advantages of lenses over reflectors.

#### References

1. Ruze, J., "The Effect of Aperture Errors on the Antenna Radiation Pattern," *Suppl. al Nuovo Cimento*, Vol. 9, pp. 364-380, 1952.
2. Robieux, J., "Influence de la précision de fabrication d'une antenne sur ses performances," *Ann. Radioélectr.*, Vol. 11, pp. 29-56, 1956.
3. Bracewell, R. N., "Tolerance Theory of Large Antennas," *IRE Trans. on Antennas and Propagation*, Vol. AP-9, pp. 49-58, 1961.
4. Bracewell, R. N., "Radioastronomy Techniques," in *Handbuch der Physik*, Vol. 54, pp. 76-82, Springer-Verlag, Berlin, 1962.
5. Ruze, J., "Antenna Tolerance Theory—A Review," *Proc. IEEE*, Vol. 54, pp. 633-640, 1966.
6. Zarghame, M. S., "On Antenna Tolerance Theory," *IEEE Trans. on Antennas and Propagation*, Vol. AP-15, pp. 777-781, 1967.

## VII. Spacecraft Telecommunications Systems

### TELECOMMUNICATIONS DIVISION

#### A. On the Stability of Second-Order Tracking Loops With Arbitrary Time Delay, *M. K. Simon*

##### 1. Introduction

For many years researchers have been answering questions related to the design and performance of tracking loops employing the phase-lock principle. The two primary directions of study have been toward (1) an understanding of the loop's transient behavior in terms of its acquisition time and cycle slipping performances, and (2) the determination of the steady-state loop performance as expressed in terms of the moments of the probability density function of the phase error. Another area of investigation is that of loop stability which relative to the above studies has received less attention. The main reason for this is that the techniques which are available for performing a stability analysis (i.e., the Nyquist criterion, root-locus diagram, Routh-Hurwitz criterion, phase-plane, Lyapunov function, etc.) have thus far shown applicability only to a noise-free condition. Furthermore, for standard second-order phase-locked loops, the stability problem is not critical since a linear, noise-free analysis predicts unconditional stability, whereas the nonlinear techniques of above demonstrate the existence of an adequate pull-in range.

With the introduction of the data-aided loop (SPS 37-60, Vol. III, pp. 46-56), the problem of loop stability

becomes more poignant due to the presence of a delay element in the open-loop transfer function. This delay, which is ideally equal to the reciprocal of the data rate, will at sufficiently low data rates cause instability problems. With this motivation in mind, we address ourselves in this article to the general question of second-order tracking loop stability in the presence of arbitrary time delay. A specific application of the results will yield an approximate answer to the question of how low in data rate one can go before the data-aided loop becomes unstable.

The principal tool used in the linear stability analysis which follows is the Nyquist diagram (and the accompanying Nyquist criterion) as applied to the open-loop, noise-free, transfer function of a generalized second-order tracking loop. For the nonlinear stability performance, we resort to a method based upon an approximation to the phase-plane behavior.

##### 2. Linear Stability Analysis

Consider a generalized phase-locked tracking system whose noise-free differential equation of operation is given by

$$\phi = \theta - AK \frac{F(p)}{p} e^{-pT_d} g(\phi) \quad (1)$$

where  $\theta$  is the input phase to be tracked,  $\phi = \theta - \hat{\theta}$  is the loop phase error,  $\hat{\theta}$  is the tracker's estimate of  $\theta$ ,  $F(p)$  is the loop filter,  $g(\phi)$  is the system nonlinearity normalized to unit slope at the origin,  $AK$  is the open-loop gain, and  $T_d$  is an arbitrary delay, assumed ideal. An expression such as Eq. (1) is typical of a data-aided loop with

$$g(\phi) = \sin \phi \left[ \frac{1 - 2P_E(\phi)}{1 - 2P_E(0)} \right]$$

and  $T_d$  equal to the reciprocal of the data rate. The quantity  $P_E(\phi)$  is the data detector's error rate conditioned on the tracking loop's phase error. The open-loop transfer function  $T(p) = \hat{\theta}(p)/\theta(p)$  is from Eq. (1),

$$T(p) = AK \frac{F(p)}{p} e^{-pT_d} \quad (2)$$

where  $g(\phi)$  has been linearized to  $\phi$ . For a second-order tracking system, it is quite common to assume a loop filter of the form

$$F(p) = \frac{1 + \tau_2 p}{1 + \tau_1 p} \quad (3)$$

Then, substituting Eq. (3) into Eq. (2) and letting  $p = j\omega$ , we get

$$T(j\omega) = \frac{AK(1 + j\omega\tau_2)}{j\omega(1 + j\omega\tau_1)} e^{-j\omega T_d} \quad (4)$$

Introducing the normalization  $x = \omega\tau_1$  and letting  $r = AK\tau_2^2/\tau_1$ , Eq. (4) simplifies to

$$T(jx) = \frac{r \left( \frac{\tau_1}{\tau_2} \right)^2 \left( 1 + jx \frac{\tau_2}{\tau_1} \right) \exp \left( -jx \frac{T_d}{\tau_1} \right)}{jx(1 + jx)} \quad (5)$$

The quantity  $r$  is related to the loop damping,  $\xi$ , at zero delay by

$$r = 4\xi^2 \quad (6)$$

It is desired to study the magnitude and phase characteristics of  $T(jx)$  as defined in Eq. (5) and from these, using the Nyquist criterion, determine the set of normalized frequencies,  $\{x_n\}$ , at which the phase character-

istic,  $\arg T(jx)$ , equals  $(2n - 1)\pi$  radians,  $n = 0, 1, 2, \dots$ . Thus the set  $\{x_n\}$  is the solution to

$$\tan^{-1} \left[ x_n \left( \frac{\tau_2}{\tau_1} \right) \right] - \tan^{-1} x_n - \frac{\pi}{2} - x_n \frac{T_d}{\tau_1} = (2n - 1)\pi \quad (7)$$

In particular, the solution of Eq. (7) for  $n = 0$  (i.e., the first crossing of the horizontal axis by the Nyquist plot) must have a corresponding magnitude of  $T(jx)$  which is less than unity in order to insure stability. Thus, if the delay  $T_d$  is such that

$$-\tan^{-1} \left[ x_0 \left( \frac{\tau_2}{\tau_1} \right) \right] + \tan^{-1} x_0 + x_0 \frac{T_d}{\tau_1} = \frac{\pi}{2} \quad (8a)$$

and

$$|T(jx)|^2 = \frac{r^2 \left( \frac{\tau_1}{\tau_2} \right)^4 \left[ 1 + x_0^2 \left( \frac{\tau_2}{\tau_1} \right)^2 \right]}{x_0^2 \left[ 1 + x_0^2 \right]} < 1 \quad (8b)$$

Then the second-order tracking loop will be absolutely stable in the linear sense. Assuming  $r\tau_1 \gg \tau_2$ , then the solution of Eq. (8b) for  $x_0$  is

$$x_0 > \left( \frac{r\tau_1}{\tau_2} \right) \sqrt{\frac{1 + \sqrt{1 + \frac{4}{r^2}}}{2}} \quad (9)$$

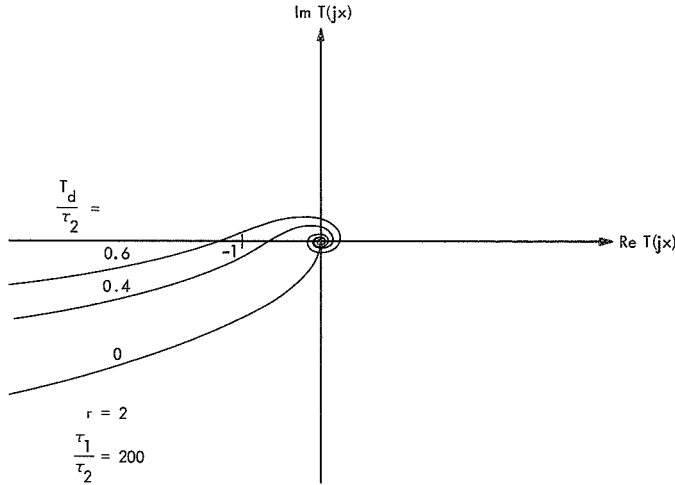
and hence

$$\tan^{-1} x_0 \approx \frac{\pi}{2} \quad (10)$$

Substituting expressions (9) and (10) in Eq. (8a) gives

$$\frac{T_d}{\tau_2} < \frac{\text{P.V.} \left\{ \tan^{-1} \left[ r \sqrt{\frac{1 + \sqrt{1 + \frac{4}{r^2}}}{2}} \right] \right\}}{r \sqrt{\frac{1 + \sqrt{1 + \frac{4}{r^2}}}{2}}} \quad (11)$$

where P.V. denotes principal value. Equation (11) then represents a bound on normalized loop delay in order to guarantee stability. As an example for  $r = 2$  (i.e.,  $\xi = 0.707$ ),  $T_d/\tau_2$  must be less than 0.52 (see Fig. 1).



**Fig. 1. Nyquist plot for second-order tracking loop with delay**

The quantitative results given above rely heavily on the ability to make the small phase error assumption which allowed us to use the linear model and a linear stability criterion. In fact, we notice that the linear stability analysis determines only whether the steady-state phase error becomes zero or unbounded and hence does not predict the existence of stable and unstable limit cycles. Thus, one should regard the above results as a qualitative guide (possibly even as a loose upper bound) to the real behavior which must be predicted by a non-linear model.

### 3. Nonlinear Stability Analysis

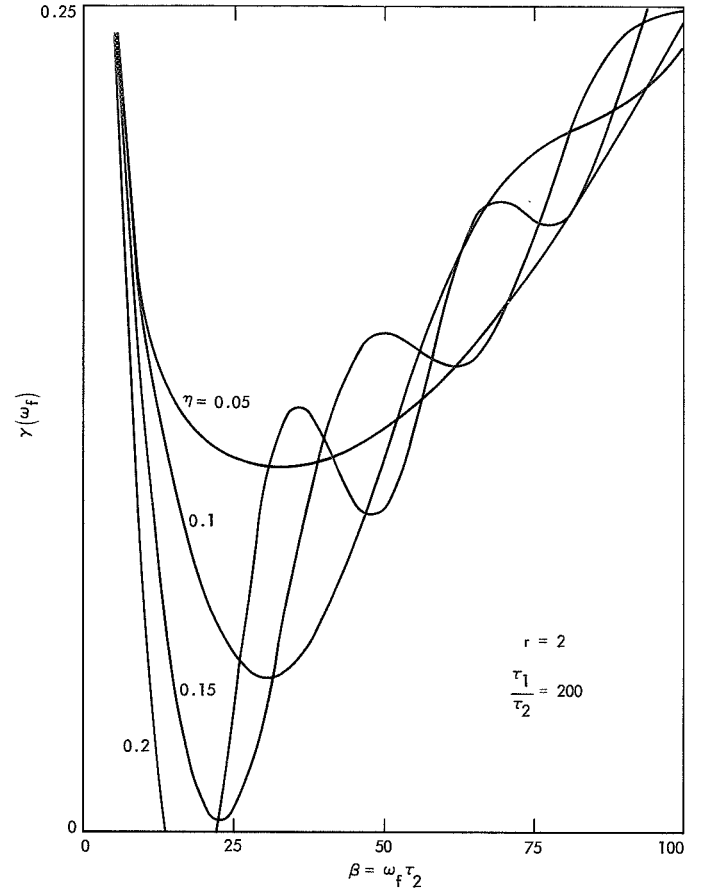
The approach taken here is based upon a recent result communicated to the author by W. C. Lindsey<sup>1</sup> for the synchronization boundary of a nonlinear tracking loop in the presence of arbitrary delay. In particular, the function

$$\gamma(\omega_f) \simeq \frac{\omega_f}{AK} + \frac{rT_d}{2\tau_2} \left[ \frac{\cos \omega_f T_d}{\omega_f T_d} + \frac{\sin \omega_f T_d}{\omega_f T_d} \right] \quad (12)$$

defines the synchronization boundary where  $\omega_f$  is the fundamental beat note in a Fourier series expansion of  $\phi(t)$ . Setting

$$\frac{\partial \gamma(\omega_f)}{\partial \omega_f} = 0 \quad (13)$$

<sup>1</sup>Electrical Engineering Department, University of Southern California.



**Fig. 2. Synchronization boundary of a second-order phase-locked loop with arbitrary delay**

gives in general several values of  $\omega_f$ , say  $\{\omega_{f_n}\}$  (see Fig. 2), which satisfy Eq. (13). Calling the smallest of these values  $\omega_{f_0}$ , and denoting the initial frequency offset by  $\Omega_0$ , then the following conclusions may be drawn:

- (1) If  $\Omega_0/AK < \gamma(\omega_{f_0})$ , the loop will lock.
- (2) If  $\Omega_0/AK > \gamma(\omega_{f_0})$ , then frequency pushing occurs and the loop goes into a stable limit cycle.

It is interesting to note that for the special case  $T_d = 0$ ,  $r\tau_1/\tau_2 \gg 1$ , the synchronization boundary defined above checks with a result given by Viterbi (Ref. 1, equation 3.39), for the pull-in range of a second-order phase-locked loop; i.e.,

$$\frac{\Omega_0}{AK} < \sqrt{\frac{2\tau_2}{\tau_1}} \quad (14)$$

In order to make the two results agree, Viterbi's loop gain  $AK$  must be multiplied by  $\tau_1/\tau_2$  to correspond to

our definition of loop gain since we define our filter,  $F(s)$ , such that it has unity gain at zero frequency.

The most stringent requirement placed on  $T_d$  is that value at which the loop will not lock even if initially the voltage-controlled oscillator is at its quiescent frequency and the offset in frequency is zero. In this situation, the loop will always become unstable. To arrive at this value of  $T_d$ , we must solve the equation

$$\gamma(\omega_{f_0}) = 0 \quad (15)$$

for  $T_d$ . Defining the normalized parameters

$$\beta = \omega_f \tau_2 \quad \eta = \frac{T_d}{\tau_2} \quad (16)$$

then Eq. (13) may be rewritten as

$$\gamma(\beta) = \frac{\beta}{r \frac{\tau_1}{\tau_2}} + \frac{r}{2} \left[ \frac{\cos \beta \eta}{\beta} + \frac{\sin \beta \eta}{\beta} \right] \quad (17)$$

Differentiating  $\gamma(\beta)$  with respect to  $\beta$ , and equating to zero, we get

$$\frac{2}{r \left( r \frac{\tau_1}{\tau_2} \right)} = \frac{(1 + \beta_n \eta) \sin \beta_n \eta + (1 - \beta_n \eta) \cos \beta_n \eta}{\beta_n^2} \quad (18)$$

where  $\beta_n$  corresponds to the normalized value of  $\omega_{f_n}$ ; i.e.,  $\beta_n = \omega_{f_n} \tau_2$ . Letting  $n = 0$  in Eq. (18), we have from Eqs. (14) through (16),

$$\gamma(\beta_0) = \frac{\beta_0}{r \frac{\tau_1}{\tau_2}} + \frac{r}{2} \left[ \frac{\cos \beta_0 \eta}{\beta_0} + \frac{\sin \beta_0 \eta}{\beta_0} \right] = 0 \quad (19)$$

Solving Eqs. (18) and (19) simultaneously (with  $n = 0$  in Eq. 18) gives an equation for the critical value of normalized delay,  $\eta_c$ , i.e.,

$$\tan \beta_0 \eta_c = - \frac{(2 - \beta_0 \eta_c)}{(2 + \beta_0 \eta_c)} \quad (20)$$

whose principal solution is

$$\beta_0 \eta_c = 3.394 \quad (21)$$

Substituting Eq. (21) in Eq. (18) and solving for  $\eta_c$  gives

$$\eta_c = \sqrt{\frac{19.34}{r \left( r \frac{\tau_1}{\tau_2} \right)}} \quad (22)$$

and the absolute instability region is then given by  $\eta > \eta_c$ . Conversely for values of  $\eta < \eta_c$ , the loop will or will not be stable depending on the value of initial frequency offset as already discussed. Again, for  $r = 2$ , and  $\tau_1/\tau_2 = 200$ ,  $\eta_c = 0.156$ . We note that this bound depends on  $\tau_1/\tau_2$  whereas that predicted by the linear theory did not. Similar results have been obtained by Tausworthe (SPS 37-46, Vol. IV, pp. 226-234), Develet (Ref. 2), and Gruen (Ref. 3) using a relation for the synchronization boundary slightly different from Eq. (2).

Strictly speaking, the results of this section apply only to a second-order sinusoidal phase-locked loop. However, we anticipate from experimental results that the above solutions are reasonably close for data-aided loops with large signal-to-noise ratio. An improved result for the data-aided loop specifically will hopefully be available in the near future, along with accompanying phase-plane diagrams.

One justification for the applicability of the results obtained here to the data-aided loop comes from experimental evidence. If Eqs. (11) and (22) are expressed in terms of the loop bandwidth at zero delay,  $W_L = (r + 1)/2\tau_2$ , then for  $r = 2$  the corresponding inequalities necessary to guarantee absolute stability become

$$\left. \begin{aligned} R &> 1.3 W_L && \text{(linear theory)} \\ R &> 4.3 W_L && \text{(nonlinear theory)} \end{aligned} \right\} \quad (23)$$

where  $R \triangleq 1/T_d$  is the data rate. Experimentally it has been found that for  $R > 2W_L$  the data-aided loop is stable.

## References

1. Viterbi, A. J., *Principles of Coherent Communication*. McGraw-Hill Book Co., Inc., New York, 1966.
2. Develet, J. A., Jr., "The Influence of Time Delay on the Second-Order Loop Acquisition Range," *International Telemetry Conference*, London, England, pp. 432-437, 1963.
3. Gruen, N. J., *A Simple Derivation of Pull-In Range and Pull-In Time for Second-Order Frequency and Phase-Control Loops*. Thompson Ramo Wooldridge, Inc., Canoga Park, Calif., Mar. 7, 1962.

## B. Signal Design for Single-Sideband Phase Modulation, H. D. Chadwick

### 1. Introduction

In 1966, Springett (SPS 37-40, Vol. IV, pp. 163-169) and others showed that a signal of the form

$$s(t) = A e^{af(t)} \cos [\omega_c t + \hat{a}f(t)] \quad (1)$$

has a spectrum which is one-sided about the carrier frequency,  $\omega_c$ . In this equation  $f(t)$  is an arbitrary function of time and  $\hat{f}(t)$  is its Hilbert transform, given by

$$\hat{f}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(s)}{s-t} ds \quad (2)$$

The integral here is the so-called Cauchy principal part (Ref. 1).

While Springett's work and subsequent work by Secor<sup>1</sup> showed that a single-sideband phase-modulated (SSB-PM) scheme was practical and that bandwidth reductions on the order of 50% over standard phase modulation systems were achievable, no effort has so far been directed toward finding a good modulating signal,  $f(t)$ . In this article, a description is given of the derivation of a set of such signals, which are designed to minimize intersymbol interference.

### 2. The Detector

A correlation, or matched filter, detector of the form shown in Fig. 1 has been assumed. In a general  $m$ -ary system ( $m \geq 2$ ),  $m$  such detectors would be employed, each one correlating the input waveform with one of the possible transmitted signals, and the one showing the largest output would be chosen as corresponding to the output decision. Such a detector is known to provide the minimum probability of error for signals mixed

<sup>1</sup>Secor, D. J., *Single-Sideband Phase Modulation Using a Delay Line Hilbert Transformer*, July 4, 1967 (JPL internal document).

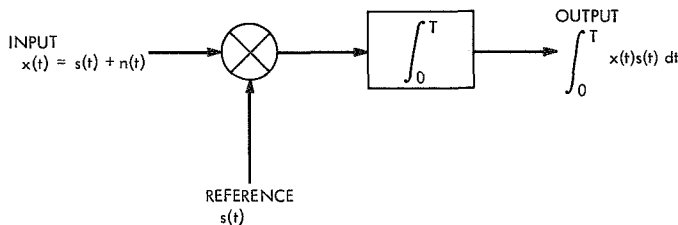


Fig. 1. Correlation detector

with additive, white, gaussian noise. Ideally, the signals would be chosen to occupy only a fixed time interval, so that a string of signals transmitted sequentially would not interfere with each other and the integrators could be reset to zero at the end of each interval. When the signals are not confined to a fixed interval, a problem of intersymbol interference exists because neighboring signals in the string will spill over or interfere with each other, degrading the detector performance.

When a time-limited pulse signal is sent over a frequency band-limited channel, intersymbol interference will always exist. This occurs because the Fourier transform of a band-limited signal cannot be time-limited. Slepian and Pollack (Ref. 2) have showed that one function, the prolate spheroidal wave function of order zero, has the optimum property that when time-limited before transmission and then band-limited in the channel, its received energy is the greatest of any function. This property has been used in the design of signals to reduce intersymbol interference (Ref. 3).

### 3. The Problem

An intersymbol interference problem similar to that of band-limited signals exists for the SSB-PM signal. The Hilbert transform of a time-limited signal is not time-limited so that it is not possible for both  $f(t)$  and  $\hat{f}(t)$  in Eq. (1) to be simultaneously time-limited. In order to use a simple correlation detector of the type of Fig. 1, it is desirable to keep this intersymbol interference to a minimum. With this objective, the following criterion was used for the design of the signals  $f(t)$  and  $\hat{f}(t)$ :

Given that  $f(t)$  is strictly time-limited to the interval  $(0, T)$ , choose that  $f(t)$  for which the energy of  $\hat{f}(t)$  in  $(0, T)$  is a maximum.

### 4. Integral Equation

Mathematically, the energy of  $\hat{f}(t)$  in the interval  $(0, T)$  is given by the relation

$$\alpha E = \int_0^T \hat{f}^2(t) dt \quad (3)$$

where  $\alpha$  is a constant  $0 \leq \alpha \leq 1$  expressing the relative amount of the total energy which is in the interval, and the total energy  $E$  is given by

$$E = \int_{-\infty}^{\infty} \hat{f}^2(t) dt \quad (4)$$

It can be shown that the total energy of the Hilbert transform  $\hat{f}(t)$  is the same as that of the original signal,  $f(t)$ , or that

$$E = \int_{-\infty}^{\infty} \hat{f}^2(t) dt = \int_0^T f^2(t) dt \quad (5)$$

because  $f(t)$  is strictly time-limited.

Substituting Eq. (2) into Eq. (3) and letting  $T = 1$  for simplicity gives

$$\alpha F = \int_0^1 \left[ \frac{1}{\pi} \int_0^1 \frac{f(u)}{u-t} du \right] \left[ \frac{1}{\pi} \int_0^1 \frac{f(v)}{v-t} dv \right] dt \quad (6)$$

Interchanging the order of integration (without regard to the consequences) gives

$$\alpha E = \frac{1}{\pi^2} \int_0^1 \int_0^1 f(u)f(v) \int_0^1 \frac{1}{(u-t)(v-t)} dt du dv \quad (7)$$

The inner integral can be evaluated except at the point  $u = v$  where it does not exist. This difficulty is overcome by the following artifice, which has not been rigorously justified, but has the advantage that it works. A  $\delta$  function has been added to represent the behavior of the integral at  $u = v$ . Using this artifice, the equation becomes

$$\alpha E = \int_0^1 \int_0^1 f(u)f(v) \times \left[ \delta(u-v) + \frac{1}{\pi^2} \left( \frac{1}{u-v} \right) \log \left( \frac{v(1-u)}{u(1-v)} \right) \right] du dv \quad (8)$$

where the  $\delta(u-v)$  has been used to represent the singularity at  $u = v$  in the integral. Rewriting the equation gives

$$\alpha E = \int_0^1 \int_0^1 K(u,v) f(u)f(v) du dv \quad (9)$$

where

$$K(u,v) = \delta(u-v) + \frac{1}{\pi^2} \left( \frac{1}{u-v} \right) \log \left( \frac{v(1-u)}{u(1-v)} \right) \quad (10)$$

Letting

$$K(u,v) = \delta(u-v) - K'(u,v) \quad (11)$$

the equation becomes

$$\alpha E = \int_0^1 f^2(t) dt - \int_0^1 \int_0^1 K'(u,v) f(u)f(v) du dv \quad (12)$$

or

$$\alpha E = E - \beta E \quad (13)$$

Hence the second integral represents the negative of the amount of energy outside the interval  $(0,1)$ , and the objective of maximizing  $\alpha$  may be achieved by minimizing  $\beta$ . Thus we have the integral equation

$$\beta E = \int_0^1 \int_0^1 K'(u,v) f(u)f(v) du dv \quad (14)$$

where

$$K'(u,v) = \frac{1}{\pi^2} \left( \frac{1}{v-u} \right) \log \left( \frac{v(1-u)}{u(1-v)} \right) \quad (15)$$

for which the objective is to find the  $f(t)$  which minimizes  $\beta$ . Letting

$$\lambda f(u) = \int_0^1 K'(u,v) f(v) dv \quad (16)$$

and substituting into Eq. (14) gives

$$\beta E = \lambda \int_0^1 f^2(u) du = \lambda E \quad (17)$$

Thus, solutions to Eq. (16) are also solutions to Eq. (14) and the solution which provides the smallest value of  $\lambda$  will be the desired solution. The problem with such an eigenvalue equation as (16) is that there is no smallest eigenvalue,  $\lambda$ . There does exist a largest eigenvalue, but all succeeding solutions produce smaller values, and there is no limit to the number of solutions. It turns out, however, when the actual solutions are examined, that the eigenvalues converge quite rapidly to a value close to zero.

## 5. Solutions to the Equation

No closed form solution has been obtained to Eq. (16). However, by using numerical techniques, a set of eigenfunction solutions has been formed in tabular form, and these solutions appear to have the desired properties. The numerical solutions were formed by using a discrete



approximation to the integral equation (16). Letting  $f_k(t)$ ,  $k = 0, 1, \dots$ , be the  $k$ th eigenfunction of the equation and  $\lambda_k$  the equivalent eigenvalue, the discrete approximation to  $f_k(t)$  is  $f_k(t_i)$ , where  $t_i$  is a discrete time point. The discrete equivalent to Eq. (16) becomes

$$\lambda_k f_k(u_i) = \sum_{j=1}^n K'(u_i, v_j) f_k(v_j) \Delta v \quad \begin{cases} i = 1, 2, \dots, n \\ k = 0, 1, \dots \end{cases} \quad (18)$$

where  $\Delta v$  is the time increment between values of  $v_j$  (a uniform spacing is assumed), and  $n$  is the number of points in the approximation. Using the matrix notation

$$\mathbf{f}_k = \begin{bmatrix} f_k(u_1) \\ f_k(u_2) \\ \vdots \\ f_k(u_n) \end{bmatrix} = \begin{bmatrix} f_k(v_1) \\ f_k(v_2) \\ \vdots \\ f_k(v_n) \end{bmatrix} \quad (19)$$

and

$$\mathbf{K}' = \begin{bmatrix} K'(u_1, v_1) & K'(u_1, v_2) & \cdot & \cdot & \cdot & K'(u_1, v_n) \\ K'(u_2, v_1) & & & & & \cdot \\ \cdot & & & & & \cdot \\ \cdot & & & & & \cdot \\ K'(u_n, v_1) & & & & & K'(u_n, v_n) \end{bmatrix} \quad (20)$$

Eq. (19) may be written

$$\lambda_k \mathbf{f}_k = \mathbf{K}' \mathbf{f}_k \quad (21)$$

which is a familiar matrix eigenvector problem. The discrete solutions  $\mathbf{f}_k$  are the eigenvectors, and the constants  $\lambda_k$  the eigenvalues of the matrix  $\mathbf{K}'$ . Obviously, the larger the number of points used,  $n$ , the more accurate will be the discrete approximation, but, as  $n$  becomes large, the difficulty in evaluating the eigenvectors becomes greater.

## 6. Computed Solutions

Solutions to Eq. (21) were calculated on a digital computer for  $n = 64$  and  $n = 128$ . The first six eigenvectors obtained are illustrated for  $n = 128$  in Fig. 2. These discrete solutions were then Hilbert transformed using a discrete Hilbert transform technique to obtain the six

functions,  $\hat{f}_k(t)$ , illustrated in Fig. 3. As a check on the reliability of this process, the energy contained in the interval (0,1) of the Hilbert transformed functions,  $\alpha E$ , can be compared to the calculated eigenvalues of the matrix, since by Eqs. (13) and (17)

$$\alpha E = (1 - \lambda) E \quad (22)$$

The solutions have been normalized so that  $E = 1$ . The results are summarized in Table 1.

**Table 1. Comparison of computed eigenvalues with energy of Hilbert transform in interval**

k	n = 64		n = 128	
	$\alpha$	$1 - \lambda$	$\alpha$	$1 - \lambda$
0	0.197	0.224	0.177	0.180
1	0.576	0.656	0.530	0.603
2	0.841	0.893	0.800	0.857
3	0.954	0.974	0.933	0.958
4	0.990	0.994	0.984	0.989
5	0.999	0.999	0.998	0.998

## 7. Discussion of Results

The results show that the eigenfunction solutions to Eq. (16) converge quite rapidly to the desired result. There is no single optimum solution, because the higher the order of the eigenfunction, the better the energy of the Hilbert transform is concentrated in the desired interval. This is offset, however, by the increased complexity of the higher-order eigenfunctions and by their apparently increased bandwidth. They also have the undesirable property that they go to infinity at the end points of the interval. In a practical system this is obviously not permissible, and a finite approximation must be substituted.

Further work is now being pursued in the following areas:

- (1) The bandwidth properties of the modulated signal  $s(t)$  when the functions  $f_k(t)$  and  $\hat{f}_k(t)$  are used for modulation are being investigated. Reduced bandwidth is the primary motive for investigating SSB-PM.
- (2) The performance of the correlation detector in terms of probability of error versus signal-to-noise ratio for modulated signals of this type is being

determined. This performance should be equivalent to that of double-sideband systems.

### 8. Conclusions

A set of time-limited functions whose Hilbert transform has maximum energy within a limited time interval has been obtained by discrete approximation. These functions are useful in minimizing the problem of intersymbol interference in single-sideband phase-modulated communication systems.

### References

1. Titchmarsh, E. C., *Introduction to the Theory of Fourier Integrals*, Chapter V. Oxford University Press, London, 1962.
2. Slepian, D., and Pollack, H. O., "Prolate Spheroidal Wave Functions, Fourier Analysis and Uncertainty," *Bell System Technical Journal*, Vol. 40, Jan. 1961.
3. Salzberg, B., and Kurz, L., *Design of Bandlimited Signals for Binary Communication Using Simple Correlation Detection*, Technical Report 400-90. New York University Laboratory for Electrosience Research, Feb. 1964.

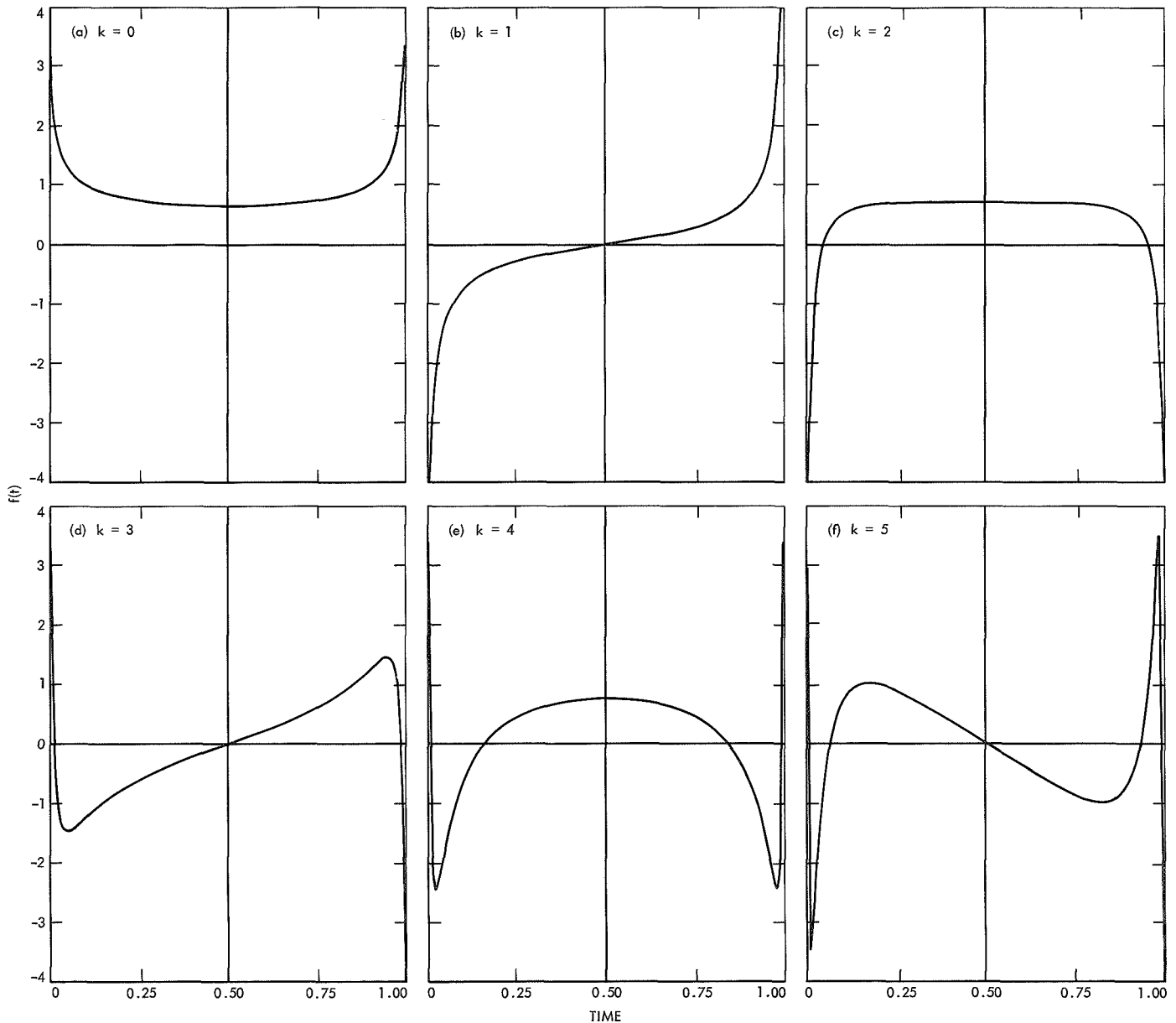
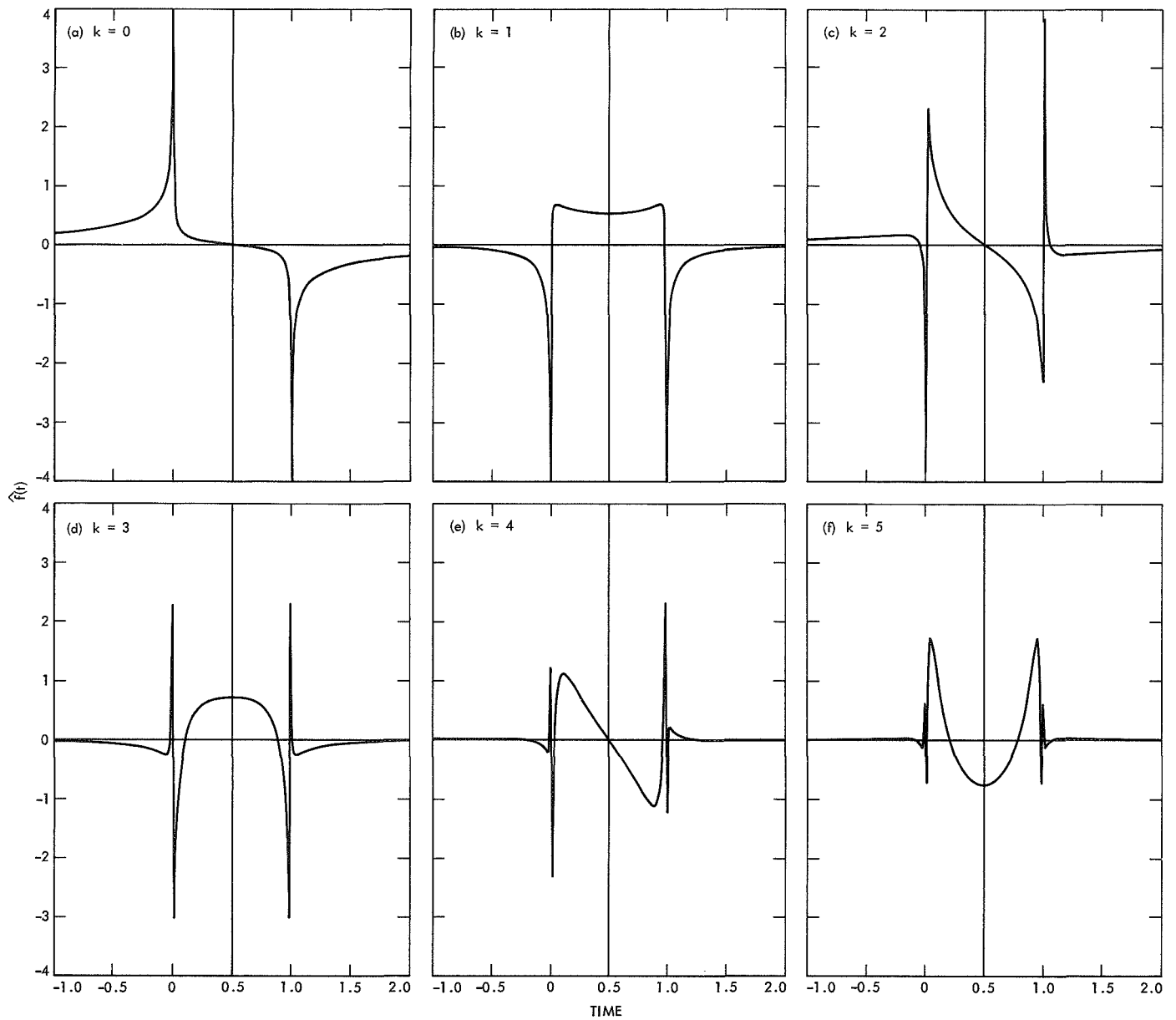


Fig. 2. Solution to integral equation



**Fig. 3. Hilbert transform of solution**

## C. The Steady-State Performance of a Data-Transition Type of First-Order Digital Phase-Locked Loop, M. K. Simon

### 1. Introduction

Recently, a fully suppressed carrier data-aided carrier tracking loop and attendant data detector were implemented<sup>1</sup> and proposed for possible application to future planetary spacecraft programs, communication satellites, and the Manned Space Station. Several unanswered problems existed at the time of the proposal which prevented the idea from being immediately committed to a specific project. This article is the first in a series which attempts to answer some of these questions.

One of the significant advantages of the proposed data-aided receiver is its ability to give satisfactory tracking (phase-noise) and bit error probability performance, in the presence of a "coarse" estimate of bit sync timing. In order to understand how this is possible, we must first study the behavior and performance of the bit synchronizer itself, which may be done apart from the total receiver configuration.

Hence, the purpose of this article is to develop a mathematical model which describes the coarse bit synchronizer used in the data-aided receiver and then proceed to analyze its steady-state performance. The results are obtained through a generalization of a recent contribution by Holmes (SPS 37-64, Vol. III, pp. 17-22) on first-order digital phase-locked loops and from previous work of the author (SPS 37-55, Vol. III, pp. 54-62).

<sup>1</sup>A paper delineating the design and implementation is presently being written.

### 2. A Description of the Operation of the Bit Synchronizer

The method whereby the phase of the bit synchronizer's local oscillator is discretely updated to match that of the incoming bit sequence makes use of the digital phase-lock loop principle. In particular, the phase detector topology of the data-transition tracking loop (SPS 37-55, Vol. III) is used to provide an error signal which is used to either bump the phase of the voltage-controlled oscillator (VCO) clock  $\pm \Delta T$  from its present position, or leave the VCO at its present phase position. The probabilities with which these three possible phase updates occur depend on the magnitude of the phase error between the incoming bit sequence and the clock, and the presence or absence of a transition in the data.

Referring to Fig. 1, the input signal  $s(t, \epsilon)$  is a random pulse train, representing the data bits; i.e.,

$$s(t, \epsilon) = \sum_n d_n u[t - nT - \epsilon] \quad (1)$$

where  $u(t)$  is the unit rectangular pulse defined by  $u(t) = 1$  for  $0 \leq t \leq T$ ,  $u(t) = 0$  for all other  $t$ ;  $T$  is the bit period;  $d_n$  the bit amplitude, takes on the values  $\pm A$  with equal probability; and  $\epsilon$  is the random epoch to be estimated. The input additive noise process  $n(t)$  is assumed to be white gaussian with a two-sided spectral density  $N_0/2$ . The in-phase and mid-phase filters,  $F_{IP}(s)$  and  $F_{MP}(s)$  respectively, are, in principle, matched filters to the input (i.e., integrate and dump circuits), although they can be implemented in several ways to perform their function. (In the proposed implementation they are reset accumulators as the input ensemble is quantized.) The decision device is simply a hard limiter with a "sgn" function output versus input characteristic. The transition detector then examines two adjacent decisions,  $\hat{d}_{j+1}$ ,  $\hat{d}_j$ , and records an output  $I_j$  according to the following rule:

$$\left. \begin{aligned} &\text{If } \hat{d}_j = \hat{d}_{j+1}, \text{ then } I_j = 0 \\ &\text{If } \hat{d}_j = -1, \hat{d}_{j+1} = +1, \text{ then } I_j = -1 \\ &\text{If } \hat{d}_j = +1, \hat{d}_{j+1} = -1, \text{ then } I_j = +1 \end{aligned} \right\} \quad (2)$$

The output  $J_j$  of the mid-phase filter is also sampled at intervals of  $T$  and must be delayed by an amount  $\tau = T/2$  before multiplication with the appropriate  $I_j$ . The error signal,  $e_j = I_j \cdot J_j$ , is in general accumulated for an interval  $MT$  (i.e.,  $M$  values of  $e_j$  are summed)

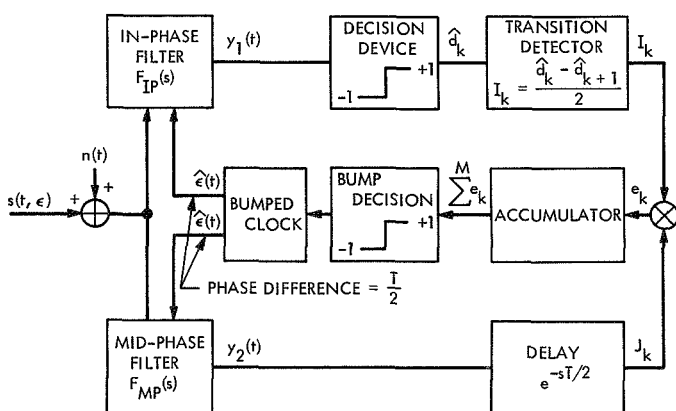


Fig. 1. System model for data-transition type of digital bit synchronizer

and the clock is bumped (i.e., the phase estimate  $\hat{\epsilon}_n$  at time  $nMT$  is updated) according to the following rule:

$$\left. \begin{aligned} \hat{\epsilon}_{n+1} &= \hat{\epsilon}_n - \Delta T, & \text{if } \sum_{j=n-M+1}^n e_j < 0 \\ \hat{\epsilon}_{n+1} &= \hat{\epsilon}_n, & \text{if } \sum_{j=n-M+1}^n e_j = 0 \\ \hat{\epsilon}_{n+1} &= \hat{\epsilon}_n + \Delta T, & \text{if } \sum_{j=n-M+1}^n e_j > 0 \end{aligned} \right\} \quad (3)$$

Letting  $2N = T/\Delta T$  denote the number<sup>2</sup> of phase positions for the bumped clock VCO, then the normalized phase error  $\lambda_n = (\epsilon - \hat{\epsilon}_n)/T$  is a discrete random variable which in the symmetrical case<sup>3</sup> ranges over the set of  $2N$  values

$$-\frac{1}{2} + \frac{1}{4N}, -\frac{1}{2} + \frac{3}{4N}, \dots, -\frac{1}{4N},$$

$$\frac{1}{4N}, \dots, \frac{1}{2} - \frac{3}{4N}, \frac{1}{2} - \frac{1}{4N}$$

Our initial goal is to determine the probability frequency function (PFF) of  $\lambda_n$  in the steady state (i.e., as  $n \rightarrow \infty$ ). Of course, the input timing epoch can fall anywhere within a bump interval with uniform probability and hence, in general, the phase error will take on a non-symmetrical set of values,

$$-\frac{1}{2} + \frac{1}{4N} + \delta, -\frac{1}{2} + \frac{3}{4N} + \delta, \dots, -\frac{1}{4N} + \delta,$$

$$\frac{1}{4N} + \delta, \dots, \frac{1}{2} - \frac{3}{4N} + \delta, \frac{1}{2} - \frac{1}{4N} + \delta$$

where  $|\delta| \leq 1/(4N)$ . Thus, the most general treatment would be to find the PFF of  $\lambda_n$  conditioned on  $\delta$  and then average over the uniform probability distribution on  $\delta$ . We shall treat only the symmetric case for simplicity.

<sup>2</sup>The procedure carried out here is not restricted to an even number of phase positions although we assume this for simplicity.

<sup>3</sup>By symmetric here we mean that the input timing offset,  $\epsilon$ , falls in the middle of a bump interval.

### 3. The Steady-State PFF of $\lambda = \lim_{n \rightarrow \infty} \lambda_n$

The method used to find the steady-state PFF of  $\lambda$  is based upon a random-walk model for a first-order digital phase-locked loop proposed by Holmes in SPS 37-64, Vol. III. If the  $2N$  allowable values for  $\lambda_n$  are looked upon as "states" in a transition diagram, then the steady-state probabilities of these states exist and satisfy a second-order difference equation of the form

$$P_k = q_{k-1} P_{k-1} + r_k P_k + p_{k+1} P_{k+1} \quad (4)$$

In Eq. (4),  $P_k$  denotes the probability that the steady-state phase error,  $\lambda$ , is in state  $k$ ;  $|k| = 1, 2, \dots$ , which is defined as the condition

$$\lambda = (\text{sgn } k) \left[ \frac{1}{4N} + \frac{(k-1)}{2N} \right]$$

Also,  $p_k$ ,  $q_k$ , and  $r_k$  are respectively the transition probabilities corresponding to the events that starting at state  $k$ , the clock phase is bumped  $\Delta T$ ,  $-\Delta T$ , and not at all. From Eq. (3), these transition probabilities may be evaluated as

$$p_k = \text{Prob} \left\{ \sum_{j=1}^M e_j > 0 \mid \lambda = (\text{sgn } k) \left[ \frac{1}{4N} + \frac{(k-1)}{2N} \right] \right\}$$

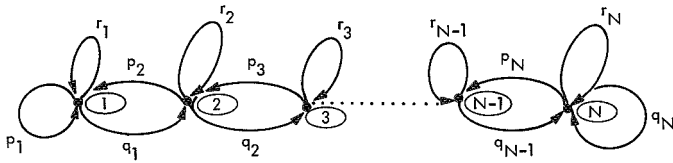
$$q_k = \text{Prob} \left\{ \sum_{j=1}^M e_j < 0 \mid \lambda = (\text{sgn } k) \left[ \frac{1}{4N} + \frac{(k-1)}{2N} \right] \right\}$$

$$r_k = \text{Prob} \left\{ \sum_{j=1}^M e_j = 0 \mid \lambda = (\text{sgn } k) \left[ \frac{1}{4N} + \frac{(k-1)}{2N} \right] \right\} \quad (5)$$

and

$$p_k + q_k + r_k = 1 \quad (6)$$

Since we are only interested in the steady-state probability distribution of  $\lambda$  reduced modulo 1, we must fold the probability frequency function of  $\lambda$  into the interval  $-1/2 \leq \lambda \leq 1/2$ . In this regard, Holmes talks about "reflected states," and, since we are considering only the symmetric case, it is sufficient to restrict ourselves to the positive states (i.e., those corresponding to positive values of  $\lambda$ ). Hence, Eq. (4) is valid for all positive values



**Fig. 2. Markov chain model for positive phase error states**

of  $k = 2, 3, \dots, N-1$ , and at the end points we have the boundary conditions (see Fig. 2)

$$\left. \begin{aligned} P_1 &= (r_1 + p_1)P_1 + p_2P_2 = (1 - q_1)P_1 + p_2P_2 \\ P_N &= q_{N-1}P_{N-1} + (r_N + q_N)P_N = q_{N-1}P_{N-1} + (1 - p_N)P_N \end{aligned} \right\} \quad (7)$$

Comparing Eqs. (4) and (7) with Eqs. (7) and (8) of Holmes' work, one can readily see that the solution for  $P_k$  satisfies the identical recursive relationship as given by Eq. (9) of Holmes; i.e.,

$$P_{k+1} = \frac{q_k}{p_{k+1}} P_k, \quad k = 1, 2, \dots, N \quad (8)$$

except that  $p_k$  and  $q_k$  sum to  $1 - r_k$  rather than unity. Solving Eq. (8) together with the normalization condition

$$\sum_{k=1}^N P_k = 1 \quad (9)$$

gives the result

$$\left. \begin{aligned} P_1 &= \frac{1}{1 + \sum_{k=2}^N \prod_{j=1}^{k-1} \frac{q_j}{p_{j+1}}} \\ P_k &= \frac{\prod_{j=1}^{k-1} \frac{q_j}{p_{j+1}}}{1 + \sum_{k=2}^N \prod_{j=1}^{k-1} \frac{q_j}{p_{j+1}}}, \quad k = 2, 3, \dots, N \end{aligned} \right\} \quad (10)$$

What remains is the computation of transition probabilities. This is the most difficult and tedious part of the problem, and in fact manageable results can only be obtained for the case  $M = 1$ , i.e., the clock is updated every bit interval, and the limiting case of large  $M$ . Since accumulation of the error signal  $e_j$  over many bit intervals produces a smoothing or filtering effect on the loop dynamics, we can anticipate that the PFF of  $\lambda$  obtained for  $M = 1$  would lead to a worst-case prediction of bit

synchronizer performance and that for large  $M$  as a best case.

#### 4. Calculation of the Transition Densities

As noted in SPS 37-55, Vol. III, the  $k$ th bit interval error signal,  $e_k$ , produced by the data transition type of phase detector topology in Fig. 1 is correlated with those error signals corresponding to the two adjacent intervals; i.e.,  $e_{k-1}$  and  $e_{k+1}$ . Hence, to find the probability density function (PDF) of the random variable

$$Y_k = \sum_{j=k-M}^k e_j \quad (11)$$

for arbitrary  $M > 1$ , one must be able to compute the three-dimensional PDF of  $e_{k-1}$ ,  $e_k$ , and  $e_{k+1}$ . As we shall see shortly, computing the one-dimensional PDF of  $e_k$  is indeed quite tedious, and hence, as mentioned above, we shall be satisfied with treating the special case  $M = 1$ . In Subsection 5, we shall also treat the limiting case of large  $M$  by making use of the theory of  $j$ -dependent sequences (Ref. 1).

With reference to Fig. 1,

$$\left. \begin{aligned} y_1(kT) &\doteq y_{1k} = c_k + v_k \\ y_2\left[\left(k + \frac{1}{2}\right)T\right] &\doteq y_{2k} = b_k + \mu_k \end{aligned} \right\} \quad (12)$$

where

$$\begin{aligned} c_k &= \int_{(k-1)T + \hat{\epsilon}_k}^{kT + \hat{\epsilon}_k} s(t) dt \\ v_k &= \int_{(k-1)T + \hat{\epsilon}_k}^{kT + \hat{\epsilon}_k} n(t) dt \\ b_k &= \int_{(k-1/2)T + \hat{\epsilon}_k}^{(k+1/2)T + \hat{\epsilon}_k} s(t) dt \\ \mu_k &= \int_{(k-1/2)T + \hat{\epsilon}_k}^{(k+1/2)T + \hat{\epsilon}_k} n(t) dt \end{aligned}$$

and

$$\left. \begin{aligned} \hat{d}_k &= \text{sgn } y_{1k} \\ I_k &= \frac{\hat{d}_k - \hat{d}_{k+1}}{2} \end{aligned} \right\} \quad (13)$$

Recalling that  $e_k = I_k \cdot J_k$ , then combining Eqs. (12) and (13), we get

$$e_k = (b_k + \mu_k) \left[ \frac{\text{sgn}(c_k + v_k) - \text{sgn}(c_{k+1} + v_{k+1})}{2} \right] \quad (14)$$

From Eq. (12), the quantities  $\mu_k$ ,  $v_k$ , and  $v_{k+1}$  are gaussian random variables with zero mean and variances given by

$$\sigma_{\mu_k}^2 = \sigma_{v_k}^2 = \sigma_{v_{k+1}}^2 \triangleq \sigma^2 = \frac{N_0 T}{2} \quad (15)$$

Their covariance matrix,  $\Lambda$ , may also be found from the definitions in Eq. (12) as

$$\Lambda = \begin{matrix} & \begin{matrix} v_k & \mu_k & v_{k+1} \end{matrix} \\ \begin{matrix} v_k \\ \mu_k \\ v_{k+1} \end{matrix} & \begin{bmatrix} \sigma^2 & \frac{\sigma^2}{2} & 0 \\ \frac{\sigma^2}{2} & \sigma^2 & \frac{\sigma^2}{2} \\ 0 & \frac{\sigma^2}{2} & \sigma^2 \end{bmatrix} \end{matrix} \quad (16)$$

with determinant

$$\det[\Lambda] = \frac{\sigma^6}{2} \quad (17)$$

Hence, the joint PDF of  $v_k$ ,  $\mu_k$ ,  $v_{k+1}$  is three-dimensional gaussian and is given by

$$p(v_k, \mu_k, v_{k+1}) = \frac{1}{(2\pi\sigma^2)^{3/2} \frac{1}{\sqrt{2}}} \exp \left\{ - \frac{[v_k, \mu_k, v_{k+1}] \begin{bmatrix} \frac{3}{2} & -1 & \frac{1}{2} \\ -1 & 2 & -1 \\ \frac{1}{2} & -1 & \frac{3}{2} \end{bmatrix} \begin{bmatrix} v_k \\ \mu_k \\ v_{k+1} \end{bmatrix}}{2\sigma^2} \right\} \quad (18)$$

Our immediate goal is to find the conditional distribution function of  $e_k$  given  $\lambda$ ; i.e.,

$$F(V|\lambda) = \text{Prob}\{e_k \leq V|\lambda\} = 1 - \text{Prob}\{e_k > V|\lambda\}; \quad -\infty \leq V \leq \infty \quad (19)$$

from which the transition probabilities  $p_k$ ,  $q_k$ ,  $r_k$  as defined in Eq. (5) can be found. From Eq. (14),

$$\text{Prob}\{e_k > V|\lambda\} = \begin{cases} \text{Prob}\{c_k + v_k > 0; c_{k+1} + v_{k+1} < 0; b_k + \mu_k > V\} \\ + \text{Prob}\{c_k + v_k < 0; c_{k+1} + v_{k+1} > 0; b_k + \mu_k < -V\}, \\ \text{for } V \geq 0 \\ \\ \text{Prob}\{c_k + v_k > 0; c_{k+1} + v_{k+1} < 0; b_k + \mu_k > V\} \\ + \text{Prob}\{c_k + v_k < 0; c_{k+1} + v_{k+1} > 0; b_k + \mu_k < -V\} + \text{Prob}\{e_k = 0|\lambda\}, \\ \text{for } V < 0 \end{cases} \quad (20)$$

where

$$\begin{aligned} \text{Prob}\{e_k = 0|\lambda\} &= \text{Prob}\{c_k + v_k > 0; c_{k+1} + v_{k+1} > 0; -\infty < b_k + \mu_k < \infty\} \\ &+ \text{Prob}\{c_k + v_k < 0; c_{k+1} + v_{k+1} < 0; -\infty < b_k + \mu_k < \infty\} \end{aligned} \quad (21)$$

In terms of the joint probability density of  $v_k, \mu_k, v_{k+1}$ , Eqs. (19) and (20) become

$$\text{Prob} \{e_k > V | \lambda\} = \begin{cases} \int_{-c_k}^{\infty} \int_{V-b_k}^{\infty} \int_{-\infty}^{-c_{k+1}} p(v_k, \mu_k, v_{k+1}) dv_{k+1} d\mu_k dv_k \\ + \int_{-\infty}^{-c_k} \int_{-\infty}^{V-b_k} \int_{-c_{k+1}}^{\infty} p(v_k, \mu_k, v_{k+1}) dv_{k+1} d\mu_k dv_k \\ \text{for } V \geq 0 \\ \\ \int_{-c_k}^{\infty} \int_{V-b_k}^{\infty} \int_{-\infty}^{-c_{k+1}} p(v_k, \mu_k, v_{k+1}) dv_{k+1} d\mu_k dv_k \\ + \int_{-\infty}^{-c_k} \int_{-\infty}^{V-b_k} \int_{-c_{k+1}}^{\infty} p(v_k, \mu_k, v_{k+1}) dv_{k+1} d\mu_k dv_k + \text{Prob} \{e_k = 0 | \lambda\}, \\ \text{for } V < 0 \end{cases} \quad (22)$$

and since  $v_k$  and  $v_{k+1}$  are independent,

$$\text{Prob} \{e_k = 0 | \lambda\} = \int_{-c_k}^{\infty} p(v_k) dv_k \int_{-c_{k+1}}^{\infty} p(v_{k+1}) dv_{k+1} + \int_{-\infty}^{-c_k} p(v_k) dv_k \int_{-\infty}^{-c_{k+1}} p(v_{k+1}) dv_{k+1} \quad (23)$$

The integrals in Eqs. (22) and (23) can be somewhat simplified with the result

$$\text{Prob} \{e_k > V | \lambda\} = \text{Prob} \{e'_k > V' | \lambda\} = \begin{cases} \int_{V'}^{\infty} h(e'_k; b'_k, c'_k, c'_{k+1}) de'_k, & \text{for } V \geq 0 \\ \int_{V'}^{\infty} h(e'_k; b'_k, c'_k, c'_{k+1}) de'_k + \text{Prob} \{e_k = 0 | \lambda\}, & \text{for } V < 0 \end{cases} \quad (24)$$

$$\text{Prob} \{e_k = 0 | \lambda\} = \frac{1}{4} \text{erfc} [-c'_k] \text{erfc} [-c'_{k+1}] + \frac{1}{4} \text{erfc} [c'_k] \text{erfc} [c'_{k+1}] \triangleq g(c'_k, c'_{k+1}) \quad (25)$$

where the prime on  $V, b_k, c_k$ , and  $c_{k+1}$  denotes normalization with respect to  $\sqrt{2}\sigma$ ; e.g.,  $V' = V/\sqrt{2}\sigma$  and

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \exp(-y^2) dy \quad (26)$$

Furthermore, the function  $h(e'_k; b'_k, c'_k, c'_{k+1})$  is defined by

$$h(e'_k; b'_k, c'_k, c'_{k+1}) = \frac{1}{2\pi \sqrt{\frac{3}{4}}} \left[ \int_{-c'_k}^{\infty} \text{erfc} \left[ \sqrt{\frac{3}{2}} \left\{ c'_{k+1} + \frac{2}{3} \left( e'_k - b'_k - \frac{z}{2} \right) \right\} \right] \exp \left[ -\frac{4}{3} \{ z^2 + (e'_k - b'_k) - z(e'_k - b'_k) \} \right] dz \right. \\ \left. + \int_{c'_k}^{\infty} \text{erfc} \left[ \sqrt{\frac{3}{2}} \left\{ -c'_{k+1} + \frac{2}{3} \left( e'_k + b'_k - \frac{z}{2} \right) \right\} \right] \exp \left[ -\frac{4}{3} \{ z^2 + (e'_k + b'_k) - z(e'_k + b'_k) \} \right] dz \right] \quad (27)$$



The functional dependence of the parameters  $b'_k$ ,  $c'_k$ , and  $c'_{k+1}$  on  $\lambda$  was derived in SPS 37-55, Vol. III, with the result

$$\left. \begin{aligned} b'_k &= \sqrt{R_s} \left[ \frac{d_{k-1}}{A} \left( \frac{1}{2} + \lambda \right) + \frac{d_k}{A} \left( \frac{1}{2} - \lambda \right) \right] \\ c'_k &= \sqrt{R_s} \left[ \frac{d_{k-2}}{A} \lambda + \frac{d_{k-1}}{A} (1 - \lambda) \right] \\ c'_{k+1} &= \sqrt{R_s} \left[ \frac{d_{k-1}}{A} \lambda + \frac{d_k}{A} (1 - \lambda) \right] \end{aligned} \right\} 0 \leq \lambda \leq \frac{1}{2} \quad (28)$$

In the above,

$$R_s = \frac{A^2 T}{N_0} = \frac{A^2 T^2}{2\sigma^2} \quad (29)$$

is the data signal-to-noise ratio.

Substituting Eq. (28) in Eqs. (24), (25), and (27), and averaging over all possible equally likely sequences formed from the bits  $d_{k-2}$ ,  $d_{k-1}$ , and  $d_k$ , we get

$$\text{Prob} \{e_k > V \mid \lambda\} = \begin{cases} \frac{1}{4} \int_{V'}^{\infty} [h(e'_k; \sqrt{R_s}, \sqrt{R_s}, \sqrt{R_s}) + h(e'_k; 2\lambda \sqrt{R_s}, \sqrt{R_s}, -(1-2\lambda) \sqrt{R_s}) \\ + h(e'_k; 2\lambda \sqrt{R_s}, (1-2\lambda) \sqrt{R_s}, -(1-2\lambda) \sqrt{R_s}) + h(e'_k; \sqrt{R_s}, (1-2\lambda) \sqrt{R_s}, \sqrt{R_s})] de'_k, \\ \text{for } V \geq 0, \quad 0 \leq \lambda \leq \frac{1}{2} \\ \frac{1}{4} \int_{V'}^{\infty} [h(e'_k; \sqrt{R_s}, \sqrt{R_s}, \sqrt{R_s}) + h(e'_k; 2\lambda \sqrt{R_s}, \sqrt{R_s}, -(1-2\lambda) \sqrt{R_s}) \\ + h(e'_k; 2\lambda \sqrt{R_s}, (1-2\lambda) \sqrt{R_s}, -(1-2\lambda) \sqrt{R_s}) + h(e'_k; \sqrt{R_s}, (1-2\lambda) \sqrt{R_s}, \sqrt{R_s})] de'_k \\ + \text{Prob} \{e_k = 0 \mid \lambda\}, \quad \text{for } V < 0, \quad 0 \leq \lambda \leq \frac{1}{2} \end{cases} \quad (30a)$$

$$\begin{aligned} \text{Prob} \{e_k = 0 \mid \lambda\} &= \frac{1}{4} [g(\sqrt{R_s}, \sqrt{R_s}) + g((1-2\lambda) \sqrt{R_s}, \sqrt{R_s}) \\ &\quad + g(\sqrt{R_s}, -(1-2\lambda) \sqrt{R_s}) + g((1-2\lambda) \sqrt{R_s}, -(1-2\lambda) \sqrt{R_s})], \\ &\quad \text{for } 0 \leq \lambda \leq \frac{1}{2} \end{aligned} \quad (30b)$$

From the above, one can now evaluate the transition densities defined in Eq. (5) for the special case of  $M = 1$ .

Before proceeding with our discussion, it is of interest to study the limiting behavior of the transition densities for  $M = 1$  as the signal-to-noise ratio approaches infinity. From Eqs. (30b) and (25),

$$\lim_{R_s \rightarrow \infty} \text{Prob} \{e_k = 0 \mid \lambda\} = \begin{cases} 0.5, & 0 \leq \lambda < 0.5 \\ 0.625, & \lambda = 0.5 \end{cases} \quad (31)$$

and hence

$$\lim_{R_s \rightarrow \infty} r_k = 0.5, \quad k = 1, 2, \dots, N \quad (32)$$

Also, from Eqs. (30a) and (27),

$$\left. \begin{aligned} \lim_{R_s \rightarrow \infty} \text{Prob} \{e_k > 0 | \lambda\} &= \begin{cases} 0.25, & \lambda = 0 \\ 0.5, & 0 < \lambda < 0.5 \\ 0.1875, & \lambda = 0.5 \end{cases} \\ \lim_{R_s \rightarrow \infty} \text{Prob} \{e_k < 0 | \lambda\} &= 1 - \lim_{R_s \rightarrow \infty} \{\text{Prob} \{e_k > 0 | \lambda\} + \text{Prob} \{e_k = 0 | \lambda\}\} \\ &= \begin{cases} 0.25, & \lambda = 0 \\ 0, & 0 < \lambda < 0.5 \\ 0.1875, & \lambda = 0.5 \end{cases} \end{aligned} \right\} \quad (33)$$

and hence

$$\left. \begin{aligned} \lim_{R_s \rightarrow \infty} p_k &= 0.5 \\ \lim_{R_s \rightarrow \infty} q_k &= 0 \end{aligned} \right\} k = 1, 2, \dots, N \quad (34)$$

Substituting Eq. (34) into Eq. (10) gives a limiting, one-sided, steady-state distribution

$$\lim_{R_s \rightarrow \infty} P_1 = 1; \quad \lim_{R_s \rightarrow \infty} P_k = 0, \quad k = 2, 3, \dots, N \quad (35)$$

or the phase error will in the no-noise condition alternate between  $1/(4N)$  and  $-1/(4N)$  with equal probability.

## 5. Approximation of the Transition Densities for Large $M$

When  $M$ , the number of bit intervals over which the error variable is accumulated, becomes large, we can make use of a theorem relating to  $j$ -dependent sequences to obtain approximate expressions for the transition densities. To begin, a sequence of random variables,  $x_1, x_2, \dots$  is  $j$ -dependent if  $(x_1, x_2, \dots, x_r)$  is always independent of  $(x_s, x_{s+1}, \dots)$  for  $s - r > j$ . With this definition in mind, we state the following theorem (Ref. 1):

**Theorem.** *If  $x_1, x_2, \dots$  is a stationary  $j$ -dependent sequence of random variables with  $E\{x_1\} = \mu_x$ ,  $E\{|x_1|^3\} < \infty$ , then as  $M \rightarrow \infty$ , the limiting distribution of*

$$M^{-1/2} \sum_{i=1}^M x_i$$

*is gaussian with mean  $M^{-1/2} \mu_x$  and variance*

$$\text{var} \{x_1\} + 2 \sum_{i=1}^j \text{cov} \{x_1 x_{1+i}\}$$

From previous work in SPS 37-55, Vol. III, we note that the sequence of error variables  $e_k, e_{k+1}, \dots$  is stationary and  $j$ -dependent with  $j = 1$ . Applying the theorem,  $Y_k$  as defined by Eq. (11) is gaussian for large  $M$  with moments

$$\left. \begin{aligned} E\{Y_k\} &\triangleq \mu_Y = \sqrt{M} E\{e_k\} \\ E\{(Y_k - \mu_Y)^2\} &\triangleq \sigma_Y^2 = \text{var}\{e_k\} + 2E\{e_k e_{k+1}\} \end{aligned} \right\} \quad (36)$$

The mean and variance of  $e_k$  were derived in SPS 37-55, Vol. III. Using the results of Eqs. (8), (12), and (13) of this reference,

$$\left. \begin{aligned} \mu_Y &= \sqrt{M} A T g_n(\lambda) \\ \sigma_Y^2 &= \frac{N_0 T}{4} h(\lambda) \end{aligned} \right\} \quad (37)$$

where

$$g_n(\lambda) = \lambda \operatorname{erf}(\sqrt{R_s}(1-2\lambda)) - \frac{1}{8}(1-2\lambda) [\operatorname{erf}(\sqrt{R_s}) - \operatorname{erf}(\sqrt{R_s}(1-2\lambda))] \quad (38)$$

and

$$\begin{aligned} h(\lambda) = & 1 + \frac{R_s}{2} [1 + 3(2\lambda)^2] - \frac{3}{8\pi} \exp^2\{-R_s\} - \frac{3}{8\pi} \exp^2\{-R_s(1-2\lambda)^2\} - \frac{1}{4\pi} \exp\{-R_s\} \exp\{-R_s(1-2\lambda)^2\} \\ & + \frac{\operatorname{erf}^2 \sqrt{R_s}(1-2\lambda)}{16} \{4 - R_s[3(2\lambda)^2 + 10(2\lambda) + 3]\} + \frac{\operatorname{erf}^2 \sqrt{R_s}}{16} \{-4 - R_s[3(2\lambda)^2 - 6(2\lambda) + 3]\} \\ & - \frac{\operatorname{erf} \sqrt{R_s} \operatorname{erf} \sqrt{R_s}(1-2\lambda)}{8} R_s \{(2\lambda)^2 - 2(2\lambda) + 1\} + \frac{1-2\lambda}{8} \sqrt{\frac{R_s}{\pi}} \exp\{-R_s\} \operatorname{erf} \sqrt{R_s} \\ & - \frac{3(1-2\lambda)}{8} \sqrt{\frac{R_s}{\pi}} \exp\{-R_s\} \operatorname{erf} \sqrt{R_s}(1-2\lambda) + \frac{-3+6\lambda}{8} \sqrt{\frac{R_s}{\pi}} \exp\{-R_s(1-2\lambda)^2\} \operatorname{erf} \sqrt{R_s}(1-2\lambda) \\ & + \frac{-3-10\lambda}{8} \sqrt{\frac{R_s}{\pi}} \exp\{-R_s(1-2\lambda)^2\} \operatorname{erf} \sqrt{R_s} \end{aligned} \quad (39)$$

Applying the definitions of the transition probabilities as given by Eq. (5), we find that for large  $M$

where from Eqs. (38) and (39),

$$\frac{\mu_Y}{\sqrt{2} \sigma_Y} = \sqrt{2R_s M} \left[ \frac{g_n(\lambda)}{h(\lambda)} \right] \quad (41)$$

$$\left. \begin{aligned} p_k &\cong \frac{1}{2} \operatorname{erfc} \left[ -\frac{\mu_Y}{\sqrt{2} \sigma_Y} \right] \Bigg|_{\lambda = (\operatorname{sgn} k) \left[ \frac{1}{4N} + \frac{k-1}{2N} \right]} \\ q_k &\cong \frac{1}{2} \operatorname{erfc} \left[ \frac{\mu_Y}{\sqrt{2} \sigma_Y} \right] \Bigg|_{\lambda = (\operatorname{sgn} k) \left[ \frac{1}{4N} + \frac{k-1}{2N} \right]} \\ r_k &\cong 0 \end{aligned} \right\} \quad (40)$$

The reason that  $r_k = 0$  is that as  $M$  gets large, the probability of no transition occurring  $M$  times in succession becomes vanishingly small.

In the limit as  $M \rightarrow \infty$  at finite  $R_s$ ,

$$\left. \begin{aligned} p_k &\rightarrow 1 \\ q_k &\rightarrow 0 \\ r_k &\rightarrow 0 \end{aligned} \right\} \quad (42)$$

and hence from Eq. (10), the steady-state distribution becomes

$$\left. \begin{aligned} P_1 &= 1 \\ P_k &= 0, \quad k = 2, 3, \dots, N \end{aligned} \right\} \quad (43)$$

Of course, in practice one could never let  $M$  approach infinity, and hence the limiting performance obtained under this condition is purely of academic interest.

## 6. Synchronizer Performance

The PFF of  $\lambda$  as given by Eq. (10) together with Eqs. (5) and (30) is plotted in Fig. 3a for a signal-to-noise ratio  $R_s = 3$  and  $N = 8$ . Since this PFF is an even function of  $\lambda$ , we have actually divided the values in Eq. (10) by 2 and plotted the function for both positive and negative  $\lambda$ . The corresponding approximate PFF for large  $M$  (e.g.,  $M = 64$ ) is plotted in Fig. 3b.

Of practical interest is the mean square timing jitter of the bit synchronizer. Denoting this quantity by  $\sigma_s^2$ , we have

$$\sigma_s^2 = \frac{2}{(2N)^2} \sum_{k=1}^N \left( k - \frac{1}{2} \right)^2 P_k \quad (44)$$

In Fig. 4 we plot  $\sigma_s^2$  as a function of signal-to-noise ratio,  $R_s$ , for values of  $M = 1$  and  $M = 64$ .

The error probability performance of this bit synchronizer used in conjunction with a matched filter detector can be obtained by averaging the conditional error probability of the detector for a fixed timing offset over the probability distribution of the phase error; i.e.,

$$P_E = \sum_{k=1}^N P_{E_k} P_k \quad (45)$$

where

$$P_{E_k} = \frac{1}{4} \left\{ \operatorname{erfc}(\sqrt{R_s}) + \operatorname{erfc}(\sqrt{R_s}(1 - 2\lambda)) \right\} \bigg|_{\lambda = \frac{1}{4N} + \frac{k-1}{2N}} \quad (46)$$

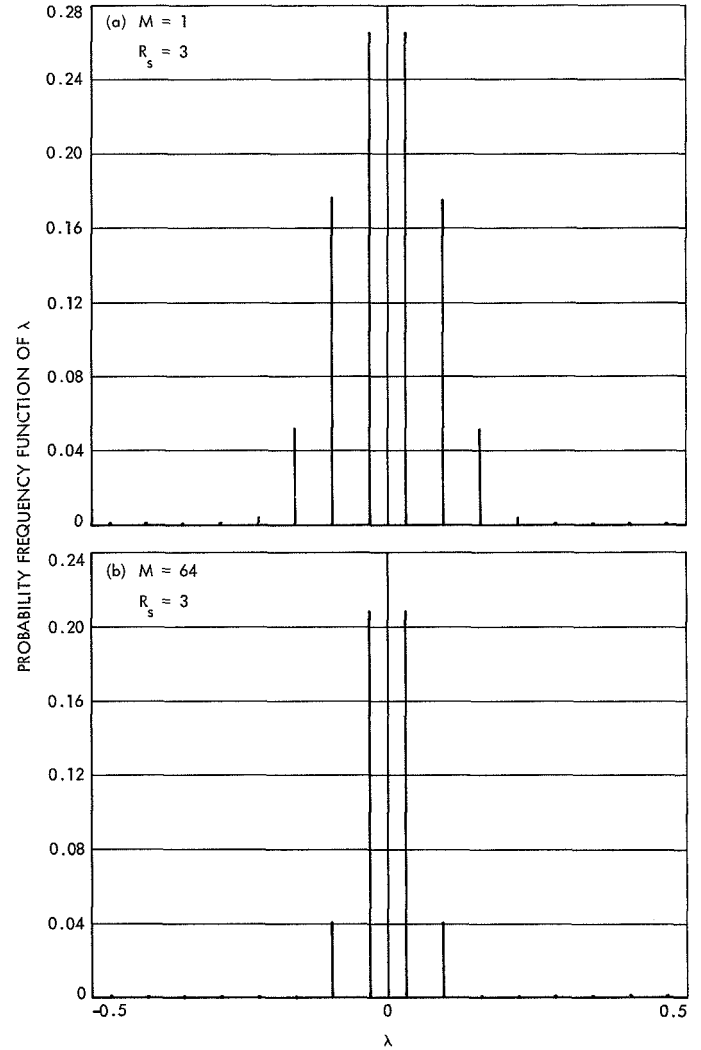


Fig. 3. Probability frequency function of  $\lambda$

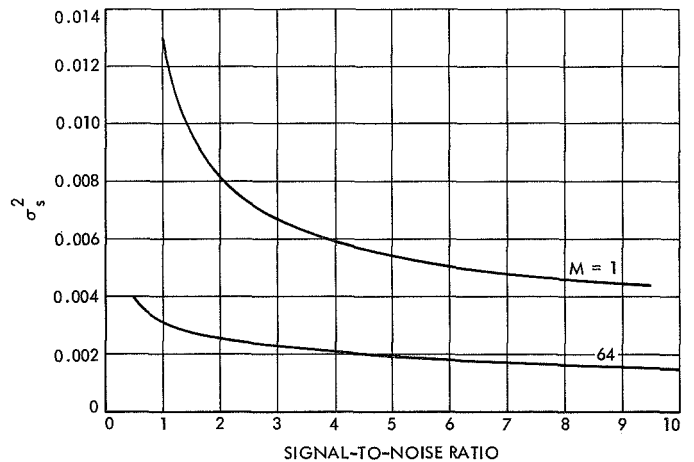


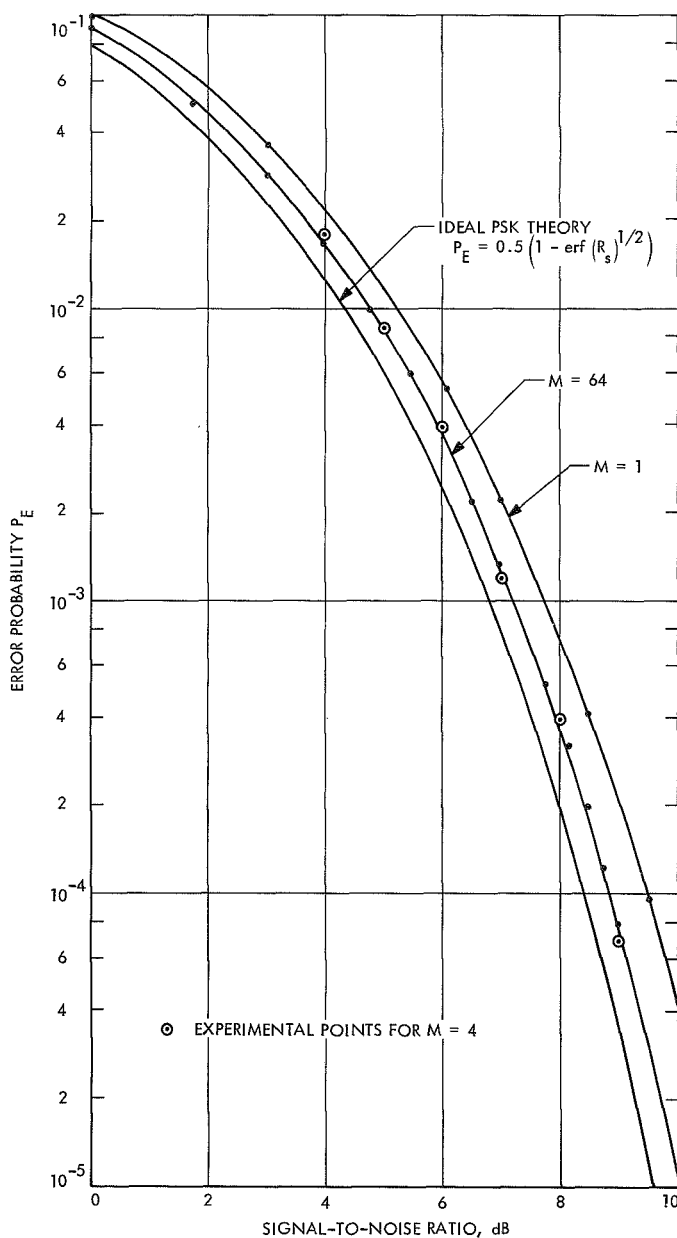
Fig. 4. Mean square timing jitter as a function of signal-to-noise ratio

In Fig. 5 we plot this error probability performance as a function of signal-to-noise ratio in dB for the extreme cases of  $M = 1$  and  $M = 64$ . The error probability performance of an ideal phase-shift-keyed (PSK) system is indicated on this figure for comparison. Also included is a set of experimental points taken for  $M = 4$ . Note that they fall closer to the  $M = 64$  curve than the  $M = 1$  curve, which indicates that rapid improvement with

respect to  $M = 1$  is possible for small  $M$ , and that, practically,  $M$  on the order of 4 or greater is sufficient.

#### Reference

1. Fraser, D. A. S., *Nonparametric Methods in Statistics*, p. 215. John Wiley & Sons, Inc., New York, 1957.



**Fig. 5. Error probability as a function of signal-to-noise ratio**

# D. Optimum Modulation Index for a Data-Aided, Phase-Coherent Communication System, M. K. Simon

In SPS 37-63, Vol. III, pp. 63-66, results were presented for optimally choosing a modulation factor in a data-aided, phase-coherent communication system. It was

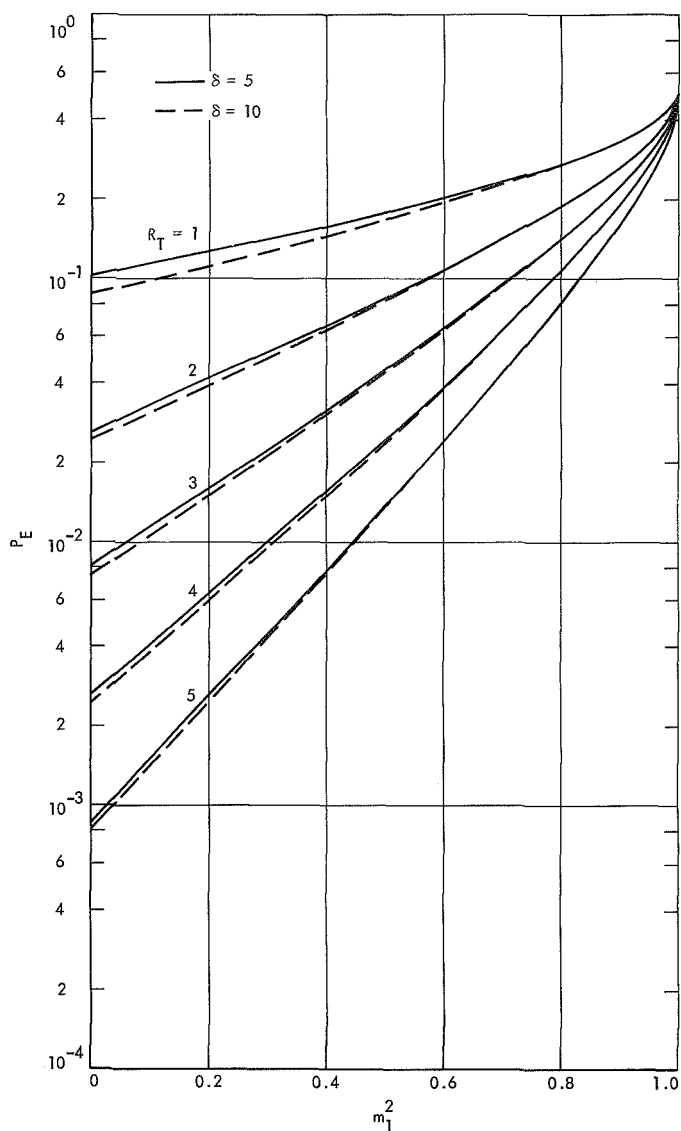


Fig. 1. Bit error probability vs signal modulation factor

shown there that for a fixed total energy-to-noise ratio  $R_T$  and loop bandwidth-symbol time product  $\delta$ , a value of modulation factor  $m_1^2$  does indeed exist that minimizes the data detector's error probability performance. Since the case treated there assumed that the ambiguities in the phase error probability density function were not resolved, the bit error probability took on the value 0.5 at both extreme values of  $m_1^2$  (i.e.,  $m_1^2 = 0$  and  $m_1^2 = 1$ ). Hence, the optimum value of  $m_1^2$  referred to above was for any finite  $R_T$  different from zero (see Fig. 9 on p. 65 of SPS 37-63, Vol. III).

It is clear that regardless of whether the phase ambiguities are resolved, the bit error probability equals 0.5 when  $m_1^2 = 1$  since at that point there exists no power in the data stream and hence the effect of the noisy detection reference bears no effect on performance. At the other extreme, however, one should attempt to resolve the phase error ambiguities since the error probability performance at low values of  $m_1^2$  (where the additional ambiguities are significant) can be greatly improved by doing so. In fact, at  $m_1^2 = 0$ , the probability density function of the phase error,  $\phi$ , has equal humps at  $\phi = 0$  and  $\phi = \pi$ . Thus, our intent here is merely to present the curves (Fig. 1) corresponding to Fig. 9 of SPS 37-63, Vol. III, for the case where the phase error ambiguities are resolved at each value of  $m_1^2$ . We note, as one might intuitively expect, that the best error probability performance under these conditions is achieved for  $m_1^2 = 0$  (i.e., a suppressed-carrier system) regardless of the values of  $R_T$  and  $\delta$ .

One further note is that Eq. (4) on p. 64 of SPS 37-63, Vol. III, should read

$$U_0(\phi) = \frac{4\Omega_0}{N_0 K_u^2 [1 + M^2 \operatorname{erf}^2(R^{1/2})]} \phi + \rho \left\{ [1 + M^2 \operatorname{erf}(R^{1/2}) \operatorname{erf}(R^{1/2} \cos \phi)] \cos \phi + \frac{M^2}{(\pi R)^{1/2}} \operatorname{erf}(R^{1/2}) \exp(-R \cos^2 \phi) \right\}$$

The curves in Fig. 9 are, however, correct since the correct potential function as above was used in obtaining them.

## VIII. Spacecraft Power

### GUIDANCE AND CONTROL DIVISION

#### A. Review of Radioisotope Thermoelectric Generators for Outer Planet Missions, R. S. Caputo

##### 1. Introduction

A decision to use a particular type of radioisotope thermoelectric generator (RTG) for the outer-planet missions must be based on a comprehensive review of the capabilities of all the RTG programs which could be applicable. For the most part, the RTG technology that is available is based on lead telluride (PbTe) thermoelectric material, and the status of readiness varies from preliminary design to flight qualified. The nominal performance is given for each RTG system, and an estimate is made of long-term performance of the more suitable RTG systems. The RTG characteristics which are of greatest interest are specific power (W/kg), and the degree of confidence in acceptable performance for the long mission times associated with outer-planet missions. There are many other factors of interest in evaluating an RTG, but these two are the most critical.

The RTG programs which are reviewed are: SNAP 19 *Nimbus*, SNAP 27 *Apollo*, *Transit* (Isotec), *Pioneer* (SNAP 19-TAGS), SNAP 27 Integral, and the Multi-Hundred Watt (MHW). The performance is derived from flight data or current manufacturer's design esti-

mates. The long-term performance projections are based on the most useful module data with extrapolation to longer time periods.

##### 2. Discussion and Results

The results of a review of current RTG programs is shown in Table 1 where the values for several design parameters are shown for all RTG systems of interest. The generator data is divided into two areas, the converter and heat source. All the systems except the MHW are based on PbTe thermoelectric materials. Of the PbTe systems, only the Isotec design is not sealed to limit PbTe sublimation. The Isotec design reduces the temperature level so that the rate of material sublimation is acceptable. The remaining systems are sealed and have a cover gas to suppress thermoelectric material sublimation.

The two current flight programs, SNAP 19 *Nimbus* and SNAP 27, cannot be directly applied to an outer-planet mission. The SNAP 19 performance is too low ( $< 2.2$  W/kg), and SNAP 27 does not have an integral heat source required for an unmanned system. The *Pioneer* program is heavily based on the *Nimbus* generator, with the significant converter modification being the substitution of the TAGS material for the 2P material for better performance and life characteristics. In

Table 1. RTG performance characteristics

Parameters	Nimbus SNAP 19	Apollo SNAP 27	Transit (Isotec)	Pioneer (SNAP 19- TAGS)	SNAP 27 Integral	MHW <sup>e</sup>
Converter						
Specific power, W/kg						
BOL	2.15	3.57	2.93	3.11	2.82	4.68
EOL	1.81	3.54	2.37	2.55	2.46	3.72
Power, W						
BOL	28.1	71.0	37	40.9 <sup>a</sup>	62.4	155
EOL	23.5	68.8	30	33.5 <sup>b</sup>	54.5	123
Weight, kg	8.5	12.4	5.6	8.3	11.5	15.2
RTG	13.0	19.3	12.6	13.1	22.1	33.0
Mission time, yr	1	1	5	2.5	5	12
Efficiency						
BOL	4.5	4.8	4.62	6.4	4.46	7.05
EOL	3.76	4.65	3.75	5.23	3.83	5.6
Hot-junction temperature, °K	794	850 (day) 816 (night)	673	805	833 <sup>c</sup> 811 <sup>d</sup>	1373
Size diam × length, cm	48.3 × 53.3	39.9 × 45.7	58.4 × 43.2	39.9 × 28.4	30.7 × 46.7	35.3 × 50.8
Heat source						
Heat load, W(t)	625	1480	800	640	1400	2200
Weight, kg	4.51	6.92	6.18	4.93	10.58	17.86
Specific power, W(t)/kg	139	215	129	130	132	123
Operational temperature, °K						
Surface	866	1022	778	866	NA	1450
Structure	1108	1022	1005	1300	NA	NA
Status	Flown	Flown	Prototype	Prototype	Study	Preliminary design
<sup>a</sup> At fueling. <sup>b</sup> 3.25 yr after fueling. <sup>c</sup> For 1-yr mission. <sup>d</sup> For 5-yr mission. <sup>e</sup> SiGe thermoelectric material. NA = not applicable.						
BOL = beginning of life. EOL = end of life. MHW = Multi-hundred Watt						

addition, the heat source is based on the Atomics International (AI) advanced capsule technology. Data is also shown for an integral SNAP 27, i.e., one with an integral heat source. The current reference design for the MHW program is also listed and is the only generator using silicon germanium (SiGe) thermoelectric materials. The evaporation rate for the SiGe materials is considerably less than for PbTe, and it may not be necessary to seal the generator to suppress evaporation even at high temperatures (1373°K) and long missions (10 yr). Other factors, such as prevention of oxidation of multi-foil insulation on the pad, may require the RTG to be sealed initially.

The RTG technologies of interest for further consideration are Isotec, SNAP 19-TAGS, and SNAP 27 Integral as well as the MHW. The long-life performance of each of these systems has been evaluated for possible

use in outer-planet missions; the results of these evaluations are discussed in *Subsection 3*.

### 3. Isotec Technology

The Gulf General Atomic (GGA) Company is developing the Isotec technology for use in the *Transit* program. There has been a significant amount of testing in the Isotec program, including two-couple tests (1900 hr), *p*-element tests (24000 hr), and *n*-element tests (16,000 h). There have been two tests of panels which incorporate current design techniques; these are the PG05, which is a single panel tested for 10,000 h, and GC04, a ten-panel module tested for 5300 h. A prediction of long-life performance is based on these two tests extrapolated to 5 yr. Since the tests are conducted at constant heat input, a correction is made for the effects of isotope decay.



The test data shows an average 6-mo degradation at the rate of 10% per year. The longer test (PG05) shows a reduced rate of degradation of 4% per year during the second 6 mo of testing. Assuming the material effects continue at this rate, the total degradation at 5 yr would be about 23%. The full effects of 5 yr of isotope decay ( $Q/Q_0 = 0.96$ ) could reduce the power to 72% of the original amount ( $P/P_0 = 0.72$ ).<sup>1</sup> Increased thermal resistance of the thermoelectric material during life tends to maintain the beginning of life (BOL) temperature drop across the thermopile. This tends to compensate for the isotope decay effects on power reduction, and the total power loss at 5 yr is estimated to be 75% of the BOL power ( $P/P_0 = 0.75$ ), compared to TRW Systems' estimate of 0.80 and GGA's estimate of 0.86. Using the BOL specific power of 2.93 W/kg, the 5-yr value reduces to 2.20 W/kg. This is shown in Fig. 1 along with estimates for the other systems. Although a 5-yr useful life is usually assumed for PbTe systems, the curves for Isotec, SNAP 19-TAGS, and SNAP 27 Integral are extended to 12 yr (dashed line) for the purpose of comparison with the SiGe performance. The basis for long-life predictions for the remaining systems is shown in the following subsections.

The use of an unsealed PbTe system even at low temperatures for mission times greater than 5 yr is questionable. In fact, the basis for 5-yr performance at anticipated levels is not well documented. Test data from the performance module, assembled with improved process control, will give greater confidence in long-life performance when sufficient test time has been accumulated.

#### 4. SNAP 19-TAGS Technology

Isotopes, a Teledyne subsidiary, is developing the SNAP 19-TAGS technology for the *Pioneer F* and *G* missions. There are many SNAP 19 TAGS-85/2N converters that have been assembled and placed on test by Isotopes (SN 26 to SN 33). Only SN 26 has sufficient test time (1 yr) to be very useful in making a long-life performance predictions. Most of the test time on this converter is at a greater heat load ( $Q = 675$  W) and as a result, a greater hot-junction temperature ( $T_{HJ} = 841^\circ\text{K}$ ) than the *Pioneer* reference design ( $Q = 640$  W and  $T_{HJ} = 805^\circ\text{K}$ ). SN 29 provides some information on the power loss when changing the converter cover gas from argon to helium. This effect is important, since the helium

<sup>1</sup> $Q/Q_0 = (\text{thermal power input at time } t) / (\text{initial thermal power input}).$

$P/P_0 = (\text{electrical power output at time } t) / (\text{initial electrical power output}).$

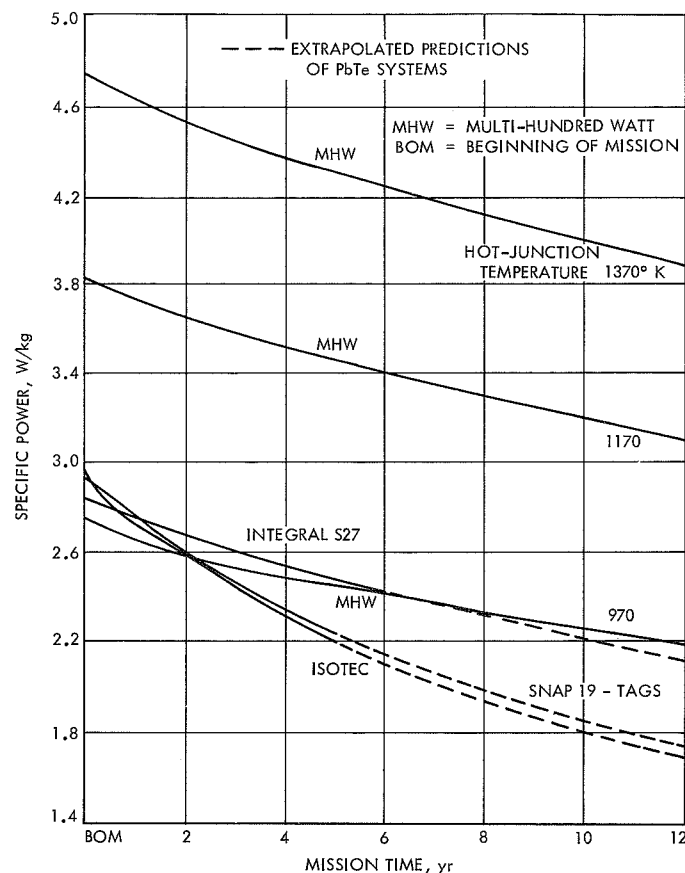


Fig. 1. RTG specific power 12-yr performance prediction

released by this RTG is vented into the thermopile region. Due to helium's high thermal conductivity, compared to the original cover gas, argon, there is a reduction in the temperature drop across the thermopile, which causes a power reduction. Some additional data is available from SN 31 being tested at JPL at  $T_{HJ}$  equal to  $819^\circ\text{K}$ .

Two methods are used to estimate the long-life performance. The first is based on SN 26 test data, the Isotopes analytical model at lower temperatures ( $T_{HJ} < 794^\circ\text{K}$ ), the power loss caused by helium release, and the power reduction due to isotope decay. This approach indicates 71% of the original power remaining 5 yr after launch (5.75 yr after fueling) based on 100% helium release. When considering end-of-mission (EOM) power, a more conservative design approach would be based on a 100% helium release rate. This accounts for about half of the difference between the Isotopes 5-yr estimate ( $P/P_0 = 0.805$ ) and the previous estimate of  $P/P_0 = 0.71$ . The remaining half is due to differences in consideration of material effects.

The second method used to estimate long-life performance is based on the percent material degradation versus temperature relationship from S/N 26, 29 and 31, and the power reduction due to helium release and isotope decay. The temperature versus time and percent degradation versus temperature relationships are integrated to yield a 7.5% power loss due to material degradation over 5 yr. The 100% helium release rate and isotope decay added to the material effects result in a reduction in power to 72% of the original value ( $P/P_0 = 0.72$ ). This compares very well to the first method estimate of 0.71. Based upon these results and the specific power at fueling of 3.11 W/kg, the EOM value is 2.24 W/kg. This curve is also shown in Fig. 1. The beginning of mission (BOM) is 0.75 yr after fueling and some power loss has already occurred to reduce the specific power to 2.93 W/kg at BOM.

## 5. SNAP 27 Technology

There is an extensive amount of test data on the Minnesota Mining and Mfg. Co. (3M) couples used in the SNAP 27 program. There have been twenty-one 10-couple module tests that lasted as long as 3 yr. In addition, there have been fifteen 104-couple modules with test durations of nearly 3 yr. These tests investigated such design parameters as hot-junction temperature, cover gas pressure, and leg spring pressure. Also, complete engineering generators were placed on test, Models 5, 8B, and 10, with a Model 10 test time of 10,000 hr. Model 15 is on test at JPL at a reduced temperature ( $T_{HJ} = 811^\circ\text{K}$ ), which is an operating condition of interest for a long-term mission.

A model of generator performance has been made by the 3M Company based upon the 10-couple module tests. The best agreement between the data and the model exists when  $\gamma$  equals  $-0.4$ ,

where

$$\frac{k}{k_0} = \left( \frac{R_0}{R} \right)^\gamma$$

$k$  = thermal conductivity

$R$  = electrical resistivity

zero subscript = beginning of life value

In addition, over half a year of performance is available from the flight unit used on *Apollo 12*. The power decay rate of this unit is less than 2% per year at an average temperature of  $833^\circ\text{K}$ ; this coincides exactly

with the 3M model. Based on these data and the model, the remaining power would be 89% of the original after 5 yr with the  $T_{HJ} = 811^\circ\text{K}$ , and 85% after 8 yr.

Depending on the mission duration, the design BOM  $T_{HJ}$  would be varied to optimize performance at EOM. For short missions ( $\sim 1$  yr) the  $T_{HJ}$  would be set at  $833^\circ\text{K}$ , while missions of 5 yr or longer would have a  $T_{HJ}$  of  $811^\circ\text{K}$ . This type of relationship is used to show performance over a range of mission times.

The SNAP 27 generator used on *Apollo* did not have an integral heat source that could survive reentry conditions. A *Pioneer* study was used to include an integral heat source in a modified SNAP 27 converter. The heat source design used was similar to the AI technology, except that solid-solution fuel form is used; the specific power of the heat source is 177 W(t)/kg. This is higher than the heat sources being designed for *Pioneer* and *Transit* missions. Although it is normally possible to produce a lighter specific weight heat source as the fuel loading is increased, it is considered more appropriate if the heat source weight would better represent those on the committed missions. Therefore, the heat source specific power was changed from 177 to 132 W(t)/kg, and the heat source weight became 10.6 kg for  $Q = 1400$  W(t). Based on this change, the BOM specific power for the RTG is 2.82 W/kg at  $T_{HJ}$  equal to  $833^\circ\text{K}$ . Based on the long-life effects established earlier, this reduces to 2.46 W/kg at 5 yr for  $T_{HJ}$  equal to  $811^\circ\text{K}$ . This relationship is also shown in Fig. 1.

The long test times on the SNAP 27 program give good confidence for a 5-yr capability, and the substitution of 2N thermoelectric material for the 3N in one leg of the couple would increase power by 15% without any weight increase. The five-year specific power would increase to 2.82 W/kg on this basis.

## 6. MHW Technology

The SiGe Air-Vac couples produced by the Radio Corporation of America (RCA) have a stable long-life capability which has been demonstrated by long-term testing. There have been a great number of couple-hours ( $5 \times 10^6$ ), and module tests (panels) have continued for as long as 49,400 hr with an average test time of 47,600 hr. The 68.8% silicon material has been used for much of this testing at hot-junction temperatures from 1120 to  $1320^\circ\text{K}$ . Therefore, there is a high degree of confidence in performance prediction for 5 yr. The material change phenomenon of dopant precipitation seems

to be well understood and gives credibility to a prediction for an even longer time such as 10 or 12 yr.

Based on a current evaluation of material aging effects and the Resalab, Inc., SiGe conductivity data, SiGe RTGs are optimized for various mission times. Using a 123 W(t)/kg capsule (which is typical of the AI technology as well as the MHW Helipak design), the RTC specific power is shown in Fig. 1. Curves are shown for  $T_{HJ}$  equal to 970, 1170, and 1370°K for the 80% material based on generators with multi-foil insulation. These results can also be used to represent the 63.5% material at lower temperatures, ( $\sim 1270^\circ\text{K}$ ). The 970°K temperature level gives performance comparable to PbTe system performance over lifetimes associated with PbTe, but there is much greater confidence in long-time mission capability for the SiGe MHW.

The  $T_{HJ}$  level that may be used is entirely a function of the limiting temperature of the heat source. The capsules for the *Transit* and *Pioneer* programs are based on the AI technology which has a 5-yr mission design goal of 1366°K for the capsule surface temperature. The  $T_{HJ}$  of a SiGe RTG is expressed as a function of the capsule temperature in Fig. 2 for an AI-type capsule. A band is shown due to possible variations in heat source design. Based on this relationship, the specific power of the SiGe RTG is shown in Fig. 3 as a function of possible capsule temperature. At the 1366°K design goal for the AI capsule, the resulting  $T_{HJ}$  is approximately 1170°K with a three-year mission specific power of 3.36 W/kg, compared to about 2.46 W/kg for the PbTe systems. This design could be accomplished using 63.5% silicon material and the same heat source developed for the *Transit* and *Pioneer* missions.

The MHW program with a Helipak design heat source could extend performance so that an 1370°K  $T_{HJ}$  is possible (with 80% thermoelectric material); the 3-yr mission specific power is nearly 4.5 W/kg. The long-term mission of 12 yr would have performance of greater than 3.81 W/kg. Therefore, using the SiGe generator at temperatures acceptable for the 63.5% material (up to 1270°K) in a neutral environment such as Kaowool or Min-K insulation seems quite satisfactory for a long-term mission.

As the MHW program develops, the specific power estimate may vary somewhat from the values shown in Fig. 1. Also, as of this date, a complete analysis has not been performed on the use of the Atomics International-

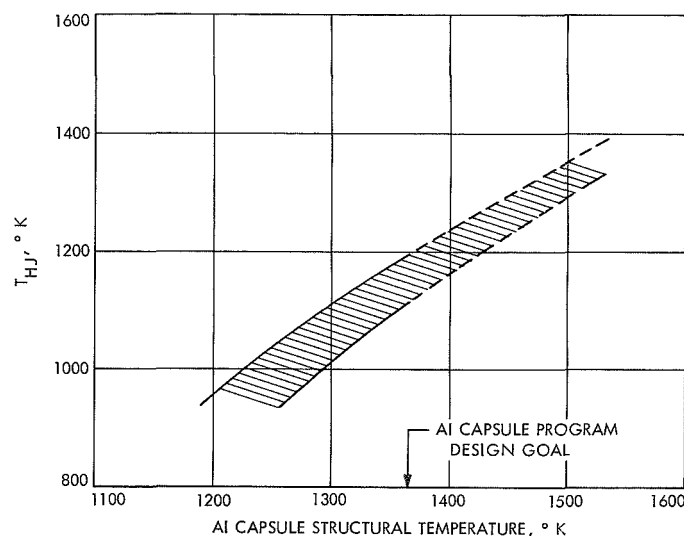


Fig. 2.  $T_{HJ}$  of MHW versus capsule  $T_a$  111 temperature

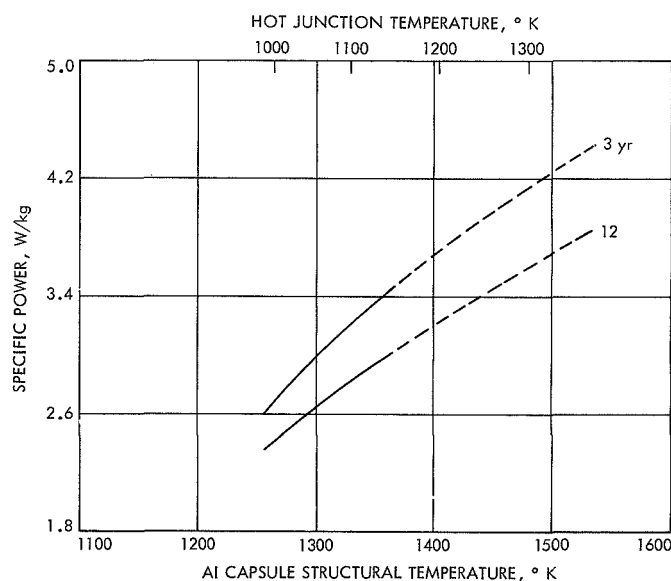


Fig. 3. MHW specific power versus possible capsule temperature

type heat source in the MHW converter. The temperature relationship shown in Figs. 2 and 3 may change somewhat as more design information becomes available. The use of alternate insulation, 80% material at 1370°K, is not well verified at this time. Only as material testing occurs and compatibility with other environments is verified can a long-term mission seriously be based on the 80% material used at 1370°K with foil insulation.

## 7. Conclusions

The SiGe Air-Vac converter has the strongest position of all thermoelectric technologies, especially for long-term missions ( $\gtrsim 3$  yr) based on the following:

- (1) Long-term test time to date with relative confidence in long-term mission performance predictions.
- (2) Basic stability of the SiGe material and predictability of long-term aging effects.
- (3) Ability to accept whatever heat source technology is available for a particular mission.
- (4) High specific power, even with current state-of-the-art heat source design.
- (5) Flexibility of final design adjustments in the converter.
- (6) Possible use of a nonsealed converter or, if a sealed unit is used, a converter that does not require a seal after launch.

## IX. Guidance and Control Analysis and Integration

### GUIDANCE AND CONTROL DIVISION

#### **A. Optimization and Reliability Calculations for Multi-Thermionic-Converter Systems,**

*R. Szejn and K. Shimada*

##### **1. Introduction**

Power sources that make use of nuclear energy are one means of producing the electrical power necessary for deep space probes. Among such sources are thermionic energy converters, which have been shown to offer a large output power with a conversion efficiency as large as 15%. Although the converter technology has reached a point where it is feasible to implement these converters as a power source, the reliability of converters and systems using these devices has not been investigated well enough to arrive at a flyable system.

One method of developing a power supply having a high reliability is to use a large number of moderately reliable converters connected in a matrix configuration. An optimized configuration should be developed for each system, depending upon the system's power output, the component reliability, and the required system reliability.

In this article, approaches to optimization and reliability calculations are presented along with an analysis of one particular system. There are also some general qualitative remarks in regard to optimization and reli-

ability as a function of system configuration, followed by a reliability analysis of a 150-W power system.

##### **2. Approaches to Optimization and Reliability Calculations**

Reliability has two meanings in the field of thermionics, depending on whether converter reliability or system reliability is being referred to. In the former case, a reliability of 0.95, for example, indicates that life tests on similarly designed units have shown 95 out of 100 to be still operating above some given performance level at the end of a life test. In the latter case, one speaks of having a certain system reliability at a specified output power. For example, one might have a system reliability of 0.8 for a relative power output of 0.9. This means that there is an 80% probability that the converter system will be delivering at least 90% of the original undegraded output power at the end of a mission. System reliability will be the topic of concern in this article.

There are three methods for determining a functional dependence between reliability and relative power. One method requires that the relative power output is calculated for every possible combination of failed diodes. This assumes two modes of failure—open or short—with some probability for an open failure  $P_o$  and some probability

for a shorted failure  $P_s$ . (Note that only abrupt failures are considered.)

This method gives the distinct power levels obtainable. Obviously, the number of failure combinations that need be calculated can be reduced if the symmetry involved is taken into account. For example, in a 2-series, 3-parallel cylindrical network, only one calculation is required to determine the power output one to one open failure, while there are six possible locations where the failure can occur. The reliability  $R(W)$  at any one possible relative power output,  $W (\leq 1.0)$ , is given by

$$R(W) = 1 - \sum [\text{all probabilities that yield power degradation larger than } (1 - W)]$$

(Ref. 1)

This method is practical only for a very small system, since the number of distinct cases that must be treated grows quite rapidly with the number of diodes being used.

The second available method for determining reliability employs a Monte Carlo sampling scheme. The approach is to generate a random number between zero and one using a uniform pseudo-random number generator, and then, depending on the assigned converter reliability  $P$ , label the diode as good or bad according to whether the corresponding random number is less than or greater than  $P$ . If the diode is bad, then a second random number is generated to determine whether the failure is to be called an open or short. This simulated converter system is then analyzed to give a relative power value. This process is repeated a large number of times, resulting in a distribution curve from which the reliability curve can be deduced.

The final method also uses a sampling scheme, but in a more restricted sense (Ref. 2). Here the number of failures is specified and the random number generator merely picks the locations to be labeled as failures and at the same time dictates whether they are to be open or shorted failures. In this way, a distribution of relative power values is obtained for a given number of failures. The reliability curve is then determined by plotting the average relative power for  $r$  failures against the probability  $P(r)$  of having  $r$  or less failures:

$$P(r) = \sum_{i=0}^r {}_nC_i P^{n-i} (1-P)^i$$

where  $n$  is the number of converters in the system and  ${}_nC_i = n!/i!(n-i)!$  These are the present methods being used in reliability studies; the next area of concern is optimization.

There are a number of factors to be considered in optimizing a system, but the one which is of interest here is the value of the interconnection resistance in a matrix of converters (Fig. 1). The optimized resistance is a function of the diode configuration and the probabilities of having open and shorted failures. The user will specify whether the system reliability is to be maximized at a particular power level, perhaps the minimum power level usable; or whether the reliability should be maximized at the upper power levels.

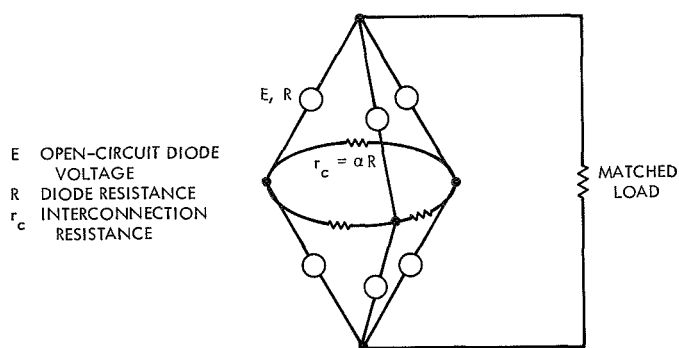


Fig. 1. Converter system configuration

The method for optimizing a converter system uses an iteration technique employing one of the methods mentioned above for determining the reliability as a function of the interconnection resistances. In the case where the optimization is with respect to one failure, an alternate method may be used based on maximizing the quantity:

$$P_s W_s(r_c) + P_o W_o(r_c)$$

where  $P_s$  and  $P_o$  are as before,  $W_s$  and  $W_o$  are the relative powers for short and open failures respectively, and  $r_c$  is the interconnection resistance.

The computer is essential in the above analysis in all but the simplest cases, where no more than six diodes are involved. Programs have been written for generating and solving the loop equations corresponding to flat, cylindrical, and star-shaped matrix systems; for calculating the reliability of a system using either of the approximate schemes mentioned; and for maximizing the system reliability in some limited cases.

### 3. An Analysis of a Particular System

For purposes of elucidating the remarks in Subsection 2, the results from an analysis of a converter system employing two converters in series and three converters in parallel is presented in a series of graphs.

Figure 1 is a simplified drawing of the configuration being considered.  $E$  is the open-circuit voltage of the individual converters,  $R$  is the resistance of a single converter, and  $r_c (= \alpha R)$  is the resistance of the interconnecting resistors. For convenience,  $r_c$  will be expressed as a factor  $\alpha$  of  $R$ , the diode resistance. In all calculations, a matched load is used to maximize the power output.

In Fig. 2, the functional relationship between the maximizing factor  $\alpha$  for the interconnection resistance and the probability of a failure being a short  $P_s$  is shown for the case in which the maximizing resistance is calculated assuming one failure and the case where it is calculated assuming two failures. It is seen that the two curves are very close together, and it was shown that a 0.1 variation in  $\alpha$  produces a 0.001 change in the maximum relative power for either one or two failures. This implies that there is little difference in maximizing for one failure or two failures. Notice that for  $P_s \leq 0.14$ ,  $\alpha = 0$ , which means physically that the interconnecting resistors would be

completely shorted. Likewise, note that for  $P_s \geq 0.64$ ,  $\alpha$  is infinite, meaning that the interconnecting resistors would be completely removed.

Figure 3 shows how system reliability changes as a function of the relative power. This figure also shows how reliability changes with the type of failure that is most probable to occur. These curves are not smooth since there are only discrete power levels available. Also, the reliability for a system with either all open ( $P_s = 0$ ) or all shorted ( $P_s = 1.0$ ) failures is better than that for a system in which both types of failures can occur, since in the latter case the optimization value is a "compromise" between optimizing for just an open failure or just a shorted failure. The effect of having open failures in conjunction with a small number of shorted failures is quite detrimental for the configuration being studied, as Fig. 3 shows. Table 1 presents a more accurate picture of how the reliability varied with the relative power  $W$  and the probability of having shorted failures  $P_s$ .

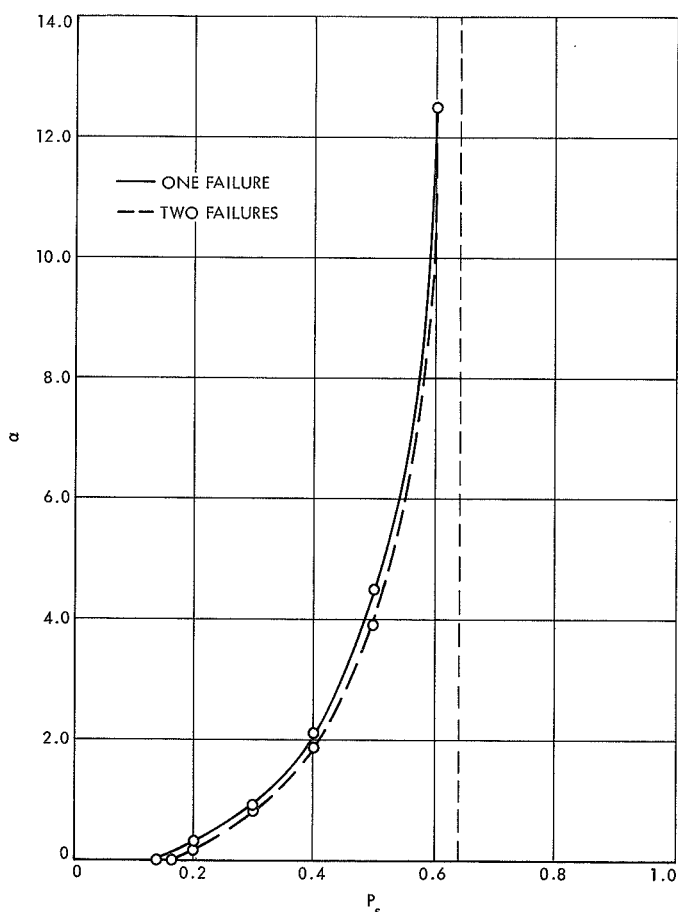


Fig. 2. Maximizing factor for interconnection resistance vs probability of shorted failures

#### 4. System Application

The results of an optimization and reliability analysis, using the techniques presented in Subsection 2, are presented for a generator system having the following specifications:

Terminal voltage, V	3.0
Nominal power, W	150
Contingency factor, %	20
DC-DC conditioner efficiency, %	80
Converter reliability, %	95
Probability of short, %	60
Probability of open, %	40

The components, other than the converters, are assumed to have reliabilities of 100%. One more constraint is that the reliability is maximized in the peak power range, meaning that the system will be optimized for one failure.

A network of five converters in series and three in parallel is selected to be the configuration capable of meeting the above specifications. The diodes in this system must deliver 0.6 V at 25.00 A per converter, producing a gross power output of 225.0 W for this 15-converter system.

The curves for the maximizing factor  $\alpha$  versus the relative power output for one failure occurring in the first, second, and third levels of the five-level configuration are shown in Fig. 4. By weighting the curves obtained for

Table 1. System reliability<sup>a</sup> for values of  $P_s$  and  $W$

$P_s$	$W$								
	0.0	0.333	0.444	0.5	0.6	0.667	0.7	0.75	0.8
0.0	1.0	0.99975	0.99924	0.99732	—	0.98508	—	—	0.96672
0.25	1.0	0.99797	0.99184	0.97807	0.91993	0.91239	—	0.777/0.90896	—
0.50	1.0	0.99700	0.4479/0.99257	0.97730	—	0.97116	0.714/0.96658	—	—
0.75	1.0	0.99509	0.99305	0.98364	0.98328	0.96954	—	0.90883	—
1.00	1.0	—	—	0.99251	0.99165	—	—	0.97717	—

<sup>a</sup>The reliability at a relative power of 1.0 is 0.73509 in all cases.

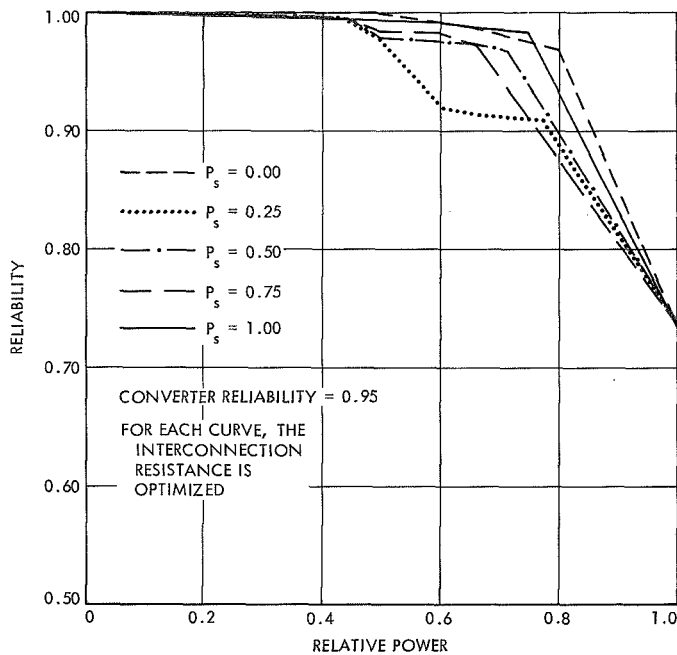


Fig. 3. System reliability vs relative power

each level (as described in *Subsection 2*), three values for the maximizing factor were obtained, one for each level. These values were then weighted according to the probability of a failure occurring in the corresponding level. The result was an optimization factor of 1.90. This factor was then used in conjunction with a diode resistance of  $0.024 \Omega (= 0.6/25.00)$  to obtain a reliability curve. A sampling scheme (as described in *Subsection 2*) was used in calculating the reliability data which appears in Fig. 5. As this figure shows, the reliability at the required nominal power is approximately 0.985.

In summary, it has been shown that a 225-W system is necessary to achieve a nominal power output of 150 W, given a 20% contingency factor and 80% DC-DC conditioner efficiency. The generator system used to produce such a power output employed 15 converters in a 5-series, 3-parallel cylindrical configuration. Assuming a converter

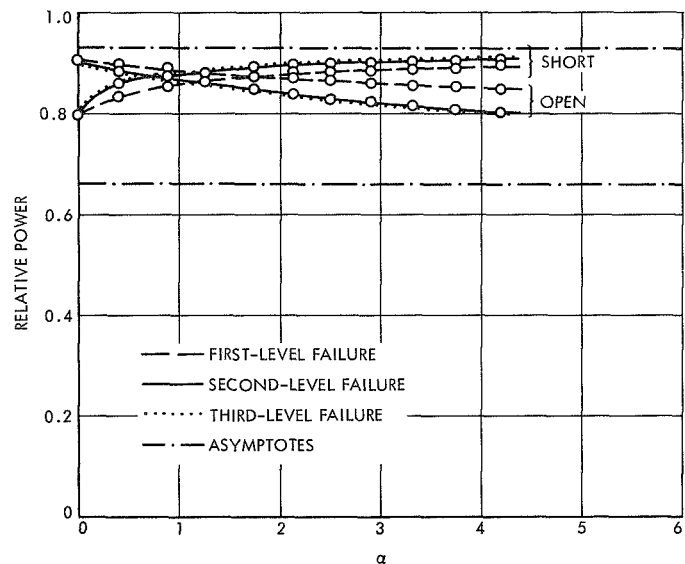


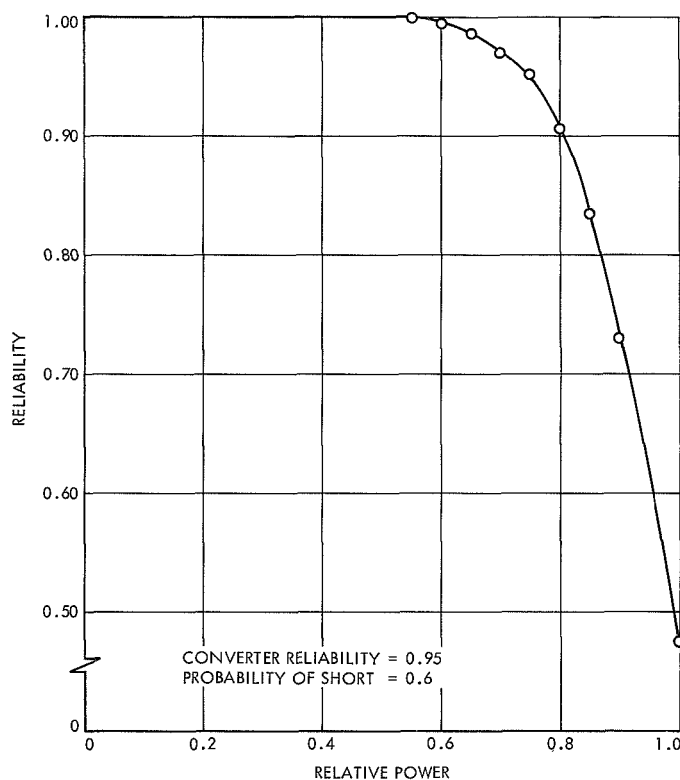
Fig. 4. Maximizing factor for interconnection resistance vs relative power

reliability of 0.95 and a shorted failure probability of 0.6, it was shown that the system reliability at the nominal power output (150 W) is 0.985.

## 5. Conclusions

There are some general conclusions that can be drawn concerning reliability and optimization as functions of system configuration and open-short failure probabilities. Figure 2 described the relationship between the maximizing factor  $\alpha$  and the shorted failure probability  $P_s$  for a 2-series, 3-parallel system. The general shape of the curve ( $\alpha$  monotonically increasing with  $P_s$ ) will hold for all system configurations, but the points at which  $\alpha$  goes to zero and infinity will shift. For systems in which there are a large number of series connections, the curve will shift to the right, while for systems in which there are a large number of parallel connections, the curve will shift to the left. At  $P_s = 0$  and 1.0,  $\alpha = 0$  and  $\infty$ , respectively; this is true regardless of the matrix dimensions.





**Fig. 5. System reliability vs relative power (sampling scheme)**

In general, the most reliable systems have diode configurations that are dependent upon the probabilities of open and shorted failures. Thus, if there is a high probability of having open failures, the system should be designed to have a large number of parallel branches and

vice versa for a high shorted failure probability. It was also noted that the effect of opens in a system optimized for this type of failure is identical to the effect of shorts in a system optimized for shorted failures, if the number of diodes in parallel equals the number in series.

Of the two methods presented for calculating reliabilities, the Monte Carlo method offers the greater generality and flexibility. This method was successfully applied in optimizing the design of a 150-W thermionic generator system. For generators requiring a large number of converters, however, the computation time quickly increases, requiring the development of new computational schemes.

All the calculation schemes presented here are based upon a linearized model for thermionic converters which, in reality, are not linear devices. Modification of the existing method to include nonlinear systems is being considered. Also, experimental work on a nonlinear system will be performed to estimate the accuracy of the reliability calculations for a linear system.

## References

1. Shimada, K., deWinter, F., *Optimization of Thermionic Generator Systems of High Reliability*, presented at the Fifth Space Congress, Cocoa Beach, Florida, Mar. 11-14, 1968.
2. Holland, J. N., Bosseau, D. L., *Thermionic Converter Network Reliability*, General Atomic (John Jay Hopkins Laboratory) Report 6A-6102. Bureau of Ships, Department of the Navy, San Diego, Calif., Mar. 19, 1965.

## B. A Proposed Laser Obstacle Detection Sensor for a Mars Rover, W. L. Kuriger<sup>1</sup>

### 1. Introduction

The exploration of Mars is expected to require the use of an unmanned surface roving vehicle. Because of the large and variable time delay involved in earth-Mars communications, such a vehicle would be required to traverse the Martian surface with minimal external con-

trol; specifically, the vehicle should be able to detect obstacles and avoid them by making small deviations from an otherwise pre-planned trajectory. It is the purpose of this article to analyze the requirements for an obstacle detection system and propose a specific configuration for such a sensor.

The surface of Mars has not yet been observed with anywhere near the resolution necessary to distinguish objects on the scale of meters; thus, possible distributions of obstacles are speculative. The highest resolution pictures of the Martian surface presently available (Refs. 1, 2,

<sup>1</sup>University of Oklahoma (NASA summer faculty fellowship program).

and 3) are those taken by *Mariners VI* and *VII*, and show features down to about 300 m in extent. The Martian terrain can be classified conveniently into four general types: heavily cratered terrain, featureless terrain, chaotic terrain, and polar ice caps. The heavily cratered terrain covers the largest portion of the Martian surface, and is, in many respects, similar to but somewhat smoother than the surface of the moon. The featureless terrain, of which the largest known example is the "desert" *Hellas*, consists of regions which appear absolutely smooth in photographs capable of resolving detail down to about 300 m. The chaotic terrain consists of jumbled short ridges and valleys, and is nearly devoid of craters. This type of surface was first discovered in the *Mariner VI* and *VII* photographs; its extent, causes, etc., are not yet known. The polar ice caps appear to be heavily cratered terrain with added drifts of "snow," probably frozen carbon dioxide. An extensive amount of work has been done with regard to probable obstacle distributions (Refs. 1 and 4), but this data is undergoing continuous revision as the *Mariner VI* and *VII* photographs are analyzed, and will undoubtedly be considerably revised when the much more complete and higher resolution data from *Mariner Mars 1971* becomes available. By the time an exploratory vehicle is landed on Mars, the surface will be continuously mapped to a resolution of a few kilometers, and selected small areas will be photographed with a resolution of perhaps 50 m. At present, it appears that, with the possible exception of the chaotic terrain, the Martian surface will be sufficiently smooth that a roving vehicle will only occasionally encounter an obstacle.

The Martian atmosphere is some two orders of magnitude less dense than that of earth, with a mean surface pressure of about 7 mbar, and is considered to consist of about 80% CO<sub>2</sub> (by mass) with the remainder being largely inert gas, most likely nitrogen (Ref. 5). The Martian atmosphere is probably too tenuous to be a factor in the obstacle sensor design.

Moore (Ref. 6) has performed a preliminary analysis of requirements for a Martian surface navigation capability. He envisions a vehicle equipped with both a short-range obstacle detector, to serve a purely protective function, and a relatively long-range obstacle detector which would permit the vehicle to sense obstacles such as boulders, small craters, and steep slopes in time to change its heading and so avoid them. We consider only the second function.

Present plans are that the exploratory vehicle will be of six-wheeled articulated construction and will traverse the

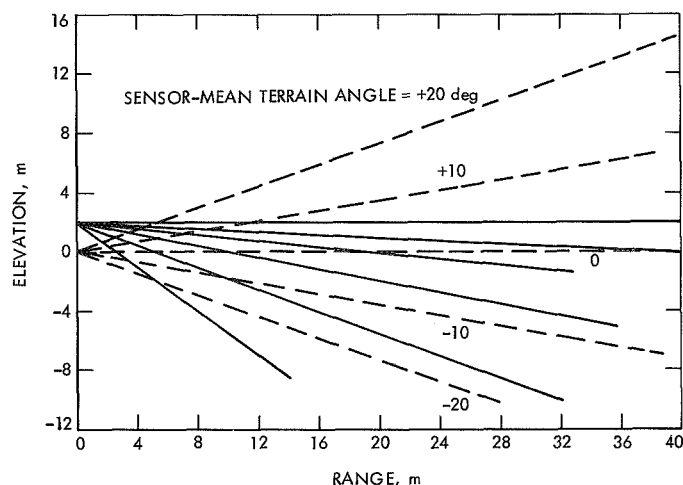
Martian surface at a typical rate of ½ km/h, with a maximum speed of perhaps 1½ km/h. The obstacle detection sensor should mount directly on the vehicle and be capable of detecting obstacles with dimensions as small as 0.5 m. Design constraints include that the sensor should be capable of operating in bright sunlight and, if possible, the sensor should weigh no more than 10 lb, contain no moving parts, and consume at most 10 W of power.

Since the identification of obstacles is greatly facilitated if range information is known, the only form of obstacle detector considered is a rangefinder, specifically one using electromagnetic radiation in the infrared-optical region to meet the resolution requirements and to best utilize the transmission medium and available technology. Optical rangefinders may be active or passive, that is, they may direct a beam of energy to the target or use a natural source such as sunlight, and they may utilize a number of different principles, which include time-delay measurements, triangulation, interferometry, and optical effects such as focussing and parallax in various guises. Several of these approaches are discussed in a report by Biernson, Euling, and Jarnagin (Ref. 7). The weight, power, and sunlight constraints imposed strongly favor an active rangefinder utilizing a laser, with a semiconductor injection laser being a particularly attractive choice because of its small size and weight and its compatibility with miniaturized electronics. The strictures on moving parts, coupled with the necessity to scan a significant area, militate against the use of triangulation or focussing techniques; thus, the sensor configuration which appears to be best adapted to the requirements of the Mars rover application is one utilizing an injection laser in the pulse time-delay measurement or radar mode. With the choice of rangefinder type made, the obstacle detector design divides into three major components: choice of scan configuration and design of the scan mechanism, the rangefinder design, and the devising of methods for interpreting the information obtained by the rangefinder.

## 2. Scan Requirements and Mechanisms

An upper bound to the range at which the rangefinder should operate is determined by the geometry of the situation. If we assume the sensor is located at a height of 2 m above the surface of the terrain, then, at ranges much greater than about 20 m, even gentle surface undulations would serve to mask pits and depressions entirely and would make it difficult to unambiguously discern the presence of a positive obstacle such as a boulder. Let us assume that it is desirable to scan an area extending from a few meters distant out to 20 or 30 m. This requires only

a small scan angle if the vehicle attitude and terrain surface are level, as is shown graphically in Fig. 1, but a considerable increase in scan angle is necessary to compensate for more realistic angles between vehicle attitude and the mean surface in front of the vehicle.

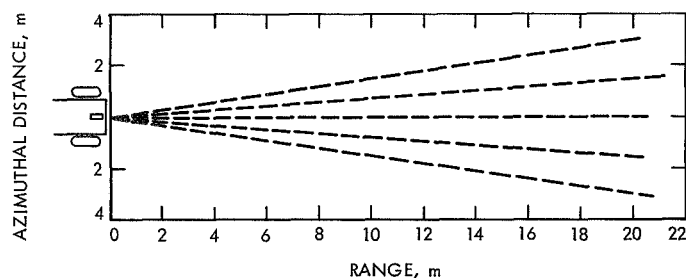


**Fig. 1. Effect of angle between sensor and surface on range for fixed intervals of vertical scan (sensor height assumed to be 2 m)**

A reasonable azimuthal scan requirement is that the observed area be at least as wide as the swath of the roving vehicle at useful elevation scan positions. A wider azimuthal scan would be desirable if large numbers of obstacles were expected, but since this is not the case, alternate vehicle headings can be explored simply by steering the vehicle.

Achieving the desired scan without the luxury of moving parts poses a difficult problem indeed. A large variety of electro-acoustic, electro-optic, and piezoelectric beam deflection techniques exist (Refs. 8 and 9), but none are capable of achieving large scan angles without considerable complexity and perhaps are not at all capable of this with the multimode output characteristic of an injection laser. Additionally, these techniques are best adapted to beams of very small diameter, while both transmitted and received beam diameters must be on the order of tens of millimeters in this application. If the expression no "moving" parts is interpreted to mean no wearing parts, a suitable elevation scan can be achieved by use of a resonant vibrating mirror of the fork or taut band type developed by the American Time Products Division of the Bulova Watch Company, and the much less stringent azimuth requirement can be met by use of a transmitting

array. The specific scheme proposed employs an array of five laser diodes (operated one at a time) properly positioned along the focal plane of a convergent lens to give the azimuth pattern shown in Fig. 2, while the receiver beam is shaped into a fan pattern by use of a cylindrical lens and is not scanned in azimuth. Only one of the lasers in the array is operated during each elevation scan, so five elevation scans are required to completely scan the pattern. As Fig. 2 shows, the use of only five laser diodes at the particular spacing chosen is marginal in that the azimuthal beamwidth does equal the vehicle width (assumed to be 2 m) at ranges greater than 7 m, but the azimuthal resolution is better than 0.5 m only for ranges less than about 7 m. This situation is considerably alleviated by the fact that a particular portion of a scene will be scanned many times from slightly different aspects as the vehicle progresses.



**Fig. 2. Azimuthal beam paths for array of 5 laser diodes (diodes assumed spaced at 6-mm intervals in focal plane of lens with 80-mm focal length)**

A specification for a vibrating mirror which should be within or at the state of the art includes a 25-Hz vibration frequency,  $\pm 12.5$ -deg peak mirror excursion (50-deg total beam deflection), and a 30- $\times$  60-mm mirror size, with the long axis parallel to the axis of rotation. This size mirror permits the transmitter and receiver beams to be side by side; it might also be possible to use two separate vibrating mirrors if they can be synchronized sufficiently closely. The elevation scan is sinusoidal, but can be partially linearized by utilizing only a portion of the mirror excursion. If the mirror attitude is set so that its equilibrium position is looking down from the horizontal by 0.15 rad (8.6 deg) and its usable scan angle is  $\pm 0.25$  rad ( $\pm 14$  deg), then it will scan a range from 5 m out to infinity when the vehicle attitude and surface are level, from 3 m out to about 13 m when the vehicle-mean surface angle is  $-0.25$  rad, and from 13 m out to infinity when the vehicle-mean surface angle is  $+0.25$  rad. Since the vehicle-mean surface angle should be this large for only brief periods, this range of scan angle appears to be adequate.

The beam swept out by the vertical scan moves at an angular rate of 68.5 rad/s in the vicinity of the mirror equilibrium point, so a transmitter pulse repetition rate of 10 kHz means that the terrain is sampled at 0.61-m intervals at the midrange distance of 13.3 m for the level vehicle attitude-level terrain case. The geometry of the situation automatically results in finer resolution being attained at closer ranges—the samples are spaced at approximately 6 cm at a range of 4 m—and less fine resolution at greater ranges, an eminently sensible arrangement. Since only one laser diode is operated on each vertical scan, the entire field is scanned once each 0.2 s, a time which is sufficiently short that vehicle pitching and rolling should only occasionally interfere with a scan. At the nominal condition of level vehicle attitude and level terrain, a vehicle traveling at a velocity of 0.139 m/s (0.5 km/h) will detect a given obstacle 900 times in the 5- to 30-m search range. This redundancy will be somewhat reduced by vehicle pitching, but is still of such a magnitude as to suggest spending some of it in combatting false alarms, that is, requiring that an obstacle be sensed several times in succession before the sensor signals the presence of an obstacle.

### 3. Rangefinder Design

The next topic considered is the design of the rangefinder used in the sensor. Major components of the sensor are shown in block diagram form in Fig. 3. The trans-

mitter portion of the rangefinder consists of an array of five RCA TA7610 gallium-arsenide laser diodes, each equipped with its own driving circuitry. This laser diode has a typical output of 13 W, and the emitting area is on the order of  $3 \times 225 \mu\text{m}$ . The diffraction-limited beam-spread for the smaller laser dimension is on the order of 0.3 rad, and since the target area should be illuminated with a beam of the smallest possible diameter to minimize target-induced pulse spreading and split returns, a collimating lens is necessary. It is difficult to predict precisely the effect of a spherical lens on the laser beam since it is multimode and not circularly symmetric, but a 25-mm  $f/3.2$  simple lens should collect at least 80% of the radiation emitted by any one of the five diodes and should project it onto a spot never larger than 0.1 m in diameter for ranges up to 30 m. The diodes must be spaced at 6-mm intervals along the lens focal plane to achieve the scan coverage of Fig. 2, and individual adjustments of diode orientation will probably be necessary to achieve the desired results. The function of the pulser/timer is to sequentially operate one laser of the array on each elevation sweep and to fire the laser diodes at a 10-kHz rate and with a pulse duration of about 4 ns.

The receiver portion of the rangefinder consists of collecting optics, a photodetector, an amplifier, and a threshold detector. The receiver optics consists of a cylindrical lens followed by a 25-mm-diameter simple lens to produce a receiver beamwidth of 3 mrad in elevation and 0.4 rad in azimuth. This particular beamshape includes in its field of view all the azimuthal area which can be illuminated by the transmitter, but is otherwise as small as possible to reduce the amount of background solar radiation intercepted. The receiver optical system also contains a 50-Å optical filter to further discriminate against background radiation. The optical bandwidth is made as narrow as laser characteristics permit. Room temperature gallium-arsenide injection lasers emit at a nominal wavelength of 9020 Å, with an output wavelength-spread on the order of angstroms caused by multimoding and by heating within the duration of a pulse. There are differences between diodes on the order of tens of angstroms because of material and processing variations, and the output wavelength shifts with diode temperature at a rate of approximately  $1.75 \text{ Å}/^\circ\text{C}$ ; thus, a 50-Å optical bandwidth is a compromise between obtaining good background rejection and requiring tight wavelength control.

Sensor performance is calculated for three different types of detectors—silicon PIN photodiode, avalanche photodiode, and photomultiplier tube—for the case that detection is based on a threshold criterion. Parameter

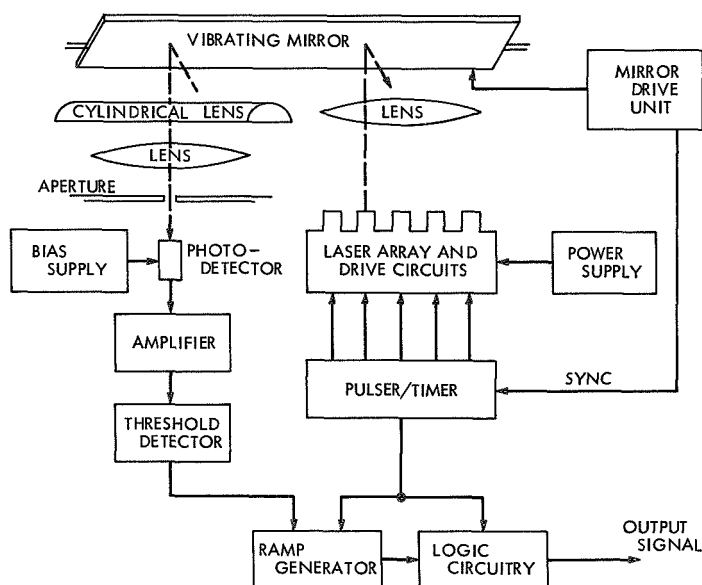


Fig. 3. Obstacle-detecting sensor

**Table 1. Parameter symbols and values used in the calculation of  
Mars roving vehicle rangefinder performance**

Symbol	Parameter	Value
$A_r$	Receiver antenna area	$0.49 \times 10^{-3} \text{ m}^2$
$A_v$	Voltage gain of postdetection amplifier	—
$B$	Optical bandwidth	50 Å
$c$	Velocity of light	$3 \times 10^8 \text{ m/s}$
$e$	Electronic charge	$1.602 \times 10^{-19} \text{ C}$
$\Delta f$	Postdetection bandwidth	250 MHz
$F$	Noise factor of postdetection amplifier	2
$G$	Internal gain of photodetector	1 for PIN diode; $5 \times 10^4$ for photo-multiplier; 20, 100, or 200 for avalanche photodiode
$h$	Planck's constant	$6.625 \times 10^{-34} \text{ J-s}$
$I_{av}$	Mean value of photocathode current	—
$I_d$	Photodetector dark current (at photocathode)	0.15 nA for PIN diode; $3.25 \times 10^{-16} \text{ A}$ for photomultiplier; $10^{-8}/G + 0.5 \times 10^{-12} \text{ A}$ for avalanche photodiode
$k$	Boltzmann's constant	$1.38 \times 10^{-23} \text{ J/}^\circ\text{K}$
$L$	Ground range	—
$m$	Exponent of internal gain for noise power	2 for photomultiplier; 2.3 for avalanche photodiode
$M$	Solar spectral irradiance at 0.9 $\mu\text{m}$	$0.04 \text{ W/m}^2\text{-Å}$
$P_b$	Background power incident on detector photocathode	—
$P_r$	Signal power incident on detector	—
$P_t$	Laser power output (peak)	13 W
$Q, Q_s, Q_n$	Quantum efficiency, photoelectrons/photon	0.5 for PIN diode; $3 \times 10^{-3}$ for photomultiplier; $Q_s = 0.3$ , $Q_n = 0.9$ for avalanche photodiode
$R$	Range	—
$R_L$	Photodetector load resistance	50 $\Omega$
$T$	Absolute temperature	290°K
$T_a$	Martian atmospheric transmission at 0.9 $\mu\text{m}$	1
$V_{sig}$	Signal voltage at output of postdetection amplifier	—
$V_n$	Noise voltage at output of postdetection amplifier	—
$\alpha_r$	Receiver antenna azimuthal beamwidth	0.4 rad

Table 1 (contd)

Symbol	Parameter	Value
$\alpha_v$	Receiver antenna vertical beamwidth	3 mrad
$\Gamma_r$	Efficiency of receiver optics	0.5
$\Gamma_t$	Efficiency of transmitter optics	0.8
$\xi$	Martian surface reflectivity at 0.9 $\mu\text{m}$	0.3
$\Phi$	Angle between direction of sun and surface normal	—
$\lambda$	Operating wavelength	0.902 $\mu\text{m}$

symbols and values used in the calculations are given in Table 1. The choice of postdetection bandwidth is dictated by the requirement of maintaining a sufficiently fast risetime to preserve the pulse leading edge, and the bandwidth requirement in turn places an upper bound on the allowable load resistance. The value of solar spectral irradiance is determined from quantities given in Allen (Ref. 10), while the value for Martian surface reflectivity is from Meisenholder (Ref. 11). The receiver antenna area is selected to be small but comparable to the area required for the transmitter optics, and all other parameter values are dictated by characteristics of available components. The detector specifications are based on a Hewlett-Packard HP 4207 PIN photodiode, a Texas Instruments TIXL55 avalanche photodiode, and an RCA C70102B photomultiplier tube.

Using the symbols and values given in Table 1, the signal power delivered to the detector's photocathode is

$$P_r = \frac{P_t \Gamma_t \Gamma_r \xi T_a A_r}{2\pi R^2} = \frac{0.1219 \times 10^{-3}}{R^2} \quad (1)$$

which is 0.135  $\mu\text{W}$  at a range of 30 m and 4.87  $\mu\text{W}$  at a range of 5 m.

The solar background power at the detector photocathode is given by

$$P_b = \frac{M \alpha_v \alpha_r T_a B \xi \Gamma_r A_r \cos \Phi}{2\pi} \quad (2)$$

and taking the worst-case situation of the sun directly overhead, so  $\Phi = 0$ ,  $P_b$  is found to be 0.028  $\mu\text{W}$  and is independent of range.

If the photodetector is characterized by a quantum efficiency  $Q$ —we must use “fast” and “slow” values  $Q_s$  and  $Q_n$  for the case of an avalanche photodiode—and a dark

current  $I_d$ , the mean value of current leaving the photocathode is

$$I_{av} = I_d + \frac{e\lambda (Q_s P_r + Q_n P_b)}{hc} \quad (3)$$

and the standard deviation is given by  $(2e\Delta f I_{av})^{1/2}$ . The distribution of photoelectrons is Poisson, but since there are many electrons per independent observation interval, the process is adequately modeled by gaussian statistics. In the latter form, the shot noise due to incident signal and background radiation and internal effects may be conveniently combined with the thermal noise of the detector load resistance and the noise contributed by a post-detection amplifier. If the photodetector has internal gain  $G$  and is loaded by a preamplifier with input impedance  $R_L$ , noise factor  $F$ , and voltage gain  $A_v$ , and if the preamplifier passband is set to pass the high frequencies associated with a fast pulse risetime but not the low frequency fluctuations induced by scanning terrain of varying albedo, etc., the rms noise voltage output of the preamplifier is

$$V_n = A_v (G^m R_L^2 2e\Delta f I_{av} + F 4kT\Delta f R_L)^{1/2} \quad (4)$$

The mean value of preamplifier output voltage during a signal pulse is given by

$$V_{sig} = \frac{A_v R_L G Q_s e \lambda P_r}{hc} \quad (5)$$

and the voltage signal-to-noise ratio is obtained by combining the various equations:

$$\frac{V_{sig}}{V_n} = \frac{\left(\frac{e\lambda}{hc}\right) R_L G Q_s P_r}{\left\{ 2e\Delta f G^m R_L^2 \left[ I_d + \frac{e\lambda}{hc} (Q_s P_r + Q_n P_b) \right] 4kT\Delta f R_L F \right\}^{1/2}} \quad (6)$$

**Table 2. Voltage signal-to-noise ratios at the postdetection amplifier output for various types of detectors**

Detector	$\frac{V_{sig} \text{ (20-m range)}}{V_n \text{ (signal off)}}$	$\frac{V_{sig} \text{ (20-m range)}}{V_n \text{ (signal on)}}$	$\frac{V_{sig} \text{ (30-m range)}}{V_n \text{ (signal off)}}$	$\frac{V_{sig} \text{ (30-m range)}}{V_n \text{ (signal on)}}$
PIN photodiode	0.276	0.276	0.123	0.123
Photomultiplier	9.41	2.75	4.18	1.88
Avalanche photodiode				
G = 20	3.29	3.24	1.46	1.45
G = 100	13.25	9.76	5.89	5.02
G = 200	17.05	10.34	7.58	5.71

Numerical values are summarized in Table 2 for ranges of 20 and 30 m for the worst-case situation of the sun directly overhead.

The dominant source of noise in the case of the PIN photodiode detector is thermal noise in the load resistance, and the  $V_{sig}/V_n$  results (Table 2) indicate that an increase of transmitted power or receive antenna area of 2 orders of magnitude would be necessary in order to use this particular device (a signal-to-noise ratio of 10 is the approximate requirement). The photomultiplier's high internal gain makes load thermal noise negligible, but its low quantum efficiency and the relatively large value of background illumination limit its performance. The avalanche photodiode is a semiconductor with properties that fall inbetween a silicon photodiode and a photomultiplier, but it has two peculiarities which degrade its signal-to-noise performance somewhat. One is that all incident photons are not immediately absorbed, and as a result the device has a substantially lower value of ac quantum efficiency than dc quantum efficiency. The other undesirable characteristic is that the avalanche internal gain mechanism is not noiseless; this is accounted for in the device model by the exponent of the gain being different for noise power than for signal power. This latter characteristic causes an optimum value of gain which maximizes the signal-to-noise ratio for any particular set of circumstances; the optimum value of gain for the particular set of parameter values used in the calculations is beyond the "typical" attainable value of 200 but does not result in performance significantly better than that achieved with a gain of 200. The results given in Table 2 indicate that an avalanche photodiode operated at a gain on the order of 100 or 200 yields an adequate signal-to-noise ratio at a range of 20 m and a marginal one at 30 m; thus, this is the device that would be incorporated into the sensor.

For the case that signal detection is based on a threshold criterion, the signal-to-noise ratio necessary in a par-

ticular application is determined by the need to set a threshold voltage level which is sufficiently high that it is only rarely exceeded by noise peaks in the absence of a pulse but not so high that it fails to detect signal pulses. To calculate the threshold level necessary to ensure that false alarms occur no more than, say, once a month, we observe that there are on the order of 60 signal pulses/line scan, 300 pulses/field, 1,500 pulses/s, and  $2 \times 10^9$  pulses/mo (assuming 12 h of operation per day). From sampling theory, independent voltage samples occur every 2 ns for a 250-MHz bandwidth, and if range gating is employed (this would occur automatically if the ramp scheme to analog-process the range data is used) only to consider times for a pulse to return from up to a 30-m range (200 ns), there are 100 independent observation intervals/pulse, and on the average half of these would occur before receipt of a signal pulse. Thus, a false alarm rate of 1 pulse/mo requires that the probability of the signal-absent noise exceeding the threshold be about  $10^{-11}$ , but this number is not particularly critical because of the rapid exponential decline of the wings of gaussian functions. If we define a missed pulse as one which is not detected in one observation interval of 2 ns, then the probability of a missed pulse is equal to the probability that the signal-present noise exceeds the difference between the mean signal level and the threshold voltage. Since the sensor might detect a given obstacle as many as 900 times as the vehicle progresses, and since this calculation will result in a threshold which will be increasingly exceeded by the signal level as the range to the obstacle is reduced, a reasonable requirement is that a missed pulse should occur no oftener than once a minute, which in turn requires that the probability of the signal-present noise exceeding the mean signal-threshold differences be about  $10^{-5}$ . From tables of gaussian probabilities, we can write

$$\left. \begin{aligned} V_T &\geq 6.71 V_n \text{ (signal off)} \\ V_{sig} - V_T &\geq 4.26 V_n \text{ (signal on)} \end{aligned} \right\} \quad (7)$$

where  $V_T$  signifies the threshold voltage. Combining these equations yields an equation that constrains possible values of signal-to-noise ratios.

$$\frac{6.71}{\frac{V_{sig}}{V_n(\text{signal off})}} + \frac{4.26}{\frac{V_{sig}}{V_n(\text{signal on})}} \leq 1 \quad (8)$$

A similar analysis shows that the signal-to-noise requirements can be relaxed if it is required to sense the presence of an obstacle a number of times before triggering an output. If it is required to detect an obstacle on four successive frames, the equation that constrains possible values of signal-to-noise ratios becomes

$$\frac{2.91}{\frac{V_{sig}}{V_n(\text{signal off})}} + \frac{4.55}{\frac{V_{sig}}{V_n(\text{signal on})}} \leq 1 \quad (9)$$

Equations (8) and (9) are plotted in Fig. 4. Note that the requirement of a multiple indication of an obstacle not only eases the performance requirements but also greatly reduces the probability of false alarms caused by anomalies not considered in this analysis.

The ranging accuracy of the proposed rangefinder is difficult to calculate, but can be estimated from experi-

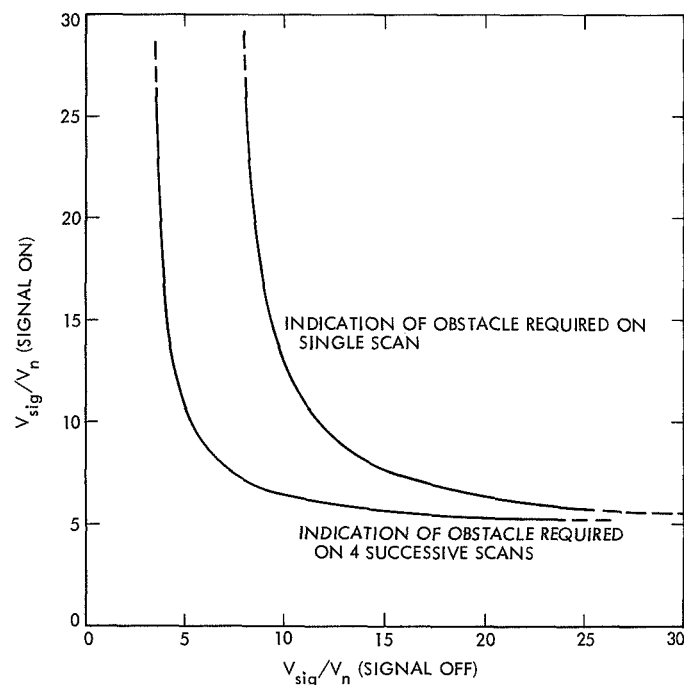


Fig. 4. Curves relating minimum required signal-to-noise ratios for proper sensor operation

mental results on other systems. A laser profilometer reported by Hopson (Ref. 12) uses the same type of threshold detection system considered here with about the same overall pulse risetime, and experimental results for that system indicate a range accuracy of  $\pm 10$  cm.

#### 4. Interpretation of Rangefinder Returns

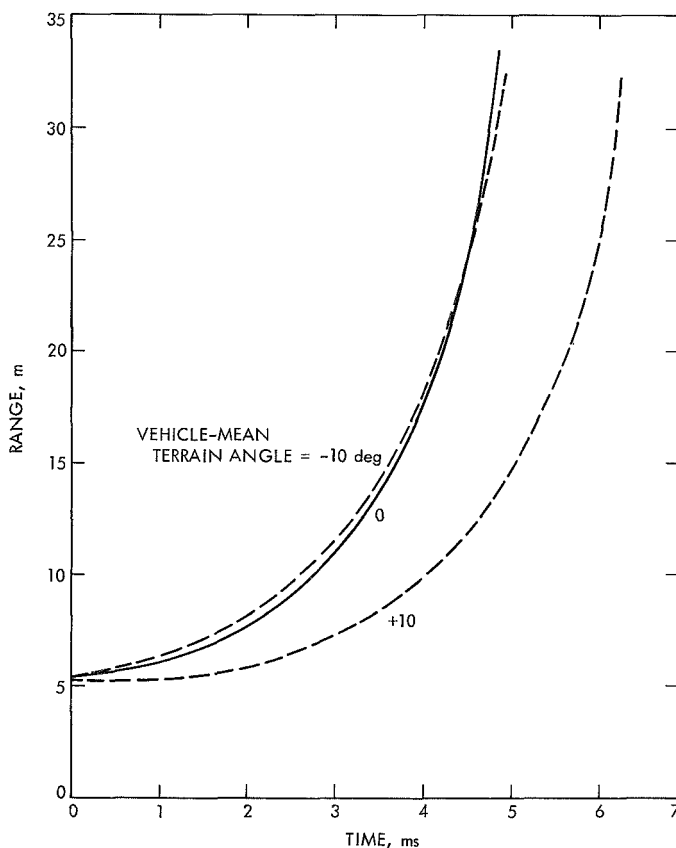
The interpretation of returned signals to determine whether or not an obstacle is present can be accomplished in a variety of ways. One possible method is to use a fast timer to measure the elapsed time between the emission of a pulse and the detection of the return. This information is then fed to the vehicle computer along with vehicle attitude information. The angular coordinates of each pulse are known from timing considerations, so the computer can be programmed to distinguish obstacles to any desired degree consistent with the sensor resolution. Another possible method, which might not require the use of the vehicle computer at all, is to convert each elevation scan into an analog waveform containing the desired information. If a fast ramp is initiated when a pulse is transmitted and terminated upon receipt of a return signal or the passage of 200 ns (the time for a signal to travel 30 m and return), whichever comes first, the range information will be preserved as a sequence of equally spaced pulses whose amplitudes are proportional to range. In the case of level terrain, the pulse train for each individual sweep would increase approximately quadratically in amplitude with time on a scale determined by the angle between vehicle and terrain. The presence of a positive obstacle, such as a boulder, would be manifested by a smaller than normal or even negative increase in amplitude between successive pulses, while a larger than expected increase would signal the presence of a negative obstacle, such as a fresh crater. The technique of comparing each pulse's amplitude with that of its immediate predecessor ignores nearly all the information in the return but looks specifically for the most characteristic indication of an obstacle—the presence of a discontinuity in range as a line-of-sight is swept. Unfortunately, it does not appear feasible to detect obstacles on the basis of pulse-to-pulse differences only. As shown in Fig. 5, the angle between vehicle and mean terrain is a significant factor. The apparent asymmetry between the three curves of Fig. 5 is caused by the nonlinearity of the sinusoidal scan compensating the geometric range nonlinearity in one case and complementing it in the other. Range differences between successive pulses vary considerably, as summarized in Table 3.

An adequate test can be mechanized using the range at each point within an elevation sweep. Dividing the



**Table 3. Range differences between successive pulses at several ranges and vehicle-mean terrain angles for the proposed sensor operating on flat terrain**

Nominal ground range, m	$\Delta R$ , m		
	Vehicle-mean terrain angle = -10 deg	Vehicle-mean terrain angle = 0 deg	Vehicle-mean terrain angle = +10 deg
5	0.090	0.077	0.035
10	0.327	0.359	0.289
20	1.178	1.394	1.300
30	2.558	3.192	2.964



**Fig. 5. Envelopes of range vs time for 3 different vehicle-mean terrain angles (range is distance from 2-m high sensor to flat surface; time is measured from instant that 5-m ground range (5.385-m slant range) is achieved)**

range difference between successive pulses by the range squared ( $\Delta R/R^2$ ) results in a number which is approximately constant for flat terrain, and can thus form the basis for generating a detection criterion. For proper interpretation of the rangefinder returns, it is necessary that the angle between the vehicle and mean terrain be within  $\pm 10$  deg for each sweep. It is assumed that a 0.5-m step is an obstacle, while a 0.2-m step is not.

Because of the geometry, positive (step up) and negative (step down) obstacles may best be discerned by different criteria. A distinctive feature of a positive obstacle is that the range increase is small or zero for several successive pulses, as the scan travels from the bottom of the obstacle to the top. Thus, one method for determining the presence of a positive obstacle is to count the number of pulses for which  $\Delta R/R^2$  falls below a fixed threshold. The return of greater than some predetermined number of successive pulses (e.g., 3 or 4) for which  $\Delta R/R^2$  was below threshold would signal an obstacle. As an indication of the sensitivity of this criterion, the number of successive pulses necessary for the sensor to "see" the top of positive 0.2- and 0.5-m step obstacles is given in Table 4. The figures indicate that this particular criterion fails at the 5-m range, because it cannot distinguish a 0.2-m step from a 0.5-m step over the complete range of vehicle-mean terrain angles, but can be used at greater ranges. The data of Table 4 show that it would be desirable to make the number of pulses required to signal an obstacle dependent on the range.

**Table 4. Pulses required for sensor to scan from bottom to top of 0.2- and 0.5-m positive step obstacles at various ranges and vehicle-mean terrain angles**

Nominal ground range $L$ , m	Vehicle-mean terrain angle, deg	Number of pulses	
		Positive 0.2-m step at range $L$	Positive 0.5-m step at range $L$
5	+10	15	37
	0	7	17
	-10	6	14
10	+10	4	9
	0	3	8
	-10	4	8
20	+10	2	4
	0	2	4
	-10	2	5
30	+10	1	3
	0	1	3
	-10	2	4

On the other hand, we suggest that the criterion for detection of a negative obstacle be the occurrence of a value of  $\Delta R/R$  larger than a predetermined threshold. Table 5 gives values of the range difference between successive pulses divided by the range,  $\Delta R/R$ , and indicates that this criterion is adequate to detect negative obstacles if the threshold is set at approximately 0.180. The use of the quantity  $\Delta R/R$  rather than  $\Delta R/R^2$  yields a criterion that is more independent of range and terrain angle.

**Table 5. Range differences between successive pulses divided by range for 0.2- and 0.5-m negative step obstacles for various ranges and vehicle-mean terrain angles**

Nominal ground range $L$ , m	Vehicle-mean terrain angle, deg	$\Delta R/R$		
		No obstacle	Negative 0.2-m step at range $L$	Negative 0.5-m step at range $L$
5	+10	0.006	0.097	0.205
	0	0.014	0.104	0.211
	-10	0.017	0.106	0.213
10	+10	0.028	0.117	0.223
	0	0.034	0.122	0.227
	-10	0.032	0.120	0.226
20	+10	0.063	0.148	0.250
	0	0.066	0.151	0.253
	-10	0.057	0.143	0.246
30	+10	0.095	0.177	0.276
	0	0.098	0.180	0.279
	-10	0.080	0.164	0.264

Thus, it appears entirely feasible to discern the presence of obstacles by relatively simple processing, but additional effort would be required to thoroughly evaluate this aspect of the problem.

## 5. Conclusions

The proposed sensor contains relatively small, light, and few optical components and the electronic components are all of the miniature solid-state type; thus, the weight of the overall sensor is primarily that of the mechanical packaging. It appears entirely feasible to hold the overall weight to 10 lb or less. If we assume a power efficiency of 0.1% for the laser diodes, a pulse duration of 4 ns, and a pulse repetition rate of 10 kHz half of the time (no transmission during vibrating-mirror scan retrace), the power required by the laser diodes is about 0.26 W. The vibrating-mirror drive power is on the order of 10 mW; thus, the power requirements are primarily those of the auxiliary electronics, and an overall power requirement of a few watts should be attainable. It

appears feasible to discern whether or not the rangefinder returns indicate the presence of an obstacle by use of relatively simple signal-processing techniques, thus allowing the sensor to be a self-contained unit.

## References

1. *Mariner-Mars 1969: A Preliminary Report*, NASA SP-225. National Aeronautics and Space Administration, Washington, D.C.
2. Leighton, R. B., *The Surface of Mars*, pp. 26-41. Scientific-American, New York, May 1970.
3. Leighton, R. B., et al, "Mariner 6 and 7 Television Pictures: Preliminary Analysis," *Science*, Vol. 166, pp. 49-67, Oct. 3, 1969.
4. *Mars Surface Models (1968)*, NASA SP-8020. National Aeronautics and Space Administration, Washington, D.C., May 1969.
5. *Models of Mars Atmosphere (1967)*, NASA SP-8010. National Aeronautics and Space Administration, Washington, D.C., May 1968.
6. Moore, J. W., *Surface Navigation on Mars*, paper presented at the National Space Meeting of the Institute of Navigation, Moffett Field, Calif., Feb. 17-19, 1970.
7. Biernson, G. A., Euling, R., and Jarnagin, W. S., *Range Sensor Techniques for Automatic Object Identification*, Final Report AFRL Contract F 19628-69-C-0071. Air Force Cambridge Research Laboratory, Bedford, Mass., Nov. 1969.
8. Fowler, V. J., and Schlafer, J., "A Survey of Laser Beam Deflection Techniques," *Proc. IEEE*, Vol. 54, pp. 1437-1444, Oct. 1966.
9. W. Kulcke, et al, "Digital Light Deflectors," *Proc. IEEE*, Vol. 54, pp. 1419-1429, Oct. 1966.
10. Allen, C. W., *Astrophysical Quantities*, pp. 140 and 155. University of London Athlone Press, London, England, 1955.
11. Meisenholder, G. W., *Planet Illuminance*, Technical Report 32-361. Jet Propulsion Laboratory, Pasadena, Calif., 1962.
12. Hopson, J. E., *Low Altitude Laser Measurement System*, Technical Report AFFDL-TR-67-176. Air Force Flight Dynamics Laboratory, Wright-Patterson AFB, Ohio, Dec. 1967.

## C. Automated Test Techniques for Guidance and Control Subsystems, R. E. Williamson

### 1. Introduction

The long-range objective of this effort is the development of guidance and control (G&C) support equipment technology to meet the requirements of future planetary missions. The specific near-term goals, which are directed toward the support of the thermoelectric outer planet spacecraft (TOPS) project, include the following: (1) development of test requirements for the TOPS G&C subsystems, (2) development of automated test techniques to implement the TOPS requirements, and (3) demonstration of the test techniques with a generalized support equipment approach centered around a small general-purpose computer.

### 2. Test Requirements for TOPS G&C Subsystems

One of the major tasks for this fiscal year is the development of a baseline test requirements document for the TOPS G&C subsystems (electrical power, attitude control, scan, and approach guidance). The objectives of this effort are twofold: (1) to assure that a testing capability is implemented in the subsystem designs, and (2) to identify the types of test functions which must be implemented in future support equipment designs.

The outer planet missions require advanced design approaches for the G&C subsystems which result in new elements to be tested and new complexities in testing requirements. Following are examples of some of the new designs planned for the G&C subsystems.

*Approach guidance.* This is a new subsystem required for navigation to the outer planets. Although the test technology will be similar to that required for the spacecraft science vidicon and imaging systems, there will be unique test requirements for the approach guidance subsystem.

*Attitude control and scan.* The attitude control subsystem is completely different from those previously employed due to increased accuracy and reliability requirements. Some of the new elements include: digital inertial and celestial sensors, digital attitude control electronics, hydrazine thrusters for coarse attitude propulsion, reaction wheels or pulsed plasma jets for fine attitude corrections, and multi-redundant circuits and components.

*Electrical power.* The radioisotope thermoelectric generators (RTGs) planned as the power source for the outer planets missions represent new elements in checkout. Associated with the RTGs is a new design for the power

conditioning equipment. As with other spacecraft subsystems, considerable redundancy is required to achieve the long-life requirements.

*Spacecraft computer interfaces.* The G&C subsystems designs include a strong dependency upon the spacecraft computer. In addition to the normal functions of subsystem mode control, the computer is used to detect failures in the subsystems, confirm them, and take corrective action to reestablish a desired functional state whenever possible. Simulation of the computer is required when testing the subsystems at the subassembly and subsystem level.

### 3. Automated Test Techniques

A second major task for this fiscal year is the development of test techniques to implement the baseline test requirements. Automated techniques, using a computer-oriented test system approach, are being planned for most tests. Development of the techniques will include preparation of functional specifications for: (1) the software routines and algorithms required to perform the test functions, (2) the hardware required to interface the spacecraft subsystem and the computerized test set, and (3) any special-purpose hardware required with the test set, such as signal sources.

The reasons for the emphasis upon an automated test system are as follows:

*Spacecraft complexity.* A comparison of the complexity of the TOPS spacecraft with the *Mariner* Mars 1971 spacecraft shows a small increase in the number of actual parts. However, due to the use of large-scale integration on TOPS, the equivalent number of parts is greatly increased (estimated equivalent parts on TOPS = 900k, *Mariner* Mars 1971 = 100k). Associated with the parts increase are expansions of the spacecraft functions with the consequent escalation in testing requirements. One significant area of increased complexity is in the expansion of spacecraft modes associated with the implementation of multi-redundancy throughout the subsystems. The amount of data to be handled is such that automation is required to assure thorough system validation.

*Technical requirements.* Simulation of the spacecraft computer during subassembly and subsystem testing will require an automated test set which can operate synchronously with the spacecraft subsystem. Execution of test sequences which have time limitations also will call for automated testing. The increased use of digital components on the TOPS spacecraft provides an advantage for

the automated approach in that the digital circuits readily adapt to the interfaces of a computerized test system.

*Test reliability.* Reliability considerations for the long duration outer-planet missions will have a significant influence upon testing operations. Control of the test environment and of all pre-launch spacecraft operations will become more important than has been necessary for present spacecraft. With automated testing, the capability exists to provide greater control of testing and to make tests less subject to human errors. Test repeatability is an example where improved test control is possible with automated techniques. One of the weakest characteristics of manual testing, yet one of the strongest points of automated testing, is the ability to repetitively perform tests in precisely the same way. Identification of potential failures through trend analysis requires test repeatability.

#### 4. Computer-Oriented Test System

A computer-oriented test system (CTS) will be used to develop and demonstrate the automated test techniques required to support the TOPS G&C subsystems. During this fiscal year, a functional specification will be prepared for the system design—including software and hardware—of the overall automated test set. Software development will be started using a leased computer, a Honeywell DDP-124, which is shared with the Astrionics Division.

Following is a discussion of past CTS technology development and future plans.

*Past CTS technology development.* The initial emphasis in developing a CTS technology was directed towards testing spacecraft computer subsystems. To demonstrate the basic capabilities of an automated test system, a prototype CTS was assembled and interfaced with a *Mariner* Mars 1969 central computer and sequencer (CC&S). The prototype system was used to develop and demonstrate the following:

- (1) Capability for automatic control and verification of the performance of the test hardware in a real-time environment. The test system receives, time tags,

decodes, and processes all test data and prints all changes and errors. Test command sequences are output at pre-determined times. In the automatic mode, the test data is compared with reference data generated by a software simulator which runs concurrently with the CC&S.

- (2) Capability for operational flexibility and operator interaction. The necessary controls, displays, and software are provided such that the operator can apply stimuli and verify the hardware performance in a manual mode. New tests can be set up from operator instructions for entry into the test library.

In December 1969, CTS development for flight computer subsystems, including responsibility for the prototype CTS, was transferred to the Astrionics Division. The Guidance and Control Division retained responsibility for development of the CTS technology required to support G&C spacecraft subsystems. The technology base developed with the prototype CTS is being utilized in applying the automated testing approach to the more complex TOPS G&C subsystems.

*Future CTS technology development plans.* It is currently planned to assemble a CTS in fiscal year 1972 to demonstrate the automated test techniques required for the TOPS G&C subsystems with breadboards or simulations of the TOPS subsystems. In addition to developing the test techniques, consideration will be given to expanding the current CTS technology in the following areas:

- (1) *Operational flexibility.* The same basic test system design is to be adaptable to testing all G&C subsystems, and adaptable to changes in the flight hardware and test procedures. The test system developed for the outer planet missions is to have the capability for reuse on follow-on missions.
- (2) *Man-machine interaction.* Proper data presentation is required to permit human review and judgment. Operator control of the test sequences will be necessary with safeguards to prevent damage to the spacecraft. Hardcopy documentation of all test anomalies and all operator actions must be included.

## X. Spacecraft Control

### GUIDANCE AND CONTROL DIVISION

#### A. The Attitude Control of a Flexible Solar Electric Spacecraft, E. L. Marsh

##### 1. Introduction

In keeping with recent advances in solar-electric-powered spacecraft design, a study of the attitude control of such a vehicle during its thrust phase was carried out. The major result of the study was that the influence of large flexible solar arrays was not detrimental to controlling the attitude. It was also determined that "tail-wags-dog" effects on attitude control were minor.

The nonlinear equations of motion for a particular version of a solar-electric-powered spacecraft were detailed in SPS 37-63, Vol. III, pp. 115-122. Figure 1 shows the model for which the equations are valid.  $B_1$  is the rigid spacecraft bus,  $A$  is a set of two flexible roll out arrays, and  $E$  is the electric engine cluster. The vehicle is designed so that the cluster can translate parallel to either  $b_1$  or  $b_2$ , and so that the two outer engines  $B_3$  and  $B_4$  can gimbal about the diagonal along which they lie. Equations of motion for the entire vehicle, the appendage  $A$ , the gimballed engines, the translating masses, and the motors that produce gimbaling or translating appear in SPS 37-63, Vol. III, pp. 115-122, in Eqs. (2), (5), and (8)-(15). The hybrid coordinates procedure of Likins (Ref. 1) was utilized to account for the dynamics of the flexible solar arrays. A solution of the nonlinear equations of motion is presented in Eq. (16) of SPS 37-63, Vol. III, pp. 115-122.

##### 2. Linear Equations of Motion

The linear variational equations of motion for the system are derived in Ref. 2. For the composite vehicle, the equations are

$$\left. \begin{aligned} I_1 \ddot{\theta}_1 + \bar{\kappa}_3 r \ddot{\alpha}_6 - \delta_{51} \ddot{\eta}_5 &= 3 \frac{(\mathcal{M} - \bar{m})}{\mathcal{M}} r F \alpha_6 - \frac{3F \Delta_{42}}{\mathcal{M}} \eta_4 \\ I_2 \ddot{\theta}_2 - \kappa_3 r \ddot{\alpha}_5 - \delta_{12} \ddot{\eta}_1 &= 3 \frac{(\mathcal{M} - m)}{\mathcal{M}} r F \alpha_5 \\ I_3 \ddot{\theta}_3 + (\kappa_2 r + I_3^5) \ddot{\alpha}_5 - (\bar{\kappa}_1 r - I_3^6) \ddot{\alpha}_6 - \delta_{23} \ddot{\eta}_2 &= 2\sqrt{2} p_1^{3*0} F \alpha_3 \end{aligned} \right\} \quad (1)$$

The symbols in these equations are defined in SPS 37-63, Vol. III, or Ref. 2. For the modal coordinates  $\eta_i$ ,  $i = 1, \dots, 6$ , the following equations are relevant.

$$\left. \begin{aligned} \ddot{\eta}_1 + 2\zeta_1 \sigma_1 \dot{\eta}_1 + \sigma_1^2 \eta_1 &= \delta_{12} \ddot{\theta}_2 \\ \ddot{\eta}_2 + 2\zeta_2 \sigma_2 \dot{\eta}_2 + \sigma_2^2 \eta_2 &= \delta_{23} \ddot{\theta}_3 \\ \ddot{\eta}_3 + 2\zeta_3 \sigma_3 \dot{\eta}_3 + \sigma_3^2 \eta_3 &= 0 \\ \ddot{\eta}_4 + 2\zeta_4 \sigma_4 \dot{\eta}_4 + \sigma_4^2 \eta_4 &= -\Delta_{42} r \ddot{\alpha}_6 \\ \ddot{\eta}_5 + 2\zeta_5 \sigma_5 \dot{\eta}_5 + \sigma_5^2 \eta_5 &= \delta_{51} \ddot{\theta}_1 \\ \ddot{\eta}_6 + 2\zeta_6 \sigma_6 \dot{\eta}_6 + \sigma_6^2 \eta_6 &= 0 \end{aligned} \right\} \quad (2)$$

Either of the gimbaling engine's equations has the form

$$I_1^3 \left[ \ddot{\alpha}_3 + \frac{1}{\sqrt{2}} (\ddot{\theta}_1 + \ddot{\theta}_2) \right] + d_3 \ddot{\alpha}_3 - k_3 (\alpha_{G1} - \alpha_3) = 0 \quad (3)$$

While for the translator,

$$\left. \begin{aligned} \left( I_3^5 + mr^2 + \frac{m^3 r^2}{\mathcal{M}^2} - 2 \frac{m^2 r^2}{\mathcal{M}} \right) \ddot{\alpha}_5 + r^2 d_1 \dot{\alpha}_5 + k_2 \alpha_5 + \frac{m^2 \bar{m}}{\mathcal{M}^2} r^2 \ddot{\alpha}_6 - k_2 \alpha_{T2} + \left( \frac{mr\mu}{\mathcal{M}} - \mu r \right) \ddot{\theta}_2 + I_3^5 \ddot{\theta}_3 &= 0 \\ \left( I_3^6 + \bar{m} r^2 + \frac{m \bar{m}^2}{\mathcal{M}^2} r^2 - 2 \frac{\bar{m}^2}{\mathcal{M}} r^2 \right) \ddot{\alpha}_6 + r^2 d_1 \dot{\alpha}_6 + k_1 \alpha_6 + \frac{m^2 \bar{m}}{\mathcal{M}^2} r^2 \ddot{\alpha}_5 - \frac{1}{\mathcal{M}} \left( \frac{m \bar{m}}{\mathcal{M}} r - \bar{m} r \right) \Delta_{42} \ddot{\eta}_4 - k_1 \alpha_{T1} \\ - \left( \frac{\bar{m} r \mu}{\mathcal{M}} - \bar{\mu} r \right) \ddot{\theta}_1 + I_3^6 \ddot{\theta}_3 &= 0 \end{aligned} \right\} \quad (4)$$

### 3. Control

Differential equations associated with the control system are

$$\tau_{Si} \dot{V}_{Si} + V_{Si} = K_{Si} \theta_i, \quad i = 1, 2, 3 \quad (5)$$

$$\left. \begin{aligned} r \dot{\alpha}_{T1} &= K_{M1} K_{V1} (V_{S1} - V_{F1}) \\ r \dot{\alpha}_{T2} &= K_{M2} K_{V2} (-V_{S2} - V_{F2}) \\ \dot{\alpha}_G &= K_{M3} K_{V3} (V_{S3} - V_{F3}) \end{aligned} \right\} \quad (6)$$

$$\left. \begin{aligned} \ddot{V}_{F1} &= \frac{K_{F1} K_{11} K_{M1} K_{V1}}{\tau_{11} \tau_{21}} \left[ \frac{\tau_{21}}{\tau_{S1}} K_{S1} \theta_1 + \left( 1 + K_{21} - \frac{\tau_{21}}{\tau_{S1}} \right) V_{S1} - \left( \tau_{21} + \frac{\tau_{11} + \tau_{21}}{K_{F1} K_{11} K_{M1} K_{V1}} \right) \dot{V}_{F1} \right. \\ &\quad \left. - \left( 1 + K_{21} + \frac{1}{K_{F1} K_{11} K_{M1} K_{V1}} \right) V_{F1} \right] \\ \ddot{V}_{F2} &= - \frac{K_{F2} K_{12} K_{M2} K_{V2}}{\tau_{12} \tau_{22}} \left[ \frac{\tau_{22}}{\tau_{S2}} K_{S2} \theta_2 + \left( 1 + K_{22} - \frac{\tau_{22}}{\tau_{S2}} \right) V_{S2} + \left( \tau_{22} + \frac{\tau_{12} + \tau_{22}}{K_{F2} K_{12} K_{M2} K_{V2}} \right) \dot{V}_{F2} \right. \\ &\quad \left. + \left( 1 + K_{22} + \frac{1}{K_{F2} K_{12} K_{M2} K_{V2}} \right) V_{F2} \right] \\ \ddot{V}_{F3} &= \frac{K_{F3} K_{13} K_{M3} K_{V3}}{\tau_{13} \tau_{23}} \left[ \frac{\tau_{13}}{\tau_{S3}} K_{S3} \theta_3 + \left( 1 + K_{23} - \frac{\tau_{23}}{\tau_{S3}} \right) V_{S3} - \left( \tau_{23} + \frac{\tau_{13} + \tau_{23}}{K_{F3} K_{13} K_{M3} K_{V3}} \right) \dot{V}_{F3} \right. \\ &\quad \left. - \left( 1 + K_{23} + \frac{1}{K_{F3} K_{13} K_{M3} K_{V3}} \right) V_{F3} \right] \end{aligned} \right\} \quad (7)$$

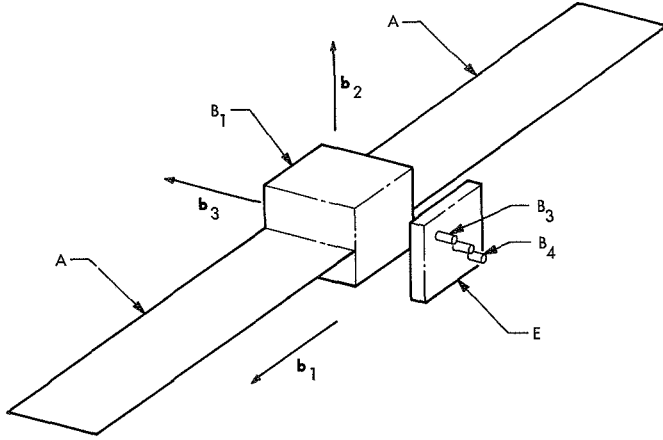


Fig. 1. Solar-electric-powered spacecraft

New symbols introduced here are defined in Ref. 2. In the parametric stability analysis which follows,  $K_{Si}$ ,  $i = 1, 2, 3$ , was the only parameter varied. Figure 2 is the block diagram representation of Expressions (1)–(7).

The open loop transfer functions for the three axes are

$$G_1 H_1 = -\bar{\kappa}_3 K_{S1} K_{M1} K_{V1} (s^6 + N_5 s^5 + N_4 s^4 + N_3 s^3 + N_2 s^2 + N_1 s + N_0) \left( s + \frac{1}{\tau_{21}} \right) \left( s + \frac{1}{\tau_{11}} \right) \left( I_1 - \delta_{11}^2 \right) \times \tau_{S1} s^3 (s^4 + D_3 s^3 + D_2 s^2 + D_1 s + D_0) \left( s + \frac{1}{\tau_{S1}} \right) \left\{ s^2 + \frac{1}{\tau_{21} \tau_{11}} [\tau_{11} + (1 + K_{11} K_{M1} K_{V1} K_{F1}) \tau_{21}] \right. \\ \left. \times s + \frac{1}{\tau_{21} \tau_{11}} + \frac{K_{11} K_{M1} K_{V1} K_{F1}}{\tau_{21} \tau_{11}} (1 + K_{21}) \right\} \quad (8)$$

$$G_2 H_2 = -\frac{\kappa_3 K_{S2} K_{M2} K_{V2}}{(I_2 - \delta_{12}^2)} \left[ s^2 + 3 \frac{(\mathcal{M} - m)}{\kappa_3 \mathcal{M}} F \right] (s^2 + 2s\zeta_1\sigma_1 + \sigma_1^2) \left( s + \frac{1}{\tau_{22}} \right) \left( s + \frac{1}{\tau_{12}} \right) \left( s^2 + \frac{2I_2}{I_2 - \delta_{12}^2} \right. \\ \left. \times \zeta_1\sigma_1 s + \frac{I_2\sigma_1^2}{I_2 - \delta_{12}^2} \right) \left( s + \frac{1}{\tau_{S2}} \right) \left\{ s^2 + \frac{1}{\tau_{22} \tau_{12}} [\tau_{12} + (1 + K_{21} K_{M2} K_{V2} K_{F2}) \tau_{22}] \right. \\ \left. \times s + \frac{1}{\tau_{22} \tau_{12}} + \frac{K_{12} K_{M2} K_{V2} K_{F2}}{\tau_{22} \tau_{12}} (1 + K_{22}) \right\} \quad (9)$$

$$G_3 H_3 = -\frac{\sqrt{8} p_1^{3*0} F K_{S3} K_{M3} K_{V3}}{(I_3 - \delta_{23}^2) \tau_{S3}} (s^2 + 2\zeta_2\sigma_2 s + \sigma_2^2) \left( s + \frac{1}{\tau_{23}} \right) \left( s + \frac{1}{\tau_{13}} \right) \left( s^2 + \frac{2I_3\zeta_2\sigma_2}{I_3 - \delta_{23}^2} + \frac{I_3\sigma_2^2}{I_3 - \delta_{23}^2} \right) \\ \times \left( s + \frac{1}{\tau_{S3}} \right) \left\{ s^2 + \frac{1}{\tau_{23} \tau_{13}} [\tau_{13} + (K_{M3} K_{F3} K_{13} K_{V3} + 1) \tau_{23}] s + \frac{1}{\tau_{23} \tau_{13}} + \frac{K_{F3} K_{M3} K_{13} K_{V3}}{\tau_{23} \tau_{13}} (1 + K_{23}) \right\} \quad (10)$$

The dashed lines indicate coupling terms. Solid lines are associated with the uncoupled equations. The mode shapes corresponding to the modes excited by rotation about each axis are drawn at the right of the block diagram. The third and sixth modes are absent. This is predictable from Expression (2), which indicates that the third and sixth modes of structural vibration decouple from the remaining equations and have no effect on the attitude motion.

#### 4. Stability Analysis

Only the uncoupled equations were considered in the first part of the stability analysis. This uncoupling transforms the attitude control problem into one that is amenable to root locus analysis. In addition to uncoupling, the high-frequency translator and gimballing dynamics were omitted from the initial stability analysis. This system where coupling and translator and gimballing dynamics are omitted will be referred to as the modified system.

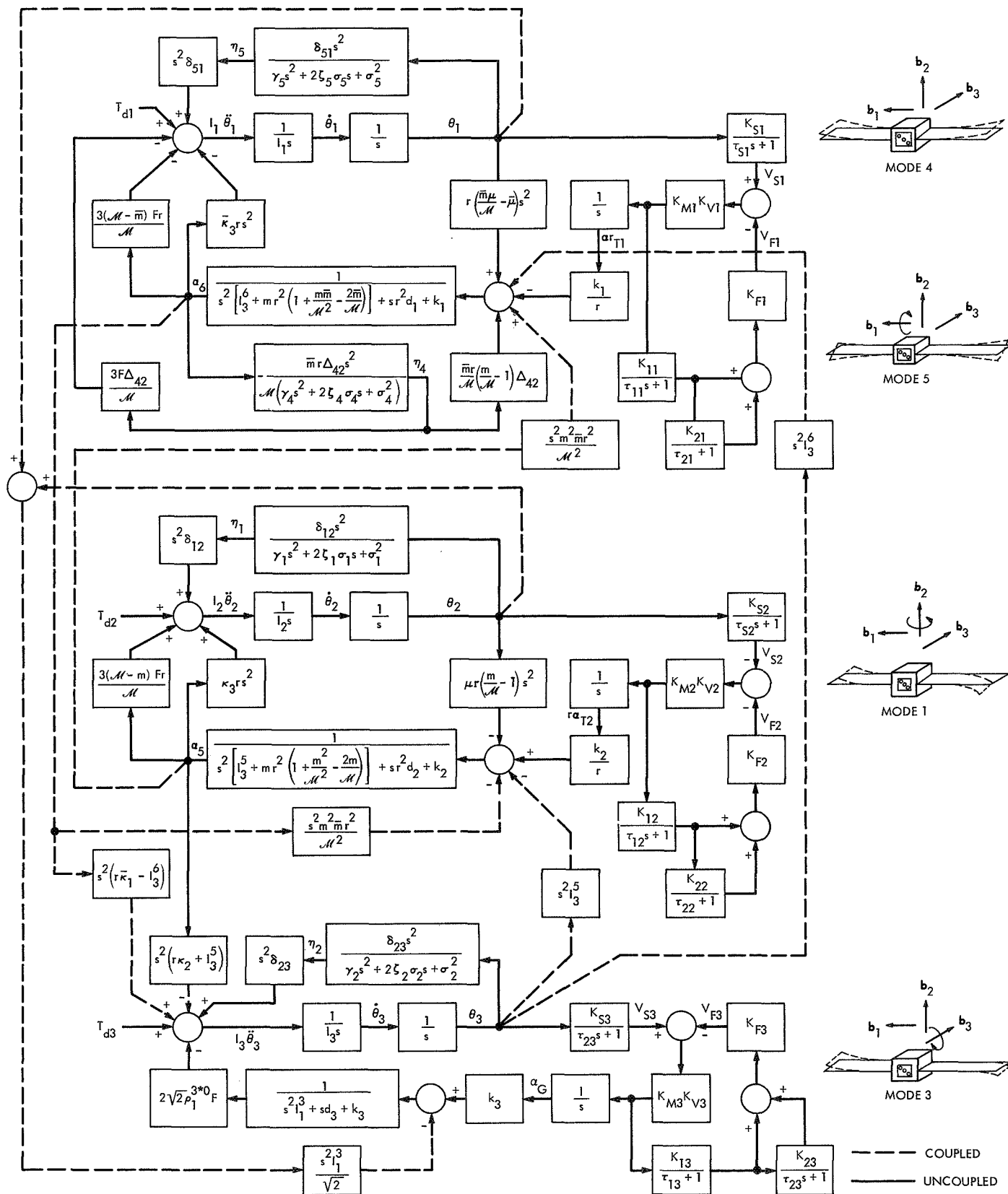


Fig. 2. Block diagram for coupled and uncoupled equations, including mode shapes



where  $G_i$  is the forward loop transfer function and  $H_i$  is the feedback loop transfer function for the axis parallel to  $\mathbf{b}_i$ . Expressions for  $N_i$ ,  $i = 0, \dots, 5$ , and for  $D_j$ ,  $j = 0, \dots, 3$ , are

$$\left. \begin{aligned} N_5 &= 2(\zeta_4\sigma_4 + \zeta_5\sigma_5) \\ N_4 &= \sigma_4^2 + \sigma_5^2 + 4\zeta_4\zeta_5\sigma_4\sigma_5 + \frac{3(\mathcal{M} - \bar{m})F}{\bar{\kappa}_s\mathcal{M}} - \frac{3F\bar{m}\Delta_{42}^3}{\bar{\kappa}_s\mathcal{M}^2} \\ N_3 &= 2\sigma_4\sigma_5(\zeta_4\sigma_5 + \zeta_5\sigma_4) + \frac{6(\mathcal{M} - \bar{m})}{\bar{\kappa}_s\mathcal{M}}(\zeta_4\sigma_4 + \zeta_5\sigma_5) \\ &\quad - \frac{6F\bar{m}\Delta_{42}^2\zeta_5\sigma_5}{\bar{\kappa}_s\mathcal{M}^2} \\ N_2 &= \sigma_4^2\sigma_5^2 + \frac{3(\mathcal{M} - \bar{m})}{\bar{\kappa}_s\mathcal{M}}(\sigma_4^2 + \sigma_5^2 + 4\zeta_4\zeta_5\sigma_4\sigma_5) \\ &\quad - \frac{3F\bar{m}\Delta_{42}^2\sigma_5^2}{\bar{\kappa}_s\mathcal{M}^2} \\ N_1 &= \frac{6(\mathcal{M} - \bar{m})F\sigma_4\sigma_5}{\bar{\kappa}_s\mathcal{M}}(\zeta_4\sigma_5 + \zeta_5\sigma_4) \\ N_0 &= \frac{3(\mathcal{M} - \bar{m})F\sigma_4^2\sigma_5^2}{\bar{\kappa}_s\mathcal{M}} \end{aligned} \right\} \quad (11)$$

and

$$\left. \begin{aligned} D_3 &= \frac{I_1}{I_1 - \delta_{51}^2} \left[ 2(\zeta_4\sigma_4 + \zeta_5\sigma_5) - \frac{2\delta_{51}^2\zeta_4\sigma_4}{I_1} \right] \\ D_2 &= \frac{I_1}{I_1 - \delta_{51}^2} \left( \sigma_4^2 + \sigma_5^2 + 4\zeta_4\zeta_5\sigma_4\sigma_5 - \frac{\delta_{51}^2\sigma_4^2}{I_1} \right) \\ D_1 &= \frac{2I_1\sigma_4\sigma_5}{I_1 - \delta_{51}^2}(\zeta_4\sigma_5 + \zeta_5\sigma_4) \\ D_0 &= \frac{I_1\sigma_4^2\sigma_5^2}{I_1 - \delta_{51}^2} \end{aligned} \right\} \quad (12)$$

Values selected for the parameters in these expressions, as well as for the parameters in Expressions (1)–(10), are

$$\begin{aligned} K_{1i} &= 100 \text{ V-s/ft}, & i &= 1, 2, 3 \\ K_{2i} &= 4.5 & i &= 1, 2, 3 \\ K_{Fi} &= 80 & i &= 1, 2, 3 \\ K_{M1} &= K_{M2} = 2.2 \times 10^{-4} \text{ ft/pulse} \\ K_{M3} &= 10^{-4} \text{ rad/pulse} \\ K_{Vi} &= 40 \text{ pulses/V-s}, & i &= 1, 2, 3 \end{aligned}$$

$$\begin{aligned} \delta_{23} &= 94.51 \text{ slug-ft}^2 \\ \delta_{12} &= 94.60 \text{ slug-ft}^2 \\ \delta_{51} &= 7.084 \text{ slug-ft}^2 \\ \Delta_{42} &= 3.077 \text{ slug-ft} \\ \zeta_i &= 0.005 \text{ slug}^{1/2}\text{-ft}, & i &= 1, 2, 3, 4, 5 \\ \sigma_1 &= 0.02129 \text{ slug}^{1/2}\text{-ft/s} \\ \sigma_2 &= 0.02141 \text{ slug}^{1/2}\text{-ft/s} \\ \sigma_4 &= 0.02376 \text{ slug}^{1/2}\text{-ft/s} \\ \sigma_5 &= 0.02838 \text{ slug}^{1/2}\text{-ft/s} \\ \tau_{1i} &= 10^3 \text{ s}, & i &= 1, 2, 3 \\ \tau_{2i} &= 5 \times 10^2 \text{ s}, & i &= 1, 2, 3 \\ \tau_{Si} &= 1 \text{ s}, & i &= 1, 2, 3 \\ F &= 0.02 \text{ lb} \\ I_1 &= 326.4 \text{ slug-ft}^2 \\ I_2 &= 9420.0 \text{ slug-ft}^2 \\ I_3 &= 9148.0 \text{ slug-ft}^2 \\ M &= 41.09 \text{ slug} \\ m &= 3.727 \text{ slug} \\ \bar{m} &= 3.106 \text{ slug} \\ \bar{\kappa}_s &= -9.318 \text{ slug-ft} \\ \kappa_s &= -7.765 \text{ slug-ft} \end{aligned}$$

The sensor gains  $K_{Si}$  have not been specified and will serve as variable parameters upon which the attitude stability depends.

A root locus diagram for the  $\mathbf{b}_1$  axis appears in Fig. 3. Values for the open loop poles and zeros are given. Poles and zeros having non-zero imaginary parts are associated with the fourth and fifth modes. A real zero at 0.084324 is due to the so-called "tail wags dog" effect, or the term  $\bar{\kappa}_s r \ddot{a}_6$  in the first of Expression (1). This is an inertia torque exerted on the vehicle when the engine cluster unit translates parallel to the  $\mathbf{b}_2$  vector. Such motion excites fourth-mode oscillations. Figure 4 is an enlarged version of the  $\mathbf{b}_1$  axis root locus and shows detail near the origin. For very small gain, i.e., very small  $K_{S1}$ , the locus emanating from the multiple pole at zero initially passes into the right half plane. Thus, instability is implied. As  $K_{S1}$  is increased, the locus crosses back into the left half plane and finally returns to the right half plane for sufficiently large  $K_{S1}$ . The range of  $K_{S1}$  for which all parts of the locus lie in the left half plane, i.e., for which stability is guaranteed for the  $\mathbf{b}_1$  axis, is  $0.5 < K_{S1} < 23$ .

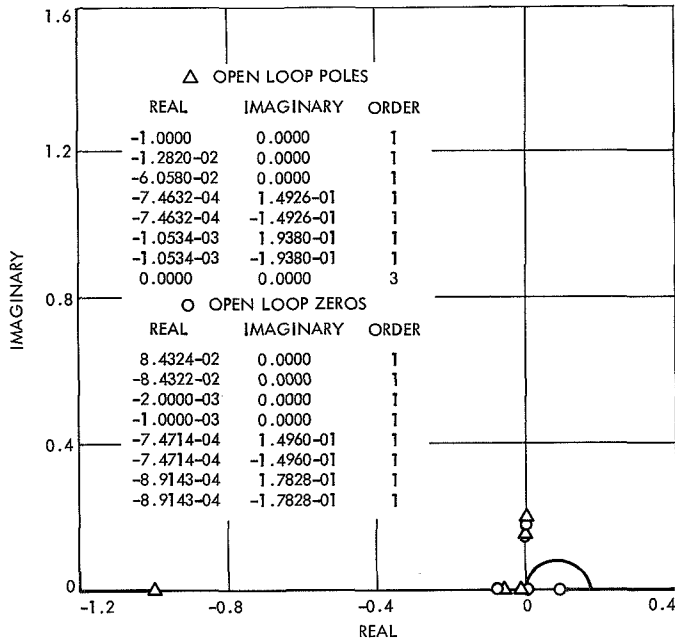


Fig. 3.  $b_1$ -axis root locus

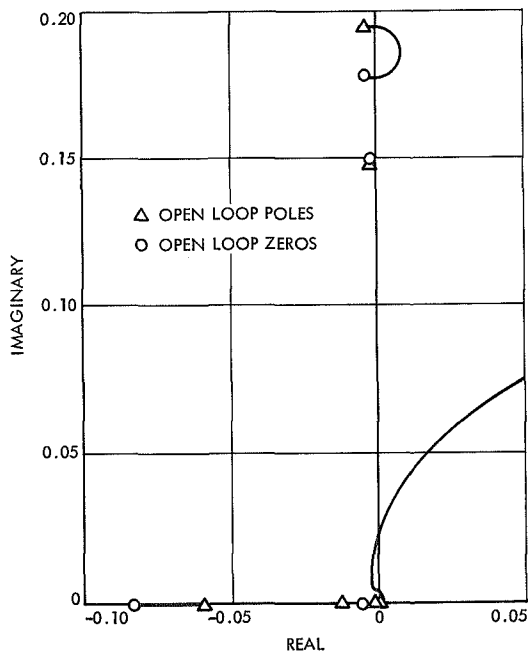


Fig. 4. Enlarged  $b_1$ -axis root locus

Similar analyses were conducted for the  $b_2$  and  $b_3$  axes. The stable ranges for  $K_{s2}$  and  $K_{s3}$  are

$$15 < K_{s2} < 201$$

$$11 < K_{s3} < 221$$

The effect of the flexible solar arrays was assessed by performing root locus analyses for the rigid spacecraft models. The resulting stable  $K_{si}$ ,  $i = 1, 2, 3$ , ranges were

$$0.52 < K_{s1} < 23.52$$

$$16 < K_{s2} < 276$$

$$11 < K_{s3} < 221$$

Comparison with the values for the non-rigid case shows that only the  $b_2$  axis results are appreciably modified when flexibility is taken into account.

Time histories of the attitude angles  $\theta_1$ ,  $\theta_2$ , and  $\theta_3$  appear in Figs. 5 and 6. The continuous system simulation language (CSSL), developed through support of the Guidance and Control Branch of the Office of Advanced Research and Technology, was used in programming the linear equations for the Univac 1108 digital computer. For the axis parallel to  $b_i$ , the initial attitude angle rate  $\dot{\theta}_i$  is  $10^{-4}$  rad/s. This rate is considerably larger than those expected to be encountered from disturbance torques for a deep space mission. All other variables, i.e.,  $\theta_i$ ,  $V_{si}$ , etc., have an initial value of zero.

Figure 5 shows the attitude behavior for cases when  $K_i$ ,  $i = 1, 2, 3$ , is in the stable range stated earlier, in particular,  $K_{s1} = 20$ ,  $K_{s2} = 100$ ,  $K_{s3} = 100$ . At approximately 100 s,  $\theta_1$  reaches its maximum amplitude of  $6.4 \times 10^{-3}$  rad. For a time of 2800 s, the amplitude has diminished to nearly  $7 \times 10^{-4}$  rad or one-tenth the maximum amplitude. For  $\theta_2$  and  $\theta_3$ , the response curves are of a lower frequency than for  $\theta_1$ .

Values for  $K_{s1}$ ,  $K_{s2}$ , and  $K_{s3}$  in Fig. 6 are 50, 350, and 350, respectively. These fall outside of the stable range. The  $\theta_i$  responses are then expected to be unstable. This is seen to be the case. However, the curves are not reliable as true indicators of the behavior of  $\theta_i$  since the linear equations are not really valid for large  $\theta_i$ .  $\theta_i$  is indeed unstable, but its time history probably departs markedly from the curves of Fig. 6.

Following the root locus analysis, a stability analysis for the coupled system of equations (including the translator and gimbaling dynamics) was undertaken. For linear systems, stability is guaranteed whenever the eigenvalues are distinct and have negative real parts. For the case at hand, the system of first- and second-order linear differential equations was converted to a system of first-order differential equations of the form

$$\dot{X} = AX \quad (13)$$

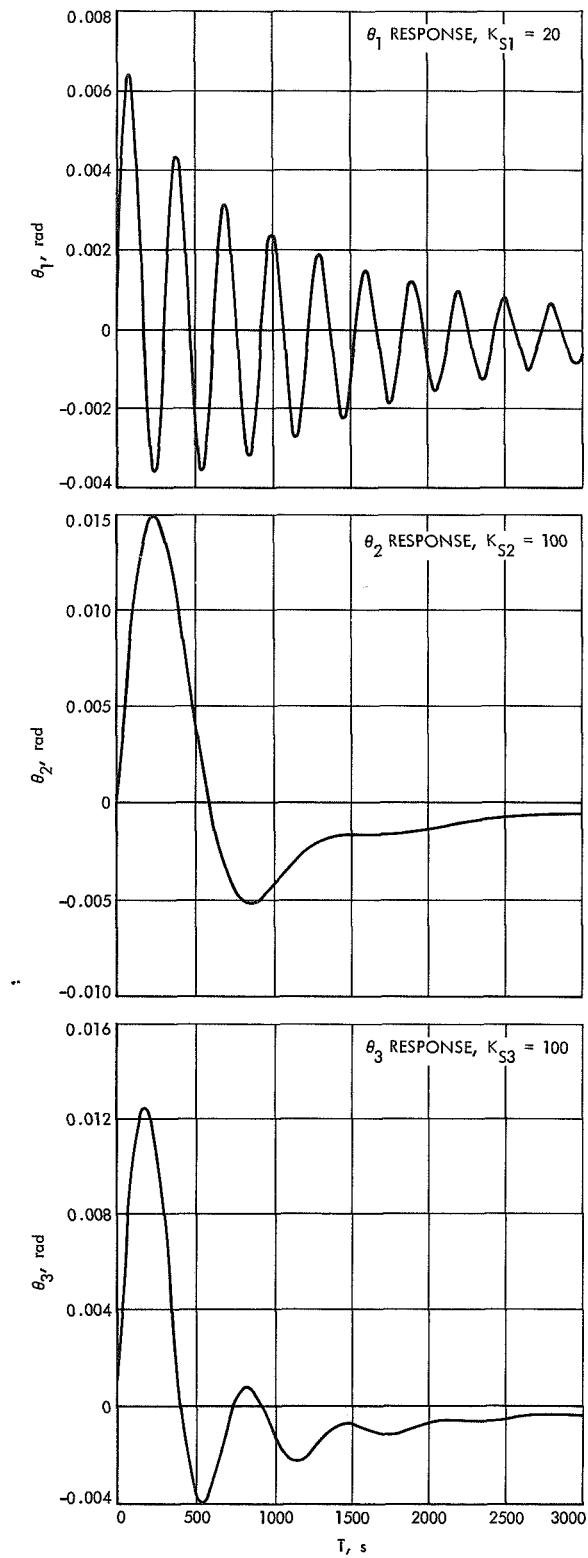


Fig. 5. Response curves in stable range

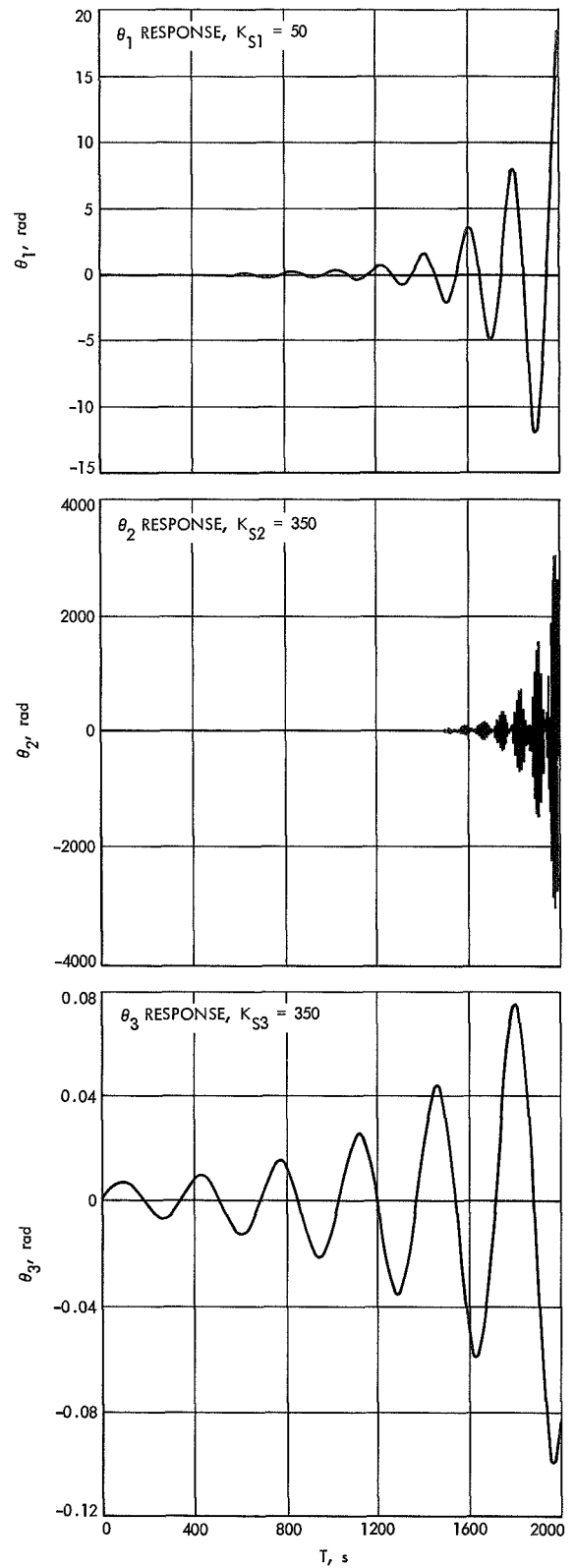


Fig. 6. Response curves outside stable range

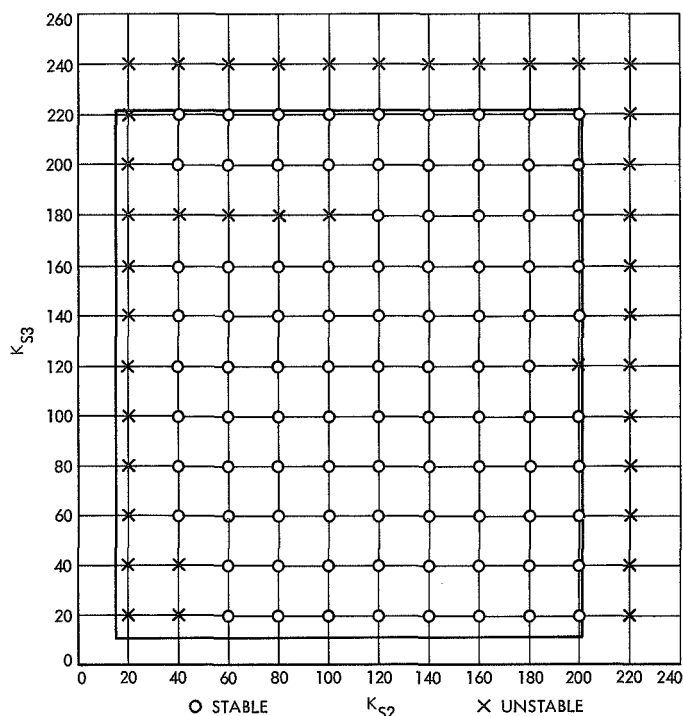


Fig. 7. Stability chart for modified and unmodified systems ( $K_{s1} = 10$ )

where  $X$  is a  $32 \times 1$  vector and  $A$  is a  $32 \times 32$  matrix. Determination of the eigenvalues for the system is equivalent to determining the eigenvalues for  $A$ . Figure 7 shows the results of the eigenvalue analysis. The value for  $K_{s1}$  was fixed at 10. Figure 7 presents stability information for both the modified and the unmodified systems. The region enclosed by the rectangle, i.e.,  $15 < K_{s2} < 201$  and

$11 < K_{s3} < 221$ , is, for  $K_{s1} = 10$ , the stability zone for the modified system. Any  $K_{s2}$  or  $K_{s3}$  value falling in that zone is such that stability for the modified system is guaranteed. The eigenvalue analysis determined that some minor alteration of the boundaries is required when coupling and translator and gimbaling dynamics are accounted for. Each grid intersection point in Fig. 7 was investigated for stability for the unmodified system. For small values of  $K_{s2}$ , it was found that configurations which were formerly stable for the modified system are unstable. Instabilities in the vicinity of  $K_{s3} = 180$  also appear. A more detailed analysis of this region showed that instabilities were not restricted to  $K_{s3} = 180$ .

## 5. Conclusions

The work presented here has demonstrated the feasibility of using electric propulsion engines for attitude control as well as for propulsion during the low-thrust cruise phase of a deep space mission. In addition, it has been shown that flexibility associated with large roll-out solar arrays poses no great threat to the success of such missions. Furthermore, the detrimental effects from torques due to "tail-wags-dog" terms can be offset.

## References

1. Likins, P. W., *Dynamics and Control of Flexible Space Vehicles*, Technical Report 32-1329 (Revision 1). Jet Propulsion Laboratory, Pasadena, Calif., Jan. 15, 1970.
2. Marsh, E. L., *The Attitude Control of a Flexible Solar Electric Spacecraft*, paper presented at the AIAA 8th Electric Propulsion Conference, Stanford, Calif., Aug. 31-Sept. 2, 1970.

## B. Finite Element Modeling for Appendage Interaction With Spacecraft Control, P. W. Likins<sup>1</sup> and E. L. Marsh

### 1. Introduction

Useful analyses of the spacecraft attitude control/structural dynamics interaction problem can be carried out when accurate models of the attitude control system and the spacecraft dynamics are available. For the spacecraft dynamics, a hybrid coordinates model, utilizing discrete coordinates for rigid body kinematics and modal coordinates related to the elastic structure's deformation coordinates, has been suggested by Likins.

The basic hybrid coordinates philosophy is adhered to in this article. However, a detailed description of an improved model<sup>2</sup> for the flexible components is discussed. For the model currently in use at JPL (Ref. 1), flexible appendages are approximated by an array of rigid finite bodies connected by massless elastic beams. The improved model consists of a collection of finite elastic elements possessing mass and interconnected at points where finite rigid masses may be located. The points of connection and the masses at these points are referred to as nodes and nodal bodies respectively. In the following discussion, equations of motion for the finite elements, the rigid bodies, and the composite appendage will be stated. These must be used in conjunction with the vehicle equations and the control system equations in order to properly analyze the spacecraft dynamics and control interaction problem.

### 2. Appendage Idealization

Figure 1 shows the two basic types of units considered in the analysis. Body  $b$  is a rigid body and  $a$  is a flexible appendage. For a real vehicle, these may represent the spacecraft bus and a solar panel, respectively. The tetrahedral object having dotted lines for boundaries and with parallelepipeds at its vertices represents the typical finite element  $s$  and nodal bodies as described above. Shapes other than tetrahedrons and parallelepipeds may be assumed for the finite elements and nodal bodies. The deformed tetrahedron having dashed line boundaries indicates the steady-state deformation which may be imposed by a constant rotation of  $a$  in inertial space. Finally, the deformed version of the finite element having solid line

boundaries represents the configuration for an excited state where both steady-state and oscillatory deformations occur.

Unit vectors fixed in  $b$  are  $\mathbf{b}_i$ ,  $i = 1, 2, 3$ . Unit vectors fixed in the undeformed appendage are  $\mathbf{a}_i$ .  $Q$  is a point on the  $a$ - $b$  interface.  $O$  and  $O'$  locate the vehicle mass center in the steady state and undeformed state, respectively. The mass center for a general oscillatory state is  $CM$ . The meanings of other symbols in Fig. 1 are either self evident or are described in Section D, Ref. 1, or will be described in the sequel.

### 3. Finite Element Equation of Motion

This equation is established by equating the work  $W^*$  done by external forces in the course of a virtual displacement of the nodes to the strain energy  $U^*$  stored in the deforming element. The nodal displacement vector  $\bar{\mathbf{y}}$  for the element in consideration is defined by

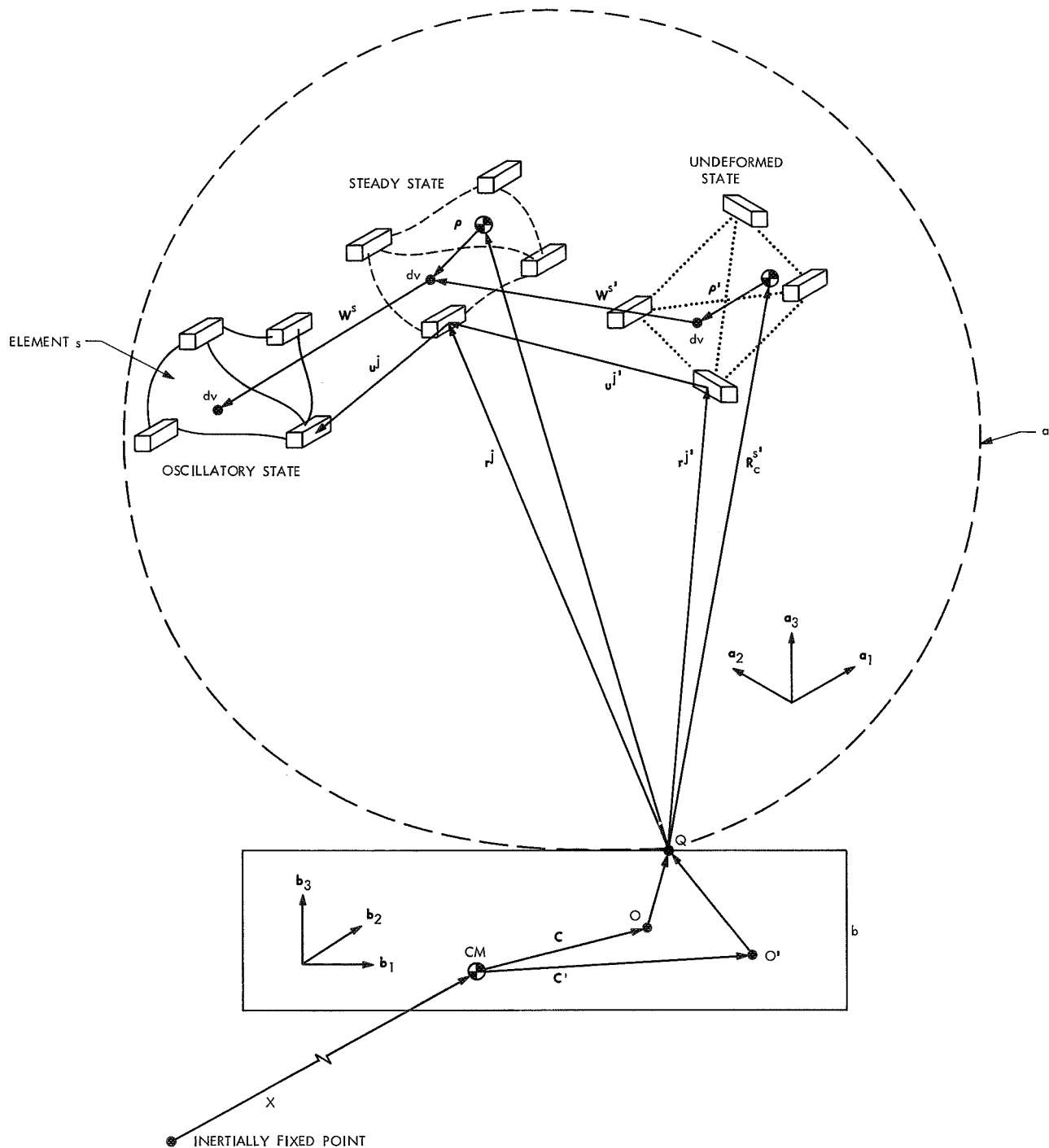
$$\bar{\mathbf{y}} = \begin{bmatrix} \bar{\mathbf{u}}^k \\ \bar{\beta}^k \\ \bar{\mathbf{u}}^i \\ \bar{\beta}^i \\ \cdot \\ \cdot \\ \bar{\mathbf{u}}^j \\ \bar{\beta}^j \end{bmatrix} \quad (1)$$

where  $\bar{\mathbf{u}}^k$  and  $\bar{\beta}^k$  are the  $3 \times 1$  translational and rotational deformation vectors, respectively, for the  $k$ th node of element  $s$  and expressed in terms of local coordinates  $\xi, \eta, \zeta$  parallel to  $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ , which are unit vectors fixed in the steady-state configuration for the element. Whenever a vector is expressed in terms of local coordinates, it is so denoted by an overbar. When expressed in terms of  $\mathbf{a}_i$ , the bar is absent. The sequence  $k, i, \dots, j$  is used to indicate that the nodes are numbered neither consecutively nor sequentially. The symbol  $\bar{\mathbf{y}}^*$  will be used to denote an arbitrary virtual displacement of the nodes.

Determination of  $U^*$  and  $W^*$  requires a knowledge of the oscillatory deformation field for the element. A standard procedure is to select a deformation function  $\bar{\mathbf{w}}$  of  $\xi, \eta$ , and  $\zeta$ . The elements of the  $3 \times 1$  vector  $\bar{\mathbf{w}}$  are the deformations parallel to axes  $\xi, \eta$ , and  $\zeta$  and usually are chosen as polynomials in  $\xi, \eta$ , and  $\zeta$ . Coefficients  $\Gamma_i$  for these polynomials are  $6N$  in number where  $N$  is the number of nodes for the element. It is seen then that the number of coefficients matches the number of translational

<sup>1</sup>UCLA, Los Angeles, California.

<sup>2</sup>Likins, P. W., *Finite Element Appendage Equations for Hybrid Coordinate Dynamic Analysis* (JPL Technical Report to be published).



**Fig. 1. Spacecraft model**

and rotational displacements of the nodes. The deformation vector has the form

$$\bar{w} = P \Gamma \quad (2)$$

where

$$P = \begin{bmatrix} 1 & \xi & \eta & \zeta & \xi^2 & \xi\eta & \xi\zeta & \eta^2 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 1 & \xi & \eta & \zeta & \xi^2 & \cdots & 0 & 0 & 0 & \cdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & \xi & \eta & \zeta & \cdots \end{bmatrix} \quad (3)$$

$$\Gamma = [\Gamma_1 \Gamma_2 \cdots \Gamma_{6n}]^T \quad (4)$$

The partitioned sub-matrices of  $P$  are of size  $3 \times 2N$ . Before proceeding to the calculation of  $W^*$  and  $U^*$ ,  $\Gamma$  must be determined. Equation (2) must apply to the nodes as well as to points within the element. Thus, for node  $j$ ,

$$\bar{u}^j = \bar{w}(\xi_j, \eta_j, \zeta_j) = P(\xi_j, \eta_j, \zeta_j) \quad (5)$$

and

$$\bar{\beta}^j = \frac{1}{2} \tilde{\nabla} \bar{w}|_{\xi_j, \eta_j, \zeta_j} = \frac{1}{2} \tilde{\nabla} P|_{\xi_j, \eta_j, \zeta_j} \quad (6)$$

where

$$\tilde{\nabla} = \begin{bmatrix} 0 & -\partial/\partial\zeta & \partial/\partial\eta \\ \partial/\partial\zeta & 0 & -\partial/\partial\xi \\ -\partial/\partial\eta & \partial/\partial\xi & 0 \end{bmatrix} \quad (7)$$

Expressions of the type detailed in Eqs. (5) and (6) can be combined as follows:

$$\begin{bmatrix} \bar{u}^k \\ \bar{\beta}^k \\ \bar{u}^i \\ \bar{\beta}^i \\ \cdot \\ \cdot \\ \cdot \\ \bar{u}^j \\ \bar{\beta}^j \end{bmatrix} = \begin{bmatrix} P|_k \\ 1/2 \tilde{\nabla} P|_k \\ P|_i \\ 1/2 \tilde{\nabla} P|_i \\ \cdot \\ \cdot \\ \cdot \\ P|_j \\ 1/2 \tilde{\nabla} P|_j \end{bmatrix} \Gamma \quad (8)$$

where  $|_j$  implies evaluation at  $\xi_j, \eta_j, \zeta_j$ , etc. If the matrix containing  $P|_j$  and  $\tilde{\nabla} P|_j$  terms is denoted  $F$ , then<sup>3</sup>

$$\bar{y} \underset{(1,8)}{=} F \Gamma \Rightarrow \Gamma = F^{-1} \bar{y} \quad (9)$$

<sup>3</sup>Numbers beneath equal signs indicate the equations referenced in writing the equation to which the equal sign corresponds.

and

$$\bar{w} \underset{(2,9)}{=} P F^{-1} \bar{y} \quad (10)$$

Define  $W$  as the matrix  $P F^{-1}$ . Then

$$\bar{w} = W \bar{y} \quad (11)$$

The deformation function is then a function of  $\xi, \eta, \zeta$  and the nodal displacement vector.

A final step preparatory to the evaluation of the strain energy  $U^*$  is the determination of expressions for the variational stress and strain fields for the element. Variational stress and strain mean the increments in stress and strain beyond the steady-state value. It can be shown that the incremental strain in the element beyond any steady-state strain can be related to the deformation function  $\bar{w}$  by

$$\bar{\epsilon} = D \bar{w} \quad (12)$$

where

$$\bar{\epsilon} = \begin{bmatrix} \bar{\epsilon}_{11} \\ \bar{\epsilon}_{22} \\ \bar{\epsilon}_{33} \\ \bar{\epsilon}_{12} \\ \bar{\epsilon}_{23} \\ \bar{\epsilon}_{13} \end{bmatrix}, \quad D = \begin{bmatrix} \partial/\partial\xi & 0 & 0 \\ 0 & \partial/\partial\eta & 0 \\ 0 & 0 & \partial/\partial\zeta \\ \partial/\partial\eta & \partial/\partial\xi & 0 \\ 0 & \partial/\partial\zeta & \partial/\partial\eta \\ \partial/\partial\zeta & 0 & \partial/\partial\xi \end{bmatrix} \quad (13)$$

The variational thermal strain  $\bar{\epsilon}_T$  is

$$\bar{\epsilon}_T = \bar{\alpha} T [1 \ 1 \ 1 \ 0 \ 0 \ 0]^T \quad (14)$$

where  $T$  is the deviation in temperature from the steady state at a given point and  $\bar{\alpha}$  is the coefficient of thermal

expansion. An expression for the variational stress is

$$\bar{\sigma} = S \bar{\epsilon} - \bar{\sigma}_T \quad (15)$$

where the stress vector  $\bar{\sigma}$  is

$$\bar{\sigma} = [\bar{\sigma}_{11} \bar{\sigma}_{22} \bar{\sigma}_{33} \bar{\sigma}_{12} \bar{\sigma}_{23} \bar{\sigma}_{31}]^T \quad (16)$$

and

$$S = \frac{E}{(1+2)(1-2\nu)} \begin{bmatrix} 1-\nu & \nu & \nu & 0 & 0 & 0 \\ \nu & 1-\nu & \nu & 0 & 0 & 0 \\ \nu & \nu & 1-\nu & 0 & 0 & 0 \\ 0 & 0 & 0 & (1-2\nu)/2 & 0 & 0 \\ 0 & 0 & 0 & 0 & (1-2\nu)/2 & 0 \\ 0 & 0 & 0 & 0 & 0 & (1-2\nu)/2 \end{bmatrix} \quad (17)$$

$$\bar{\sigma}^T = S \bar{\epsilon}_T \quad (18)$$

$E$  is the modulus of elasticity and  $\nu$  is Poisson's ratio.

The expression for  $U^*$  is then

$$U^* = \int \bar{\epsilon}^{*T} \bar{\sigma} dv = \bar{y}^{*T} \int W^T D^T S D W dv \bar{y} - \bar{y}^{*T} \int W^T D^T \bar{\sigma}_T dv \quad (19)$$

where  $dv$  is the differential element of volume of the finite element.

Work in deforming the element is done by inertial forces, body forces, forces exerted by nodal bodies, and surface forces. Distributing the surface forces to the nodes, as is ordinarily done, has the effect of eliminating them from the element equation. Denote the inertial force exerted on a differential element of the finite element by  $-A_\mu dv$  where  $A$  is the acceleration of the mass center of the differential element, and  $\mu$  is the mass density. Let  $\bar{G}(\xi, \eta, \zeta)$  denote the body force for a generic point of the

element. Finally, let  $\bar{L}^s$  be the nodal body force vector for elements where

$$\bar{L}^s = [\bar{F}^{ks} \bar{T}^{ks} \dots \bar{F}^{js} \bar{T}^{js}]^T \quad (20)$$

and  $\bar{F}^{ks}$  and  $\bar{T}^{ks}$  are  $3 \times 1$  matrices in the local vector basis representing, respectively, the force and torque applied by the  $k$ th nodal body to the  $s$ th element. An expression for the work done is then

$$W^* = \bar{y}^{*T} \bar{L} + \int \bar{w}^{*T} \bar{G} dv - \int \bar{w}^{*T} \bar{A}_\mu dv \quad (21)$$

where integration is carried out over the volume of the finite element and reference to element  $s$  is omitted. Equating  $W^*$  to  $U^*$  and use of the expression in Eq. (11) implies

$$\bar{L} = \int W^T D^T S D W dv \bar{y} + \int W^T [\bar{A}_\mu - \bar{G} - D^T \bar{\sigma}_T] dv \quad (22)$$

After the evaluation of an expression for the acceleration  $\bar{A}$ , as in Ref. 1, the finite element equation takes the form

$$\begin{aligned} \bar{L}^s = & \bar{m}^s \ddot{\bar{y}} - \mathcal{M}_s W_c^{sT} \bar{C}^{sT} \left[ \frac{\sum_{r=1}^{\epsilon} \bar{C}^{rT} \mathcal{M}_r W_c^r \ddot{\bar{y}}^r + \sum_{i=1}^n m_i \ddot{u}^i}{\mathcal{M}} \right] + \bar{g}^s \dot{\bar{y}}^s - 2 \mathcal{M}_s W_c^{sT} \bar{\omega}^a \left[ \frac{\sum_{r=1}^{\epsilon} \bar{C}^{rT} \mathcal{M}_r W_c^r \dot{\bar{y}}^r + \sum_{i=1}^n m_i \dot{u}^i}{\mathcal{M}} \right] \\ & + (\bar{k}_0^s + \bar{k}_\Delta^s + \bar{\kappa}^s + \bar{\alpha}^s) \bar{y}^s - \mathcal{M}_s W_c^{sT} \bar{C}^{sT} (\bar{\omega}^a + \bar{\omega}^a \bar{\omega}^a) \left[ \frac{\sum_{r=1}^{\epsilon} \bar{C}^{rT} \mathcal{M}_r W_c^r \bar{y}^r + \sum_{i=1}^n m_i u^i}{\mathcal{M}} \right] + \int_s W^T \bar{C} (\bar{\omega}^a + \bar{\omega}^a \bar{\omega}^a) \bar{C}^T \bar{\rho}_\mu dv \\ & + \mathcal{M}_s W_c^{sT} \{ \bar{C}^s C [\odot \ddot{X} + (\bar{\omega} + \bar{\omega} \bar{\omega}) R] + \bar{C}^s (\bar{\omega}^a + \bar{\omega}^a \bar{\omega}^a) R_c^s \} - \int_s W^T (\bar{G} - D^T \bar{\sigma}_T) dv \end{aligned} \quad (23)$$



New symbols introduced above have the following definitions:

- $\bar{C}$  a  $3 \times 3$  matrix relating  $\mathbf{e}_i$  to  $\mathbf{a}_i$  i.e.,  $\{\mathbf{e}\} = C\{\mathbf{a}\}$
- $g^s$  the element gyroscopic coupling matrix, i.e.  $2 \int W^T C \omega^a C^T W_\mu dv$
- $\bar{k}_0^s$  stiffness matrix for an unloaded element
- $\bar{k}_\Delta^s$  preload stiffness matrix
- $\mathcal{M}$  total vehicle mass
- $\mathcal{M}_s$  mass of the  $s$ th element
- $m_i$  mass of the  $i$ th nodal body
- $\bar{m}^s$   $\int_s W^T W_\mu dv$ , the consistent mass matrix
- $n$  number of nodal bodies
- $W_c^s$  the value of  $W^s$  when evaluated at the mass center  $\xi_c, \eta_c, \zeta_c$  of element  $s$
- $\epsilon$  the number of finite elements

- $\Theta$  a  $3 \times 3$  matrix which transforms inertial coordinates to  $\mathbf{a}_i$  coordinates
- $\bar{\kappa}^s$  centrifugal stiffness matrix for finite element
- $\bar{\rho}$  position vector of a differential element of volume of the finite element with respect to the element's mass center
- $\omega^a$  the angular velocity of  $a$  relative to  $b$

#### 4. Nodal Body Equations of Motion

For nodal body  $j$  having mass  $m_j$ , inertial acceleration  $\mathbf{A}^j$ , inertia dyadic  $\mathbf{I}^j$ , and inertial angular velocity  $\boldsymbol{\omega}^j$ , the translational and rotational equations are

$$\mathbf{F}^j = m_j \mathbf{A}^j \quad (24)$$

$$\mathbf{T}^j = \mathbf{I}^j \cdot \boldsymbol{\omega}^j + \boldsymbol{\omega}^j \times \mathbf{I}^j \cdot \boldsymbol{\omega}^j \quad (25)$$

By using procedures outlined in Section IIIB of Ref. 1, it can be shown that Eqs. (24) and (25) have the more detailed forms

$$f^j - \sum_{s \in \epsilon_j} \bar{C}^{st} \bar{F}^{js} = m^j \left\{ C \Theta \ddot{\mathbf{X}} + C [\ddot{\mathbf{e}} + 2\tilde{\omega} \dot{\mathbf{e}} + (\tilde{\omega} + \tilde{\omega} \tilde{\omega}) \mathbf{e}] + (\tilde{\omega}^a + \tilde{\omega}^a \tilde{\omega}^a) \mathbf{r}^j + \ddot{\mathbf{u}}^j - \frac{\left( \sum_{i=1}^n m^i \ddot{\mathbf{u}}^i + \sum_{r=1}^{\epsilon} \bar{C}^{rt} \mathcal{M}_r W_c^r \ddot{\mathbf{y}}^r \right)}{\mathcal{M}} \right. \\ \left. + 2\tilde{\omega}^a \left[ \dot{\mathbf{u}}^j - \frac{\left( \sum_{i=1}^n m^i \dot{\mathbf{u}}^i + \sum_{r=1}^{\epsilon} \bar{C}^{rt} \mathcal{M}_r W_c^r \dot{\mathbf{y}}^r \right)}{\mathcal{M}} \right] + (\tilde{\omega}^a + \tilde{\omega}^a \tilde{\omega}^a) \left[ \mathbf{u}^j - \frac{\left( \sum_{i=1}^n m^i \mathbf{u}^i + \sum_{r=1}^{\epsilon} \bar{C}^{rt} \mathcal{M}_r W_c^r \mathbf{y}^r \right)}{\mathcal{M}} \right] \right\}, \quad j = 1, \dots, n \quad (26)$$

$$\tau^j - \sum_{s \in \epsilon_j} \bar{C}^{st} \bar{T}^{js} = \mathbf{I}^j \dot{\boldsymbol{\omega}}^a + \tilde{\omega}^a \mathbf{I}^j \boldsymbol{\omega}^a + \mathbf{I}^j \ddot{\boldsymbol{\beta}}^j + [\tilde{\omega}^a \mathbf{I}^j - (\mathbf{I}^j \tilde{\omega}^a) + \mathbf{I}^j \tilde{\omega}^a] \dot{\boldsymbol{\beta}}^j + [\mathbf{I}^j \tilde{\omega}^a - (\mathbf{I}^j \dot{\boldsymbol{\omega}}^a) + \tilde{\omega}^a \mathbf{I}^j \tilde{\omega}^a - \tilde{\omega}^a (\mathbf{I}^j \boldsymbol{\omega}^a)] \boldsymbol{\beta}^j \quad (27)$$

Reference 1 or a report to be published (Footnote 2) should be consulted for the definition of any new symbols. Substitution of the expressions in Eqs. (20) and (23) for  $\bar{F}^{js}$  and  $\bar{T}^{js}$  into Eqs. (26) and (27) produces a complete set of dynamical equations for the appendage.

Combining of Eqs. (20), (23), (26), and (27) produces an equation of the form

$$\mathbf{M}' \ddot{\mathbf{q}} + \mathbf{D}' \dot{\mathbf{q}} + \mathbf{G}' \dot{\mathbf{q}} + \mathbf{K}' \mathbf{q} + \mathbf{A}' \mathbf{q} = \mathbf{L}' \quad (28)$$

where

$$\mathbf{q} = [u_1^1 u_2^1 u_3^1 \beta_1^1 \beta_2^1 \beta_3^1 \dots \beta_3^n]^T \quad (29)$$

is the matrix of nodal deformation coordinates.  $\mathbf{M}'$ ,  $\mathbf{D}'$ , and  $\mathbf{K}'$  are symmetric while  $\mathbf{G}'$  and  $\mathbf{A}'$  are skew symmetric and  $\mathbf{L}'$  is a  $6n \times 1$  matrix.

In more detail,

$$\mathbf{M}' = \mathbf{M} + \mathbf{M}^c - \mathbf{M}^* \quad (30)$$

where  $\mathbf{M}$  is null except for the  $3 \times 3$  matrices  $\mathbf{m}^1$ ,  $\mathbf{I}^1$ ,  $\mathbf{m}^2$ ,  $\dots$ ,  $\mathbf{I}^n$  associated with the nodal bodies along its principal diagonal,  $\mathbf{M}^c$  is the consistent mass matrix associated with the finite elements, and  $-\mathbf{M}^*$  accommodates the reduction of the effective inertia matrix due to mass center shifts within the vehicle induced by deformation.

Actually, the term  $D'\dot{q}$  does not result through combining Eqs. (20), (23), (26), and (27) but is an artifice introduced to account for structural damping. This procedure is common practice in structural dynamics analyses and its validity will not be discussed here.  $G'$  is a function of  $\omega^a$  and is referred to as the gyroscopic coupling term. Three types of terms contribute to  $K'$ . They are  $\bar{k}_0^s$ ,  $k_\Delta$ , and  $\bar{k}^s$ . All remaining terms in  $q$  comprise  $A'$ .  $L'$  is the matrix of externally applied forces and torques.

## 5. Summary

A brief synopsis of the dynamics analysis for an improved model of a spacecraft flexible appendage has been presented. The appendage matrix equation of motion, Eq. (28), is similar in form to Eq. (84) of Ref. 1. However, new interpretations must be given to each term. Equation (28) must be augmented by the vehicle equations of

motion as described in *Section IIIC* of Ref. 1 and by the control system equations described in *Section V* of Ref. 1.

Equation (28) is rather unwieldy as it stands. Simplifications must be made in order for the appendage equations to have real utility. Such simplifications are discussed in the latter part of a report to be published (Footnote 1). In particular, coordinate transformations are suggested which will permit the replacement of the homogeneous form of Eq. (28) with a set of completely uncoupled differential equations. Reducing the number of these equations by truncation, when justified, permits further simplification of the appendage equations:

## Reference

1. Likins, P. W., *Dynamics and Control of Flexible Space Vehicles*, Technical Report 32-1329. Jet Propulsion Laboratory, Pasadena, Calif., Jan. 15, 1970.

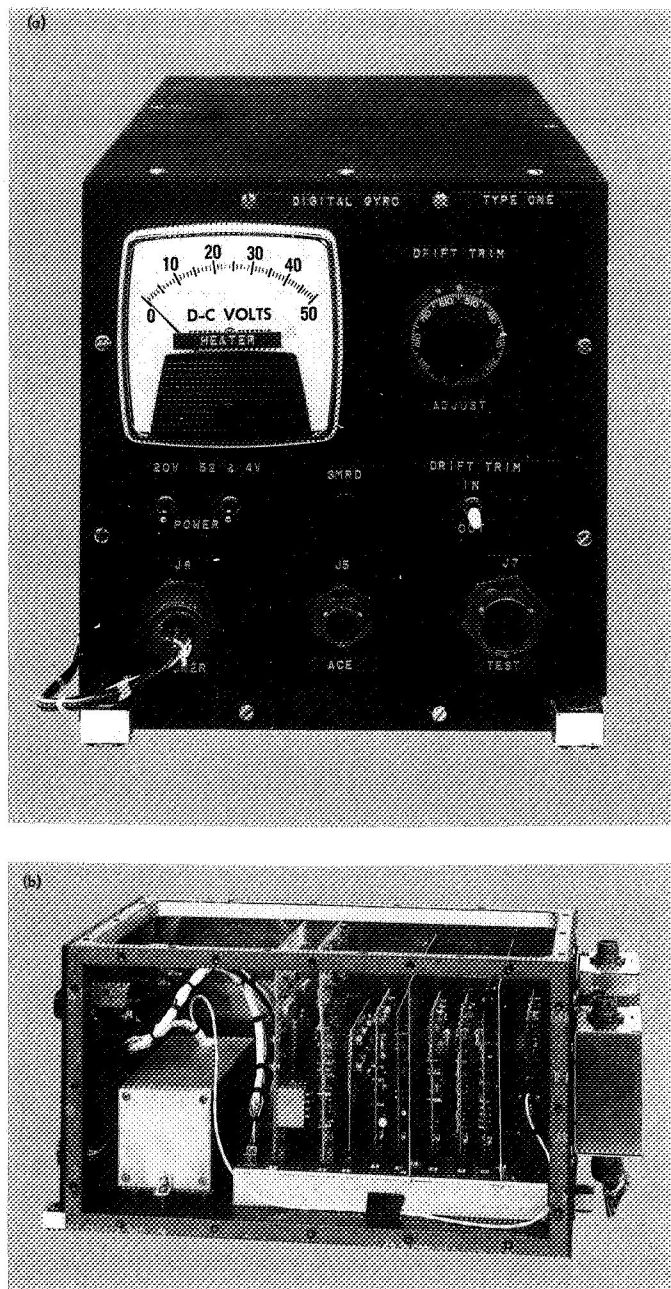
## C. Digital Gyro System (Phase II), P. J. Hand

A digital gyro system has been developed, built, and tested by the Inertial Sensors Group for the thermoelectric outer planet spacecraft (TOPS) single-axis simulator. The serial digital position information from this system will be used to stabilize the digital attitude-control system and will provide accurate commanded turn capability for the single-axis simulator.

The breadboard electrical design described in SPS 37-59, Vol. III, pp. 130-132, has been modified and packaged into a single unit containing the gyro, the electronic rebalance loop, and the dc-to-dc power converters. The system is self-contained and requires only 30-V, 0.7-A dc power. The output interface is a two-wire (positive  $\Delta\theta$  and negative  $\Delta\theta$ ) line to the attitude-control electronics package. The output circuits utilize low-power transistor-transistor logic integrated circuits with a pulse scale factor of 0.001 deg/pulse. Gyro capture capability is maintained up to 1.2 deg/s of input rate.

The gyro, described in SPS 37-59, Vol. III, is a Honeywell Inc., type GC334S, and was recently rebuilt by Honeywell for use with this digital system. The complete digital gyro system for installation on the single-axis simulator is shown in Fig. 1. The top and side covers were removed (Fig. 1b) to show the location of the internal parts. The gyro is in the rectangular box on the left, the plug-in electronics cards are in the center, and the two dc-to-dc converters are on the right, outside the main chassis. The dc-to-dc converters were designed and supplied by the Power Subsystems Group.

This complete gyro system has been extensively tested since August 1970. During the testing, the gyro was oriented as it will be in the simulator, as an azimuth gyro, i.e., with the gyro input axis vertical. As of October 15, 1970, 930 h of operating time had been accrued. Daily runs have been made to determine the drift characteristics of the gyro and the pulse scale factor of the gyro system. A plot of these data is shown in Fig. 2. The points



**Fig. 1. Digital gyro system: (a) front view, (b) internal view, top and side covers removed**

represent the average for one day's run, which may contain from 2 to 30 readings of the gyro's output with vertical earth's rate input. The curve plotted as composite drift rate error is the difference between the measured gyro output and the actual value of 8.45497 deg/h (JPL vertical earth rate component). This drift is actually due to the composite effects of gyro spin axis mass unbalance, fixed torques, and pulse scale factor variations. It is indicative of the stability of this gyro system on a day-to-

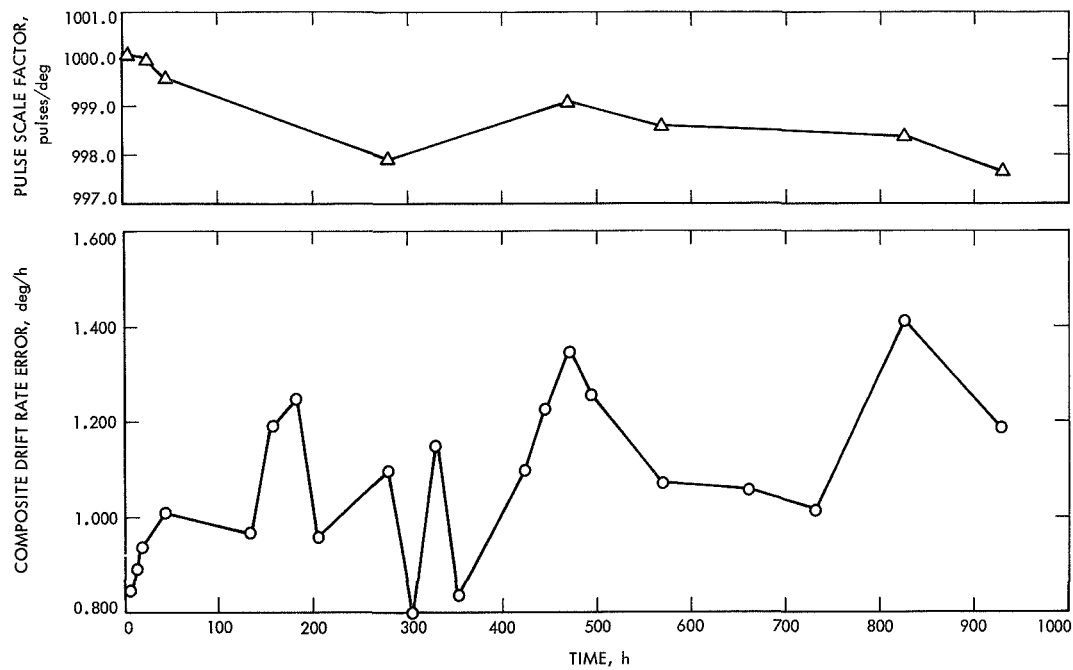
day basis. The data has a range of 0.60 deg/h, which is representative of the performance of the GC334S model gyros. An increasing trend is also present in the drift data, which was also noted in previous testing of other gyros of this model.

Observation of Fig. 2 indicates that the pulse scale factor is slowly decreasing with increasing operating time; however, this trend appears to be stabilizing. The pulse scale factor stability is affected by both the gyro command rate scale factor stability and the digital loop electronics stability. The digital loop parameters were monitored carefully during this test period and did not change, so it is probable that the pulse scale factor trend is due to changes in the gyro command rate scale factor.

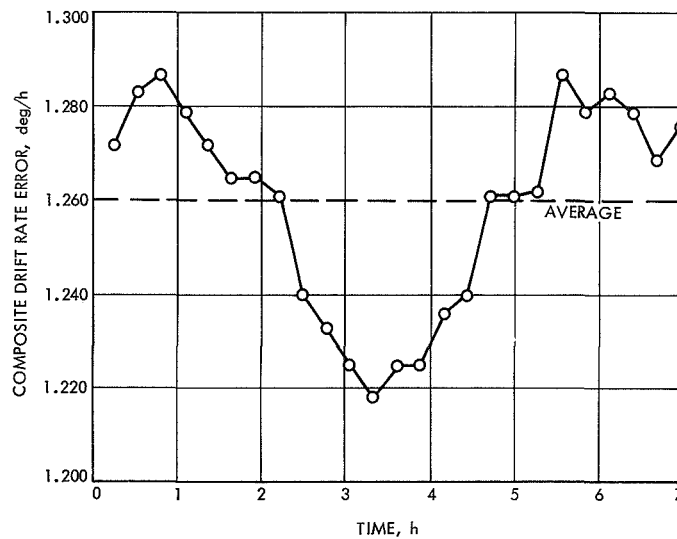
When used on the single-axis simulator, the gyro system will be trimmed to null-out both the effect of vertical earth's rate and the composite drift rate term. This will be accomplished by injecting a small analog current into the gyro torquer. The gyro will then only be sensing motions of the simulator. However, this nulling of the major drift input cannot take care of the short-term random drift effect of the gyro. Figure 3 is a plot of one typical day of drift, showing the random drift effect. The data were obtained by totalizing  $\Delta\theta$  pulses for 1000-s periods during a normal working day. Random drift is a non-systematic time varying component of drift rate measured under controlled operating conditions. It is normally expressed as the standard deviation of the gyro output over a specified period of time. The standard deviation of the one-day run shown in Fig. 3 is 0.021 deg/h. The data range is 0.069 deg/h and the average value is +1.260 deg/h.

It is planned that the gyro system will run continuously, even when the simulator is not operating, to eliminate the long warm-up time characteristics associated with this model gyro. The composite drift will then be adjusted to as near zero as is practical, just before a simulator run is started. In this manner, only the gyro short-term random drift will affect the simulator system. A random drift of 0.021 deg/h is acceptable for the simulator system application.

The Phase I system was developed to demonstrate this simplified digital torquing technique on the single-axis simulator. Continued development of this concept is planned, using newer, high reliability, low-power micro-circuitry. Simpler circuit designs will be investigated and the final circuit will be tailored to function with a smaller, lower-powered, and more accurate gyro than the GC334S used in the Phase I system.



**Fig. 2. Digital gyro system performance vs running time**



**Fig. 3. Digital gyro system drift vs time (1-day run time at start = 495.3 h; 1000-s integration time)**

## 1. Introduction

Commercially available stepper motor driver electronics have three characteristics that make them unacceptable for the thrust vector control. These are:

- (1) The windings are continuously powered, therefore, wasteful of power.
- (2) The electronics has a latch-up mode that is correctable only by removing the power supply voltage.
- (3) The reliability of the electronic components is unknown.

## 2. Circuit Description

- (1) Power driver switches to the motor windings.
- (2) Clock pulse generator.
- (3) Four-bit ring counter.
- (4) Logic for the ring counter.

The function of the clock circuit is to delay the clock pulse to the ring counter for a few microseconds, long enough for the ring counter logic to set.

The four-bit ring counter, along with the ring counter logic, transfers a logical "1" either clockwise or counterclockwise each time a clockwise or counterclockwise pulse appears at the input. The outputs of the ring counter are fed into the motor winding driver circuits in order to drive the stepper motor in either direction.

The ring counter and ring counter logic are complex in order to overcome the "latch-up" feature of commercial stepper motor drivers. This circuit has an error correcting logic in case an unacceptable state appears in the ring counter. Unacceptable states can appear in the ring counter as a result of turn-on transients and power supply transients. Table 1 shows the logical behavior of the ring counter for both acceptable and unacceptable states. The ring counter is comprised of series 54L73 dual J-K flip flops. The formulas for the J-K inputs are shown below:

$$\begin{aligned} J_A &= B\bar{D}E + D\bar{E}, & K_A &= 1 \\ J_B &= AD + A\bar{E} + \bar{A}\bar{D}E, & K_B &= \bar{A} + \bar{D} \\ J_C &= B\bar{D}\bar{E} + \bar{A}DE, & K_C &= \bar{J}_C \\ J_D &= \bar{A}\bar{B}\bar{E} + AE, & K_D &= 1 \end{aligned}$$

**Table 1. Ring counter logic**

$N$	Clockwise $N + 1$	Counterclockwise $N + 1$
<p><u>ABCD</u></p> <p>Acceptable states</p> <p>1000</p> <p>0100</p> <p>0010</p> <p>0001</p> <p>Unacceptable states</p> <p>0000</p> <p>1111</p> <p>1110</p> <p>1101</p> <p>1011</p> <p>0111</p> <p>0011</p> <p>1001</p> <p>1100</p> <p>0110</p> <p>1010</p> <p>0101</p>	<p><u>ABCD</u></p> <p>0001</p> <p>1000</p> <p>0100</p> <p>0010</p> <p>Acceptable states</p> <p>0100</p> <p>0100</p> <p>0001</p> <p>0100</p> <p>0100</p> <p>0100</p> <p>0010</p> <p>0010</p> <p>0100</p> <p>0001</p> <p>1000</p> <p>0001</p> <p>0010</p>	<p><u>ABCD</u></p> <p>0100</p> <p>0010</p> <p>0001</p> <p>1000</p> <p>Acceptable states</p> <p>0010</p> <p>0100</p> <p>0010</p> <p>0100</p> <p>0100</p> <p>0100</p> <p>1000</p> <p>1000</p> <p>0100</p> <p>0010</p> <p>0010</p> <p>0100</p> <p>1000</p>

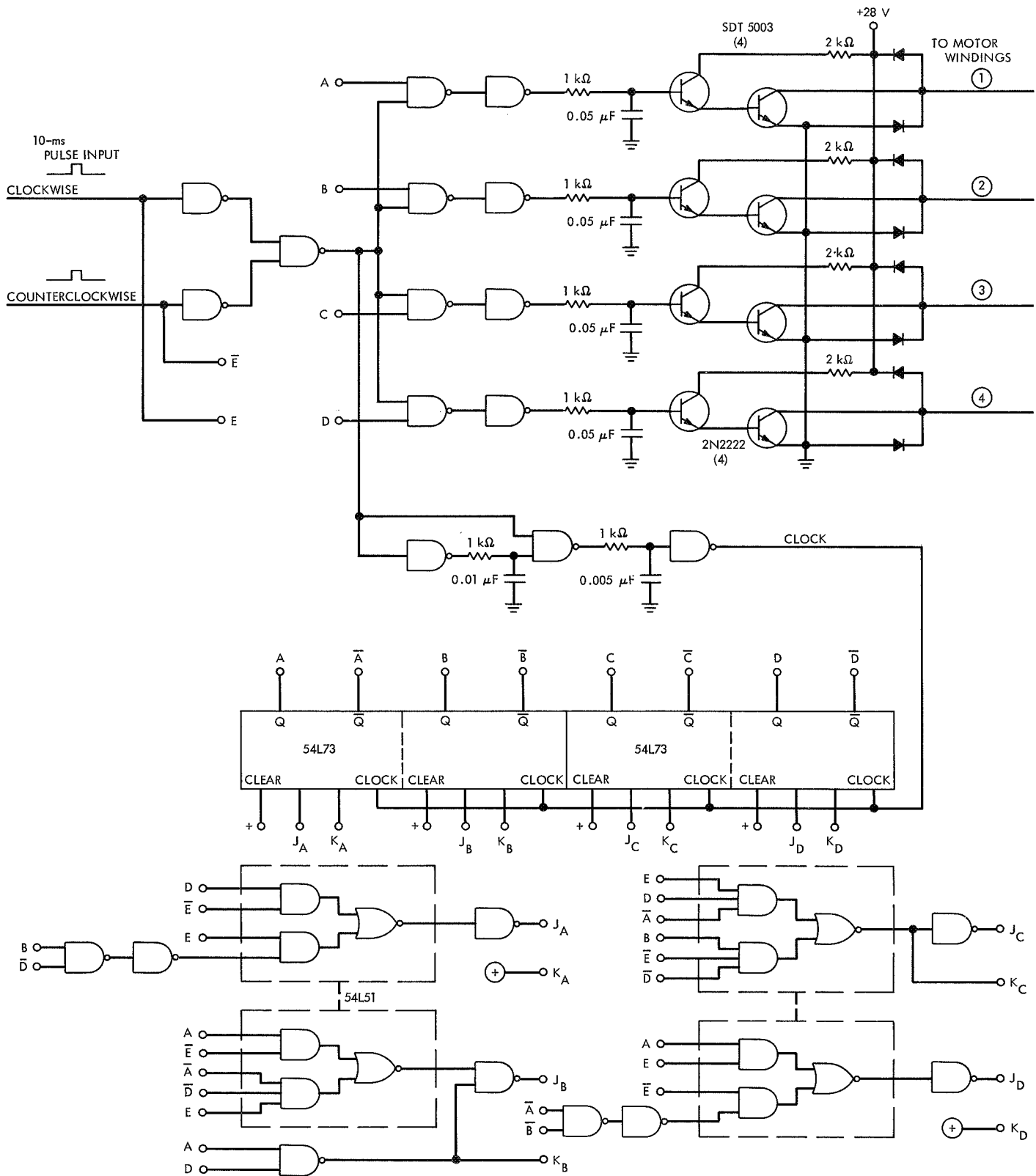


Fig. 1. Stepper motor driver electronics

### 3. Conclusion

The circuits described in this article have been breadboarded and tested and are now incorporated in the solar electric closed-loop thrust vector control system. This system uses six stepper motor driver electronics, two for translator actuators and four for the gimbal actuators.

Figure 2 shows two stepper motor driver electronics packaged on a printed circuit board. As can be seen from Fig. 2, each driver requires nine flat-packs, eight transistors, and a few additional discrete parts. The circuit design, as shown in Fig. 1, uses very little power, is free from "latch-up" states, and uses high reliability components.

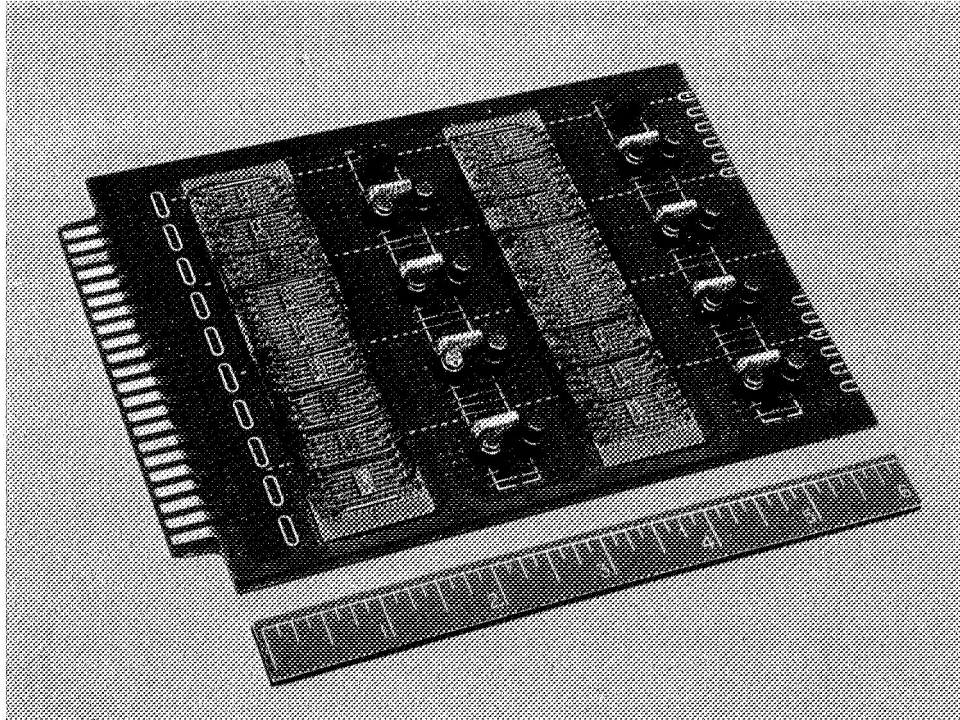


Fig. 2. Printed circuit board of two stepper motor driver electronics

# XI. Materials

## ENGINEERING MECHANICS DIVISION

### A. Spacecraft Adhesives for Long Life and Extreme Environment, W. D. Roper

#### 1. Introduction

The present state-of-the-art of high performance adhesives is being reviewed to establish those new adhesive materials which show promise as spacecraft adhesives for use in future planetary missions. In these future missions it is anticipated that structural adhesives will be required to function for extensive periods in environmental extremes ranging from above 500°F to -375°F. The performance requirements on adhesive materials will, therefore, be quite high and will go beyond what has thus far been required of spacecraft adhesives.

The studies performed have been divided into two phases. The first phase consisted of a state-of-the-art literature review in order to determine all the promising new materials which could be considered for future spacecraft use. This review was completed and reported previously (SPS 37-63, Vol. III, pp. 128-133). The second phase of the work unit consisted of a limited laboratory evaluation of the candidate adhesives which were selected from the Phase 1 review. The results of some of this laboratory evaluation are now being reported.

The polymeric materials that were selected for laboratory evaluation as adhesives were the polyimide (PI),

polybenzimidazole (PBI), and the polyquinoxaline (PQ) polymers. As adhesives these materials were evaluated for thermal shock resistance and for long-term aging at both elevated and cryogenic temperatures. The thermal shock testing was recently completed and the long-term aging tests are currently in process.

#### 2. Adhesive Thermal Shock Evaluation

*a. Test specimen preparation.* The two commercially available adhesives selected were the polyimide adhesive, FM-34 (American Cyanamid Co., Bloomington Dept.) and the polybenzimidazole, Imidite 850 (Whittaker Corp., Narmco Materials Division). Both of these adhesives were supplied as films supported by fiberglass carriers. The polyquinoxaline material was purchased as a resin solution (Whittaker Corp., Research and Development Division). This material required further processing in the laboratory in order to prepare it as a fiberglass-supported adhesive film. For comparison purposes, a typical epoxy paste-type adhesive, Epon 913 (Shell Chemical Co.), was also selected for testing along with the newer materials. This particular epoxy adhesive was selected because it had been used frequently in the past in spacecraft fabrication.

The selected adhesives were prepared into standard lap shear specimens and were tested according to the



ASTM D1002 (American Society for Testing and Materials) test procedure. The testing was performed to evaluate the metal-to-metal bond strength of each and the effect of high thermal shock on this bond strength. Since the cure temperatures and the anticipated service temperatures of the adhesives were well above 500°F, high temperature adherends were necessary. Consequently the adhesives were evaluated for the metal-to-metal bonding of titanium. The lap shear specimens were prepared using titanium (6Al-4V) adherends. In the case of the epoxy adhesive, which was used only for comparison purposes, a typical spacecraft aluminum alloy (6061-T6) was used.

**b. Thermal shock cycling.** The thermal shock testing performed in these studies was designed to simulate a condition typical of what might be expected in a near-sun type of mission (Mercury or Venus flyby). In this case, rapid structural temperature changes between 400°F and -100°F may be typical during a midcourse maneuver (Ref. 1). The thermal cycling used in these studies was therefore as shown in Fig. 1. In this exposure, specimens of each adhesive were given two thermal cycles between the 400°F and -100°F temperature extremes. Some *in situ* adhesion testing of the materials was also performed at the -100°F extreme as well as before and after the cycling. As seen from Fig. 1, the thermal shock rates experienced by the adhesives ranged between 25 to 55°F/min. A thermal shock of this order could be considered typical of a Venus-type mission (Ref. 1).

The procedure used in the cycling was as follows:

- (1) Lap shear test specimens of each adhesive were tested at room temperature in order to establish the room temperature strength of each of the adhesives. The specimens were tested according to the ASTM D1002 test method.
- (2) Approximately nine specimens of each adhesive were then placed in a preheated 400 ± 10°F oven.

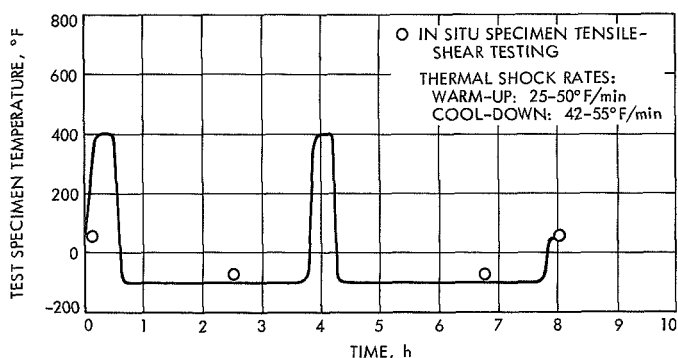


Fig. 1. Adhesive thermal shock testing

The warm-up of the specimens was monitored by a temperature recorder using thermocouples mounted on the specimens.

- (3) When the specimens reached 400 ± 10°F, they were allowed to soak at that temperature for 10-15 min. They were then removed and quickly placed inside a prechilled -100 ± 10°F chamber. The specimen cool-down was again continuously recorded during the cooling process.
- (4) After the specimens reached -100 ± 10°F, specimens of each material were pulled *in situ* on a tensile testing machine according to the ASTM D1002 test method.
- (5) Following the -100°F testing, the remaining specimens were again rapidly heated to 400 ± 10°F in the preheated oven. The temperature of the specimens was again continuously recorded.
- (6) After soaking at 400 ± 10°F for 10-15 min, the specimens were then removed from the oven and again rapidly chilled in the -100 ± 10°F chamber. Specimens of each material were then tested *in situ* at the -100°F temperature.
- (7) Following the second chilling and *in situ* testing, the last remaining specimens were then warmed to room temperature and tested at this temperature.

In addition to the preceding specimen testing, an attempt was also made to evaluate the creep resistance of the FM-34 and the PQ adhesives. To accomplish this, one creep test specimen of each adhesive was prepared according to the ASTM D2294 test procedure. The specimens were mounted in creep test fixtures and were each placed under a 750-psi tensile load. The specimens were then subjected to the same thermal shock as given to the lap shear specimens. The creep test specimens were observed for evidence of creep or bond failure during the exposure to the thermal shock.

### 3. Results and Discussion

The lap shear adhesion data obtained on each adhesive during the thermal shock testing are shown in Table 1. The data include lap shear strength of the adhesives at room temperature prior to cycling, at -100°F during cycling, and again at room temperature at the end of the two thermal cycles. These data permit some evaluation of the overall effect of the cycling as well as some evaluation of the adhesives at the lower temperature extreme. Resources precluded testing at the 400°F temperature extreme. Table 1 also shows the results obtained with the creep test specimens which were subjected to the same thermal cycling as the lap shear specimens.

Table 1. Thermal shock test data

Adhesive	Adherend	Tensile-shear strength, psi				Creep, mils
		Cycle 1		Cycle 2		
		75°F	-100°F	-100°F	75°F	
Imidite 850 (PBI)	Titanium 6Al-4V	1630	1160	1040	1100	Not tested
		1350	1130	1250	1100	
		1440	1380		1360	
		<u>1540</u>	—	—	—	
		Av 1490	1160	1150	1190	
FM-34 (PI)	Titanium 6Al-4V	2560	2720	2150	1850	0
		1620	2450	2640	1610 <sup>a</sup>	
		<u>2780</u>	<u>2780</u>	<u>3160</u>	<u>1420<sup>a</sup></u>	
		Av 2320	2650	2650		
Polyquinoxaline (PQ)	Titanium 6Al-4V	1690	1880	1220	1790	0
		2110	1430	1520	2310	
		<u>1260</u>	—	—	—	
		Av 1690	1660	1370	2050	
Epon 913 (epoxy)	Aluminum 6061-T6	3090	1720	1540	3920	Not tested
		4170	2060	1930	3860	
		<u>4360</u>	<u>1830</u>	—	<u>3790</u>	
		Av 3870	1870	1740	3860	

<sup>a</sup>Defective specimen.

<sup>a</sup>Defective specimen.

Of the newer adhesives tested, the polyimide, FM-34, gave the highest bond strength at room temperature. The polyquinoxaline was only slightly higher in strength than the polybenzimidazole, Imidite 850. The epoxy adhesive, Epon 913, had the greatest strength of all. However, it should be noted that the epoxy bonding was on aluminum, whereas the new high temperature adhesives were on titanium. Bond strengths are generally significantly lower on titanium substrates as compared to aluminum or steel alloys.

The data obtained on the polybenzimidazole adhesive (Imidite 850) were particularly interesting. The two thermal cycles resulted in an overall bond strength loss of approximately 20%. The data taken during the cycling at -100°F indicates that this loss occurred during the first cycle. After the first cycle the bond strength remained essentially constant at the 20% lower level. Statistical treatment of the data indicated that this strength loss was real to a better than 95% confidence level. The strength loss after the first cycle was apparently nonrecoverable when rewarmed to room temperature. No difference in the mode of specimen failure could be observed during the pulling of the specimens. The specimen failure was in all instances 90% or more cohesive (failure within the adhesive).

The cause of the strength loss can only be speculated. It is suspected that bond damage occurred during the

chilling phase of the first cycle due to thermal contraction effects. As noted earlier, in the adhesives review, some polybenzimidazoles may exhibit more than one glass transition temperature ( $T_g$ ). One PBI polymer has been found to have three, one of which is in the region of -94°F (Ref. 2). The PBI adhesive evaluated in these studies may also exhibit some low  $T_g$  value. If this is the case, this may account for the apparent loss in bond strength which occurred on chilling the adhesive to -100°F. If the material had passed through a low temperature  $T_g$ , then some bond damage may have occurred due to resultant physical property changes of the adhesive. Another cause may have simply been that the PBI structure was too rigid to adequately respond to the rapid thermal shock imposed. A complete investigation of the cause of the PBI strength loss during thermal cycling was beyond the scope of these studies.

The data obtained on the polyimide adhesive, FM-34, showed that the adhesive tensile-shear strength at room temperature and -100°F were essentially the same. No significant embrittlement of the adhesive, therefore, occurred at the low temperature. The mode of specimen failure with all the specimens was 90% or more cohesive. No creep was exhibited by the creep test specimens which were also thermal cycled with a 750-psi tensile load. Investigation of the two low tensile-shear values (1610 and 1420 psi) in the final room temperature tests revealed that these specimens were defective due to panel misalignment during bonding. These values were not, therefore, averaged as they were not representative of good specimens. Since all the other specimens tested throughout the cycling showed essentially the same strength, it was apparent that the thermal cycling had not affected the adhesive. This observation was also in line with the fact that all reported  $T_g$  values for polyimides are in excess of 500°F. Therefore, it did not seem likely that the chilling of the polyimide adhesive to -100°F should affect the adhesive performance due to passage through a  $T_g$  region.

The polyquinoxaline adhesive also did not seem affected by the thermal shock cycling. Its low temperature strength was essentially the same as that at room temperature. No creep was obtained with the creep test specimens which were thermal cycled with a 750-psi tensile load. The specimen failures during testing were approximately 40% cohesive. This indicated that the bond strength obtained was not maximized. Inexperience in the processing and formulation of the resin into an adhesive material was very likely a cause. No attempt was made to optimize the processing of this material into high strength bonds. It was conceivable that higher bond strengths can be

achieved with further development of this material as an adhesive.

The epoxy adhesive, Epon 913, was given the thermal shock cycling to compare its performance with that of the newer adhesive materials. As can be seen from the test data, the adhesive has high room temperature strength but it experiences considerable embrittlement at the lower ( $-100^{\circ}\text{F}$ ) temperature. This is not too surprising as epoxy adhesives are generally reported to have brittle points in the region of  $-70^{\circ}\text{F}$ . The mode of failure for the specimens pulled at room temperature was almost 100% cohesive. The mode was significantly different with the  $-100^{\circ}\text{F}$  specimens. At this low temperature, the specimens pulled only 30–40% cohesively. Embrittlement of the adhesive at the low temperature was, therefore, very evident. However, this embrittlement was reversible as the adhesive regained its room temperature strength after the cycling. The specimens pulled at room temperature after the two thermal cycles also pulled about 100% cohesively. The thermal cycling, therefore, did not apparently damage the adhesive.

#### 4. Conclusions

The thermal shock testing of the adhesives provided an effective screening for the determination of suitable spacecraft adhesives. On the basis of the shock testing, the order of decreasing preference of the adhesives for future spacecraft application was: polyimide, polyquinoxaline, and polybenzimidazole. The shock testing showed that the polyimide was by far the better adhesive. Its tensile-shear strength remained relatively unchanged within the range of  $75$  to  $-100^{\circ}\text{F}$ . It was also unaffected by the high rates of thermal shock to which it was subjected. The polyimide was also the most attractive on the basis of handling ease and required only moderately high cure temperatures and pressures. The polyimide could, therefore, be considered in structural bonding ap-

plications in future spacecraft where high thermal shock will be encountered.

The polyquinoxaline withstood the high thermal shock exposure. However, its bond strength on titanium was not particularly impressive. The PQ requires development into a fully compounded adhesive. Optimization of its formulation may then result in improved bond strengths. At present, it is the most difficult to process into adhesive bonds as it requires carrier prepregging and lengthy high temperature cures. Part of this handling could be reduced if the material ultimately achieves commercial availability as a fiberglass-supported film. The polybenzimidazole was not particularly attractive as a potential spacecraft adhesive. Its bond strength on titanium was found to be quite low and would not be sufficient for structural bonding. At the same time the material appeared to be affected by the thermal shock used in these studies. On the basis of the testing, the PBI showed no particular advantage over either the polyimide or polyquinoxaline adhesives.

As stated earlier, these same adhesives will be evaluated for their long-term aging at elevated and at cryogenic temperatures. Consequently, they are currently being exposed to  $500^{\circ}\text{F}$  at  $10^{-6}$  torr. It is planned that this exposure will be continued for a period of 5000 h. This testing will provide further characterization of these materials. Long-term aging tests at cryogenic ( $-375^{\circ}\text{F}$ ) temperature levels are planned for the future.

#### References

1. Ragsdale, G. C., and Mesnard, D. C., *Mariner Venus 67 Spacecraft Environmental Test Results*, Technical Report 32-1249. Jet Propulsion Laboratory, Pasadena, Calif., June 15, 1968.
2. Gillham, J. K., "Thermomechanical Behavior of a Polybenzimidazole System by Torsional Braid Analysis," *Am. Chem. Soc. Polymer Preprint*, Vol. 7, pp. 513–519, 1966.

## B. Cracking of Filter Layers in a High Performance Solar Cell Filter, W. Jaworski

### 1. Introduction

For spacecraft flights in the vicinity of Mercury or even closer to the sun, high performance filter covers are required to protect solar cells from exceeding the allowable temperature. Temperatures substantially higher than  $140^{\circ}\text{C}$  cause a steep degradation of solar cell power conversion; they also may induce some disintegration of cell interconnecting structure. The purpose of the filter is, therefore, to reduce as much as possible the solar energy outside of the visible (power convertible) wavelengths region (mostly infrared), and, in this way, keep the cell temperature within desired limits.

A filter of the described characteristic is generally an interference-type filter, normally of a complex design. It contains several stacks of filter layers, grouped to reflect certain predetermined bandwidths of solar irradiation. These layers are vacuum-deposited on a fused silica glass to form a filter cover. Each stack may consist of a considerable number of such layers that are composed of different materials. Thus, the compatibility of multilayer interference filters with the space environment (particularly thermal) becomes one of the major considerations.

In the Advanced Research and Development program related to solar arrays suitable for swingby Venus-Mercury missions, a number of solar cell-filter combinations have been experimentally investigated. Among them, a modified 4026 filter cover<sup>1</sup> mounted on a 2- $\Omega$ -cm solar cell was found to be the most promising. This filter has three multilayer stacks superimposed on one another, each containing a number of different layer materials.

During the exploratory 500-h test for ultraviolet and particle irradiation effects, several modified 4026 filter covers and solar cell-filter combinations were subjected to vacuum-temperature exposure. First, vacuum of approximately  $5 \times 10^{-7}$  torr was applied to the samples; then the temperature of the samples was raised from room temperature (approximately 15 to  $20^{\circ}\text{C}$ ) to about  $140^{\circ}\text{C}$ . As a result, mosaic-pattern cracks appeared. These cracks were more severe in the cell-filter assemblies than in the filter covers alone.

The purpose of this article is to discuss a probable cause of the damage and possible environmental incompatibility of the modified 4026 interference filter.

<sup>1</sup>Optical Coating Laboratory, Inc., designation.

### 2. Description and Examination of the Damage

The irradiation test setup in the test chamber consisted of eighteen samples; three samples were modified 4026 filter covers, three modified 4026 filter covers on 2- $\Omega$ -cm cells, and the remainder was an assortment of other filter covers and solar cell-filter combinations. As stated above, all of these samples were subjected to a vacuum and elevated temperature to establish the test conditions for starting the 500-h irradiation exposure.

Shortly before exposure to the irradiation fluxes, it was observed that the solar cell assemblies with modified 4026 filter exhibited local cracks (seen through the chamber window) which, by appearance, remained unaltered until the end of 500 h of testing. The actual in-depth location of the cracks within the cell-filter assembly could not be determined by visual inspection after the test, and further investigation was necessary to identify the surface of the cracks and their penetration. Figures 1a and b show the cell-filter assembly before and after the cracks appeared.

To determine the position of the crack surface in the solar cell-filter assembly (see Fig. 2), a Brinkmann vernier microscope was used. By precision focusing the following distances were measured:

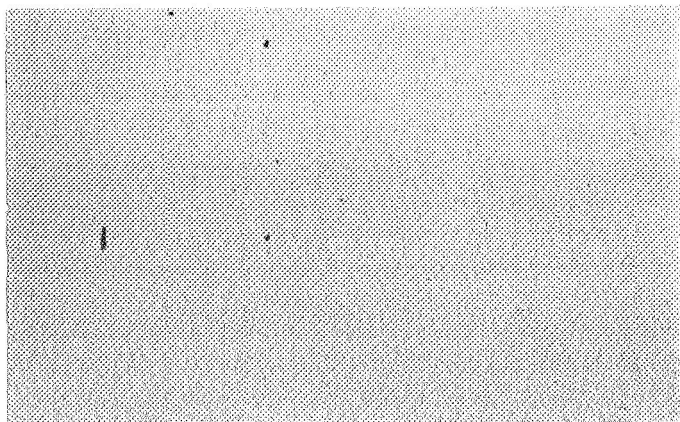
Glass top to apparent crack surface	350 $\mu\text{m}$
Glass top to cell surface	589 $\mu\text{m}$

Assuming an average index of refraction for the fused silica glass (wavelength between 0.4 and 0.7  $\mu\text{m}$ ) to be 1.46 (see Ref. 1), the real distance of crack surface from the top of the glass becomes  $y = 350 \times 1.46 = 511 \mu\text{m}$  ( $\sim 0.0201$  in.). This is exactly the thickness of the 20-mil glass used (considering tolerance). Thus, it was established that the surface of the crack was adjacent to the glass-filter interface. The difference of  $589 - 511 = 78 \mu\text{m}$  was the thickness of filter layer stacks and the adhesive (normally about 20 to 30  $\mu\text{m}$ ).<sup>2</sup>

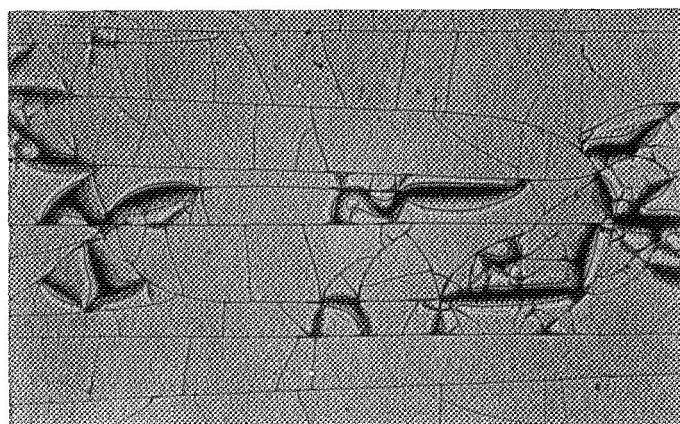
To establish crack penetration, an attempt was made to separate the filter cover from the cell by using an adhesive softener. This turned out to be a successful operation, and the separation was accomplished without disturbing the filter layers. Figure 1c shows the underside of the filter cover after separation. It is clearly seen that the crack pattern is identical with that shown in Fig. 1b.

<sup>2</sup>Distance from the glass top to the cell surface was measured directly, since the cell's edge projected outside the glass (see Fig. 2).

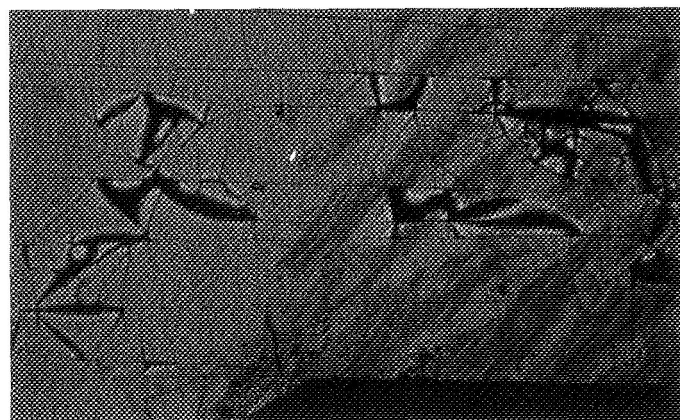
(a) CELL-FILTER BEFORE EXPOSURE (20 $\times$ )



(b) CELL-FILTER AFTER EXPOSURE (20 $\times$ )

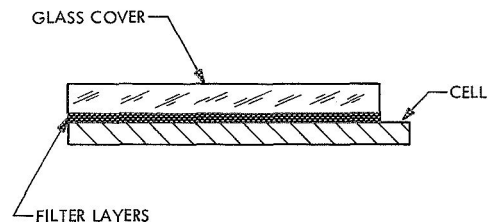


(c) UNDERSIDE OF FILTER COVER AFTER SEPARATION (20 $\times$ )



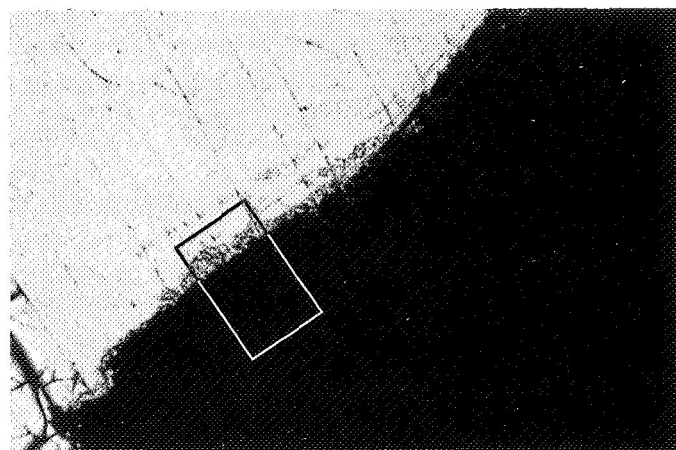
**Fig. 1. Cracking of filter layers in a cell-filter assembly during vacuum-temperature exposure**

By visual inspection of the underside of the filter cover, it was ascertained that all of the filter stacks cracked through along the pattern lines. However, it was still of some interest to know whether the surface of the glass

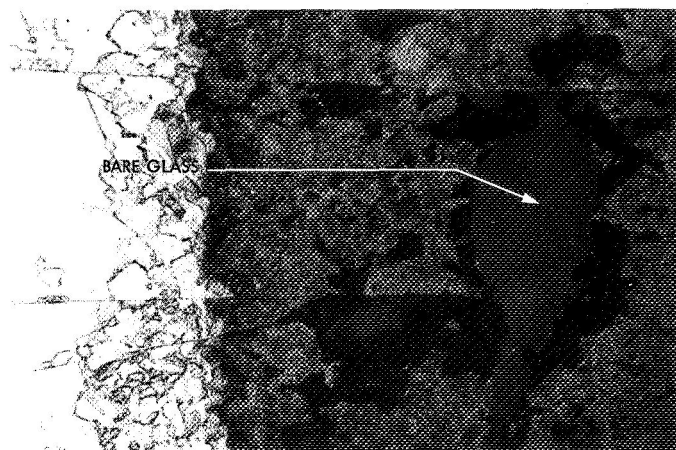


**Fig. 2. Section through solar cell-filter assembly**

(a) FILTER LAYERS PARTIALLY REMOVED (20 $\times$ )



(b) ENLARGEMENT OF BOXED AREA IN (a) (100 $\times$ )



**Fig. 3. Partial removal of filter layers to expose the bare glass**

in contact with the filter was in any way affected. To determine this, the filter layers were partially removed from the glass by scraping. Figure 3a shows a portion of the underside of the filter cover (shown in Fig. 1c) over which filter layers were partially scraped (very dark part). In this scraped dark area there are several spots in which the filter layers were completely removed from the glass,

some of them extending across the crack pattern lines seen in the unscrapped part (light portion of the picture).

Figure 3b shows a magnified view of the boxed area indicated on Fig. 3a which shows one of the bare glass spots (uniform dark grey pear-shaped area on the right side of the picture). It is seen that where this spot extends across the filter crack line, the line does not extend through the bare glass, signifying no cracking in the glass.

Similar filter cracking was also discovered in the samples representing modified 4026 filter cover alone. The crack lines, however, were less pronounced and probably caused by lower stresses than those present in a cell-filter assembly. Figure 4 shows these cracks in a filter cover sample.

### 3. Evaluation and Discussion

The presented cracking phenomena of modified 4026 filter layers under vacuum-temperature exposure reveal an undesirable design characteristic. Although the cracks did not grow as exposure continued, there is no guarantee that when thermal shocks, such as those occurring during planet occultation or spacecraft maneuvers, are applied, extensive damage to a solar array would not occur. The cracking was probably a result of initial stresses set in during manufacture of this very complex filter, aggravated by an exposure to substantial temperature change.

On the plus side of this evaluation is some evidence that the cracking of filter layers does not always take place. Two subsequent simulation tests, made in different facilities, have not produced any cracking in modified 4026 cell-filter assemblies. This would imply that initial manufacturing stresses or some imperfections in vacuum deposition of filter layers could be primarily responsible

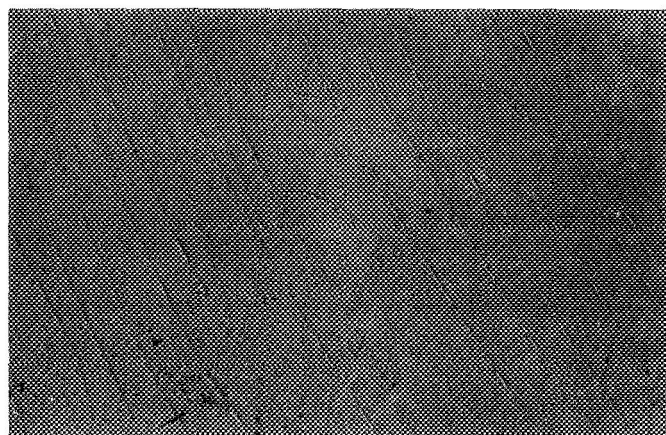


Fig. 4. Cracking of filter layers in a filter cover during vacuum-temperature exposure (20 $\times$ )

for the cracking. However, filter layers, cover glass, and solar cell are all made of different materials or groups of materials; thus the compatibility problem appears to be rather complex.

This filter has very desirable optical characteristics which do not change significantly when cracking, such as described here, does take place. Therefore, the stability of the extent of cracking would probably be of more interest to solar array designers than a complete prevention of cracking. In any event, more research is needed to clearly pinpoint the cause of cracking, and to find a solution to this problem.

### Reference

1. Malitson, T. H., "Interspecimen Comparison of the Refractive Index of Fused Silica," *J. Opt. Soc. Am.*, Vol. 55, No. 10, pp. 1205-1209, Oct. 1965.



## **C. Development of a Conical-Gregorian High-Gain Antenna, J. G. Fisher**

### **1. Introduction**

In SPS 37-63, Vol. III, pp. 124-128, initial efforts in the construction of a furlable conical-Gregorian high-gain antenna are described. The first furlable model used 0.005-in.-thick hard-rolled stainless steel as the reflective membrane, which must bend to a conical surface in the furled state (see Fig. 7 on p. 127 of SPS 37-63, Vol. III). As was noted in the referenced SPS, the 0.005-in.-thick reflective membrane would not retain the required conical surface when the antenna was rotated to an axis-horizontal position for RF testing in a one-g field. Several schemes for supporting the membrane were considered. In general, these employed circumferential stiffening members which would deploy to lie perpendicular to the reflective surface when the antenna was opened. Practical fabrication difficulties left each of these schemes open to question. It was decided to try using a thicker, and thus stiffer, material for the membrane itself, and thus avoid the use of extra members. The problems associated with the design, construction, and testing of an antenna built in this way are discussed.

### **2. Design Considerations**

To make the reflective surface essentially self-supporting, it appeared advisable to use as thick a material as would accept the bending required for furling without exceeding the elastic limit of the alloy chosen. A low modulus of elasticity/yield strength ratio would thus be desirable, with low density a secondary characteristic. A sample of 0.020-in.-thick 2024-T3 aluminum alloy was bent to the radius observed in the 0.005-in. stainless steel model, and it returned to flat on release. It was decided to use this material for the test.

### **3. Construction Details**

The required eight panels were cut using the same template used for the stainless steel model. However, at the request of the Telecommunications Division, extra material was left at the inner edges to pick up RF spill-over which had been observed during testing of the earlier configuration. As in the later-model stainless membrane antenna, the hinge pins were located in the plane of the reflective surface, to avoid opening at the ribs. Also, in an effort to avoid bending loads across the hinge-ribs, and thus retain a more precise surface, the panels were pre-curved to the deployed conical shape. Final assembly was made on the previously used plaster mandrel using epoxy adhesive (Scotchweld 2216).

### **4. Furling Test Results**

The first furling test was made manually, since a mechanism was not available. The antenna, with its subdish and supports in position, was oriented face up, and the ribs raised simultaneously. About half-way to the furled position, the epoxy adhesive bond between the membrane panels and the hinge-rib flanges began to break loose, especially at the inside ends of the panels. The test was stopped, and the antenna redeployed. Examination of the structure showed no plastic deformation, and it was concluded that the thicker aluminum membrane was merely introducing more stress into the bond joints than the adhesive could hold. The panels were therefore removed from the hub and rivets used to join the membranes to the hinge flanges. During the initial furling, it had been noticed that the ribs tended to flex in the direction of furling, as though being overpowered by the aluminum membrane. It was decided to use the same rivets to fasten angles to the hinge flanges and thus stiffen them against this bending. Because the hinges had to close in the furled position, only small angles, about a quarter inch high, could be used, but by offsetting them slightly, they could be made to nest in the furled position.

With these modifications, the furling test was resumed. This time it was found that considerable force was required to close the antenna, especially as the hinge-ribs were pushed between the subdish supports. The aluminum panels did not curve in the same way as the stainless steel ones. The latter had bent easily to a horseshoe shape, and had slid without difficulty between the subdish supports. The thicker aluminum curved in a more circular bend, and could be pushed between the legs only with considerable force. Also, in the stainless steel model, the sharpest curve had been at the outer periphery of the panels. The aluminum panels, having been extended toward the center of the antenna, were forced into a tighter bend at the inner edges. Once forced into furled position, the antenna would not deploy itself, even in the one-g field. The panels were tightly held by the subdish supports. Pulled past this binding point, considerable strain energy was released and, without holding the ribs as they opened, the assembly almost certainly would have been damaged.

After redeployment it was found that the inner portions of the panels had plastically deformed, and upward bumps were seen at the inner edges of each panel. An effort to relieve these by slitting along each hinge flange was unsuccessful. It was also found that, despite the small angles which had been incorporated into the hinges, the

aluminum panels were overpowering the ribs, producing a generally upward bend. Some thought was given to perforating the panels to weaken them and, at the same time, reduce weight. It was decided that too many other problems existed, and that a new design approach would be needed.

## 5. New Design

Experience with the thick, self-supporting membrane concept had not been promising. While early work with the 0.005-in.-thick stainless steel membrane had indicated it might work in zero *g*, testing in earth environment is highly desirable. Several concepts, including membranes intermediate in thickness to those already tried, and incorporating various stiffening members, were considered and rejected as too complex. However, the chances of success appeared good if a furlable outer ring could be provided, supported from the center hub by a minimum of radial ribs. Now, a compliant membrane made of strips to eliminate circumferential loads could be used as the reflector surface. The strips would be stretched between the center hub and the outer support ring. Experimentation with small models showed such a design concept is feasible. In this new design, the outer ring has

a channel cross section, and is supported from the center hub by three radial ribs located 120 deg apart. In furling, the ribs raise to a vertical position, and the portions of the channel between the ribs loop downward. The channel legs are "hinged" at each corner of the channel section, to permit this deformation without plastic yielding. For the membrane reflector surface, aluminized Kapton was considered. It has better resistance to ultraviolet radiation than Mylar, but for the first experiment Mylar was chosen since it was readily available. A preliminary test series on  $\frac{1}{4}$ -,  $\frac{1}{2}$ -, and 1-mil Mylar films showed that 5-gram/cm loading was adequate to stretch the material. It was decided that the  $\frac{1}{2}$ -mil material showed most promise for the first antenna construction. Tensile tests of 2.5-cm-wide samples showed no tendency to develop wrinkles when stressed an order of magnitude over the 5-gram/cm loading. A sample was loaded unidirectionally to 15 grams/cm, and clamped in an aluminum frame  $38 \times 7.6$  cm with the stress in the long direction. The assembly was immersed in liquid nitrogen to see if the material would stretch or rupture. It did not, and there was no change in apparent condition upon return to room temperature. Design of the new antenna has been completed and construction of a prototype is expected to begin shortly.



## XII. Applied Mechanics

### ENGINEERING MECHANICS DIVISION

#### A. Creep Failure of Randomly Excited Structures,

*J.-N. Yang*

##### 1. Introduction

Many structures are intended to sustain, in addition to static loadings, dynamic excitations of random nature during their service life. In addition to the interior pressure, spacecraft pressure vessels are subjected to steady-state random disturbances as well as a number of severe and highly transient random excitations resulting from each flight event, such as booster engine ignition, booster engine cut-off, etc., during the mission. The investigation of creep failure under random excitations is, therefore, of practical importance as well as theoretical interest. Unfortunately, little has been done in this area, because of the lack of a unified theory to characterize the complicated phenomenon of creep rupture. Thus, in practice, in designing a spacecraft pressure vessel using the fracture mechanics approach, one takes into account the static pressure loading only, and allows for the dynamic random loadings by introducing a safety factor.

It is the purpose of this article to develop a method by which one can predict quantitatively the reliability of

pressure vessels under both static loadings and random dynamic disturbances. This article establishes the relationship between the safety factor and the structural reliability, which is essential in a rational design of pressure vessels. Thus, by specifying the required level of reliability, one can determine, through this relationship, an appropriate safety factor for the design.

From the fracture mechanics viewpoint, the process of creep rupture is the slow growth of flaws in a structure, called subcritical flaw growth, until the critical fracture stress associated with the largest flaw is exceeded by the applied stress and then catastrophic failure occurs. Indeed, fracture mechanics is probably the most appropriate criterion for designing a pressure vessel (e.g., Refs. 1 and 2) and it is employed in this study. Under static sustained pressures, the characteristics of subcritical flaw growth associated with a particular material and a particular environmental condition can be obtained from the experimental results (Refs. 1 and 2). Cumulative flaw growth (or damage) hypothesis, which is analogous to the Palmgren-Miner rule, is employed to generalize the sustained flaw growth characteristics under dynamic random excitation.

The mean value and the variance of the relative flow growth under both stationary and nonstationary random excitations are obtained where creep failure is assumed to take place when the relative flow growth reaches unity. The probability of creep failure is then evaluated using the Janey's principle of maximum entropy (Ref. 3).

Two numerical examples, one for stationary random excitations and another for nonstationary random excitations, are worked out to illustrate the general results. It is shown that the probability of creep failure decreases rapidly as the safety factor increases.

## 2. Sustained Flow Growth Characteristics

Let  $A$  be the flaw size at some critical points of the structure. According to the well-known Griffith-Irwin equation,

$$A = Q (K_{IC}/R)^2 \quad (1)$$

$$A = Q (K_I/\underline{S})^2 \quad (2)$$

where  $K_{IC}$  is the critical stress intensity factor,  $R$  is the critical fracture stress or the resisting stress associated with the flaw size  $A$ ,  $K_I$  is the stress intensity factor associated with the flaw size  $A$  and the applied stress  $\underline{S}$ , and  $Q$  is an appropriate state parameter (Refs. 1, 2).

Equations (1) and (2) are the basic equations relating the flaw size  $A$  to the applied stress  $\underline{S}$  or the critical fracture stress  $R$ . Many structures, such as pressure vessels, fracture not only when a certain critical stress is reached but also at a relatively low stress after being subjected to a certain time of sustained loading. Such cases involve subcritical flow growth, i.e., the slow growth of the flaw until it reaches its critical value, and then catastrophic failure occurs. The subcritical flow growth under sustained loading can be conveniently described by curves of  $K_I/K_{IC}$  versus time to failure  $t_f$ . These curves are determined from tests on preflawed specimens by varying both the initial flaw size and the applied stress (Refs. 1, 2). Figure 1 gives typical experimental results and curves, which are available for many materials under different environmental conditions.

These experimental curves have characteristic trends indicating that, for a given material and environmental condition, practically no flow growth occurs until a threshold value of  $K_I/K_{IC}$  is reached. In Fig. 1, this threshold value is about 0.755 for the aluminum in a nitrogen environment. It is further observed that these curves have a characteristic shape which can be approximated by

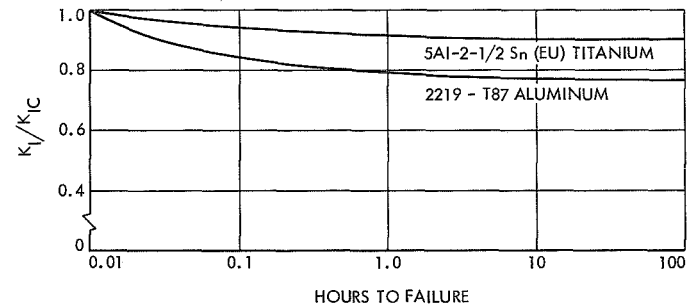


Fig. 1. Sustained load flow growth curves for 2219-T87 aluminum and 5Al-2-1/2 Sn (EU) titanium at  $-320^\circ\text{F}$  and  $-423^\circ\text{F}$  (Ref. 1)

an exponential decay to the threshold value (log scale in time). Thus

$$K_I/K_{IC} = b + \gamma \exp \left[ -\frac{1}{\beta} \log t_f \right] = b + \gamma t_f^{-1/\alpha} \quad (3)$$

in which  $b$  is the threshold values,  $\beta$  is the characteristic time in logarithmic scale,  $\alpha = \beta/0.43429$ ,  $\gamma = K_I/K_{IC} - b$  for  $t_f = 1$ , and  $t_f$  indicates the time to fracture (failure). Note that  $b$ ,  $\gamma$ , and  $\alpha$  depend not only on the material properties but also on the environmental conditions, such as temperature and surrounding medium, etc.

Substitution of Eqs. (1) and (2) into Eq. (3) yields

$$\underline{S}/R = b + \gamma t_f^{-1/\alpha} \quad (4)$$

which is the relationship among the applied stress  $\underline{S}$ , the resisting stress of materials  $R$  (critical fracture stress), and the time to failure  $t_f$ .

## 3. Cumulative Flow Growth Hypothesis

Both Eq. (4) and Fig. 1 are obtained from the experimental results under static sustained loadings. To generalize the flow growth characteristics under dynamic or random loadings, similar assumptions as the well-known Palmgren-Miner rule are made in the following:

(1) The relative flow growth  $\delta_i$  due to a sustained stress  $S_i$  for a period of time  $\Delta t_i$  is given by,

$$\begin{aligned} \delta_i &= \Delta t_i / t_f(S_i), & S_i/R &\geq b \\ &= 0, & \text{otherwise} \end{aligned} \quad (5)$$

in which  $t_f(S_i)$  is the time to failure under the constant sustained stress  $S_i$  (Eq. 4) and no subcritical flow growth

is expected when  $K_I/K_{IC}$  or  $S_i/R$  is below the threshold value  $b$ .

(2) The relative flaw growth  $\delta$  due to sustained stresses  $S_1, S_2, \dots, S_n$  for periods of time  $\Delta t_1, \Delta t_2, \dots, \Delta t_n$ , respectively, is

$$\delta = \sum_{i=1}^n \delta_i = \sum_{i=1}^n \Delta t_i / t_f(S_i) \quad (6)$$

(3) Fracture occurs when the relative flaw growth  $\delta$  reaches unity.

Let  $\underline{S}(t)$  be a random stress response process with a constant mean value  $\bar{S}$  and standard deviation  $\sigma(t)$ ,

$$\underline{S}(t) = S(t) + \bar{S} \quad (7)$$

where  $S(t)$  is the fluctuating part of  $\underline{S}(t)$  with mean zero and standard deviation  $\sigma(t)$ . The mean value  $\bar{S}$  represents the stress response of pressure vessels due to static pressure loadings, while  $S(t)$  represents the random stress fluctuation resulting from random disturbances, such as transient excitations due to booster engine cut-off. It is further assumed that the maximum standard deviation  $\sigma^*(t)$  of  $S(t)$  is small compared to the mean stress response  $\bar{S}$ , so that the fatigue failure mode is negligible and the negative stress application of  $S(t)$  can be disregarded. This assumption appears to be quite reasonable for pressure vessels.

Now define the relative flaw growth  $d\delta$  due to  $S(t)$  in the time interval  $[t, t + dt]$  as

$$d\delta = dt/t_f(\underline{S}(t)), \quad \underline{S}(t)/R \geq b \\ = 0, \quad \text{otherwise} \quad (8)$$

in which no subcritical flaw growth is expected when  $K_I/K_{IC}$  or  $\underline{S}(t)/R$  is below the threshold level  $b$ .

Then, the relative flaw growth  $\delta(T)$  due to  $\underline{S}(t)$  within the time interval  $[0, T]$  follows from Eq. (8) as

$$\delta(T) = \int_0^T d\delta = \int_0^T Z(t) dt \quad (9)$$

where

$$Z(t) = 1/t_f(\underline{S}(t)), \quad \underline{S}(t)/R \geq b \\ = 0, \quad \text{otherwise} \quad (10)$$

and fracture occurs when  $\delta(T)$  reaches unity.

With the aid of Eqs. (4) and (7), Eq. (10) can be written as

$$Z(t) = C [S(t) - S_c]^\alpha, \quad S(t) \geq S_c \\ = 0, \quad \text{otherwise} \quad (11)$$

where

$$\left. \begin{aligned} C &= (1/\gamma R)^\alpha \\ S_c &= \bar{S}(\nu - 1) \\ \nu &= bR/\bar{S} \end{aligned} \right\} \quad (12)$$

In Eq. (12),  $\nu$  is the ratio of the structural resisting stress  $R$ , multiplied by the threshold level  $b$ , to the mean applied stress  $\bar{S}$ . In the pressure vessel design, it is called a safety factor, since for  $\nu \geq 1$ , the mean applied stress  $\bar{S}$ , or the static pressure loading, does not produce the subcritical flaw growth and hence the creep rupture. In designing a pressure vessel, the safety factor  $\nu$  is usually made greater than one, and hence  $S_c \geq 0$  to ensure that no creep failure results from the static pressure loading. Therefore, the case in which  $S_c \geq 0$  is of practical importance, and it will be assumed in the following study.

The resisting stress  $R$ , of the structure will be considered as a deterministic constant in this study. This is an approximation when the statistical dispersion of the material strength is very small. If, however, the statistical dispersion of the material strength cannot be neglected, one usually has to perform the proof load test to eliminate weak strength vessels. As a result, the proof stress is regarded as  $R$  in practice. This approximation is conservative (Refs. 4-6 and Footnote 1).

#### 4. Mean Value and Variance of the Relative Flaw Growth

Consider the fluctuating part of the stress response  $S(t)$  a gaussian process with the probability density function  $f(x)$  and the joint density function  $f(x, y; \rho)$ ,

$$f(x) = \exp(-x^2/2\sigma_1^2)/\sqrt{2\pi}\sigma_1 \quad (13)$$

$$f(x, y; \rho) = \frac{1}{2\pi\sigma_1\sigma_2(1-\rho^2)^{1/2}} \exp\left\{-\frac{x^2 - 2\rho xy + y^2}{2\sigma_1\sigma_2(1-\rho^2)}\right\} \quad (14)$$

<sup>1</sup>Heer, E., and Yang, J.-N., "Structural Optimization Based on Fracture Mechanics and Reliability Criteria," to appear in *AIAA Journal*.

where

$$\begin{aligned}\sigma_1^2 &= \sigma_1^2(t_1) = E[S^2(t_1)] \\ \sigma_2^2 &= \sigma_2^2(t_2) = E[S^2(t_2)] \\ \rho\sigma_1\sigma_2 &= \rho(t_1, t_2) \sigma_1(t_1) \sigma_2(t_2) = E[S(t_1)S(t_2)]\end{aligned}$$

If  $S(t)$  is stationary,  $\sigma_1 = \sigma_2 = \sigma_s = \text{constant}$  and  $\rho(t_1, t_2) = \rho(t_1 - t_2) = \rho(\tau)$ .

Following Cramer (Ref. 7),  $f(x, y; \rho)$  can be expanded into a series,

$$f(x, y; \rho) = \frac{1}{\sigma_1\sigma_2} \sum_{n=0}^{\infty} \frac{\rho^n}{n!} \Phi^{(n+1)}\left(\frac{x}{\sigma_1}\right) \Phi^{(n+1)}\left(\frac{y}{\sigma_2}\right) \quad (15)$$

in which  $\Phi(x)$  is the standardized gaussian distribution function and  $\Phi^{(n)}(x)$  is the  $n$ th derivative of  $\Phi(x)$  with respect to  $x$ .

**a. Stationary random excitation.** When  $S(t)$  is stationary with standard deviation  $\sigma_s$ , the mean value,  $\bar{\delta}(T)$ , of the relative flaw growth,  $\delta(T)$ , within the time interval  $[0, T]$ , follows from Eqs. (9), (11), and (13) as

$$\begin{aligned}\bar{\delta}(T) &= E[\delta(T)] = \int_0^T E[Z(t)] dt \\ &= TC \int_{S_c}^{\infty} (x - S_c)^{\alpha} f(x) dx\end{aligned} \quad (16)$$

With the aid of Eq. (12), one obtains

$$\bar{\delta}(T) = T(C\sigma_s^{\alpha}) \frac{1}{\sqrt{2\pi}} \exp(-\lambda^2/4) \Gamma(\alpha + 1) D_{-(\alpha'+1)}(\lambda) \quad (17)$$

in which  $\Gamma(\cdot)$  and  $D(\cdot)$  are, respectively, the gamma function and the parabolic cylinder function, and

$$\begin{aligned}\lambda &= S_c/\sigma_s = \bar{S}(\nu - 1)/\sigma_s = (\nu - 1)/V \\ C\sigma_s^{\alpha} &= (\sigma_s/\gamma R)^{\alpha} = (Vb/\gamma\nu)^{\alpha}\end{aligned} \quad (18)$$

where

$$V = \sigma_s/\bar{S} \quad (19)$$

is the coefficient of variation of  $\underline{S}(t)$ .

For the special case in which  $\nu = 1.0$  and hence  $S_c = 0$ , one can show that

$$\bar{\delta}(T) = T(C\sigma_s^{\alpha}) (2)^{\alpha/2} \Gamma \frac{1}{2\sqrt{\pi}} T(C\sigma_s^{\alpha}) (2)^{\alpha/2} \cdot \Gamma\left(\frac{\alpha}{2} + \frac{1}{2}\right) \quad (20)$$

Therefore, the average time to fracture (creep failure)  $T^*$  follows from Eq. (17) as

$$T^* = \sqrt{2\pi}/(C\sigma_s^{\alpha}) \exp(-\lambda^2/4) \Gamma(\alpha + 1) D_{-(\alpha+1)}(\lambda) \quad (21)$$

The variance,  $\sigma_{\delta(T)}^2$ , of  $\delta(T)$  can be obtained from Eqs. (9), (11), (14) and (17) as

$$\sigma_{\delta(T)}^2 = C^2 \int_0^T \int_0^T \int_{S_c}^{\infty} \int_{S_c}^{\infty} (x - S_c)^{\alpha} (y - S_c)^{\alpha} f(x, y; \rho) dx dy dt_1 dt_2 - \bar{\delta}^2(T) \quad (22)$$

Substituting Eqs. (15) and (17) into Eq. (22) and inverting the order of integration and summation, one can show that the first term associated with  $n = 0$  is cancelled by the term  $\bar{\delta}^2(T)$ . For  $\alpha = \text{integer}$ , using the stationary condition and repeating the integration by parts with respect to each term of the series, one can show that

$$\sigma_{\delta(T)}^2 = 2T(C\sigma_s^{\alpha})^2 (\alpha!)^2 \left\{ \sum_{j=1}^{\alpha} \frac{G(j; T) F^2(j, \lambda)}{j! [(\alpha - j)!]^2} + \sum_{j=1}^{\infty} \frac{G(j + \alpha; T) H_{j-1}^2(\lambda) \phi^2(\lambda)}{(\alpha + j)!} \right\} \quad (23)$$

in which

$$F(j, \lambda) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\lambda^2}{4}\right) \Gamma(\alpha - j + 1) D_{-(\alpha-j+1)}(\lambda) \quad (24)$$

$$G(j; T) = \int_0^T \left(1 - \frac{\tau}{T}\right) \rho^j(\tau) d\tau \quad (25)$$

where  $H$  and  $\phi$  are, respectively, the Hermit polynomial and the standardized gaussian density function. The relationship

$$\Phi^{(j+1)}(x) = (-1)^j H_j(x) \phi(x) \quad (26)$$

has been employed in deriving Eq. (23).

In statistical terminology, the random variable of the relative flaw growth  $\delta(T)/C$  is called "exceedance measure" (Ref. 7). It can be observed from Eqs. (9) and (11) that, for  $\alpha = 0$ ,  $\delta(T)/C$  is the proportion of time in  $[0, T]$  that the fluctuating process  $S(t)$  spends above the level  $S_c$ . For  $\alpha = 1$ ,  $\delta(T)/C$  is just the area cut off by the process  $S(t)$  above the level  $S_c$  in  $[0, T]$  as shown in Fig. 2.

The correlation coefficient  $\rho(\tau)$  usually dies down to zero as  $\tau$  becomes large, so that  $\rho(\tau) \approx 0$  for  $\tau \geq t^*$ , say, e.g.,  $t^* = 10$  min. In practice, one is usually interested in the creep failure for  $T \gg t^*$ , say, e.g.,  $T > 1$  h. Hence,  $G(j; T)$  in Eq. (25) can be approximated by

$$G(j; T) \approx G(j) = \int_0^{t^*} \rho^j(\tau) d\tau = \int_0^\infty \rho^j(\tau) d\tau \quad (27)$$

which is not a function of  $T$ .

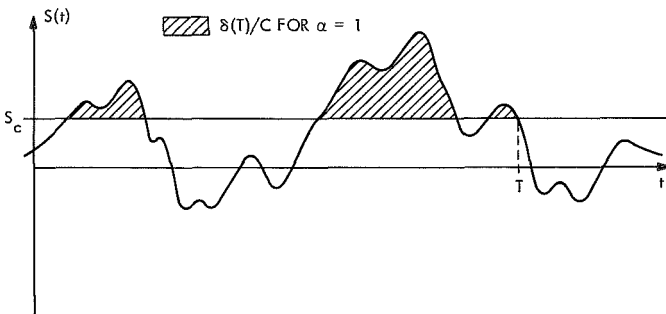


Fig. 2. Measure of relative flaw growth  $\delta(T)/C$  for  $\alpha = 1$

Therefore, for large  $T$ , the standard deviation,  $\sigma_{\delta(T)}$ , increases in proportion to the square root of  $T$ . It follows from Eqs. (17), (23) and (27) that the coefficient of variation (dispersion),  $V_{\delta(T)}$ , of the relative flaw growth  $\delta(T)$  can be expressed as

$$V_{\delta(T)} \sim \frac{W}{\sqrt{T}} \quad \text{as } T \rightarrow \text{large} \quad (28)$$

where

$$W = \frac{2\sqrt{\pi}(\alpha!)}{\exp\left(-\frac{\lambda^2}{4}\right) \Gamma(\alpha+1) D_{-(\alpha+1)}(\lambda)} \times \left\{ \sum_{j=1}^{\alpha} \frac{G(j) F^2(j, \lambda)}{j! [(\alpha-j)!]^2} + \sum_{j=1}^{\alpha} \frac{G(j+\alpha) H_{j-1}^2(\lambda) \phi^2(\lambda)}{(\alpha+j)!} \right\}^{1/2} \quad (29)$$

Equation (28) states that the dispersion of  $\delta(T)$  decreases in inversed proportion to the square root of  $T$ .

**b. Nonstationary random excitations.** When  $S(t)$  is a nonstationary gaussian process, the standard deviation  $\sigma(t)$ , is a function of time and the correlation coefficient  $\rho(t_1, t_2)$  depends on  $t_1$  and  $t_2$ . In a similar fashion as for the stationary process, one can show that the mean value of  $\delta(T)$  is

$$\begin{aligned} \bar{\delta}(T) &= C \int_0^T \int_{S_c}^{\infty} (x - S_c)^{\alpha} f(x) dx \\ &= (CS_c)^{\alpha} \frac{\Gamma(\alpha+1)}{\sqrt{2\pi}} \int_0^T \left(\frac{\sigma(t)}{S_c}\right)^{\alpha} \exp\left\{-\left(\frac{S_c}{2\sigma(t)}\right)^2\right\} \\ &\quad \times D_{-(\alpha+1)}\left(\frac{S_c}{\sigma(t)}\right) dt \end{aligned} \quad (30)$$

in which

$$(CS_c)^{\alpha} = [b(v-1)/\gamma v]^{\alpha} \quad (31)$$

The variance  $\sigma_{\delta(T)}^2$  can be obtained from Eq. (22) as follows: after substituting Eqs. (15) and (30) into Eq. (22) and inverting the order of integration and summation,

one repeats the integration by parts for each term of the series to yield, for  $\alpha = \text{integer}$ ,

$$\begin{aligned} \sigma_{\delta(T)}^2 &= (CS_c)^{2\alpha} (\alpha!)^2 \int_0^T \int_0^T \left( \frac{\sigma_1}{S_c} \right)^\alpha \left( \frac{\sigma_2}{S_c} \right)^\alpha \\ &\times \left\{ \sum_{j=1}^{\alpha} \frac{\bar{F}(j; t_1) \bar{F}(j; t_2)}{j! [(\alpha - j)!]^2} \rho^j(t_1, t_2) \right. \\ &+ \sum_{j=1}^{\infty} \frac{\rho^{j+\alpha}(t_1, t_2)}{(\alpha + j)!} H_{j-1} \\ &\times \left( \frac{S_c}{\sigma_1} \right) H_{j-1} \left( \frac{S_c}{\sigma_2} \right) \phi \left( \frac{S_c}{\sigma_1} \right) \phi \left( \frac{S_c}{\sigma_2} \right) \Big\} dt_1 dt_2 \end{aligned} \quad (32)$$

where

$$\begin{aligned} \bar{F}(j; t_1) &= \exp \left\{ - \left( \frac{S_c}{2\sigma_1} \right)^2 \right\} \Gamma(\alpha - j + 1) D_{-(\alpha-j+1)} \\ &\times \left( \frac{S_c}{\sigma_1} \right) / \sqrt{2\pi} \end{aligned} \quad (33)$$

## 5. Probability of Failure

With the mean value  $\bar{\delta}(T)$  and the variance  $\sigma_{\delta(T)}^2$  computed in the previous section, the probability of creep failure can be estimated using the Janey's principle of maximum entropy (Ref. 3). Since creep failure occurs when  $\delta(T)$  reaches unity, the probability of creep failure  $P_f(T)$  within the time interval  $[0, T]$  can be shown (Ref. 3) as

$$P_f(T) = \exp(-1/\bar{\delta}(T)) \quad (34)$$

when only the information of  $\bar{\delta}(T)$  is available. If both  $\bar{\delta}(T)$  and  $\sigma_{\delta(T)}^2$  are given,  $P_f(T)$  can be shown as

$$\begin{aligned} P_f(T) &= \frac{2}{\sqrt{\pi}} \int_1^\infty \exp(-a_0 - a_1 x - a_2 x^2) dx \\ &= \frac{1}{\sqrt{a_2}} \exp \left\{ -a_0 + \frac{a_1^2}{4a_2} \right\} \operatorname{erfc} \left( \sqrt{a_2} + \frac{a_1}{2\sqrt{a_2}} \right) \end{aligned} \quad (35)$$

with

$$\left. \begin{aligned} a_1 &= (Z - 2X^2)/\bar{\delta}(T) \\ a_2 &= a_1^2/4X^2 \\ a_0 &= (a_1^2/4a_2) - \frac{1}{2} \ln a_2 + \ln \left[ \operatorname{erfc} \left( \frac{a_1}{2\sqrt{a_2}} \right) \right] \end{aligned} \right\} \quad (36)$$

where  $X$  and  $Z$  appearing in above equations can be obtained graphically from the following equations

$$Z = 2X \exp(-x^2)/\sqrt{\pi} \operatorname{erfc}(X)$$

$$\sigma_{\delta(T)}^2/\bar{\delta}^2(T) = (2X^2 + 2X^2Z - Z^2)/(4X^4 - 4X^2Z + Z^2) \quad (37)$$

In Eqs. (35) and (37),  $\operatorname{erfc}(\cdot)$  represents the complementary error function.

## 6. Numerical Examples

Consider the displacement response  $Y(t)$ , of a structure which is related to the excitation  $g(t)$ , as follows:

$$\ddot{Y}(t) + 2\zeta\omega_0\dot{Y}(t) + \omega_0^2 Y(t) = g(t) \quad (38)$$

where  $\zeta$  and  $\omega_0$  are, respectively, the damping coefficient and the natural frequency of the structure. The stress response process  $S(t)$  is linearly related to the displacement response  $Y(t)$  as

$$S(t) = KY(t) \quad (39)$$

where  $K$  is a suitable constant. In addition to the excitation  $g(t)$ , the structure is also subjected to a deterministic static loading which produces a static stress  $\bar{S}$  in the structure.

The subcritical flaw growth characteristics of Fig. 1 for 2219-T87 aluminum is employed in this example. This curve can be fitted accurately by the following set of parameter values appearing in Eq. (3);  $b = 0.755$ ,  $\gamma = 0.038$ ,  $\alpha = 2.5$ .

**Example 1:** Let  $g(t)$  be a stationary white noise  $n(t)$  with a mean square spectral density  $S_0$ . Then the standard deviation,  $\sigma_s$ , of the stress response  $S(t)$  is

$$\sigma_s = K(\pi S_0/2\zeta\omega_0^3)^{1/2} \quad (40)$$

and the correlation coefficient  $\rho(\tau)$  is

$$\rho(\tau) = \exp(-\xi\omega_0\tau) \left[ \cos \omega\tau + \frac{\xi}{\sqrt{1-\xi^2}} \sin \omega\tau \right] \quad (41)$$

where

$$\omega = \sqrt{1-\xi^2} \omega_0$$

It is assumed that  $\omega_0 = 2 \text{ rad/s}$ ,  $\xi = 0.02$  and the coefficient of variation  $V = \sigma_s/\bar{S} = 0.05$ . It can be observed that  $\rho(\tau)$  becomes negligible for  $\tau > 10 \text{ min}$  and hence  $G(j)$  appearing in Eq. (27) can be used in approximation for  $T > 1 \text{ h}$ .

It is now possible to find the mean value  $\bar{\delta}(T)$  (Eq. 17) and the standard deviation  $\sigma_{\delta(T)}$  (Eq. 23) of the relative flaw growth as functions of safety factor  $\nu$  and time  $T$ . The average flaw growth per hour,  $\bar{\delta}(T)/T$ , and the measure of the coefficient of variation,  $W = V_{\delta(T)} \sqrt{T}$  (Eq. 28), with  $G(j)$  given in Eq. (27), are plotted against  $\nu$  in Fig. 3 as the solid and dashed curves, respectively. From Fig. 3,  $\bar{\delta}(T)$  and  $\bar{\delta}(T) \pm 3\sigma_{\delta(T)}$  can be plotted against time  $T$  for different values of  $\nu$  and is shown in Fig. 4, in which  $T^*$  represents the average time to failure. It can be observed that  $T^*$  increases rapidly with respect to the increase of the safety factor. The criterion for failure is  $\delta(T) = 1$ . Fig. 4 indicates that, given a specific time instant  $T$ , there is a distribution of  $\delta(T)$  across the ensemble of sample histories. The probability of failure can, therefore, be computed using the principle of maximum entropy (Eq. 35) and is plotted in Fig. 4 (dotted curves).

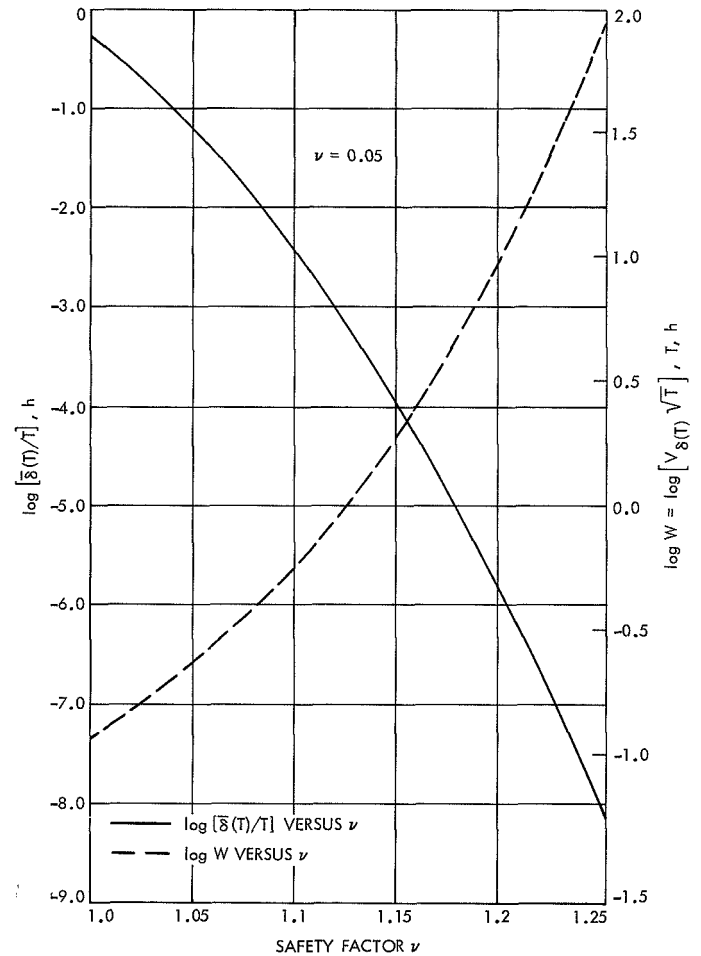


Fig. 3. Expected relative flaw growth per hour,  $\bar{\delta}(T)/T$  and measure of dispersion  $V_{\delta(T)} \sqrt{T}$  as function of safety factor  $\nu$

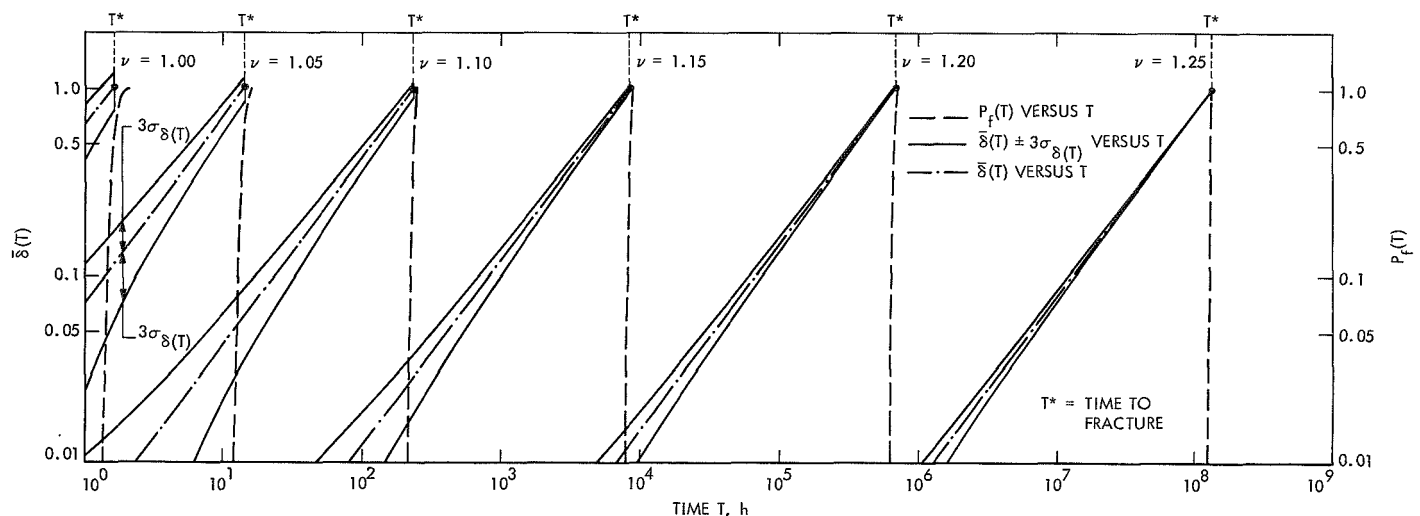


Fig. 4. Expected relative flaw growth  $\bar{\delta}(T)$ ,  $\bar{\delta}(T) \pm 3\sigma_{\delta(T)}$ , and probability of failure  $P_f(T)$  as function of time  $T$

### Example 2: Nonstationary Random Excitation

Consider the same structure (Eq. 38) where, however,  $\omega_0 = 31.4$  rad/sec  $\zeta = 0.05$ , and the excitation  $g(t)$  is a nonstationary random process obtained as an output of a filter due to a nonstationary shot noise  $\psi(t)n(t)$  with  $\psi(t)$  being a deterministic function of time,

$$g(t) = \int_{-\infty}^{\infty} h_0(t-\tau) \psi(\tau) n(\tau) d\tau \quad (42)$$

$$\psi(t) = H(t) [e^{-\beta_1 t} - e^{-\beta_2 t}] \quad (\beta_2 > \beta_1 > 0) \quad (43)$$

$$h_0(t) = H(t) [e^{-\mu_b t} \sin \omega_b t] / \omega_b \quad (44)$$

in which  $h_0(t)$  is the impulse response function of the filter,  $H(t)$  is a unit step function, and  $n(t)$  is a white noise with a mean square spectral density  $S_0$ . The significance and the application of the input  $g(t)$  given in Eq. (42) was discussed in detail in Ref. 8. It is pointed out here that by a proper choice of parameter values in-

volved,  $g(t)$  can represent a strong earthquake acceleration or a severe transient gust loading.

Using Eqs. (38) and (42) to (44), the variance  $\sigma^2(t)$  and the correlation coefficient  $\rho(t_1, t_2)$ , of the stress response process  $S(t)$  can be obtained (Ref. 9). In the present example, a choice of parameter values, which has been employed to investigate the first-excursion probability (Ref. 8) and the cumulative fatigue damage (Ref. 10), is used as follows:  $\beta_1 = 0.25$ /sec,  $\beta_2 = 0.5$ /sec,  $\omega_b = 12.3$  rad/sec,  $\mu_b = 3.86$ /sec, and  $S_0 = 7.19 \times 10^4$  in./sec. It is further assumed that  $K/R = 3.0$  per inch. With this set of parameter values,  $g(t)$  represents a transient excitation which dies down to zero after  $t \geq 15$  sec (Refs. 8, 9, Footnote 2). The mean value of the relative flaw growth (Eq. 30), denoted by  $\bar{\delta}$  for  $T \geq 15$  sec, is computed for different value of safety factor  $\nu$  and plotted in Fig. 5 as the solid curve. The probability of creep failure is then computed using Eq. (34) and plotted in Fig. 5 as the dashed curve. The results of Fig. 5 indicate that the probability of creep failure decreases rapidly as the safety factor increases. It is worthwhile to note that, in practice, a safety factor  $\nu = 1.25$  is used in designing spacecraft pressure vessels indicating a high level of reliability.

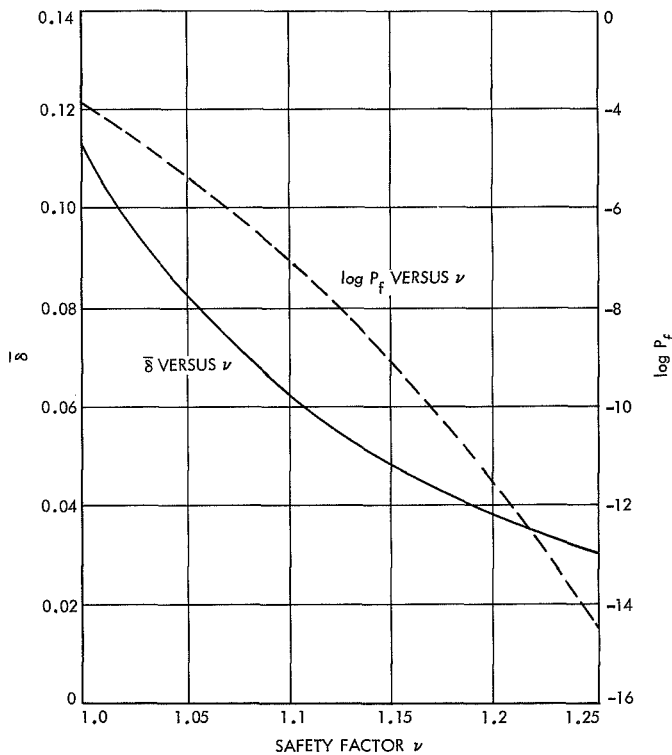


Fig. 5. Expected relative flaw growth  $\bar{\delta}$  and probability of failure  $P_f$  as function of safety factor  $\nu$  (nonstationary random excitation)

### References

1. Tiffany, C. F., Masters, J. N., and Pall, F. A., *Some Fracture Considerations in Design and Analysis of Spacecraft Pressure Vessels*, ASM Technical Report C6-2.3, 1966.
2. Masters, J. N., "Cyclic and Sustained Flaw Growth Characteristics of 6Al-4V Titanium," NASA CR-92231, July, 1968.
3. Tribus, M., "The Use of Maximum Entropy Estimate in the Estimation of Reliability," in *Recent Development in Information and Decision Processes*, pp. 102-140. Edited by R. E. Machol and P. Gray, Macmillan Co., New York, 1962.
4. Shinozuka, M., and Yang, J.-N., "Optimum Structural Design Based on Reliability and Proof Load Test," *Annals of Assurance Science*, Vol. 8, pp. 375-391, 1969.
5. Shinozuka, M., Yang, J.-N., and Heer, E., "Optimum Structural Design Based on Reliability Analysis," *Proceedings of the Eighth International Symposium on Space Technology and Sciences*, pp. 245-258, Tokyo, Japan, Aug. 1969.

<sup>2</sup>Shinozuka, M., and Yang, J.-N., "Peak Structural Response to Nonstationary Random Excitations," accepted to appear in *Journal of Sound and Vibration*.



6. Heer, E., and Yang, J.-N., "Optimum Pressure Vessel Design Based on Fracture Mechanics and Reliability Criteria," *Proceedings of the ASCE-EMD Specialty Conference on Probabilistic Concepts and Methods in Engineering*, pp. 102-106, Purdue University, Nov. 1969.
7. Cramer, H., and Leadbetter, M. R., *Stationary and Related Stochastic Process*, John Wiley and Sons, Inc., New York, 1967.
8. Shinozuka, M., and Yang, J.-N., "On the Bound of First-Excursion Probability," *J. EMD, Proc. ASCE*, Vol. 95, pp. 363-377, 1969.
9. Shinozuka, M., and Yang, J.-N., "Random Vibration of Linear Structures," *Int. J. Solids and Structures*, Vol. 5, pp. 1005-1036, Oct. 1969.

## B. Reliability of Randomly Excited Structures, J.-N. Yang and E. Heer

### 1. Introduction

Structural reliability under random excitations is a well-recognized problem in aerospace engineering. In-flight measurements taken during the launch and exit phase of recent space flights (Ref. 1) show that spacecraft excitations and responses contain to a large degree steady-state random components in addition to certain deterministic phenomena. Aircraft structural loads and excitations as a result of engine noise and boundary layer turbulence are usually also classified as stationary random. The requirement of studying structural reliability and failure under stationary random excitations is therefore apparent.

Because of mathematical expediency, three essential modes of structural failure are usually investigated separately: catastrophic, fatigue, and creep.

**a. Catastrophic failure.** This type occurs instantaneously due to some response measures, such as displacements or stresses, that exceed their limiting values. This type of failure is also referred to as the first passage failure, and the associated probability of catastrophic failure is called the first passage or the first excursion probability. Although, the exact solution for a simple structural system excited by white noise has not been obtained (Ref. 2) many approximate solutions have been proposed (Refs. 3-4, Footnote 1).

**b. Fatigue failure.** This type occurs due to the successive incremental reduction of the limiting response measures as a result of flaw extension with each load repetition. Fatigue failure estimations are usually made

using the well-known Palmgren-Miner rule, e.g., Refs. 5, 6, or using the fracture mechanics approach, e.g., Ref. 7.

**c. Creep failure.** This type occurs due to the continuous time-dependent reduction of the limiting response measures as a result of viscous flow phenomena in each sustained loading time interval. Very little information is available for the estimation of creep failures under random excitation.

Thus far, in random vibration, the material properties of structures have not been considered as random variables, and the relevancy of structural reliability to optimum structural design has not been investigated thoroughly. On the other hand, structural reliability under static or quasistatic random loadings, taking into account statistical variations of material properties, has been investigated quite extensively. The relevancy of structural reliability, coupled with proof testing and structural strength deterioration (fatigue and creep), to optimum structural design has been investigated, e.g., in Refs. 8-10 and Footnote 2.

In this article, the concepts of fracture mechanics and the point processes associated with stationary narrow-band random processes (Footnote 1) are employed to estimate the reliability of structures under stationary random excitations. Two failure modes are allowed for, i.e., catastrophic and fatigue failures, and the statistical variations of the material properties are taken into account. A quantitative investigation is conducted to determine the interaction of the failure modes and the relevancy to structural reliability estimates. The premises of this study are as follows: (1) catastrophic failure occurs as soon as the stress response process exceeds the critical

<sup>1</sup>Yang, J.-N., and Shinozuka, M., "On the First Excursion Probability in Stationary Narrow-Band Random Vibration," to appear in the *Journal of Applied Mechanics*.

<sup>2</sup>Heer, E., and Yang, J.-N., "Structural Optimization Based on Fracture Mechanics and Reliability Criteria," to appear in the *AIAA Journal*.

fracture stress (threshold level) of the structure, and (2) successive incremental reduction of the critical fracture stress (threshold level), as a result of stress repetition, increases the catastrophic failure rate in time. Therefore, the problem treated here is a first passage problem with a monotonically decreasing probabilistic barrier level that depends not only on the dispersion of the material properties but also on the history of the random response process.

## 2. Strength Deterioration of Structures

One way of explaining structural failure mechanisms and processes, such as fatigue failures is by the use of the concepts of fracture mechanics. The mechanisms of fatigue failure can then be summarized as follows: (1) flaw initiation, (2) flaw propagation, and (3) catastrophic failure. The last two mechanisms of the fatigue process are of primary concern in structural analysis and design. The flaw initiation stage is the one about which little is known at present. It can usually be assumed, however, that the material already has flaws that propagate under repetitive stress applications with sufficiently high amplitude. The fatigue failure process is thus described by the growth of flaws in a structure until the applied stress at any flawed point exceeds the critical fracture stress associated with the flaw at that critical point, and catastrophic failure occurs.

Let  $A$  represent the flaw size at the critical point of the structure. According to the well-known Griffith-Irwin equation

$$A = Q \left( \frac{K_{IC}}{R} \right)^2 \quad (1)$$

$$A = Q \left( \frac{K_I}{S} \right)^2 \quad (2)$$

where  $K_{IC}$  is the critical stress intensity factor,  $R$  is the critical fracture stress or the resisting stress associated with the flaw size  $A$ ,  $K_I$  is the stress intensity factor associated with the flaw size  $A$  and the applied stress  $S$ , and  $Q$  is an appropriate state parameter.

The flaw propagation law takes the form (Refs. 7, 11-12)

$$\frac{dA}{dn} = C K_I^b \quad (3)$$

stating that the rate of flaw extension with respect to the number of stress cycles  $n$  is proportional to the  $b$ th power

of the stress intensity factor  $K_I$ , where  $C$  is a suitable constant. By theoretical analyses as well as direct measurements of flaw propagations (Refs. 7, 11, 12) it has been shown that  $b$  ranges from 2 to 4, depending on the material used and environmental conditions such as temperature, corrosion, etc.

Substituting Eq. (2) into Eq. (3), one obtains

$$\frac{dA}{dn} = K S^b A^{b/2} \quad (4)$$

where  $K = C/Q^{b/2}$  is a constant.

In the following approach, it is assumed, for convenience and without essential loss of generality, that  $b = 2$ ; it will be shown later that this applies also to the cases  $b = 3$  or 4. Note that  $b$  is held arbitrary for  $S$ , which is a parameter in Eq. (4).

Integrating Eq. (4) successively with respect to each cycle and summing, one obtains

$$\ln A_n - \ln A_0 = K \sum_{j=1}^n S_j^b \quad (5)$$

where  $S_j$  is the  $j$ th peak of the stress response  $S(t)$ . The initial flaw size,  $A_0$ , and the flaw size after  $n$  stress cycles,  $A_n$ , are respectively related to the initial resisting stress,  $R_0$ , and the resisting stress after  $n$  stress cycles,  $R_n$ , through Eq. (1). Hence, substituting Eq. (1) into Eq. (5) yields

$$R_n = R_0 \exp(-Z) \quad (6)$$

$$Z = \frac{1}{2} K \sum_{j=1}^n S_j^b \quad (7)$$

It is clear from Eqs. (6) and (7) that the structural resisting stress after  $n$  stress cycles,  $R_n$ , decreases monotonically with respect to the number of stress cycles  $n$ .

## 3. Extreme Point Processes

Consider  $S(t)$  a stationary narrow-band gaussian stress response process with mean zero and a mean square spectral density  $\phi(\omega)$ . If  $S_m$  denotes the  $m$ th local maximum (peak) of  $S(t)$ , then the local maxima form a stationary point process called the maximum point process,  $\{S_n\}$  (Footnote 1).

The distribution function  $F_{S_n}(x)$  of the maximum point process  $\{S_n\}$  is the Rayleigh distribution

$$F_{S_n}(x) = 1 - \exp \left\{ \frac{-x^2}{2\sigma_S^2} \right\} \quad (8)$$

in which  $\sigma_S$  is the standard deviation of  $S(t)$ .

Because of the narrow-band characteristics of  $S(t)$ , the joint density function

$$f_{S_j S_{j+m}}(x, y)$$

of  $S_j$  and  $S_{j+m}$  spaced at a distance of  $m$  cycles can be approximated by the joint density function  $f_S(x, y; \tau)$  of the envelope function of  $S(t)$  spaced at a distance  $\tau = mT_0$ , where  $T_0$  is the central period of  $S(t)$ .

$$f_{S_j S_{j+m}}(x, y) = f_S(x, y; mT_0) \quad (9)$$

It has been shown in Ref. 13 that

$$f_S(x, y; \tau) = \frac{xy}{\sigma_S^4 [1 - k_0^2(\tau)]} \times I_0 \left[ \frac{xy k_0(\tau)}{\sigma_S^2 [1 - k_0^2(\tau)]} \right] \exp \left\{ \frac{-(x^2 + y^2)}{2\sigma_S^2 [1 - k_0^2(\tau)]} \right\} \quad (10)$$

in which  $I_0[\cdot]$  is the zero-order modified Bessel function of the first kind, and  $k_0(\tau)$  is a function of  $\phi(\omega)$  and its central frequency  $\omega_0$  (Footnote 1).

#### 4. Failure Rate and Reliability

The probability that a structure, having survived  $n$  stress cycles, will fail in the  $(n+1)$ th stress cycle is denoted by  $h(n)$  and is called failure rate, risk function, or hazard function, i.e.,

$$h(n) = P \left[ R_n \leq S_{n+1} \mid \bigcap_{j=1}^n (R_{j-1} > S_j) \right]; \quad n = 1, 2, \dots \quad (11)$$

$$h(0) = P[R_0 \leq S_1]$$

where  $R_n$  is given by Eq. (6).

The probability that the structure will survive  $N$  stress cycles, denoted by  $L(N)$ , is called reliability. Successive application of Eq. (11) yields

$$L(N) = \prod_{n=0}^{N-1} [1 - h(n)] \quad (12)$$

and for  $h(n) \ll 1$ , a conservative good approximation can be obtained as

$$L(N) \simeq \exp \left\{ - \sum_{n=0}^{N-1} h(n) \right\} \quad (13)$$

*a. Poisson approximation:* If the point process  $\{S_n\}$  is assumed to be of Poisson type (Footnote 1), i.e.,  $S_{n+1}$  is independent of  $S_j$ ;  $j=1, 2, \dots, n$  and the event  $\{R_n \leq S_{n+1}\}$  is independent of the past event

$$\left\{ \bigcap_{j=1}^n (R_{j-1} > S_j) \right\}$$

then the Poisson failure rate  $h_p(n)$  follows from Eq. (11) as

$$h_p(n) = P[R_n \leq S_{n+1}] = \int_0^\infty [1 - F_{S_{n+1}}(x)] f_{R_n}(x) dx \quad (14)$$

in which  $f_{R_n}(x)$  and  $F_{S_{n+1}}(x)$  are, respectively, the probability density function of  $R_n$  and the distribution function of  $S_{n+1}$  given in Eq. (8).

Eq. (6) can be written as follows:

$$\ln R_n = \ln R_0 - Z \quad (15)$$

The statistical distribution of the initial resisting stress  $R_0$  is determined by material specimen tests, the form of which is a material characteristics and reflects environmental conditions. Experience shows that the normal, log normal, gamma or Weibull distribution can be applied in most cases. Using the log-normal distribution for  $R_0$  because of simplicity and because it is quite reasonable for many engineering materials, the term  $\ln R_0$  in Eq. (15) is normally distributed.

It is observed from Eq. (7) that  $Z$  is the sum of dependent random variables  $S_j^b$ ;  $j = 1, 2, \dots, n$ . As will be

shown later, the distribution of  $Z$  is asymptotically normal. Therefore, for large  $n$ , the distribution of  $Z$  can reasonably be approximated by the normal distribution. Furthermore, we shall restrict ourselves to the case of high-cycle fatigue where the average number of cycles to failure is large. Hence, the effect of  $Z$  (fatigue failure mode) on the failure rate  $h_p(n)$ , see Eqs. (14), (6), and (7), is negligible for small values of  $n$ . Therefore, the distribution of  $Z$  for small  $n$  is irrelevant to the reliability estimate. As a result, the distribution of  $Z$  is approximated by the normal distribution. It therefore follows from Eq. (15)

that  $\ln R_n$  is also normally distributed with the mean value  $\mu_n$  and the variance  $\sigma_n^2$  given by:

$$\begin{aligned}\mu_n &= \mu_0 - \mu_z \\ \sigma_n^2 &= \sigma_0^2 + \sigma_z^2\end{aligned}\quad (16)$$

where  $\mu_0$  and  $\sigma_0^2$  are the mean value and the variance, respectively, of the normal random variable  $\ln R_0$ . The mean value  $\mu_z$  and the variance  $\sigma_z^2$  of  $Z$  can be computed using Eqs. (7), (8) and (10). The results are as follows:

$$\mu_z = \frac{n}{2} K (\sqrt{2} \sigma_s)^b \Gamma \left( 1 + \frac{b}{2} \right) \quad (17)$$

$$\sigma_z^2 = \frac{n}{4} K^2 (\sqrt{2} \sigma_s)^{2b} \left[ \Gamma(1+b) - \Gamma^2 \left( 1 + \frac{b}{2} \right) \right] + \frac{1}{4} K^2 (\sqrt{2} \sigma_s)^{2b} \quad (18)$$

$$\begin{aligned}\Gamma^2 \left( 1 + \frac{b}{2} \right) & \left\{ \sum_{m=1}^{n-1} (n-m) \left[ {}_2F_1 \left( -\frac{b}{2}, -\frac{b}{2}; 1, k_0^2(mT_0) \right) - 1 \right] \right\} \\ E[S_j S_{j+m}^b] &= (\sqrt{2} \sigma_s)^{2b} \Gamma^2 \left( 1 + \frac{b}{2} \right) {}_2F_1 \left( -\frac{b}{2}, -\frac{b}{2}; 1, k_0^2(mT_0) \right)\end{aligned}\quad (19)$$

where  $\Gamma[\cdot]$  and  ${}_2F_1(\cdot)$  are, respectively, the gamma function and the hypergeometric function, and  $E[\cdot]$  denotes the expectation. Further evaluation of  $\sigma_z^2$  requires a specific choice for the mean square spectral density of the response process  $S(t)$  in order to specify the parameter  $k_0^2(mT_0)$  and hence the hypergeometric function.

For a single-degree-of-freedom oscillator under white noise excitation, Crandall et al., (Ref. 5) obtained the following simplified expressions for  $\sigma_z^2$  when the damping coefficient  $\zeta$  is small and when  $b$  is odd:

$$\sigma_z^2 = \frac{n}{4\zeta} K^2 (\sqrt{2} \sigma_s)^{2b} \Gamma^2 \left( 1 + \frac{b}{2} \right) f_1(b) \quad (20)$$

$$V_z = \left[ \frac{f_1(b)}{n\zeta} \right]^{1/2} \quad (21)$$

where  $V_z$  is the coefficient of variation of  $Z$ , and  $f_1(b)$  is a function of  $b$  [Ref. 5]. Equation (21) indicates that the statistical dispersion of  $Z$  diminishes as  $n$  increases, and

hence the dispersion of  $R_n$ , Eq. (6), depends with increasing  $n$  mainly on the dispersion of the initial resisting stress  $R_0$ . This indicates the importance of taking into account the statistical variation of the material strength.

Thus the Poisson failure rate  $h_p(n)$  follows from Eqs. (8), (14), and (16) as

$$\begin{aligned}h_p(n) &= \int_0^\infty (x \sigma_n \sqrt{2\pi})^{-1} \exp \left\{ - \left[ \frac{x^2}{2\sigma_n^2} \right] \right. \\ &\quad \left. - \left[ \frac{(\ln x - \mu_n)^2}{2\sigma_n^2} \right] \right\} dx\end{aligned}\quad (22)$$

Once  $h_p(n)$  is computed, the reliability  $L(N)$  can be obtained from either Eq. (12) or (13) by replacing  $h(n)$  by  $h_p(n)$ .

It should be noted that the particular choice of log-normality for  $R_0$  is not essential here. Any other distribution can be used if there is reason to believe that it represents the material better.

From physical reasoning, it is appropriate to assume that the correlation coefficient  $\rho(mT_0)$  of the responses  $S(t)$  and  $S(t + mT_0)$  decreases monotonically to zero with increasing  $m$ , i.e.,

$$\rho(m_1 T_0) > \rho(m_2 T_0)$$

for  $m_2 > m_1$ , and  $\rho(mT_0) \rightarrow 0$  as  $m \rightarrow \infty$ . This assumption is obviously satisfied, for example, by the response process of a single-degree-of-freedom oscillator excited by white noise. Therefore, the following important properties of the stationary point process  $\{S_n^b\}$ , can be obtained from Eqs. (10), (18), and (19).

$$\left. \begin{aligned} \sigma_z^2 &\sim n\sigma && \sigma < \infty \\ E[|S_j^b|^{2+\delta}] &< \infty && \delta > 0 \\ E[S_j^b S_{j+m}^b] - E[S_j^b] E[S_{j+m}^b] &\rightarrow 0 && \text{as } m \rightarrow \infty \\ E[S_j^b S_{j+m}^b] &> E[S_j^b S_{j+l}^b] && \text{for } l > m \end{aligned} \right\} \quad (23)$$

Since Eq. (23) satisfies conditions for a stationary point process  $\{S_n^b\}$  to have asymptotic normality (Refs. 14 and 15), i.e., the sum of  $S_j^b; j = 1, 2, \dots, n$ , tends to be asymptotically normal for increasing  $n$ .  $Z$  is therefore asymptotically normal.

**b. Clump size approximation.** A useful concept was suggested by Lyon (Ref. 16) that the level crossings by  $S(t)$  are not independent but tend to occur in clumps of dependent crossings. For the narrow-band random process, since each positive upward level crossing by  $S(t)$  results in a peak, the events that the peaks,  $S_j; j = 1, 2, \dots$ , exceeding a threshold level are not independent but tend to occur in clumps. The Poisson assumption implies that one clump of peaks being above the threshold level constitutes only one peak (Footnote 1). Hence, if clumps are assumed to occur independently, the average failure rate denoted by  $\bar{h}(n)$  follows from Eq. (22) as

$$\bar{h}(n) = \int_0^\infty \{x\sigma_n \sqrt{2\pi} E[M|x]\}^{-1} \exp\left\{-\left[\frac{x^2}{2\sigma_n^2}\right] - \left[\frac{(\ln x - \mu_n)^2}{2\sigma_n^2}\right]\right\} dx \quad (24)$$

in which  $E[M|x]$  is the average clump size given that the barrier level is equal to  $x$ . The structural reliability based on clump size approximation can then be obtained from Eq. (12) by replacing  $h(n)$  by  $\bar{h}(n)$ .

The approximate estimation of  $E[M|x]$  has been given in Footnote 1 as follows:

$$E[M|x] \simeq 1 + \sum_{m=1}^{\infty} \frac{m(q_m - q_{m+1})}{q_0} \quad (25)$$

in which

$$q_0 = P[S_j \geq x] = \exp\left(\frac{-x^2}{2\sigma_s^2}\right) \quad (26)$$

$$q_m = P[S_j \geq x, S_{j+m} \geq x] = \int_x^\infty \int_x^\infty f_s(y, z; mT_0) dy dz \quad (27)$$

with  $f_s(y, z; mT_0)$  given by Eq. (10).

## 5. Effect of Loading History and Other Flaw Propagation Law

It can be observed from Eqs. (6), (7), and (11) that the effect of loading history (or sequence) on the structural reliability is accounted for which, however, cannot be accomplished using the cumulative damage hypothesis and the Palmgren-Miner rule. This is one of the advantages of the flaw propagation approach.

For another flaw propagation law, e.g.,  $b = 3$  or  $4$ , a similar approach (discussed above) can be applied to find failure rates  $h_p(n)$  and  $\bar{h}(n)$  and hence the structural reliability  $L(N)$ .

## 6. Numerical Example

Making the transformation  $\gamma = x/\sigma_s$ , the average failure rate  $\bar{h}(n)$  given in Eq. (24) can be written as

$$\bar{h}(n) = \frac{1}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\gamma\sigma_n E[M|\gamma\sigma_s]} \times \exp\left\{-\frac{\gamma^2}{2} - \frac{1}{2} \left[ \frac{\ln\left(\frac{\gamma \sqrt{V_0^2 + 1}}{\nu_0}\right) + \mu_z}{\sigma_n} \right]^2 \right\} d\gamma \quad (28)$$

in which

$$V_0 = [\exp(\sigma_0^2) - 1]^{1/2}$$

is the coefficient of variation of  $R_0$ , and  $\nu_0 = \bar{R}_0/\sigma_s$  is the normalized initial barrier level with  $\bar{R}_0$  the mean value of

$R_0$ ,  $v_0$  is related to a central safety factor  $v_c$  as,  $v_c = \bar{R}_0/\mu_S = v_0 \sqrt{2/\pi}$ , where  $\mu_S$  is the mean value of  $S_n$ . Since  $v_c$  is based on the central measures of location of  $R_0$  and  $S_n$ , it is called central safety factor, which plays a major role in the reliability-based optimum structural design (Refs. 8-10, Footnote 2). In deriving Eq. (28), the relationship

$$\mu_0 = \ln \left( \frac{\bar{R}_0}{\sqrt{V_0^2 + 1}} \right)$$

associated with the log-normal distribution of  $R_0$  has been employed.

With the aid of Eqs. (16), (17), and (20),  $\sigma_n$  and  $\mu_z$  appearing in Eq. (28) can be written as follows:

$$\mu_z = \frac{n}{2} C_p \left( \frac{\sqrt{2}}{v_0} \right)^b \Gamma \left( 1 + \frac{b}{2} \right) \quad (29)$$

$$\sigma_n^2 = \ln(V_0^2 + 1) + \frac{n}{4\zeta} C_p^2 \left( \frac{\sqrt{2}}{v_0} \right)^{2b} \Gamma^2 \left( 1 + \frac{b}{2} \right) f_1(b) \quad (30)$$

where  $C_p = K\bar{R}_0^b$  is a flaw propagation factor depending on the particular material and environmental conditions.

It follows from Eqs. (28), (29), and (30) that the average failure rate, and hence the structural reliability, are characterized by  $C_p$ ,  $V_0$  and  $v_0$ , where  $C_p$  and  $V_0$  are material properties associated with flaw propagation and initial strength dispersion (or initial flaw size dispersion), and  $v_0$  is the normalized initial barrier level (or a measure of the central safety factor) relating the applied load characteristics to the structural design.

Consider a simple structure, that can be idealized as a single-degree-of-freedom system, excited by white noise,  $n(t)$ , where the displacement  $Y(t)$  is related to  $n(t)$  as follows:

$$\ddot{Y}(t) + 2\zeta\omega_n\dot{Y}(t) + \omega_n^2 Y(t) = n(t) \quad (31)$$

The stress response process  $S(t)$  is related to  $Y(t)$  by  $S(t) = DY(t)$ , with  $D$  being a suitable constant.

To show the relationship of  $v_0$ , or of the central safety factor,  $v_c = v_0 \sqrt{2/\pi}$ , to the structural design and to the applied load characteristics, the simple structure consisting of a mass  $M$  supported by a beam with area  $A_1$ , length

$L$  and Young's modulus of elasticity  $E$  is considered. It can easily be shown that  $\omega_n$  appearing in Eq. (31) is  $\omega_n^2 = A_1 E / ML$  and

$$\sigma_n^2 = \frac{\phi_0 \pi E^2}{2\zeta \omega_n^3 L^2}$$

where  $\phi_0$  is the mean square spectral density of white noise  $n(t)$ . Hence,

$$v_0 = \bar{R}_0 \left( \frac{2\zeta}{\pi \phi_0} \right)^{1/2} \left( \frac{A_1}{M} \right)^{3/4} \left( \frac{L}{E} \right)^{1/4} \quad (32)$$

In the present example, it is assumed that failure is due to tension only. The damping coefficient is assumed 0.02.

Failure rates  $h_p(n)$  associated with different values of  $C_p$ ,  $V_0$  and  $v_0$  are plotted in Fig. 1. In Fig. 1(a),  $C_p = 0$ , representing the case where the fatigue failure mode is neglected, and hence failure rates are constant. In Figs. 1(b) and 1(c) failure rates are plotted versus the number of stress cycles  $n$ . It is clearly shown in Fig. 1 that for this example: (1) because of the interaction of the fatigue failure mode, the failure rate increases monotonically with respect to  $n$  and  $C_p$ ; (2) the failure rate increases as the statistical dispersion,  $V_0$ , of the initial material strength,  $R_0$ , increases; and (3) the effect of the fatigue failure mode is negligible for small  $n$  justifying the assumption of normality for the random variable  $Z$ , Eqs. (7) and (16).

In Fig. 2, the structural reliability  $L(N)$ , based on clump size approximation, is plotted as a function of stress cycles  $N$ . It is observed from Fig. 2 that neglecting the effect of the fatigue failure mode, i.e.,  $C_p = 0$ , results in an unconservative reliability prediction. This situation becomes much more critical as the flaw propagation factor  $C_p$  increases. It is further observed that the structural reliability decreases as the statistical dispersion  $V_0$  of the material strength  $R_0$  increases. This reflects an important fact that overlooking the statistical variation of material strength results in an unsafe design.

Although a particular flaw propagation law,  $b = 2$ , and a particular distribution function (log normal) for the initial material strength  $R_0$ , are used in the numerical example, above trends of observations are believed to hold for other flaw propagation laws such as  $b = 3$  or  $b = 4$ , as well as other statistical distribution functions for  $R_0$ .

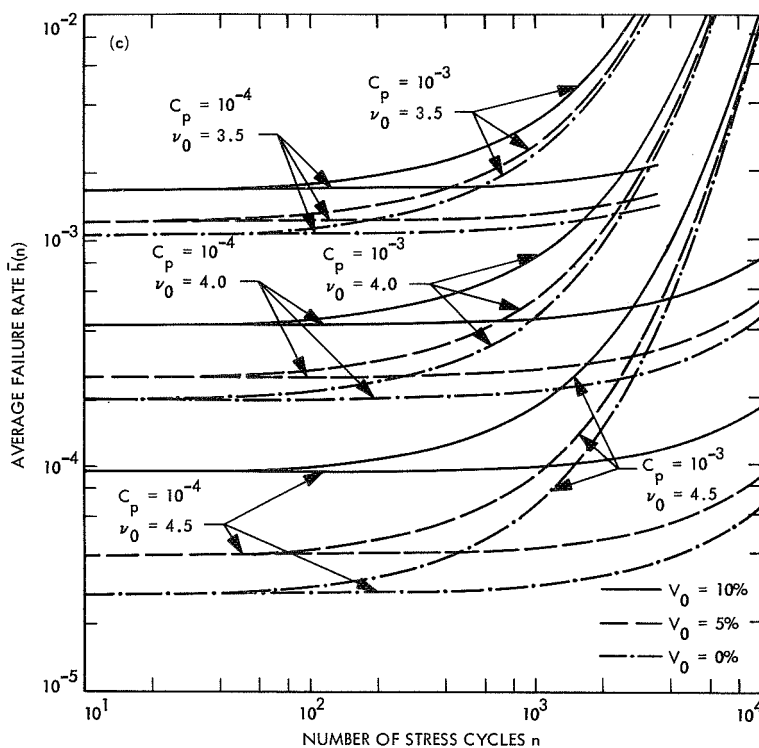
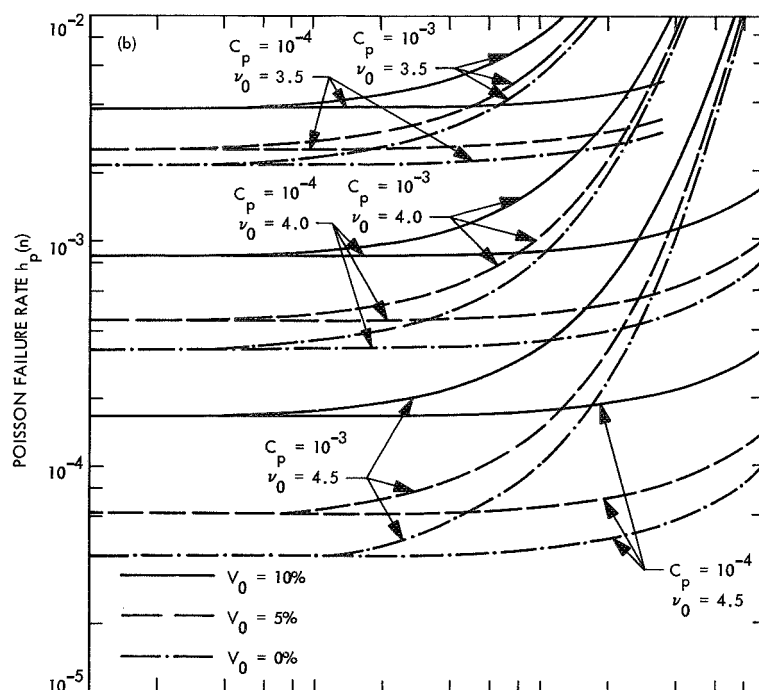
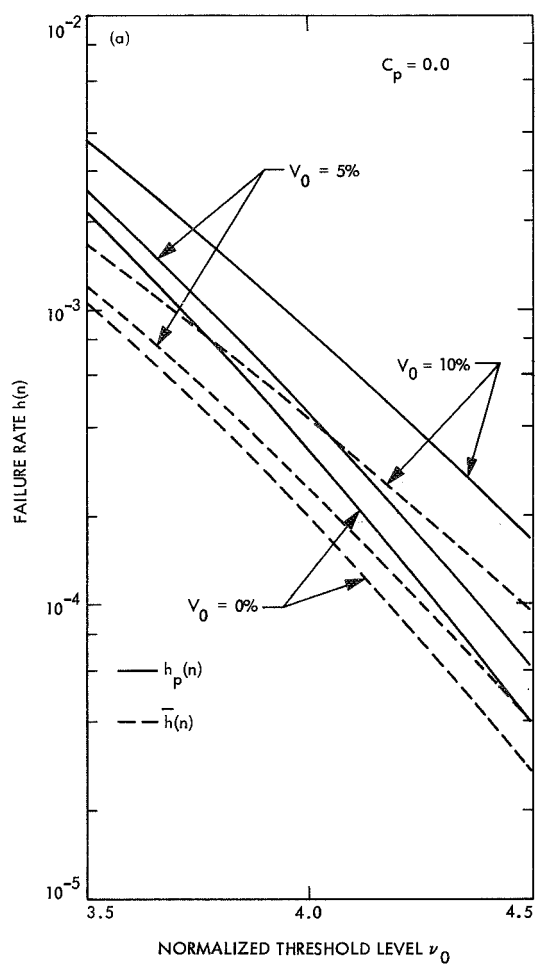
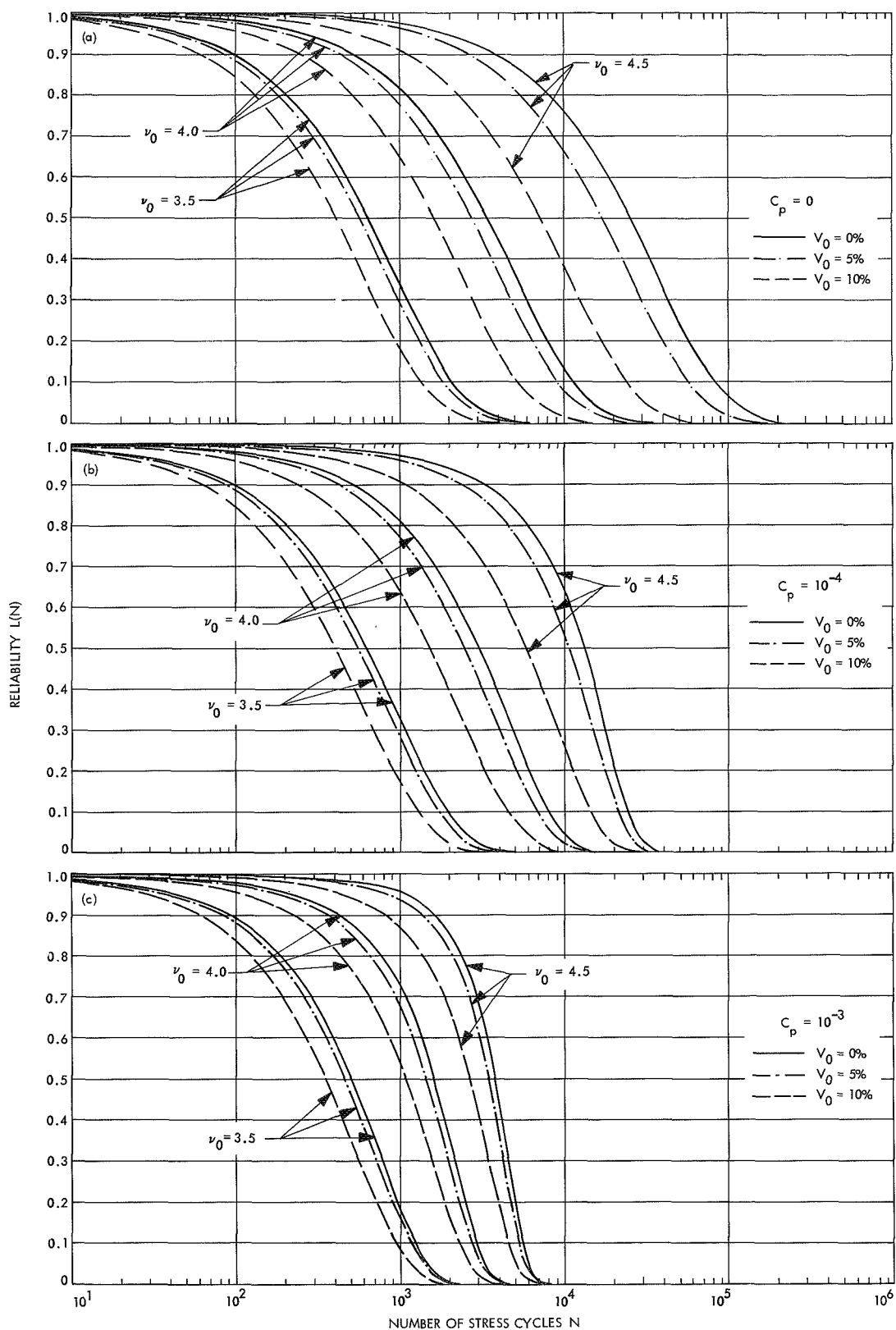


Fig. 1. Failure rates  $h_p(n)$  and  $\bar{h}(n)$ : (a) Constant failure rates  $h_p(n)$  and  $\bar{h}(n)$  for  $C_p = 0.0$ , (b) Poisson failure rate  $h_p(n)$  versus number of stress cycles  $n$ , (c) Average failure rate  $\bar{h}(n)$  based on the clump size approximation versus number of stress cycles  $n$



**Fig. 2. Structural reliability  $L(N)$  versus number of stress cycles  $N$ : (a)  $C_p = 0$ , (b)  $C_p = 10^{-4}$ , (c)  $C_p = 10^{-3}$**



## References

1. Heer, E., and Trubert, M., "Analysis of Space Vehicle Structures Using the Transfer-Function Concept," Technical Report 32-1367, Jet Propulsion Laboratory, Pasadena, Calif., April 1, 1969.
2. Yang, J.-N., and Shinozuka, M., "First-Passage Time Problem," *J. Acoust. Soc. Amer.*, Vol. 47, No. 1 (Part II), pp. 393-394, Jan. 1970.
3. Lin, Y. K., "Random Process," *J. Appl. Mech. Rev.*, Sept. 1969.
4. Crandall, S. H., "First Crossing Probabilities of the Linear Oscillator," *J. Sound Vib.*, Vol. 12, No. 3, pp. 285-299, July 1970.
5. Crandall, S. H., Mark, W. D., and Khazzaz, G. R., "The Variance in Palmgren-Miner Damage due to Random Vibration," *Proceedings of the 4th U.S. National Congress of Applied Mechanics*, Vol. I, pp. 119-126, 1962.
6. Shinozuka, M., and Yang, J.-N., "Nonstationary Peak Distribution," *Proceedings of the ASCE-EMD Specialty Conference on Probabilistic Concepts and Methods in Engineering*, pp. 169-173, Purdue University, Nov. 1969.
7. Smith, S. H., "Fatigue Crack Growth Under Axial Narrow and Broad Band Random Loading," in *Acoustical Fatigue in Aerospace Structures*, pp. 331-360. Edited by Trapp, J. W., and Forney, P. M., Syracuse University Press, 1965.
8. Shinozuka, M., and Yang, J.-N., "Optimum Structural Design Based on Reliability and Proof Load Test," *Ann. Assur. Sci.*, Vol. 8, pp. 375-391, 1969.
9. Shinozuka, M., Yang, J.-N., and Heer, E., "Optimum Structural Design Based on Reliability Analysis," *Proceedings of the Eighth International Symposium on Space Technology and Science*, pp. 245-258, Tokyo, Japan, Aug. 1969.
10. Heer, E., and Yang, J.-N., "Optimum Pressure Vessel Design Based on Fracture Mechanics and Reliability Criteria," *Proceedings of the ASCE-EMD Specialty Conference on Probabilistic Concepts and Methods in Engineering*, Purdue University, Nov. 1969, pp. 102-106.
11. Liu, H. W., "Fatigue Crack Propagation and Applied Stress Range—An Energy Approach," *J. Bas. Engin.*, ASME, Vol. 85, pp. 116-122, 1963.
12. Paris, P., and Erdogan, F., "A Critical Analysis of Crack Propagation Laws," *J. Bas. Engin.*, ASME, Vol. 85, No. 4, pp. 528-543, Dec. 1963.
13. Rice, S. O., "Mathematical Analysis of Random Noise," in *Selected Papers on Noise and Stochastic Processes*. Edited by N. Wax, Dover Pub., N.Y., 1955.
14. Sherfling, R. J., "Contribution to Central Limit Theory for Dependent Variables," Institute of Statistics Mimeo No. 514, Department of Statistics, University of North Carolina, Chapel Hill, N.C., March 1967, p. 42.
15. Ibragimov, I. A., "Some Limit Theorems for Stationary Process," *Theor. Prob. Appl.*, Vol. 7, 1962, pp. 349-382.
16. Lyon, R. H., "On the Vibration Statistics of a Randomly Excited Hard-Spring Oscillator II," *J. Acoust. Soc. Amer.*, p. 33, 1961.

## C. Boundary Layer Equations of a Heated Constrained Spherical Shell, H. E. Williams<sup>1</sup>

### 1. Introduction

The general equations governing the nonlinear, thermoelastic behavior of thin shells of revolution derived earlier by the author<sup>2</sup> can be vastly simplified when applied to specific cases. It is the purpose of this article to investigate the limiting form of these equations when the shell is thin enough for a boundary layer to develop. The range of the temperature distribution studied here

is expressly taken such that the order of magnitude of the resulting transverse deflection is comparable with the thickness of the shell. This range is well above that appropriate for small deflection theory, and should establish a range of temperatures within which these simpler theories are applicable. The boundary layer equations are developed first for a general asymmetric displacement field and then applied to the case of a uniformly heated shallow spherical shell.

### 2. Analysis of Edge Region

At this stage, it will be assumed that the shell (Fig. 1) is spherical with radii of curvature  $R$  and heated uniformly through the thickness to a temperature whose

<sup>1</sup>Consultant, Harvey Mudd College, Claremont, Calif.

<sup>2</sup>Williams, H. E., "On the Thermoelastic Equations of Thin Shells of Revolution," submitted to *Journal of Solids and Structures*.

distribution is  $T_0(\phi, \theta)$ . The order of magnitude of the temperature is measured by the parameter  $T^*$  where

$$T_0 = T^* \cdot \bar{T}_0 \quad \bar{T}_0, a \frac{\partial \bar{T}_0}{\partial r}, \frac{\partial \bar{T}_0}{\partial \theta} = 0 \quad (1)$$

The thermal strain parameter  $\alpha T^*$  is measured relative to  $h/R$  through the parameter  $H$  such that

$$H = \sqrt{3(1-\nu^2)} \cdot \frac{\alpha T^*}{h/R} = 0 \quad (1)$$

Further, it will be assumed that

$$h/R \ll 1$$

$$\frac{a}{R} < 1$$

$$\frac{Rh}{a^2} \ll 1$$

It should be noted that a uniform temperature corresponding to  $H = 0(1)$  in a shell for which  $Rh \ll a^2$  is much larger than that

$$(\alpha T_{\text{crit}} \approx h^2/a^2)$$

sufficient to cause buckling of a flat, circular plate of the same thickness  $h$  and span  $a$ . For this reason, it is expected that a boundary layer will develop with a width comparable to that predicted by linear theory. This suggests using a boundary layer coordinate  $\xi$  such that

$$\frac{r}{a} = 1 + \epsilon \xi = x$$

$$\epsilon^2 = \frac{Rh/a^2}{\sqrt{3(1-\nu^2)}}$$

and

$$\frac{\partial(\quad)}{\partial \phi} = \frac{R \cos \phi}{\epsilon a} \cdot \frac{\partial(\quad)}{\partial \xi}$$

Thus, in terms of the following dimensionless displacements,

$$(w_1, U_1, U_2) = \alpha RT^* \cdot \left[ \bar{w}(\xi, \theta), \frac{\epsilon a}{R} \cdot \bar{U}_1(\xi, \theta), \frac{\epsilon^2 a}{R} \cdot \bar{U}_2(\xi, \theta) \right]$$

$$\epsilon_m = \alpha T^* \cdot \bar{\epsilon}_m(\xi, \theta)$$

rotations,

$$\beta_1 = \frac{\alpha RT^*}{\epsilon a} \cdot \bar{\beta}_1 \quad \beta_2 = \frac{\alpha RT^*}{a} \cdot \bar{\beta}_2$$

and, stress resultants

$$N_\phi = Eh\alpha T^* \cdot \epsilon \bar{N}_r$$

$$N_\theta = Eh\alpha T^* \cdot \bar{N}_\theta$$

$$N_{\phi\theta} = Eh\alpha T^* \cdot \epsilon \bar{N}_{r\theta}$$

$$M_{\phi\theta} = \frac{Eh^2\alpha T^*}{\sqrt{3(1-\nu^2)}} \cdot \epsilon \bar{M}_{r\theta}$$

$$(M_\phi, M_\theta) = \frac{Eh^2\alpha T^*}{\sqrt{3(1-\nu^2)}} \cdot (\bar{M}_r, \bar{M}_\theta)$$

$$Q_\phi = \frac{Eh^2\alpha T^*}{\epsilon a \sqrt{3(1-\nu^2)}} \cdot \bar{Q}_r$$

$$Q_\theta = \frac{Eh^2\alpha T^*}{a \sqrt{3(1-\nu^2)}} \cdot \bar{Q}_\theta$$

$$S = \frac{Eh^2\alpha T^*}{R} \cdot \bar{S}$$

the equation of Footnote 2 reduce to the following displacement-stress resultant equations

$$\bar{\beta}_1 + \frac{\partial \bar{w}}{\partial \xi} \cdot \cos \phi_0 = 0 \quad \bar{\beta}_2 + \frac{\partial \bar{w}}{\partial \theta} = 0$$

$$\frac{\partial \bar{U}_1}{\partial \xi} \cdot \cos \phi_0 + \bar{w} + \frac{H}{2} \cdot \cos^2 \phi_0 \cdot \left( \frac{\partial \bar{w}}{\partial \xi} \right)^2 + \nu \bar{N}_\theta = \bar{T}_0(a, \theta)$$

$$\bar{w} - \bar{N}_\theta = \bar{T}_0(a, \theta)$$

$$\cos \phi_0 \cdot \frac{\partial \bar{U}_2}{\partial \xi} + \frac{\partial \bar{U}_1}{\partial \theta} + H \cdot \cos \phi_0 \cdot \frac{\partial \bar{w}}{\partial \xi} \cdot \frac{\partial \bar{w}}{\partial \theta} = a \cdot (1 + \nu) \cdot \bar{N}_{r\theta}$$

$$\frac{1-\nu^2}{4} \cdot \frac{\partial \bar{\beta}_1}{\partial \xi} \cdot \cos \phi_0 - (\bar{M}_r - \nu \bar{M}_\theta) = 0$$

$$\bar{M}_\theta - \nu \bar{M}_r = 0$$

$$\bar{\epsilon}_m + \frac{H}{2} \cdot \bar{\beta}_1^2 + \nu \bar{N}_\theta = \bar{T}_0(a, \theta)$$

$$\cos \phi_0 \cdot \frac{\partial \bar{\beta}_2}{\partial \xi} + \frac{\partial \bar{\beta}_1}{\partial \theta} = \frac{8}{1-\nu} \cdot \bar{M}_{r\theta}$$

and equations of equilibrium

$$\frac{\partial \bar{N}_r}{\partial \xi} - \bar{N}_\theta = 0$$

$$\cos \phi_0 \cdot \frac{\partial \bar{N}_{r\theta}}{\partial \xi} + \frac{\partial \bar{N}_\theta}{\partial \theta} = 0$$

$$\cos \phi_0 \cdot \frac{\partial \bar{Q}_r}{\partial \xi} - \bar{N}_\theta = 0$$

$$\cos \phi_0 \cdot \frac{\partial \bar{M}_r}{\partial \xi} - \bar{Q}_r = 0$$

$$\cos \phi_0 \cdot \frac{\partial \bar{M}_{r\theta}}{\partial \xi} + \frac{\partial \bar{M}_\theta}{\partial \theta} - \bar{Q}_\theta = 0$$

$$\sqrt{3(1-\nu^2)} \cdot \bar{S} + H \cdot \cos \phi_0 \times \left[ \bar{Q}_r \frac{\partial \bar{w}}{\partial \xi} - \cos \phi_0 \cdot \frac{\partial}{\partial \xi} \left( \bar{M}_r \cdot \frac{\partial \bar{w}}{\partial \xi} \right) \right] + \bar{M}_r + \bar{M}_\theta = 0$$

it follows that

$$\bar{N}_\theta = \bar{w} - \bar{T}_0(a, \theta)$$

$$\cos \phi_0 \cdot \frac{\partial \bar{N}_{r\theta}}{\partial \xi} = \frac{\partial}{\partial \theta} [\bar{T}_0(a, \theta) - \bar{w}]$$

$$\frac{\partial \bar{N}_r}{\partial \xi} = \bar{w} - \bar{T}_0(a, \theta)$$

$$\bar{M}_r = -\frac{\cos^2 \phi_0}{4} \cdot \frac{\partial^2 \bar{w}}{\partial \xi^2} \quad \bar{M}_{r\theta} = -\frac{1-\nu}{4} \cdot \frac{\partial^2 \bar{w}}{\partial \xi \partial \theta} \cdot \cos \phi_0$$

$$\bar{Q}_r = -\frac{\cos^3 \phi_0}{4} \cdot \frac{\partial^3 \bar{w}}{\partial \xi^3} \quad \bar{Q}_\theta = -\frac{\cos^2 \phi_0}{4} \cdot \frac{\partial^3 \bar{w}}{\partial \theta \partial \xi^2}$$

$$\frac{\partial}{\partial \xi} \cdot [\bar{U}_1 \cdot \cos \phi_0 + (1+\nu) \cdot \bar{N}_r] = -\frac{H}{2} \cdot \cos^2 \phi_0 \cdot \left( \frac{\partial \bar{w}}{\partial \xi} \right)^2$$

$$\cos^2 \phi_0 \cdot \frac{\partial^2 \bar{U}_2}{\partial \xi^2} = (1+\nu) \cdot \frac{\partial \bar{T}_0}{\partial \theta} - \frac{\partial \bar{w}}{\partial \theta} \cdot \left( 1+\nu + H \cdot \cos^2 \phi_0 \cdot \frac{\partial^2 \bar{w}}{\partial \xi^2} \right)$$

where

$$\bar{w} = \bar{T}_0(a, \theta) + e^{\bar{\xi}} \cdot [C_1(\theta) \cdot \cos \bar{\xi} + C_2(\theta) \cdot \sin \bar{\xi}]$$

in the limit  $\epsilon \rightarrow 0$ ,  $\xi$  is fixed. Note that

$$\lim_{\substack{\epsilon \rightarrow 0 \\ \xi \text{ fixed}}} T_0(r_1, \theta) = T_0(a, \theta) \quad \text{as} \quad \lim_{\substack{\epsilon \rightarrow 0 \\ \xi \text{ fixed}}} r = a$$

The solution of this set of equations depends on the form of  $\bar{T}_0(a, \theta)$  and on the particular boundary conditions at  $\xi = 0$ . Note that all variables may be expressed in terms of  $\bar{w}$ ,  $\bar{T}_0$  where  $\bar{w}(\xi, \theta)$  is governed by

$$\frac{\cos^4 \phi_0}{4} \cdot \frac{\partial^4 \bar{w}}{\partial \xi^4} + \bar{w} = \bar{T}_0(a, \theta)$$

Thus, with

$$\bar{\xi} = \frac{\xi}{\cos \phi_0}$$

Note that it is only the in-plane displacements ( $\bar{U}_1, \bar{U}_2$ ) that are nonlinearly dependent on the temperature distribution. It should be noted further that these equations

are similar to those proposed by Wittrick (Ref. 1) as applicable for small deflections.

### 3. Uniformly Heated Shell

The solution presented above is valid in the region near the edge. Before it can be completely defined, the behavior of the solution in the interior region must be established. The final form of the solution is determined from the requirement that the solution match in an intermediate region.

As an example of the use of the solution for the edge region—the inner solution—let us consider the particular case of a uniform temperature distribution in a shallow spherical shell, i.e.

$$\bar{T}_0 = 1 \quad \cos \phi_0 \approx 1$$

As the problem becomes axisymmetric, the functions of integration  $[c_1(\theta), c_2(\theta)]$  become constants and we obtain

$$\bar{w} = 1 + e^\xi \cdot (c_2 \sin \xi - \cos \xi)$$

for the inner transverse deflection satisfying the condition that  $\bar{w}(0) = 0$ . The corresponding in-plane deflection becomes

$$\begin{aligned} \bar{U}_1 = & -(1 + \nu) \cdot \bar{N}_r - \frac{H}{8} \cdot e^{2\xi} \cdot [2 \cdot (c_2^2 + 1) \\ & + (1 - 2c_2 - c_2^2) \cdot \cos 2\xi \\ & + (c_2^2 - 2c_2 - 1) \cdot \sin 2\xi] + c_3 \end{aligned}$$

Also

$$\bar{N}_r = -\frac{e^\xi}{2} \cdot [(1 - c_2) \cdot \sin \xi + (1 + c_2) \cdot \cos \xi]$$

A constant of integration is not required for the radial stress resultant  $\bar{N}_r$  as it can be shown that

$$\bar{Q}_r \cdot \cos \phi_0 = N_r$$

for a finite segment of the shell to be in vertical equilibrium in the absence of distributed surface pressure, and

$$\bar{Q}_r = -\frac{1}{4} \cdot \frac{\partial^3 w}{\partial \xi^3}$$

The constant  $c_2$  is determined by the choice of edge fixity, i.e., either

$$\frac{\partial \bar{w}}{\partial \xi}(0) = 0 \quad c_2 = 1 \quad (\text{clamped})$$

$$\bar{M}_r(0) = 0 \quad c_2 = 0 \quad (\text{simply supported})$$

With the constant  $c_2$  fixed,  $c_3$  is determined by the requirement  $\bar{U}_1(0) = 0$ , that is

$$c_3 = -\frac{1 + \nu}{2} \cdot (1 + c_2) + \frac{H}{8} \cdot (c_2^2 - 2c_2 + 3)$$

In summary, we obtain the following behavior as  $\xi \rightarrow -\infty$ :

$$\bar{w} \rightarrow 1 \quad \bar{N}_r \rightarrow 0 \quad \bar{U}_1 \rightarrow c_3 \neq 0$$

Noting that  $\bar{U}_1$  does not vanish away from the edge, the solution in the interior of the shell—the outer solution—will be assumed to be that corresponding to a uniform, stress-free dilation plus a vertical rigid body translation of magnitude  $\delta$ , i.e.,

$$\begin{aligned} U_1 = & -\delta \cdot \sin \phi \quad U_2 = 0 \quad \epsilon_m = \alpha T_0 \\ = & -\delta ax/R \end{aligned}$$

$$N_\phi, N_\theta; M_\phi, M_\theta, Q_r = 0$$

$$w = \alpha RT_0 + \delta \cdot \cos \phi \approx \alpha RT_0 + \delta$$

Note that  $\delta = \delta(\epsilon)$  at this stage.

Matching of the two solutions is accomplished by requiring

$$\lim_{\substack{\epsilon \rightarrow 0 \\ \xi \text{ fixed}}} \left[ c_3 + \frac{\delta(\epsilon) \cdot x}{\epsilon \alpha T^* R} \right] = 0$$

that is

$$\delta = -\epsilon \cdot c_3 \cdot \alpha RT^*$$

Thus, the first order composite solution for the displacements can be written

$$\bar{w} = 1 + e^{\xi} \cdot (c_2 \cdot \sin \xi - \cos \xi)$$

$$\begin{aligned} \bar{U}_1 = & -(1 + \nu) \cdot \bar{N}_r - \frac{H}{8} \cdot e^{2\xi} \cdot [2 \cdot (c_2^2 + 1) \\ & + (1 - 2c_2 - c_2^2) \cdot \cos 2\xi + (c_2^2 - 2c_2 - 1) \sin 2\xi] \\ & + x \left[ -\frac{1 + \nu}{2} \cdot (1 + c_2) + \frac{H}{8} \cdot (c_2^2 - 2c_2 + 3) \right] \end{aligned}$$

where

$$c_2 = 1 \quad (\text{clamped})$$

or

$$c_2 = 0 \quad (\text{simply supported})$$

#### 4. Discussion

It should be noted that the equations derived above are essentially linear. The in-plane deflections ( $\bar{U}_1, \bar{U}_2$ ) are exceptions, but have no effect on the stress distribution. Thus, the behavior of the stresses and transverse deflection is similar to that predicted by linear theory, even when the transverse deflection is much greater than that considered appropriate for linear theory.

That the transverse deflection is linear in the temperature should not be surprising as it has been predicted by Reissner (Ref. 2) that one should not expect nonlinearity to be important until the transverse deflection is of the order of the boundary layer thickness ( $\sqrt{Rh}$ ). The magnitude studied here was of the order of the shell thickness

as  $\alpha RT^* = h \cdot 0(1)$ . Note that Reissner's analysis assumed a set of equations shown earlier to be generally applicable, whereas the principal object of this note was to simplify equations considered unnecessarily general for the type of problem studied above.

The boundary layer equations presented above have been derived for spherical shells. In the general form, no restriction was made on the range of the opening angle ( $\phi_0$ ), i.e., the shell was not required to be shallow. Obviously, these results are not applicable for deep shells for which  $\cos \phi_0$  is small. [This range of shells has been studied by Wittrick (Ref. 3) for a non-shallow shell of arbitrary shape.]

For shallow spherical shells ( $\cos \phi_0 \approx 1$ ), the results presented above may be considered to apply to more general shells of revolution, as ellipsoidal shells of revolution are spherical within an error comparable to that incurred in writing  $\cos \phi \approx 1$ . Thus, the boundary layer equations presented above apply equally well to a wider class of shells of revolution provided the shell is shallow, or to arbitrarily deep spherical shells provided that  $\cos \phi_0$  is not small.

#### References

1. Wittrick, W. H., "Edge Stresses in Thin Shells of Revolution," *Aust. J. Appl. Sci.*, Vol. 8, p. 235, 1957.
2. Reissner, E., "On Influence Coefficients and Non-linearity for Thin Shells of Revolution," *J. App. Mech.*, pp. 69-72, March, 1959.
3. Wittrick, W. H., "A Non-linear Theory for the Edge Stresses in Thin Shells," *Int. J. Eng. Sci.*, Vol. 2, pp. 155-177.

## D. Derivation of the Equations Governing Heated Shallow Shells of Revolution, H. E. Williams<sup>1</sup>

### 1. Introduction

The general equations governing the nonlinear, thermoelastic behavior of thin shells of revolution have been derived by the author<sup>2</sup> using Reissner's Variational Theorem. When applied to specific cases, such as middle surface heating (no temperature gradient through the thickness of the shell) of shallow shells, it is possible to vastly simplify the general equations and obtain the equation generally regarded as Marguerre's equations. It is the purpose of this note to present an order of magnitude analysis of the equations derived in Footnote 2 in order to establish the range of validity of the simpler equations.

### 2. Analysis

At this stage, it will be assumed that the shell (Section C, Fig. 1) is spherical with radii of curvature  $R$  and heated uniformly through the thickness to a temperature whose distribution is given by  $T_0(\phi, \theta)$ . The order of magnitude of the temperature is measured by the parameter  $T^*$  such that

$$T_0 = T^* \cdot \bar{T}_0 \quad \bar{T}_0, a \frac{\partial \bar{T}_0}{\partial r}, \frac{\partial \bar{T}_0}{\partial \theta} = 0(1)$$

The thermal strain parameter ( $\alpha T^*$ ) is measured relative to  $(h/a)$  through the parameter  $\kappa$  such that

$$\kappa = \frac{h^2/a^2}{\alpha T^*} = 0(1)$$

<sup>1</sup>Consultant, Harvey Mudd College, Claremont, Calif.

<sup>2</sup>Williams, H. E., "On the Thermoelastic Equations of Thin Shells of Revolution," submitted to the *Journal of Solids and Structures*.

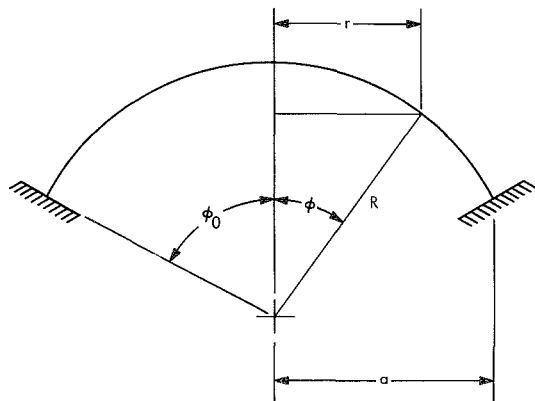


Fig. 1. Coordinate system for a spherical shell

It should be noted that a uniform temperature corresponding to  $\kappa = 0(1)$  is sufficient to cause buckling of a flat, circular plate of the same thickness ( $h$ ) and span ( $a$ ).

In terms of the variables  $x, \theta$  where  $x = r/a$ ,

$$\frac{\partial}{\partial \phi} ( ) = \frac{R \cos \phi}{a} \cdot \frac{\partial ( )}{\partial x}$$

and the following dimensionless displacements

$$w = h\bar{w}(x, \theta) \quad (U_1, U_2) = \alpha T^* \cdot (\bar{U}_1, \bar{U}_2)$$

strain

$$(\epsilon_\phi^{(0)}, \epsilon_\theta^{(0)}, \epsilon_m) = \alpha T^* \cdot (\bar{\epsilon}_\phi, \bar{\epsilon}_\theta, \bar{\epsilon}_m)$$

rotations

$$(\beta_1, \beta_2) = \frac{h}{a} \cdot (\bar{\beta}_1, \bar{\beta}_2)$$

curvatures

$$(\kappa_\phi, \kappa_\theta) = \frac{h}{a^2} (\bar{\kappa}_\phi, \bar{\kappa}_\theta)$$

and stress resultants

$$(N_\phi, N_\theta, N_{\phi\theta}) = Eh\alpha T^* \cdot (\bar{N}_r, \bar{N}_\theta, \bar{N}_{r\theta})$$

$$(M_\phi, M_\theta, M_{\phi\theta}) = Eh^4 \cdot (\bar{M}_r, \bar{M}_\theta, \bar{M}_{r\theta}) / (12a^2)$$

$$(Q_1, Q_2) = Eh^4 \cdot (\bar{Q}_r, \bar{Q}_\theta) / (12a^3)$$

the equations presented in Footnote 2 can be vastly simplified. It will be assumed further that

$$\frac{h}{a} \ll 1 \quad \frac{a}{R} < 1 \quad \frac{a^2}{Rh} = 0(1)$$

Within these limitations, the equations of (1) reduce to the following strain-displacement equations

$$\bar{\epsilon}_\phi = \frac{\partial \bar{U}_1}{\partial x} \cdot \cos \phi + \frac{\bar{w}h}{\alpha R T^*}$$

$$\bar{\epsilon}_\theta = \frac{\partial \bar{U}_2}{x \partial \theta} + U_1 \cdot \frac{\cos \phi}{x} + \frac{\bar{w}h}{\alpha R T^*}$$

curvature-displacement equations

$$\bar{\kappa}_\phi = \frac{\partial \bar{\beta}_1}{\partial x} \cdot \cos \phi \quad \bar{\kappa}_\theta = \frac{\partial \bar{\beta}_2}{x \partial \theta} + \bar{\beta}_1 \cdot \frac{\cos \phi}{x}$$

displacement-stress resultant equations

$$\bar{\epsilon}_\phi + \frac{\kappa}{2} \cdot \left[ \left( \frac{\partial \bar{w}}{\partial x} \cdot \cos \phi \right)^2 + \left( \frac{\bar{w}a}{R} \right)^2 \right] - (\bar{N}_r - \nu \bar{N}_\theta) = \bar{T}_0$$

$$\bar{\epsilon}_\theta + \frac{\kappa}{2} \cdot \left[ \left( \frac{\partial \bar{w}}{x \partial \theta} \right)^2 + \left( \frac{\bar{w}a}{R} \right)^2 \right] - (\bar{N}_\theta - \nu \bar{N}_r) = \bar{T}_0$$

$$\bar{\kappa}_\phi - (\bar{M}_r - \nu \bar{M}_\theta) = 0$$

$$\bar{\kappa}_\theta - (\bar{M}_\theta - \nu \bar{M}_r) = 0$$

$$\frac{\partial \bar{U}_1}{x \partial \theta} + \left( \frac{\partial \bar{U}_2}{\partial x} - \frac{\bar{U}_2}{x} \right) \cdot \cos \phi + \frac{\kappa}{x} \cdot \left( \frac{\partial \bar{w}}{\partial x} \cdot \frac{\partial \bar{w}}{\partial \theta} \cdot \cos \phi \right) = 2(1 + \nu) \cdot \bar{N}_{r\theta}$$

$$\frac{\partial \bar{\beta}_1}{x \partial \theta} + \left( \frac{\partial \bar{\beta}_2}{\partial x} - \frac{\bar{\beta}_2}{x} \right) \cdot \cos \phi = 2(1 + \nu) \cdot \bar{M}_{r\theta}$$

$$\bar{\epsilon}_m + \frac{\kappa}{2} \cdot (\bar{\beta}_1^2 + \bar{\beta}_2^2) + \nu \cdot (\bar{N}_r + \bar{N}_\theta) = \bar{T}$$

$$\bar{\beta}_1 + \frac{\partial \bar{w}}{\partial x} \cdot \cos \phi = 0$$

$$\bar{\beta}_2 + \frac{\partial \bar{w}}{x \partial \theta} = 0$$

and equations of equilibrium

$$-\frac{\cos \phi}{x} \cdot \frac{\partial}{\partial x} (x \bar{M}_r) - \frac{\partial \bar{M}_{r\theta}}{x \partial \theta} + \frac{\bar{M}_\theta}{x} \cdot \cos \phi + Q_r = 0$$

$$-\frac{\cos \phi}{x} \cdot \frac{\partial}{\partial x} (x \bar{M}_{r\theta}) - \frac{\partial \bar{M}_\theta}{x \partial \theta} - \frac{\bar{M}_{r\theta}}{x} \cdot \cos \phi + Q_\theta = 0$$

$$-\frac{\cos \phi}{x} \cdot \frac{\partial}{\partial x} (x \bar{N}_r) - \frac{\partial \bar{N}_{r\theta}}{x \partial \theta} + \bar{N}_\theta \cdot \frac{\cos \phi}{x} = 0$$

$$\cos \phi \cdot \frac{\partial \bar{N}_{r\theta}}{\partial x} + \frac{\partial \bar{N}_\theta}{x \partial \theta} + 2 \cdot \frac{\bar{N}_{r\theta}}{x} \cdot \cos \phi = 0$$

$$\frac{\cos \phi}{x} \cdot \frac{\partial}{\partial x} \left[ x \cdot \left( \frac{\kappa \bar{Q}_r}{12} + \bar{N}_r \cos \phi \cdot \frac{\partial \bar{w}}{\partial x} + \bar{N}_{r\theta} \cdot \frac{\partial \bar{w}}{x \partial \theta} \right) \right] + \frac{\partial}{x \partial \theta} \left( \frac{\kappa \bar{Q}_\theta}{12} + \bar{N}_{r\theta} \cdot \frac{\partial \bar{w}}{\partial x} \cdot \cos \phi + \bar{N}_\theta \cdot \frac{\partial \bar{w}}{x \partial \theta} \right)$$

$$- \frac{a^2}{Rh} \cdot (\bar{N}_r + \bar{N}_\theta) = 0$$

The transverse normal stress resultant  $S$  becomes

$$\begin{aligned} -\frac{12a^3}{Eh^4} \cdot S &= \frac{a}{R} \cdot (\bar{M}_r + \bar{M}_\theta) + \frac{h}{a} \cdot \left\{ \bar{Q}_r \cdot \frac{\partial \bar{w}}{\partial x} \cdot \cos \phi + Q_\theta \cdot \frac{\partial \bar{w}}{x \partial \theta} \right. \\ &\quad \left. - \frac{\cos \phi}{x} \cdot \frac{\partial}{\partial x} \left[ x \cdot \left( \bar{M}_r \cdot \frac{\partial \bar{w}}{\partial x} \cdot \cos \phi + \bar{M}_{r\theta} \cdot \frac{\partial \bar{w}}{x \partial \theta} \right) \right] \right. \\ &\quad \left. - \frac{\partial}{x \partial \theta} \left( \bar{M}_{r\theta} \cdot \cos \phi \cdot \frac{\partial \bar{w}}{\partial x} + \bar{M}_\theta \cdot \frac{\partial \bar{w}}{x \partial \theta} \right) \right\} \end{aligned}$$

The shallow shell equations may be obtained from those given above by writing  $\cos \phi \approx 1$  and neglecting  $(a/R)^2$  in comparison with unity. If we define

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial}{x \partial x} + \frac{\partial^2}{x^2 \partial \theta^2}$$

the governing equations for a heated, shallow shell become

$$\bar{\beta}_1 = -\frac{\partial \bar{w}}{\partial x}$$

$$\bar{\beta}_2 = -\frac{\partial \bar{w}}{x \partial \theta}$$

$$\frac{\partial \bar{U}_1}{\partial x} + \frac{\kappa a^2}{Rh} \cdot \bar{w} + \frac{\kappa}{2} \left( \frac{\partial \bar{w}}{\partial x} \right)^2 - (\bar{N}_r - \nu \bar{N}_\theta) = \bar{T}_0$$

$$\frac{\partial \bar{U}_2}{x \partial \theta} + \frac{\bar{U}_1}{x} + \frac{\kappa a^2}{Rh} \cdot \bar{w} + \frac{\kappa}{2} \left( \frac{\partial \bar{w}}{x \partial \theta} \right)^2 - (\bar{N}_\theta - \nu \bar{N}_r) = \bar{T}_0$$

$$2 \cdot (1 + \nu) \cdot \bar{N}_{r\theta} = \frac{\partial \bar{U}_1}{x \partial \theta} + \frac{\partial \bar{U}_2}{\partial x} - \frac{\bar{U}_2}{x} + \frac{\kappa}{x} \cdot \frac{\partial \bar{w}}{\partial x} \cdot \frac{\partial \bar{w}}{\partial \theta}$$

$$-(1 - \nu^2) \cdot \bar{M}_r = \frac{\partial^2 \bar{w}}{\partial x^2} + \nu \cdot \left( \frac{\partial^2 \bar{w}}{x^2 \partial \theta^2} + \frac{\partial \bar{w}}{x \partial x} \right)$$

$$-(1 - \nu^2) \cdot \bar{M}_\theta = \frac{\partial^2 \bar{w}}{x^2 \partial \theta^2} + \frac{\partial \bar{w}}{x \partial x} + \nu \cdot \frac{\partial^2 \bar{w}}{\partial x^2}$$

$$(1 + \nu) \cdot \bar{M}_{r\theta} = -\frac{1}{x} \cdot \frac{\partial^2 \bar{w}}{\partial x \partial \theta} + \frac{1}{x^2} \cdot \frac{\partial \bar{w}}{\partial \theta}$$

$$-(1 - \nu^2) \cdot \bar{Q}_r = \frac{\partial (\Delta \bar{w})}{\partial x} \quad -(1 - \nu^2) \cdot \bar{Q}_\theta = \frac{\partial (\Delta \bar{w})}{x \partial \theta}$$

$$\frac{\partial \bar{N}_r}{\partial x} + \frac{\partial \bar{N}_{r\theta}}{x \partial \theta} + \frac{1}{x} (\bar{N}_r - \bar{N}_\theta) = 0$$

$$\frac{\partial \bar{N}_{r\theta}}{\partial x} + \frac{\partial \bar{N}_\theta}{x \partial \theta} + \frac{2}{x} \cdot \bar{N}_{r\theta} = 0$$

$$-\frac{\kappa/12}{1 - \nu^2} \cdot \Delta \Delta \bar{w} - \frac{a^2}{Rh} \cdot (\bar{N}_r + \bar{N}_\theta) + \bar{N}_r \cdot \frac{\partial^2 \bar{w}}{\partial x^2}$$

$$+ \bar{N}_\theta \cdot \left( \frac{\partial^2 \bar{w}}{x^2 \partial \theta^2} + \frac{\partial \bar{w}}{x \partial x} \right) + 2 \bar{N}_{r\theta} \cdot \left( \frac{\partial^2 \bar{w}}{x \partial x \partial \theta} - \frac{\partial \bar{w}}{x^2 \partial \theta} \right) = 0$$

Often it is convenient to express this system of equations in displacement variables only. Thus, with

$$(1 - \nu^2) \cdot \bar{N}_r = -(1 + \nu) \cdot \bar{T}_0 + \frac{\partial \bar{U}_1}{\partial x} + \nu \cdot \left( \frac{\partial \bar{U}_2}{x \partial \theta} + \frac{\bar{U}_1}{x} \right) + (1 + \nu) \cdot \frac{\kappa a^2}{Rh} \cdot \bar{w} + \frac{\kappa}{2} \cdot \left[ \left( \frac{\partial \bar{w}}{\partial x} \right)^2 + \nu \cdot \left( \frac{\partial \bar{w}}{x \partial \theta} \right)^2 \right]$$

$$(1 - \nu^2) \cdot \bar{N}_\theta = -(1 + \nu) \cdot \bar{T}_0 + \frac{\partial \bar{U}_2}{x \partial \theta} + \frac{\bar{U}_1}{x} + \nu \cdot \frac{\partial \bar{U}_1}{\partial x} + (1 + \nu) \cdot \frac{\kappa a^2}{Rh} \cdot \bar{w} + \frac{\kappa}{2} \cdot \left[ \left( \frac{\partial \bar{w}}{x \partial \theta} \right)^2 + \nu \cdot \left( \frac{\partial \bar{w}}{\partial x} \right)^2 \right]$$

the in-plane equations of equilibrium become

$$\begin{aligned} & \frac{1 - \nu}{2} \cdot \left( \frac{\partial^2 U_2}{\partial r^2} + \frac{\partial U_2}{r \partial r} - \frac{U_2}{r^2} \right) + \frac{\partial^2 U_2}{r^2 \partial \theta^2} + \frac{1 + \nu}{2r} \cdot \frac{\partial^2 U_1}{\partial r \partial \theta} + \frac{3 - \nu}{2r^2} \cdot \frac{\partial U_1}{\partial \theta} \\ & + \frac{\partial w}{\partial r} \cdot \left( \frac{1 + \nu}{2r} \cdot \frac{\partial^2 w}{\partial r \partial \theta} + \frac{1 - \nu}{2r^2} \cdot \frac{\partial w}{\partial \theta} \right) \\ & + \frac{\partial w}{r \partial \theta} \left( \frac{1 - \nu}{2} \cdot \frac{\partial^2 w}{\partial r^2} + \frac{1 + \nu}{R} + \frac{\partial^2 w}{r^2 \partial \theta^2} \right) - \frac{1 + \nu}{r} \cdot \frac{\alpha \partial T_0}{\partial \theta} = 0 \end{aligned} \quad (1)$$

$$\begin{aligned} & \frac{\partial^2 U_1}{\partial r^2} + \frac{\partial U_1}{r \partial r} - \frac{U_1}{r^2} + \frac{1 - \nu}{2r^2} \cdot \frac{\partial^2 U_1}{\partial \theta^2} + \frac{1 + \nu}{2r} \cdot \frac{\partial^2 U_2}{\partial r \partial \theta} - \frac{3 - \nu}{2r^2} \cdot \frac{\partial U_2}{\partial \theta} \\ & + \frac{\partial w}{\partial r} \cdot \left[ \frac{1 + \nu}{R} + \frac{\partial^2 w}{\partial r^2} + \frac{1 - \nu}{2r} \cdot \left( \frac{\partial w}{\partial r} + \frac{\partial^2 w}{r \partial \theta^2} \right) \right] \\ & + \frac{1 + \nu}{2r} \cdot \frac{\partial w}{\partial \theta} \cdot \frac{\partial}{\partial r} \left( \frac{\partial w}{r \partial \theta} \right) - (1 + \nu) \cdot \alpha \cdot \frac{\partial T_0}{\partial r} = 0 \end{aligned} \quad (2)$$



The out-of-plane equation may also be written

$$-D\nabla^4 w + N_r \cdot \left( \frac{\partial^2 w}{\partial r^2} - \frac{1}{R} \right) + N_\theta \cdot \left( \frac{\partial^2 w}{r^2 \partial \theta^2} + \frac{\partial w}{r \partial r} - \frac{1}{R} \right) + 2N_{r\theta} \cdot \frac{\partial}{\partial r} \left( \frac{\partial w}{r \partial \theta} \right) = 0 \quad (3)$$

Equations (1), (2), and (3) have been written in dimensional form for purposes of comparison. In particular, they are identical to those given by Famili and Archer (Ref. 1) as the circular polar form of the equations derived from Marguerre (Ref. 2).

### 3. Discussion

Although the above equations were expressly derived for a spherical shell, it should be pointed out that they are more generally applicable, as an ellipsoidal shell of revolution is spherical within an error comparable to that involved in writing  $\cos \phi \approx 1$ . Thus, if  $(R_1, R_2)$  are the principal radii of curvature of an ellipsoid with semi-axes  $(a, b)$ , it may be shown that

$$\frac{R_2}{R_1} = 1 + \frac{a^2 - b^2}{b^2} \cdot \sin^2 \phi \approx 1$$

for a shallow shell for which

$$\frac{a^2 - b^2}{b^2} = 0(1)$$

As a final remark, it should be noted that the assumption  $a^2/Rh = 0(1)$  effectively precludes the formation of a boundary layer near the edge  $r = a$ . This follows from the observation that the thickness of the boundary layer is of order  $\sqrt{Rh}$ . The assumption is also consistent with the use of  $(x)$  as a variable.

### References

1. Famili, J., and Archer, R. R., "Finite Asymmetric Deformation of Shallow Spherical Shells," *AIAA Journal*, Vol. 3, No. 3, pp. 506-510.
2. Marguerre, K., "Zur Theorie der Gekrummten Platte Grosser Formanderung," *Proceedings of the Fifth International Congress of Applied Mechanics*, pp. 93-101, 1938.

## XIII. Electronic Parts Engineering

### ENGINEERING MECHANICS DIVISION

#### A. Radiation Effects on Electronic Parts: Literature Search and Data Evaluation, *K. E. Martin*

##### 1. Introduction

In support of the Thermoelectric Outer-Planet Spacecraft (TOPS) Program, a literature search and data analysis program was contracted with The Boeing Company, Seattle, Washington. The program started in July 1969 and was completed in May 1970. The objective of the program was to establish the basis for the TOPS electronic parts radiation effects support program.

Under the program, the contractor was required to search existing industry and government radiation effects literature and select and analyze pertinent data relating to the TOPS radiation criteria. Based on these findings, JPL will apply the information to anticipated TOPS electronic parts usage and recommend a program for radiation testing support to complement the testing already completed by other agencies. Radiation hardening and screening procedures will also be evaluated and recommended.

##### 2. Discussion

The radiation environments considered were (1) that produced by radioisotope thermoelectric generators

(RTGs), i.e., low-level steady-state fission neutron and gamma spectra, and (2) high energy protons, both those generated during solar events and those trapped in planetary belts. Although it was beyond the scope of this program to consider electron effects, where data was readily available, such effects were included. During the course of the program, facilities such as the following were utilized:

- (1) Radiation Effects Information Center, Battelle Memorial Institute.
- (2) NASA (computerized search performed by the Boeing Aerospace Technical Library).
- (3) Defense Documentation Center.
- (4) Defense Atomic Support Agency.
- (5) Boeing Aerospace Library.

These searches resulted in hundreds of references of which about 260 were retained after reviewing for pertinent radiation environments and part types. Analysis of the literature revealed that, in general, for most part types there are at least some neutron and gamma ray data but that for many types no proton data exists. Primarily, the analysis was directed toward the determination of permanent degradation of parameters, temporary

drifts of parameters, parameter degradation factors, and radiation hardening and screening procedures. The relative sensitivities of components as determined from the analysis are summarized in Figs. 1, 2, and 3. This information is a composite average of test data for many devices within the generic types, and it must be emphasized that specific devices within a given generic type may be far more radiation resistant than indicated in the figures. In many cases, the damage levels were different for the same part types tested at the same radiation levels by different agencies. This in part could be due to such factors as variance between manufacturer's lots, application of different power levels during radiation exposure, and variation in measurement methods and equipment. The vertical dashed lines represent the presently estimated total dose or fluence levels to be experienced by the electronic parts for the TOPS missions. A part type with the bar bisecting or a portion of the bar to the left of this dashed line is considered a problem area and will warrant special consideration. These degradation levels are generally pessimistic as they do not consider radiation screening, radiation hardening, specially selected devices, annealing effects, or circuit compromises, although it does indicate that serious problems exist.

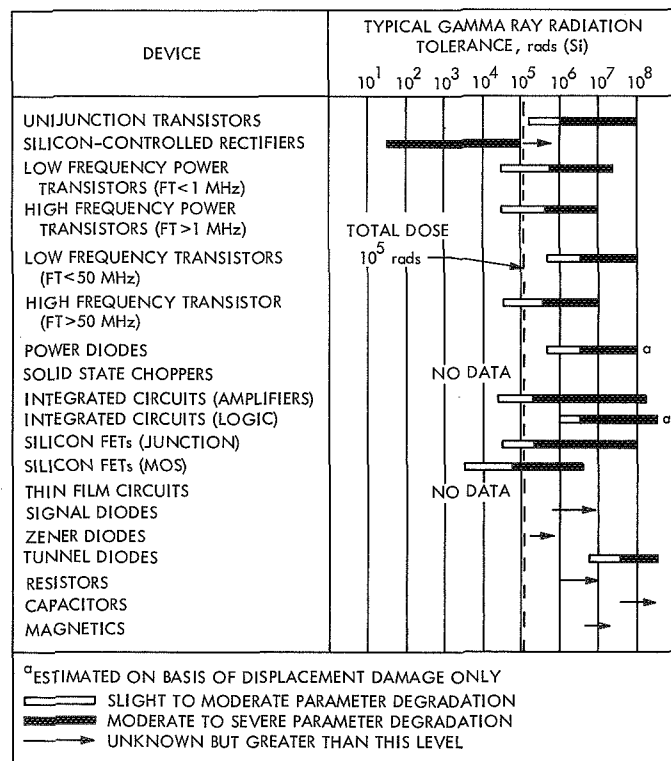


Fig. 1. Survey of component gamma degradation

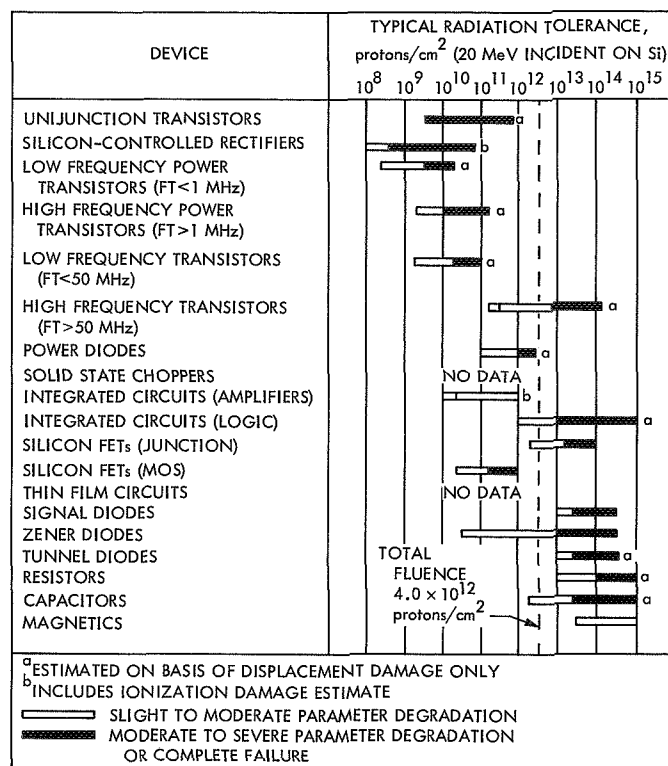


Fig. 2. Survey of component proton degradation

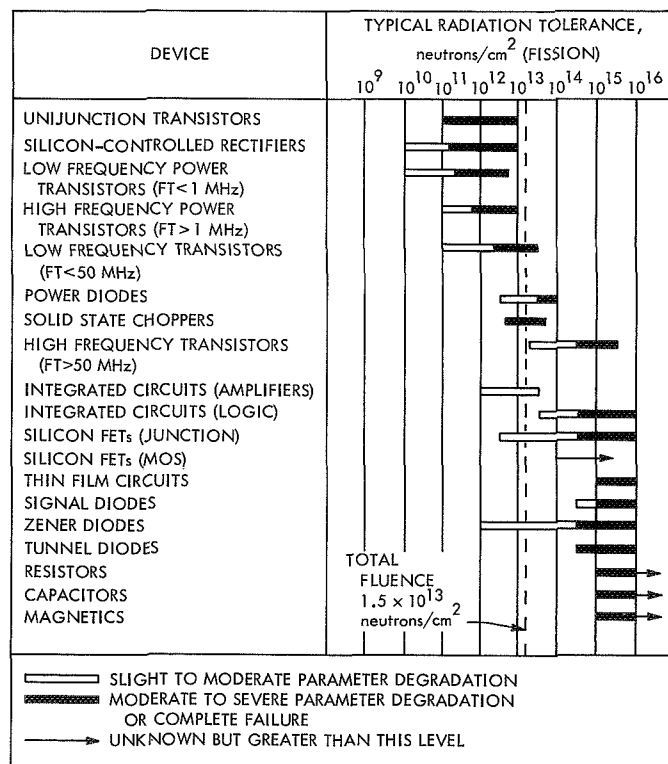


Fig. 3. Survey of component neutron degradation

In general, silicon-controlled rectifiers (SCRs) and uni-junction devices appear to be very vulnerable and SCR response is especially unpredictable. For this reason it is recommended that their usage be avoided where possible. Power and low frequency transistors appear to be the next most vulnerable group. Since their behavior is better understood, these devices can be used if sufficient gain margin is allowed or other concern is taken. It should be noted that in some cases shielding may be necessary.

Linear integrated circuits, reference diodes, and metal oxide semiconductor (MOS) devices also present a problem but, if properly characterized, should be usable. In all three cases, in order to assure reliable performance, testing of statistical samples is recommended. Resistors, magnetics, capacitors, and relays, in comparison to the active semiconductor components, do not present a serious degradation problem, although the effect of low-level radiation exposures on long-term reliability has not been fully evaluated.

### 3. Conclusions

From the review of the preliminary report, it can be concluded that:

- (1) With the present estimated TOPS radiation levels, capacitors, relays and switches, resistors, and magnetic devices should not be a problem with the possible exception of the unknown effect of combined neutron, electron, gamma ray, and proton environments.
- (2) The available information indicates that many active components will be seriously degraded by radiation during interplanetary missions.

- (3) In many cases, data is inadequate to do more than make gross estimates of degradation of part type performance.
- (4) Data evaluating proton damage is not available for many part types.
- (5) For most part types, hardening and screening procedures are not known or are in a developmental stage.
- (6) Although part degradation can be estimated for each environmental component, there is no data indicating how to assess the total degradation due to combined radiation environments.
- (7) Using currently available data, system reliability in a radiation environment will be difficult to assess, particularly for part types for which the radiation levels are near the threshold of damage. Even methods of assessing such damage needs to be more fully explored.

It is recommended that:

- (1) Evaluation testing be performed to obtain data on part types where no data exists or where lack of data is significant.
- (2) Testing with combined radiation environments be performed to obtain insight into how to assess the total threat to parts in interplanetary missions.
- (3) Methods of assessing reliability screening and hardening of irradiated components be more fully explored.

## XIV. Advanced Projects Development

### ENGINEERING MECHANICS DIVISION

#### A. TOPS Mechanical Devices, E. E. Sabelman

##### 1. RTG and Science Boom Actuator-Damper

*a. Introduction.* A torsion rod actuator and concentric rotating tube viscous damper have been proposed for deployment of the thermoelectric outer planet spacecraft (TOPS) radioisotope thermoelectric generator (RTG) and science booms (SPS 37-60, Vol. III, pp. 177-179). This combination is advantageous because the damper tubes replace a compression member between the hinges of the rigid boom to minimize the weight penalty (Fig. 1).

In the present TOPS configuration, 12L, the RTG assembly weighs 114 kg and its center of mass is 2.46 m from the boom hinge axis. The RTG boom must deploy through an angle of 70 deg, with minimum perturbation of spacecraft attitude and minimum residual velocity at latching.

A test model has been constructed to verify the deployment characteristics predicted by the analog computer program (SPS 37-60, Vol. III) by full-scale simulation

of the mass and moment arm of the RTG boom. The test setup has also been used to develop methods of low-friction deployment testing suitable for a flight-type spacecraft.

*b. Actuator-damper test model.* For fabrication reasons, the test model damper (Fig. 2) was shortened to 23 cm, half of the full-scale length. To preserve the same total shear area and thus the damping coefficient, a third tube was added, doubling the number of shear surfaces. The outer tube has a 2½-in. OD, a 0.035-in. wall, with a 0.012-in. nominal fluid film thickness. Using a one million centistoke silicone fluid, the net damping coefficient is  $10^9$  dyne-cm/(rad/s) [850 in.-lb/(rad/s)].

The torsion rod in the test model is full length, and is adjustable in preload by means of an indexing hub and pin on the rotating end. Rods ⅜ in. in diameter of 2024 aluminum and of 4340 heat-treated steel have been tested to date. The bearings on which the boom rotates are dry lubricated Acme threads cut into the aluminum hinge

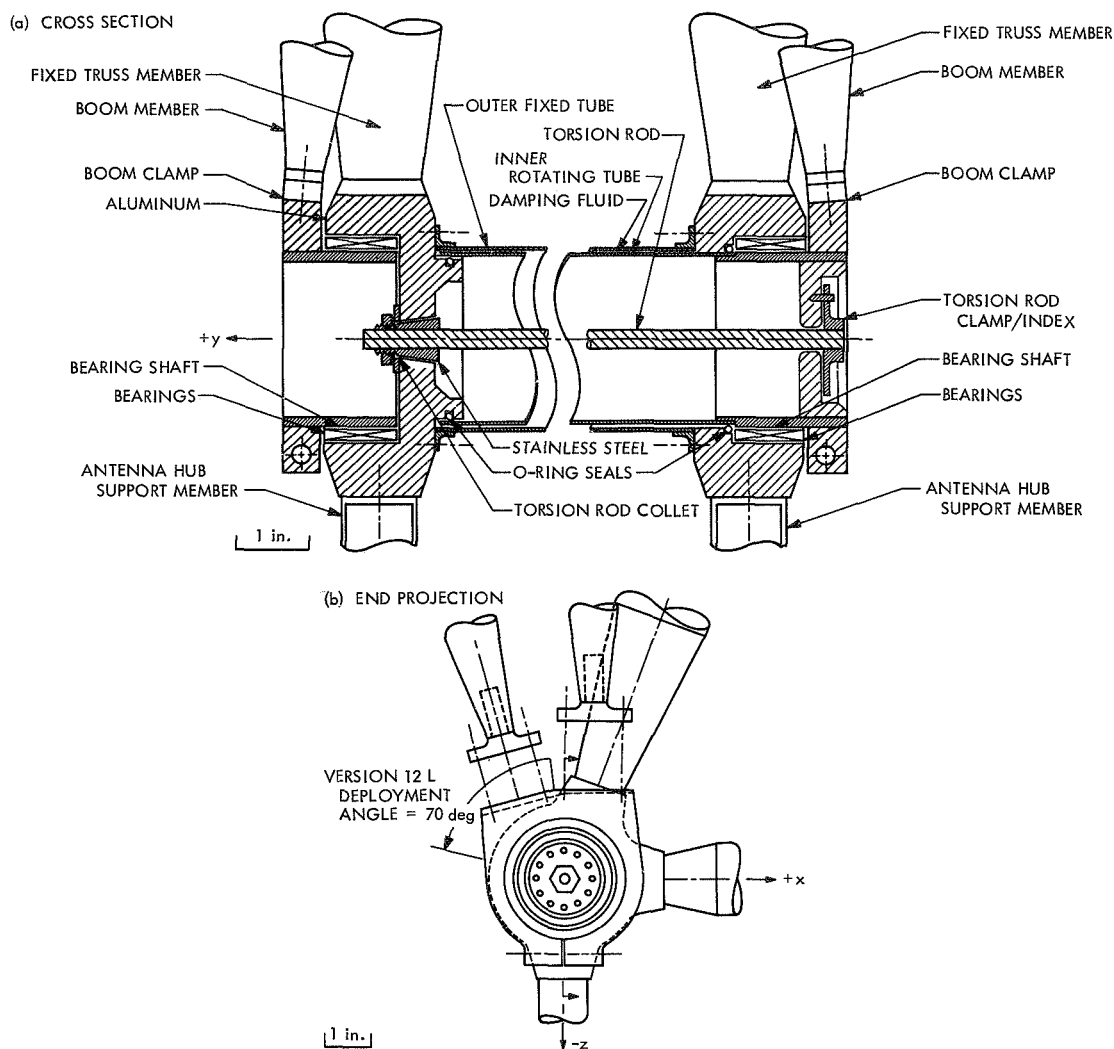


Fig. 1. Cross section and end projection of TOPS RTG actuator-damper

halves; this provides a labyrinth seal for the damping fluid, although O-ring seals would be used on a flight-type device.

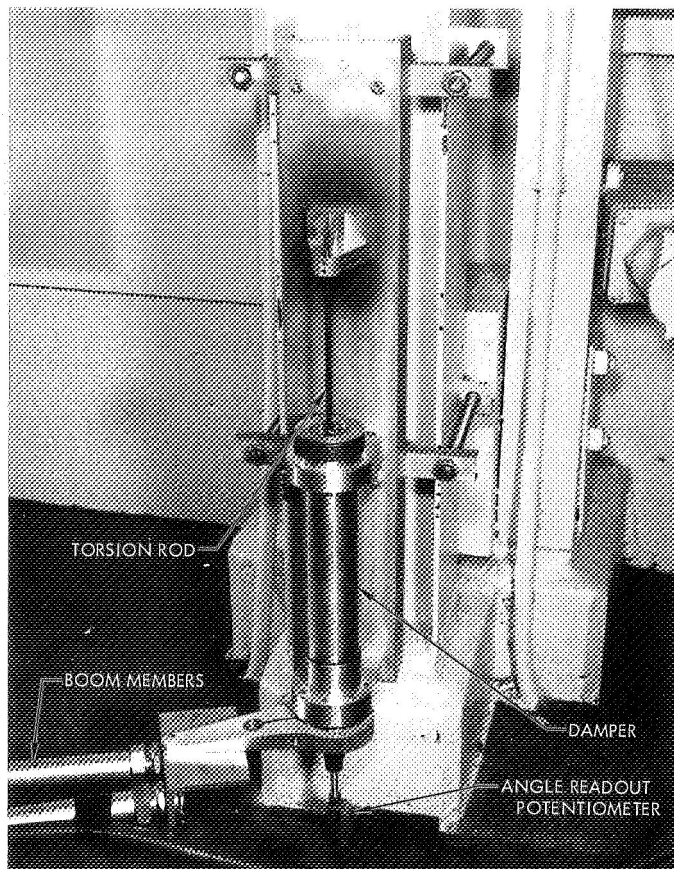
**c. RTG simulator and low-friction support.** The RTG mass is modeled by an appropriate number of lead bricks (Fig. 3). Minimal friction has been achieved by supporting the pallet of bricks on a 10-in.-diameter "Airfloat" air bearing. The bearing travels through an angle up to 100 deg on a circular track of  $\frac{3}{4}$ -in. plywood, covered with polyethylene sheeting, and supported by 42 leveling screws.

**d. Test procedures.** Most of the test runs were made to the TOPS version 12J requirements (RTG mass 100 kg at 1.95 m and 78-deg deployment angle). Because the moment of the simulated RTG is very large compared

with the actuator torque, it is quite important that the air-bearing support track is accurately leveled. The procedure for doing this is to first use a spirit level, then decouple the air-bearing pallet from the actuator-damper and use this as a much more sensitive indicator. (A thicker honeycomb sandwich construction would overcome difficulties in flexibility and unevenness of joints encountered with the plywood track.)

Testing is then initiated, with zero preload on the torsion rod and a reduced RTG mass. The mass is increased until critical damping is reached and the preload is increased if required to reach the zero angle (lock-open) point.

**e. Test results.** Performance of the damper has been as expected, but the available torque output and zero point

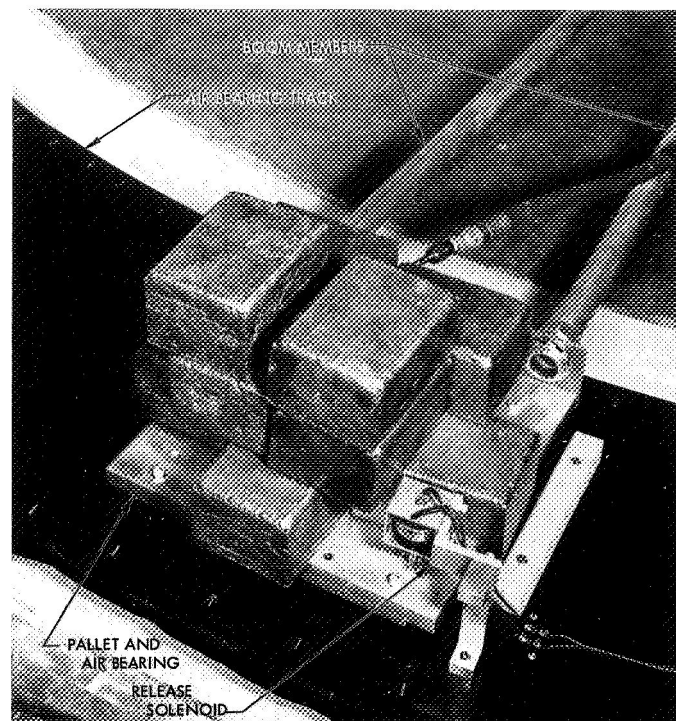


**Fig. 2. Actuator-damper test model (half-length damper)**

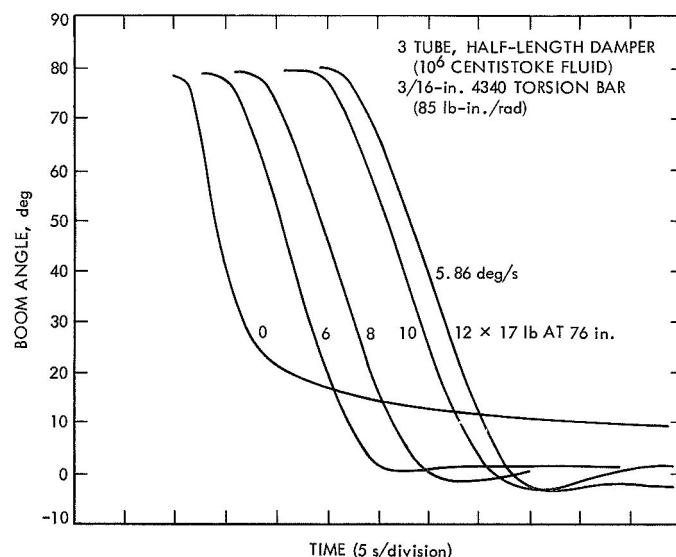
repeatability of the  $\frac{3}{16}$ -in. torsion rods have not been as great as predicted. Causes for this include: (1) residual friction in hinge bearings and air bearing (coefficient on the order of  $3 \times 10^{-4}$ ), (2) preloading of the torsion rod to within 20 deg of its yield point angle, and (3) distortions resulting from heat treating of the steel rod. The latter effects produce a torque-angle relation that appears to be a function of the square of the angle, rather than linear. Figure 4 shows typical angle versus time plots for the steel torsion rod; the peak torque coefficient of 85 in.-lb/rad corresponds to a linear coefficient of roughly 60 in.-lb/rad. The dimensionless relation

$$\frac{T_s}{C_d R_d \dot{\theta}_{\max}}$$

(where  $T_s$  is the mean torsion rod coefficient,  $C_d R_d$  is the damping torque coefficient, and  $\dot{\theta}_{\max}$  is the peak velocity in radians/second) gives an indication of the correspondence between theoretical and experimental systems; for



**Fig. 3. RTG simulator and air bearing**



**Fig. 4. RTG deployment simulator performance**

the system performance in Fig. 4, this quantity is 0.69, compared to 0.66 for the theoretical system described in SPS 37-60, Vol. III.

This result is sufficient to conclude that the concentric tube damper-torsion rod actuator is a valid approach to deployment of the high moment of inertia booms on

the TOPS spacecraft. Estimated weight for a flight-type actuator-damper is within the 2.5-lb limit: the entire assembly, including hinge fixtures, compression member, needle bearings, and boom member attachments, is 4.9 lb, while the weight penalty for inclusion of the actuator-damper is only 1.3 lb.

## 2. Magnetometer Boom Study

*a. Introduction.* A brief survey of commercially manufactured extensible booms has been conducted to determine if an advance in the state-of-the-art is necessary to meet the requirements for the vector helium magnetometer and plasma wave detector booms on TOPS. Design criteria at present are: 30-ft length; approximately 2-lb end mass; maximum weight of boom element plus deployment mechanism, 13 lb; pointing angle accuracy  $\pm 3$  deg; residual magnetization less than 0.1 gammas.

*b. Depth of study.* The available types of long extensible booms were examined for their capability of meeting the above criteria within the following constraints:

- (1) Maximum bending strength—the boom must withstand flight accelerations of 0.1 g maximum, but additionally should be capable of ground extension with minimum external support.
- (2) Torsional pointing accuracy—deflections resulting from deployment, thermal effects, and unsymmetrical acceleration loads should be within the  $\pm 3$ -deg limit.
- (3) Thermal distortion—bending due to temperature gradients around the boom circumference should be minimum. Such gradients result from solar heating and the proximity of the RTGs. As part of the study, a computer program has been devised to predict both these effects to a first-order approximation.
- (4) Maximum fundamental cantilever vibration frequency.
- (5) Minimum micrometeoroid damage or degradation.
- (6) Compatibility of materials with the minimum magnetic interference, low temperature, and zero outgassing requirements.

*c. Characteristics of tubular boom types.* Most of the common types of extensible booms are of preformed tubular metal foil, which is flattened, stored on spools,

and deployed either electrically or by stored strain energy. Variations include one, two, or three foil elements, overlapping or interlocked elements, and welded flange types. In general, all suffer from torsional weakness, high thermal distortion due to large cross section and poor circumferential heat transfer, and some minor buckling on repeated cycling.

Achieving the torsional pointing accuracy requires a heavier tubular foil boom: a typical example is the Spar Aerospace Products Bi-Stem (Fig. 5), which would be 2 in. in diameter, 0.010-in. wall, of austenitic stainless steel, with an element weight of 0.4 lb/ft, and total weight of about 21 lb.

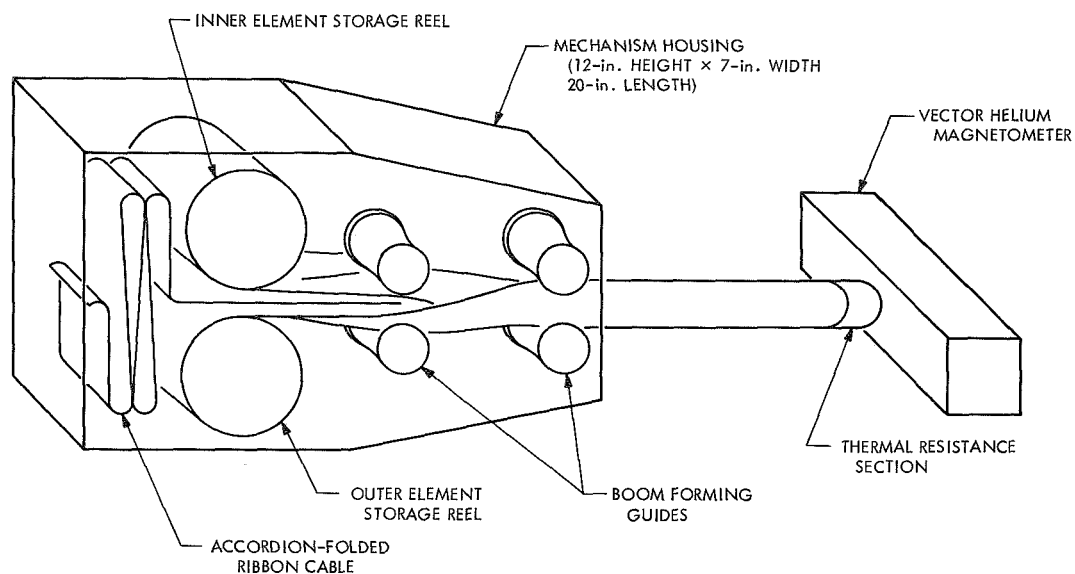
*d. Triangular truss boom.* The torsional and thermal deficiencies of the tubular boom are overcome by the Astromast, a product of Astro-Research Corporation (Fig. 6). A 6-in. boom with 0.15-in.-diameter silverplated epoxy-glass longerons would weigh 0.2 lb/ft, be torsionally rigid when fully deployed, have negligible thermal distortion since no longerons are shadowed, and have a total weight of 15 lb, including a 7-lb deployment mechanism/housing. Its major disadvantage is that cabling is exposed to the space environment, rather than being protected within the tubular boom element.

## 3. Regenerative Pump for TOPS Fluid Loop

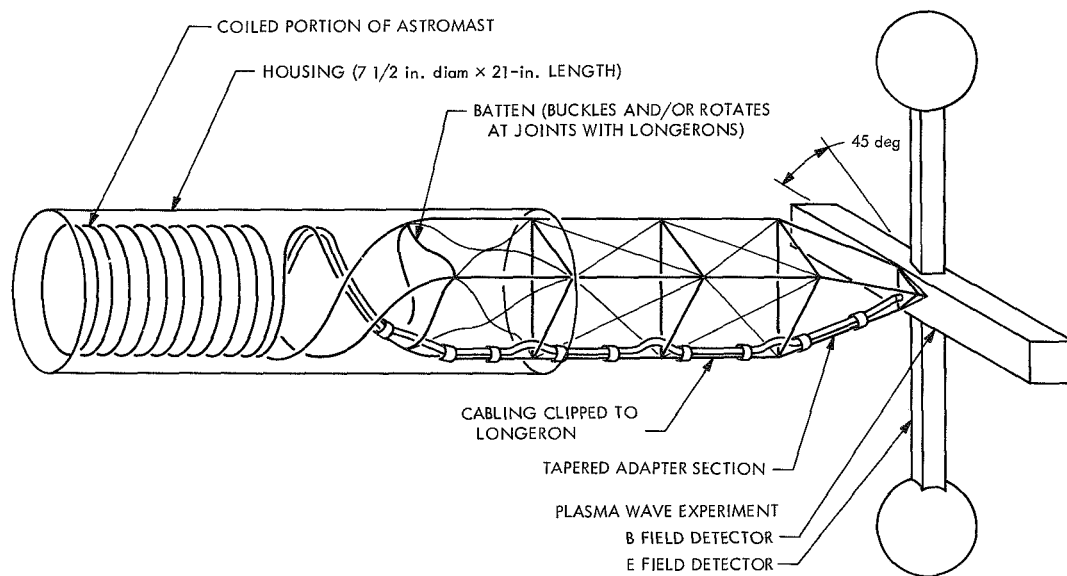
*a. Introduction.* Active temperature control of science instruments by means of a fluid loop transferring waste heat from the RTG is being contemplated. Present requirements are for a volume flow of 1.2 ml/s (8.7 lb/h) of silicate ester fluid, with a line pressure drop of about 1 psi. Redundant pumps will be carried on the spacecraft, but each should individually be capable of uninterrupted operation during the 100,000+ hour mission with minimum electric power consumption.

Candidate designs for the fluid loop pump fall into three categories: positive displacement, static, and rotary valveless. In general, positive displacement types have contacting moving parts which limit lifetime. The thermomechanical reciprocating pump conceived for the TOPS fluid loop (SPS 37-64, Vol. III, pp. 113–118) shares this fault, but its zero electric power consumption was considered attractive before the flow power requirements were reduced to the present level. Static pumps based on magnetohydrodynamic and ferrofluidic principles have been briefly investigated, but require excessive magnetic field strengths. Of the rotary pumps, centrifugal types





**Fig. 5. Two-spool Bi-Stem VHM Boom**



**Fig. 6. 30-ft Astromast boom**

have diminishing efficiencies at these flow power levels and require high-speed impellers (10,000 rpm up) (Ref. 1).

**b. Regenerative pump.** The regenerative pump (also known as "peripheral" or drag pump) operates by a combination of centrifugal and viscous effects to produce the required flow and head at very low speeds (300 rpm) with moderate efficiency (up to 10% at 0.1-W flow power). It consists of a vaned rotor and an annular cavity (Fig. 7) with the annulus constricted to minimum clearance between the outlet and inlet ports. Pressure in the outer annulus is built up by synchronized passage of the helical fluid flow through successive rotor vanes. Viscous drag on the fluid causes this induced vortex to follow the rotor until it reaches the outlet cut-off (Ref. 2).

Other advantages of the regenerative pump stem from its simplicity: the impeller can be incorporated into the rotor of a multi-pole hysteresis synchronous motor, thus requiring no penetrations of the fluid loop by shafts or electrical connections (Fig. 8). Only two bearings are required, which operate in the lubricating fluid. There are no axial bearing loads, and radial loads can be minimized by symmetrical placement of two inlet-outlet cut-offs.

**c. Test model, procedure, and results.** Commercial peripheral pumps with similar flow and head characteristics operate with higher rotor speeds (e.g., 3600 rpm); a test model with a 3-in.-diameter 76-vane rotor was therefore constructed to determine the feasibility of a 300-rpm regenerative pump (Fig. 9). Part of the stator/housing is made of Lucite plastic, to enable observation of the vanes

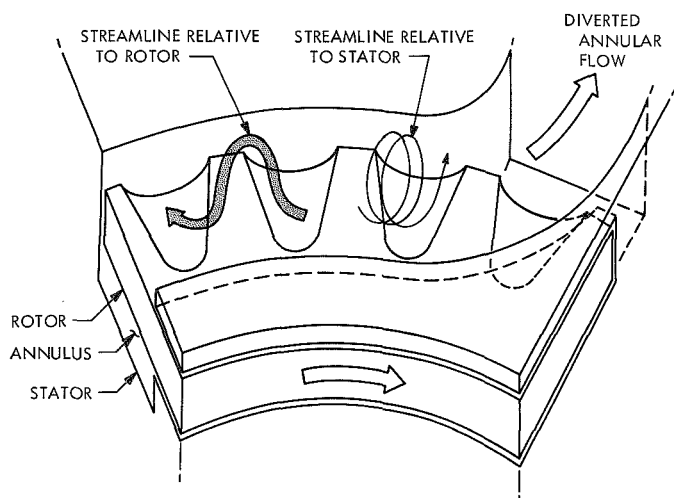


Fig. 7. Theoretical flow patterns of regenerative pump

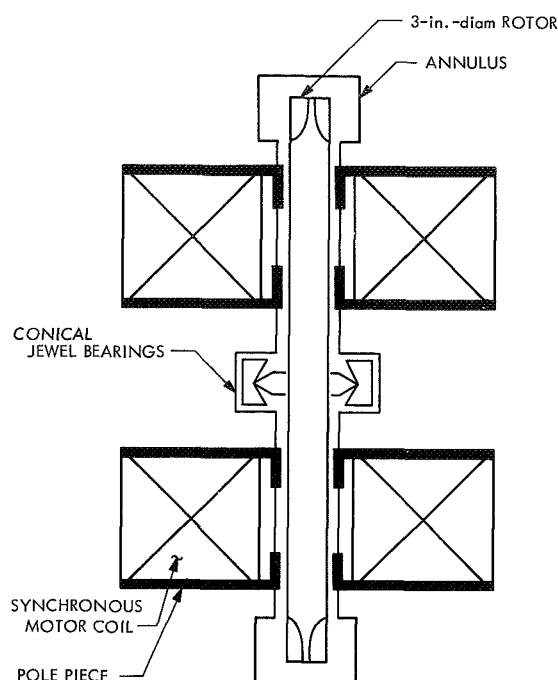
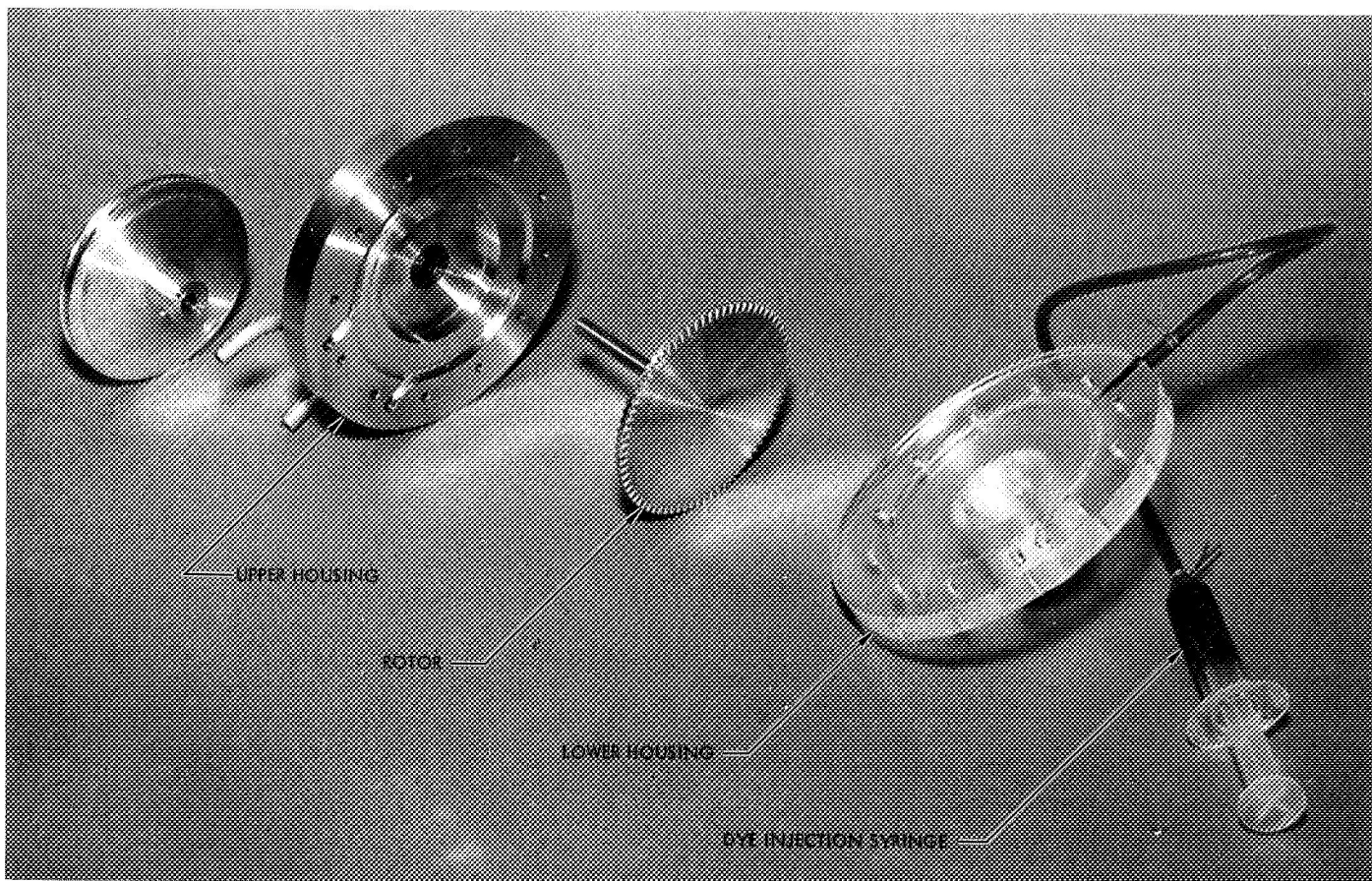


Fig. 8. Integrated pump motor

and annulus through a microscope and prism. The helical streamline pattern is marked either by bubbles entrained in the fluid or by injecting a stream of dye into the annular cavity.

Test runs were made from 180 to 1800 rpm at flow rates of 0.5 and 1.2 ml/s, driving the pump by a variable speed universal motor through a pulley. Other experimental apparatus (Fig. 10) included ammeter and voltmeter for measuring motor input power, oscilloscope and counter for rpm measurements, magnetic sensor for triggering the oscilloscope, counter and a Strobotac used for observing individual vanes, a mercury manometer, a ball-type flowmeter, and a torquemeter mounted on the drive pulley.

Efficiency (flow power/shaft power) and flow power are given in Fig. 11 for Coolanol 45 fluid and rotor speeds of 270 to 1350 rpm. Note that even this preliminary design is capable of producing the desired flow power (0.01 W) with 1.4% efficiency. The increasing efficiency with decreasing speed is not predicted by theory; efforts are now underway to extend the low range of the counter trigger and torquemeter to eliminate the possibility that this results from instrument inaccuracies. The flow power and developed pressure (Fig. 12) are nearly linear from 300 to 1500 rpm, permitting life testing of this type of pump to be accelerated as much as 5 times with easy extrapolation back to the normal operating regime.



**Fig. 9. 3-in.-diam rotor regenerative pump**

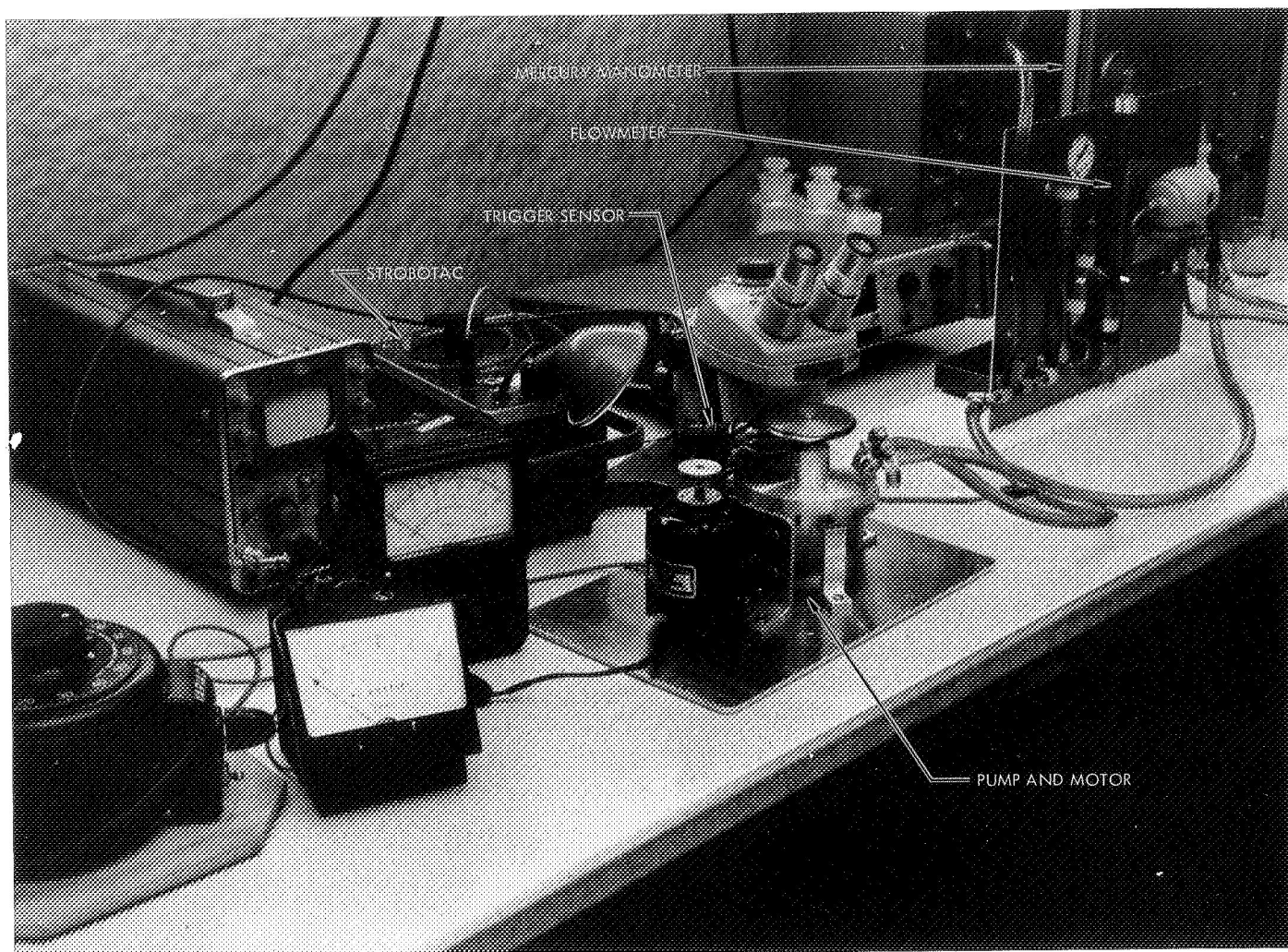
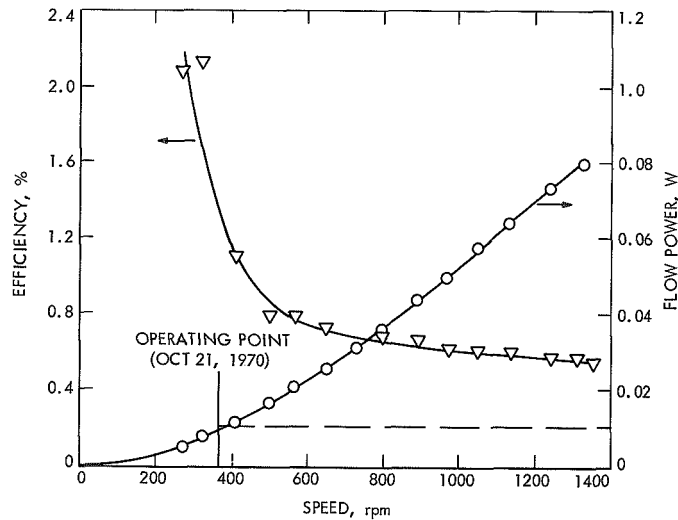
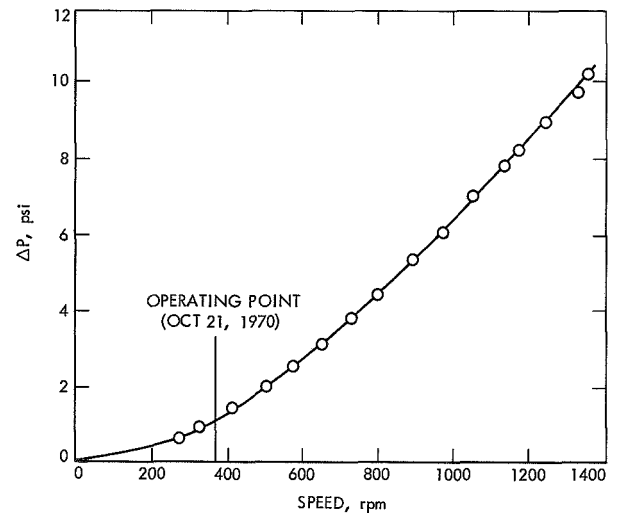


Fig. 10. 3-in.-diam pump in test assembly



**Fig. 11. Efficiency and flow power vs speed**  
(volume flow rate = 1.2 ml/s)



**Fig. 12. Pressure  $\Delta P$  vs speed at constant flow rate**  
(volume flow rate = 1.2 ml/s)

*d. Future work.* If the efficiency increase is real, it can be expected to peak at 5–10%; a second model of the regenerative pump can be designed to move this peak to the desired flow power and rpm. Alternatives to the regenerative pump are still being investigated; if this type of pump is selected, such an optimized model will be constructed and integrated into a drive motor. Possible types of bearings, and methods of determining bearing wear during accelerated and normal life testing, are being studied in anticipation of initiating such tests.

#### References

1. MacGregor, C. A., "Rotating and Positive Displacement Pumps for Low Thrust Rocket Engines," Interim Report TAMM 9115-71TU. Rocketdyne Division, North American Rockwell Corporation, Canoga Park, Calif., July 1969.
2. Wilson, W. A., Santalo, M. A., Oelrich, J. A., "A Theory of the Fluid-Dynamic Mechanism of Regenerative Pumps," *Trans. ASME*, Vol. 76, p. 1303, Nov. 1955.

## B. TOPS High-Gain Antenna,<sup>1</sup> D. J. Starkey

### 1. Introduction

The TOPS high-gain antenna employs a 4.25-m-diameter deployable main reflector. X-band efficiency requirements imply a necessary surface accuracy of approximately 1 mm rms. This antenna is being developed in-house. Although past articles (SPS 37-65, Vol. III, pp. 47-52, 122-125) have discussed several aspects of this development, this is the first comprehensive article.

### 2. History

When the TOPS requirement became known, in 1968, JPL conducted a survey of deployable high-gain antennas. Two of these, the Lockheed flexible radial rib and the Convair expandable truss, appeared most promising. Further study, however, indicated that each had certain shortcomings with respect to TOPS.

In June 1969, JPL initiated further study of a concept that had earlier been proposed for *Mariner B*. This is a rigid radial rib/soft mesh design, in which the ribs are hinged to fold up around the cassegrainian subreflector for stowage. Although this design lacks the packaging efficiency of other concepts, it appeared more straightforward and thus more likely to meet the surface accuracy requirement.

Ideally, the rib and mesh antenna would approximate the desired paraboloid by forming parabolic cylinder elements. In fact, the radial mesh tension that is necessary to prevent circumferential wrinkling of the mesh causes it to bow inward, increasing this inherent geometric error. At the time of this design study, JPL was already working with computer programs for determining geometric errors and for assessing the effects of surface errors on antenna gain.

Recognizing that the hub had to be fairly large to accommodate the hinge points around its periphery and to provide access for the feed horn, and in the belief that

doing this would tend to minimize the total rms surface error, the hub diameter was made the maximum that could be accommodated on TOPS, 1.4 m. This provides for a relatively large fixed center section and short ribs. Geometric error analysis then gave a minimum required rib number of 42; a 48-rib design was chosen somewhat arbitrarily. Preliminary layouts of the antenna were completed in September 1969.

The project status was reviewed in January 1970. This included assessment of the impact of further deployable antenna developments by industry. The conclusion was that the TOPS project objectives could best be met by continuing the in-house development of the radial rib antenna.

### 3. Design and Fabrication

The major design parameters for the TOPS high-gain antenna reflector are as follows:

Type: radial rib/mesh deployable

Deployed diameter: 4.25 m

Focal length/diameter: 0.415

Surface accuracy: 0.89 mm rms

Stowed envelope: cylindrical, 1.4-m diameter  $\times$  2.0 m

Weight (flight design goal): 15 kg

Figure 1 shows the mockup that was constructed as a part of the TOPS full-scale spacecraft model.

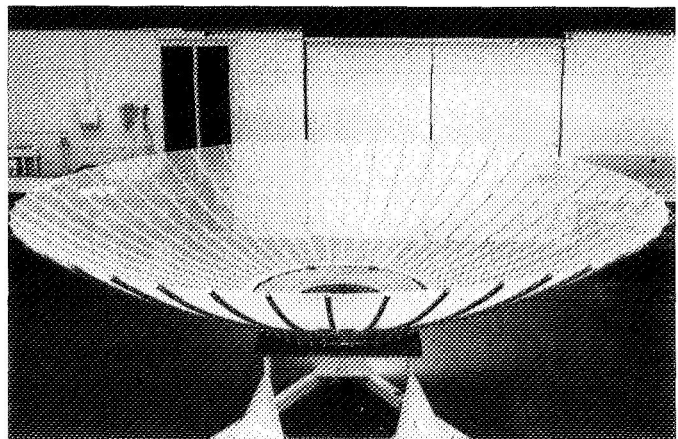


Fig. 1. TOPS high-gain antenna mockup

<sup>1</sup>In the course of this work, Radiation, Inc., disclosed to JPL that they have developed a very similar deployable rib/mesh antenna. It was designed for use with a focal point feed and incorporates a central deployment mechanism that is incompatible with the JPL cassegrainian feed. However, their experimental results served to greatly increase JPL's confidence in the correctness of the approach and the attainability of the TOPS objectives. Also, Lockheed and Convair designs have evolved; both companies have delivered antennas to NASA, so some of JPL's earlier concerns may no longer be valid.



The hub (Fig. 2) is a torus, of triangular cross section. It consists of "pork chop" stiffeners (Fig. 3), each having an integral rib hinge half, interconnected by sheet metal plates. To avoid increased fabrication costs, the development model hub was made overweight with respect to the flight design. However, this was done consistently, such that structural test results will give an accurate indication of the capabilities of the latter. The material is 6061-T6 aluminum alloy.

The rib material is 6061-T6 aluminum alloy tubing. The rib diameter tapers uniformly from 2.8 cm at the root to 0.9 cm at the tip, while the wall thickness varies from 0.5 to 0.4 mm over the same interval. The rib assembly consists of the following items, all adhesively bonded together: rib, T-section trim strip, rib extension, and elbow fitting (containing the other hinge half). The planned

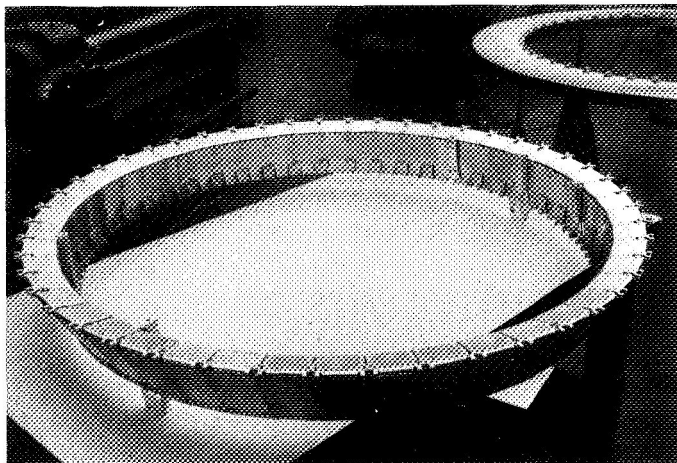


Fig. 2. TOPS high-gain antenna hub

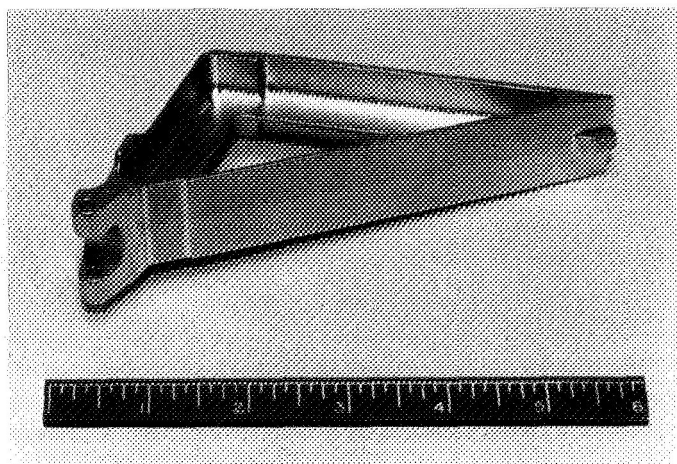


Fig. 3. TOPS high-gain antenna stiffeners

strategy is to bend the ribs to an approximately correct contour and then meet the close tolerance requirement by final machining of the trim strip. The rib will be accurately positioned with respect to the hub by line drilling of the hinge pin hole, using appropriate tooling. A stop on the rib extension, which bears against the "pork chop" stiffener, provides for adjustment of the deployed rib position. In contrast with the hub, the ribs are of substantially flight design, since there is much to be learned about their fabrication and properties.

The optimum deployed rib shape is calculated with a program which first computes the shape assumed by a mesh surface attached to parabolic ribs and then optimizes the rib shape so as to minimize the rms surface error with respect to the desired paraboloid. The as-manufactured (relaxed) rib shape is then obtained by introducing the elastic properties of the rib. A thermal offset could also be included, if it were desired to design for minimum surface error at other than the manufacturing temperature, but this was not done for the development model. The rationale is that most of the testing will be done at room temperature and that this would introduce needless complexity.

The deployment mechanism is one area wherein it is felt that the present design is less than completely satisfactory. The requirement for penetrating the center of the dish with the feed horn and its waveguide connection effectively precludes use of central deployment mechanisms which would otherwise be attractive. The present design accomplishes deployment by means of redundant Negator springs attached to the rib extensions and operating from drums contained within the hub. Pending the development of compact, reliable rib deployment rate control devices, the springs must be carefully sized with respect to the deployment environment, such that the ribs do not impact the stops with sufficient angular velocity to cause damage.

The mesh that will be used on the TOPS antenna is a rather standard one for deployables; a tricot knit, Chromel-R material having approximately 5 strands per lineal cm. It is electroless gold-plated for improved RF reflectance. While there is some concern regarding the brittleness and relatively poor adhesion of the plating relative to a long-duration outer planet mission, and efforts are being made to develop a better material, it is adequate for the development model antenna.

The strategy that is presently contemplated for fabrication and installation of the mesh surface is to stretch the

material under the proper tensions, outline the gores with glass tape (which ensures subsequent stability of the edge dimensions), and cut them out. The gores would then be machine-sewn to form the complete "skirt" and this would be attached to the ribs by sewing through holes in the T-section. This technique (except for differences in rib attachment) was used in the construction of the model previously mentioned, and worked well.

Temperature control considerations were included in the design of the hub and the rib. The present plan is to insulate the hub so as to minimize gradients therein and to provide the ribs with a grey-body finish to ensure predictability and long-term stability of properties. Preliminary rib temperatures and gradients for several points along the TOPS trajectory have been calculated, but, because of resource limitations, their total effects on surface distortion and focal point location have not yet been determined.

#### **4. Development Status**

The hub is completed; other parts (except the central dish) are either on-hand or on order. Ways of reducing the

estimated cost of the central dish are being investigated prior to ordering this item. Rib assembly and final machining is expected to be a pacing item. Completion is scheduled for June 1971.

Current tasks include: experiments in fabrication and assembly techniques, determination of mesh properties, thermal distortion analysis, development of a rib deployment rate control device, and the development of a superstructure capable of carrying the launch loads imposed by the subdish and other spacecraft equipment mounted to it, and which does not cause excessive RF blockage losses.

#### **5. Planned Tests**

Mechanical tests will include static surface accuracy, deployment repeatability, and simulated 0-g measurements. Although it would be desirable to perform thermal-vacuum tests, presently envisioned TOPS resources do not permit this, and alternate methods of confirming thermal distortion analyses are being sought.

The Telecommunications Division has developed the TOPS antenna feed and will perform RF testing.



## XV. Spacecraft Data Systems

### ASTRIONICS DIVISION

#### A. Evaluation of Recording Tape and Heads for Spacecraft Magnetic Tape Recorder

**Applications,** J. K. Hoffman, S. H. Kalfayan,<sup>1</sup> and R. H. Silver<sup>1</sup>

Phase III of a study of the characteristics of magnetic recording tapes and the tape-to-head interface in the spacecraft environment was completed. The work was performed under direction of the Spacecraft Data Systems Section with facility and personnel support from the Propulsion Division. The results of the work are reported in detail in SPS 37-65, Vol. III, pp. 168-171. Phases I and II were reported in SPS 37-63, Vol. III, p. 160, and SPS 37-64, Vol. III, p. 119, respectively.

Consistent with the objective to optimize the head-to-tape interface, recent work has concentrated on investigation of the effects of varying materials of head

fabrication. In this latest work, some previous experiments were repeated employing magnetic heads with Monel metal brackets instead of the brass bracket heads previously used.

These experiments also included examination of the behavior of a surface-treated (Vacoated) sample of a previously tested tape type (3M 20250) to evaluate the effect of the process on the stick-slip and tape-to-head adhesion phenomenon.

The knowledge obtained from this program has been instrumental in understanding and solving the tape-to-head interface problems encountered in development of the *Mariner* Mars 1971 data storage subsystem. It is of significant value relative to the design of data storage subsystems for future missions.

The investigations are continuing and further tests involving additional parameters are planned.

<sup>1</sup>Polymer Research Section, Propulsion Division.

## B. The Implementation of $m$ -ary Linear Feedback Shift Registers With Binary Devices, M. Perlman

### 1. Introduction

The binary shift register with linear logic (i.e., sum modulo 2 or complement sum modulo 2) feedback has been a subject of serious investigation since 1952. Gilbert, Golomb, Huffman, Zierler, and Peterson (Refs. 1-5) are among the notable contributors. The amenability of the linear (binary) feedback shift register (LFSR) to mathematical description, analysis, and synthesis played a significant role in its development. Mathematical theory for characterizing the LFSR actually preceded the physical existence of the LFSR (Ref. 6). However, diverse applications, coupled with advancing technology, provided underlying motivation in the study of LFSRs [as well as nonlinear FSRs (Ref. 7), which are much less tractable mathematically].

Applications include scaling, counting, prescribed sequence generation, ranging, error-correcting encoders and decoders, secure communication coding, arithmetic operations over a finite field, and load-sharing matrix switching. As noted in Ref. 7, the FSR can be used as a mathematical model for the following:

- (1) Random number generators.
- (2) Autonomous finite-state machines for research in recursive logic and automata theory.
- (3) Markov processes.

Much of the research was confined to the binary FSR because of the availability of reliable two-state devices. Linear  $p$ -ary (where  $p$  is prime) FSRs have also been treated (Ref. 5). Their behavior is related to the divisibility properties of polynomials over  $GF(p)$  ( $p = 2$  is a special case). In SPS 37-64, Vol. III, pp. 120-124, and Ref. 8, the  $m$ -ary LFSR is shown to be decomposable into parallel  $p$ -ary LFSRs.

The  $m$ -ary LFSR may be characterized by the linear recurrence relation

$$a_k = \gamma + \sum_{i=1}^r c_i a_{k-i} \bmod m \quad (1)$$

The state of the  $i$ th stage at clock pulse interval (CPI)  $k$  is denoted by  $a_{k-i}$ . Each of the  $r$ -stages can assume one of  $m$  distinct states  $0, 1, \dots, m-1$ . The digit fed back at CPI  $k$  is  $a_k$ , the output of a modulo  $m$  summer

whose inputs are  $\gamma$  and  $c_i a_{k-i}$  over all  $i$ , where  $r$  and  $c_i \in \{0, 1, \dots, m-1\}$ . The state of the  $i$ th stage at CPI becomes the state of the  $(i+1)$ th stage at CPI  $k+1$ . The sequence  $\{a_k\} = \{a_0, a_1, a_2, \dots\}$  for a given initial state  $a_{-1} a_{-2} \dots a_{-r}$  is comprised of the coefficients of ascending powers of  $x$  of  $G_\gamma(x)$ , the *generating function*. Expressed in closed form (Refs. 7 and 8),

$$G_\gamma(x) = \frac{\gamma + (1-x)g(x)}{(1-x)f(x)} \bmod m \quad (2)$$

where

$$g(x) = \sum_{i=1}^r c_i x^i (a_{-i} x^{-i} + a_{-i+1} x^{-i+1} + \dots + a_{-1} x^{-1})$$

and

$$f(x) = 1 - \sum_{i=1}^r c_i x^i$$

### 2. Implementation of $m$ -ary LFSRs with Binary Devices

An  $m$ -ary LFSR is first decomposed into  $p$ -ary parallel LFSRs where each  $p$  is a distinct prime factor of the integer  $m$ . The states of a  $p$ -ary ( $p > 2$ ) LFSR are coded in binary. The implementation of a  $p$ -ary LFSR is comprised of  $n$  parallel binary shift registers where  $2^{n-1} < p < 2^n$  with *interdependent* binary feedback functions which are in general nonlinear.

*Example 1.* Given  $a_k = 5a_{k-1} + a_{k-2} \bmod 6$ ,

$$G_0(x) = \frac{(5a_{k-1} + a_{k-2}) + a_{k-1}x}{1+x+5x^2} \bmod 6$$

The recurrence relation and its associated generating function describe the behavior of the LFSR shown in Fig. 1.

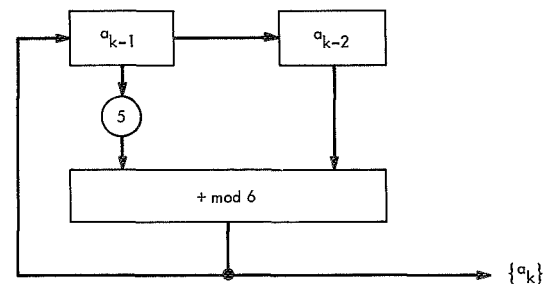


Fig. 1. LFSR characterized by  $a_k = a_{k-1} + 5a_{k-2} \bmod 6$

By decomposition (Ref. 9, Theorem 5),

$$G_0^2(x) = \frac{(b_{-1} + b_{-2}) + b_{-1}x}{1 + x + x^2} \bmod 2$$

$$G_0^3(x) = \frac{(2d_{-1} + d_{-2}) + d_{-1}x}{1 + x + 2x^2} \bmod 3$$

where

$$b_{-1} = a_{-1} \bmod 2, \quad b_{-2} = a_{-2} \bmod 2$$

$$d_{-1} = a_{-1} \bmod 3, \quad d_{-2} = a_{-2} \bmod 3$$

characterize a binary and a ternary LSFR operating in parallel.

$G_0^2(x)$  and  $G_0^3(x)$  are associated with the simultaneous recurrence relations

$$b_k = b_{k-1} + b_{k-2} \bmod 2$$

and

$$d_k = 2d_{k-1} + d_{k-2} \bmod 3$$

respectively.

The implementation of the binary LSFR is completely described by the generating function  $G_1^2(x)$  and its associated recurrence relation. It appears in Fig. 2. The states in the state table of the ternary LSFR can be represented in binary as follows:

$k$	$d_{k-1}$	$d_{k-2}$	$d_k$	$x_{k-1}$	$y_{k-1}$	$x_{k-2}$	$y_{k-2}$	$x_k$	$y_k$
0	0	1	1	0	0	0	1	0	1
1	1	0	2	0	1	0	0	1	0
2	2	1	2	1	0	0	1	1	0
3	2	2	0	1	0	1	0	0	0
4	0	2	2	0	0	1	0	1	0
5	2	0	1	1	0	0	0	0	1
6	1	0	1	0	1	1	0	0	1
7	1	1	0	0	1	0	1	0	0
0	0	0	0	0	0	0	0	0	0

The binary representations of  $d_{k-1}$  and  $d_{k-2}$  are stored  $x_{k-1}$   $y_{k-1}$  and  $x_{k-2}$   $y_{k-2}$ , respectively. The most significant binary digits reside in the  $x$  register. Thus, the ternary LSFR is implemented as two binary registers with *inter-dependent* Boolean feedback functions which are non-linear. The feedback functions (i.e.,  $x_k$  and  $y_k$ ) appear in Fig. 3 in minimum disjunctive form resulting from using unspecified binary states as "don't cares."

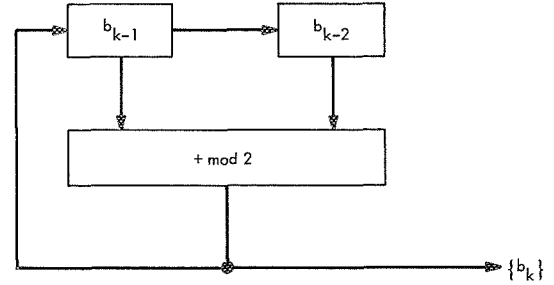
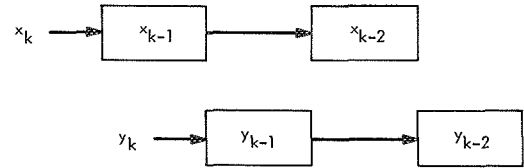


Fig. 2. LSFR characterized by  $b_k = b_{k-1} + b_{k-2} \bmod 2$



$$x_k = \bar{x}_{k-1} \bar{y}_{k-1} x_{k-2} \vee y_{k-1} \bar{x}_{k-2} \bar{y}_{k-2} \vee x_{k-1} y_{k-2}$$

$$y_k = x_{k-1} \bar{x}_{k-2} \bar{y}_{k-2} \vee \bar{x}_{k-1} \bar{y}_{k-1} y_{k-2} \vee y_{k-1} x_{k-2}$$

Fig. 3. Implementation LSFR characterized by  $d_k = 2d_{k-1} + d_{k-2} \bmod 3$  where  $d_{k-i} \longleftrightarrow x_{k-i} y_{k-i}$  for  $i = 0, 1, 2$

A one-to-one correspondence exists between the states  $a_{k-1} a_{k-2}$  and  $b_{k-1} b_{k-2} x_{k-1} y_{k-1} x_{k-2} y_{k-2}$ . Let

$$a_{-1} a_{-2} = 02$$

then

$$b_{-1} b_{-2} = 00$$

$$d_{-1} d_{-2} = 02$$

and

$$x_{-1} y_{-1} x_{-2} y_{-2} = 0010$$

The cycle (of states) containing these initial states are

$k$	$a_{k-1}$	$a_{k-2}$	$a_k$	$b_{k-1}$	$b_{k-2}$	$x_{k-1}$	$y_{k-1}$	$x_{k-2}$	$y_{k-2}$	$x_k$	$y_k$
0	0	2	2	0	0	0	0	1	0	1	0
1	2	0	4	0	0	1	0	0	0	0	1
2	4	2	4	0	0	0	1	1	0	0	1
3	4	4	0	0	0	0	1	0	1	0	0
4	0	4	4	0	0	0	0	0	1	0	1
5	4	0	2	0	0	0	1	0	0	1	0
6	2	4	2	0	0	1	0	0	1	1	0
7	2	2	0	0	0	1	0	1	0	0	0

The period 8 associated with  $a_k$  (for  $a_{-1} a_{-2} = 02$ ) can be determined from the least common multiple (LCM) of the periods of  $G_0^2(x)$  and  $G_0^3(x)$ , namely LCM (1,8) or 8.

Whenever  $m$  is a product of distinct primes, as in *Example 1*, a one-to-one correspondence exists between the cycles of  $m$ -ary states and the cartesian product of the parallel  $p$ -ary states.

*Example 2.* Given  $a_k = 1 + a_{k-1} + 3a_{k-2} \bmod 8$ , the longest cycle contains  $a_{-1} a_{-2} = 01$  as an initial state and results in a

$$G_1(x) = \frac{1}{(1-x)(1+7x+5x^2)} \bmod 8$$

The period of the longest cycle divides the period associated with

$$G_1^2(x) = \frac{1}{(1-x)(1+x+x^2)^2} \bmod 2$$

The period of the  $m$ -ary cycle ( $m = 8$ ) is 6 whereas the period associated with  $G_1^2(x)$  is 12 (Ref. 9). Decrementing the power of the exponent in the denominator of  $G_1^2(x)$  results in the generating function

$$\hat{G}_1(x) = \frac{1}{(1-x)(1+x+x^2)} \bmod 2$$

with a period of 6 equal to the  $m$ -ary cycle. Associated with  $\hat{G}_1(x)$  is

$$b_k = 1 + b_{k-2} + b_{k-4} \bmod 2$$

There are 10 cycles of length 6, one of length 3, and one of length 1 associated with  $a_k$ , whereas there are two cycles of length 6, one of length 3, and one of length 1

associated with  $b_k$ . A homomorphic mapping of cycles of states associated with  $a_k$  onto those associated with  $b_k$  exists. Thus, a binary LFSR characterized by  $b_k$  will contain at least one cycle of each length appearing in the  $m$ -ary LFSR.

## References

1. Gilbert, E. N., *Quasi-Random Binary Sequences*, Memorandum MM-53-1400-42. Bell Telephone Laboratories, Nov. 27, 1953.
2. Golomb, S. W., *Sequences With Randomness Properties*, Final Report, Contract W. 36-039SC-54-36611. Glenn L. Martin Co., Baltimore, Md., 1955.
3. Huffman, D. A., "The Synthesis of Linear Sequential Coding Networks," *Information Theory*, pp. 77-95. Edited by Colin Cherry. Academic Press, New York, 1956.
4. Zierler, N., "Linear Recurring Sequences," *J. Soc. Indust. Appl. Math.*, Vol. 7, pp. 31-48, 1959.
5. Peterson, W. W., *Error-Correcting Codes*. The M.I.T. Press, Cambridge, Mass., 1961.
6. Hall, M., "An Isomorphism Between Linear Recurring Sequences and Algebraic Rings," *Trans. Am. Math. Soc.*, Vol. 44, pp. 196-218, 1938.
7. Golomb, S. W., *Shift Register Sequences*. Holden-Day, Inc., San Francisco, Calif., 1967.
8. Perlman, M., *Pseudonoise Sequence Generation With Three-Tap Linear Feedback Shift Registers*. Technical Report 32-1492; Jet Propulsion Laboratory, Pasadena, Calif., Nov. 1969.
9. Perlman, M., "Cycle Structure Determination of an  $M$ -ary Linear Feedback Shift Register," in *Proceedings of the Fourth Hawaii International Conference on System Sciences, University of Hawaii, Honolulu, Ha.*, Jan. 1971.

## C. Extraction of Complete Subgraphs: Command Prefix Code for TOPS, N. Deo

### 1. Introduction

A method of establishing command word synchronization for the *Viking* 1975 mission was previously suggested by McEliece (SPS 37-61, Vol. III, pp. 48-50). The synchronization is accomplished by transmitting, to the spacecraft command decoder, a binary word (called prefix) sandwiched between the "idle sequence,"  $\cdots 00000 \cdots$ , and the actual command word. The prefix acts as the word-start signal and synchronizes the command decoder. As pointed out by McEliece, the criterion for selecting a prefix is that it should be a shortest possible sequence capable of being recognized by the command decoder even if the prefix (as well as the preceding idle sequence) has been corrupted (up to a certain pre-specified extent) by the noisy channel.

### 2. Command Decoding in the TOPS

A similar approach is proposed for the thermoelectric outer planet spacecraft (TOPS), with one important difference. Because of extremely long mission life, unknown and hostile environments, and long transmission distances, the spacecraft is expected to have several redundant command decoders. The central computer subsystem will also serve as a command decoder (Ref. 1). It is, therefore, proposed that the prefix, in addition to being an error-correcting synchronizer, also identify the specific command decoder for which the command word following the prefix is intended. In other words, the prefixes will act as combinations to unlock specified information-processing channels on the spacecraft. This requires a set of different prefixes, each corresponding to a unique command decoder. Depending on which of the command decoders the ground wishes to use, the corresponding prefix would be transmitted. The prefixes must, of course, be at more than a certain specified Hamming distance from each other because of the noisy channel, etc. A pair of prefixes that satisfies this distance requirement may be called *mutually compatible*. It can be shown that *mutual compatibility is a symmetric, non-reflexive, non-transitive, binary relation*. It can, therefore, be represented by a simple (i.e., free of self-loops and parallel edges) undirected graph.

Thus, the problem of finding a dictionary with the largest possible number of mutually compatible prefixes of a given length and with specified properties with respect to the idle sequence can be reduced to a graph-theoretic problem.

### 3. Graph-Theoretic Definitions

Since the basic graph-theoretic concepts are fundamental to this problem, the following definitions are presented for those unfamiliar with graph theory. (The reader is directed to Refs. 2 and 3 for a more thorough presentation.)

A *graph* consists of a finite set of points (called *vertices*) and a finite set of line-segments (called *edges*) joining pairs of these points. A graph is said to be *simple* if no edge joins a vertex to itself (i.e., no *self-loop* allowed) and no two edges join the same pair of vertices (i.e., no *parallel edges* allowed). A simple graph is said to be *complete* if there is exactly one edge between every pair of vertices. A simple graph of  $n$  vertices has  $n(n-1)/2$  edges. A *walk* is an alternating sequence of vertices and edges beginning and ending with vertices in which an edge joins the vertices preceding and following the edge. A *path* is a walk in which no vertex appears twice. A graph is said to be *connected* if there is some path between every pair of vertices. A largest connected subgraph in an unconnected graph is called a *component*. A vertex  $v$  in a connected graph  $G$  is called *articulation vertex* or *cut-vertex* if removing  $v$  (together with all edges incident on  $v$ ) leaves the remaining subgraph unconnected. An *adjacency matrix*  $A$  of graph  $G$  with  $n$  vertices is an  $n$  by  $n$  (0,1) matrix, whose  $i,j$ th ( $=j,i$ th) entry is 1 if there is an edge joining vertices  $i$  and  $j$ ; otherwise the  $i,j$ th entry is 0. The number of edges incident at a vertex is called the *degree* of a vertex. A vertex of degree one is called a *pendant* vertex.

### 4. An Example

The problem can be best explained with an example. Assume, for this example, that the length of each prefix is seven bits and that a prefix is required to be at least single-error correcting. Therefore, the Hamming distance between a prefix and all of its shifts (with zeroes filled in at the left) should be at least three. For instance, a 7-bit word  $b_1 b_2 b_3 b_4 b_5 b_6 b_7$  and its shifted version by two  $0 0 b_1 b_2 b_3 b_4 b_5$  must be different in at least three positions. Since the idle sequence consists of all zeroes, a prefix must not start with zero, i.e.,  $b_1 = 1$ .

It is not difficult to write a program for generating all 7-bit words, satisfying the constraints just pointed out. Such a program turned out a list of 33 words, each 7 bits long (out of a possible  $2^7 = 128$  words).

To obtain a largest set of mutually compatible words, we generate a graph  $G$  in which a vertex represents a

prefix word, and if two prefixes are compatible, an edge is drawn between the corresponding pair of vertices. Clearly, a pendant vertex (or a vertex of degree 1) in  $G$  represents a prefix which is compatible with only one other prefix. Since we are searching for codes with more than two prefixes, we can immediately delete all pendant vertices. This results in a graph  $G_1$  of 21 vertices, as shown in Fig. 1. The following are 21 prefixes, corresponding to the 21 vertices in  $G_1$ :

```

1 1 1 1 1 0 1
1 1 1 1 0 1 1
1 1 1 0 1 1 1
1 1 1 0 1 0 0
1 1 1 0 0 1 1
1 1 1 0 0 1 0
1 1 1 0 0 0 1
1 1 0 1 1 1 1
1 1 0 1 1 1 0
1 1 0 1 1 0 0
1 1 0 1 0 1 1
1 1 0 1 0 1 0
1 1 0 1 0 0 1
1 1 0 1 0 0 0
1 1 0 0 1 1 1
1 1 0 0 1 0 1
1 1 0 0 1 0 0
1 1 0 0 0 1 0
1 0 1 1 0 1 0
1 0 1 1 0 0 0
1 0 1 0 1 1 0

```

The vertices in Fig. 1 are labeled with the octal equivalent of the corresponding 7-bit words.

A *largest complete subgraph* (usually there is more than one largest complete subgraph in a graph) in  $G_1$  is the largest possible dictionary of mutually compatible prefixes we have been looking for. It turns out that in this graph there are no complete subgraphs of more than three vertices. Moreover, the graph contains eight different complete subgraphs of three vertices (called  $K_3$ ). One such dictionary, for example, consists of the following three words:  $\{(1110111), (1100100), (1011010)\}$ . Thus, with 7-bit, single-error-correcting prefixes, we can have up to three different command decoders.

Similarly, searching for a code of 8-bit prefixes with single-error-correcting capability, one finds that (out of a possible  $2^8$  words), graph  $G$  consists of 87 vertices and  $G_1$  of 85. The largest complete subgraphs in  $G_1$  consist

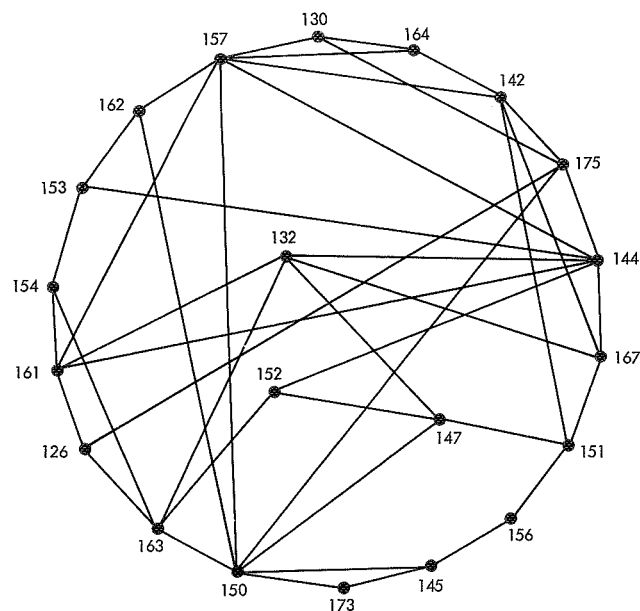


Fig. 1. Graph  $G_1$  of 7-bit, 1-bit error-correcting prefixes

of five vertices ( $K_5$ ); this allows a maximum of five command decoders.

## 5. Graph-Theoretic Algorithms

Developing an efficient algorithm for identifying (and thus extracting) all largest complete subgraphs in a given graph is of central importance here. (The futility of combinational brute force in trying to find all compatible words among 85 words taking three at a time, four at a time, and so forth, is obvious.) Algorithms available for computerized graph-theoretic operations are presently few and far between (Refs. 4 and 5).

There are several ways of representing graphs for computer storage and manipulation (Ref. 4); in this case, the adjacency matrix was used.

If a graph is unconnected, it is inefficient to manipulate the entire subgraph in looking for largest subgraphs, because no complete graph can have vertices in two different components of an unconnected graph. Therefore,  $G_1$  must be first checked for connectedness. Each component should be identified and looked at separately, if  $G_1$  is unconnected.

Similarly, since an articulation vertex (or cut vertex) cannot belong entirely to any one complete subgraph (except for the trivial cases of  $K_1$  and  $K_2$ ), a connected component of  $G_1$  should be checked for separability. If separable, the graph should be broken into its different

blocks and then each block should be searched for the largest complete subgraph.

Thus, the connectedness and separability algorithms are essential subalgorithms for extraction of complete subgraphs. The details of different algorithms and programs for identification of all largest complete subgraphs in a connected, nonseparable graph are being prepared.

The programs were written in APL/360 and run on a time-shared terminal connected to an IBM 360/50.<sup>1</sup>

---

<sup>1</sup>The author gratefully acknowledges the contribution of Harold Baugh, the cognizant engineer for the TOPS command decoder subsystem, who originally suggested this problem and helped in writing the programs.

## References

1. Martin, B. D., "Data Subsystems for 12-Year Mission," *Astronaut. Aeronaut.*, Vol. 8, No. 9, p. 56, Sep. 1970.
2. Harary, F., *Graph Theory*. Addison-Wesley Publishing Co., New York, 1969.
3. Berge, C., *The Theory of Graphs and Its Applications*. John Wiley & Sons, Inc., New York, 1962.
4. Read, R. C., "Teaching Graph Theory to Computer," in *Recent Progress in Combinations*, pp. 161-173. Edited by W. T. Tutte. Academic Press, New York, 1969.
5. Read, R. C., "Graph Theory Algorithms," in *Graph Theory and Its Applications*, pp. 51-78. Edited by B. Harris. Academic Press, New York, 1970.

## D. Jupiter's Electron Dose Calculations on Metal Oxide Semiconductor Structures,

S. P. Li and J. B. Barengoltz

### 1. Introduction

Experimental data of radiation doses on metal oxide semiconductor field-effect transistors (MOSFETs) in the literature are usually reported in terms of rad (Si). For electron irradiation, laboratory energies of several keV, 1 MeV, and 3 MeV are common and, in some cases, up to 20 MeV (Ref. 1). The relative radiation damage may be studied experimentally by using equal electron fluences of different energies, but the experiments become more difficult as the energy is increased. As an additional complication to adequate testing, a MOSFET device inside a spacecraft is not directly irradiated by electrons as in the laboratory, but by the electrons which penetrate a certain thickness of the spacecraft wall with reduced energy plus the ensuing gamma rays. Since the radiation damage on MOSFETs is dominantly due to ionization (with the creation of new surface states causing a flat-band shift being secondary [Ref. 2]), a measure of the

energy deposited in the insulator layer of a MOSFET is a direct measure of its ionization damage. In this paper, approximate calculations are performed and reported on the relative weights of damages produced by electrons of different energies with particular reference to the Jupiter radiation belt. Once this is accomplished, a model electron environment, in the form of a fluence-energy distribution, is integrated to yield the overall dose. Considerations discussed in this paper provide a means to obtain direct damage assessments from laboratory measurements corresponding to one particular energy.

A similar situation is found in assessing the total dose that might be received by an astronaut. The difference is that here we are concerned with the insulator material instead of human tissue. With respect to the Jupiter mission, we have chosen to deal with electrons because the total integrated fluxes of neutrons and the gamma rays from inside the spacecraft (due to the radioisotope thermoelectric generators) are of secondary importance (Ref. 3 and footnote 1). The proton damage is more

easily assessed (but is not necessarily smaller) and is deferred because there are presently great uncertainties of its prediction around Jupiter.<sup>1</sup>

## 2. Method of Calculation

Calculations are made on the following assumptions:

- (1) A uniform monoenergetic electron stream of "large" extent and kinetic energy  $E_0$  is incident normally on the spacecraft magnesium wall of thickness  $t$ , from which electrons of mean kinetic energy  $E'_0$  and photons of certain energy distribution emerge in the same direction and hit normally a composite Al-Si plate of thickness  $x_1$  in Al and  $x_2$  in silicon. In the treatment of a beam of infinite extent, the neglect of angular scattering underestimates the effect of shielding. Therefore, this assumption will lead to a conservative estimate of the dose.
- (2) Silicon is used to simulate the silicon dioxide ( $\text{SiO}_2$ ) layer of the MOSFET. Since their densities are almost equal (2.33 and 2.27, respectively) and  $\text{SiO}_2$  is predominately silicon,  $\text{SiO}_2$  is assumed to have the same electron stopping powers and photon cross-sections as silicon, for which data are readily available.

On the basis of the above, calculations of the dose per unit incident electron fluence can be broken down in the following sequence of steps, with some appropriate assumptions added:

*a. Step 1.* Consider one electron of a uniform beam of kinetic energy  $E_0$  passing a magnesium plate of thickness  $t$ . The mean energy  $E'_0$  of the emerging electron is found by an interpolation procedure from range-energy data for Al (Ref. 4) and is taken as that of an emerging mono-energetic beam. Range straggling is taken into account by assigning a probability of transmission for the incident electron of energy  $E_0$  through the magnesium. Thus, an incident electron fluence of energy  $E_0$  is assumed to degrade to a fluence with energy  $E'_0$  and a smaller value, decreased by the transmission factor. Values of this factor, as given by Perkins (Ref. 5), are unity for all except very small values of  $E_0$ .

*b. Step 2.* After the effect on the electrons due to the aluminum metallization layer is neglected compared to the magnesium layer, the dose absorbed by the silicon due to the transmitted electrons is calculated. This calculation is standard; the collision stopping power for Al

(Ref. 4) was employed. The absorption of energy in the silicon due to bremsstrahlung occurring in the silicon is neglected.

*c. Step 3.* Next, differential cross-sections are calculated for bremsstrahlung, which occurs as the result of the electrons losing energy by radiation in penetrating the magnesium wall. The formulas used are labelled 3BN and 3BS(e) (originally due to Schiff in Ref. 6) in the review paper of Koch and Motz (Ref. 7), for initial electron energies of 1 MeV and higher than 1 MeV, respectively. The significant assumptions involved are the Born approximation, which is valid for

$$\frac{2\pi Z}{137\beta} \ll 1$$

and the approximation in electron angle integration (not accurate for  $\phi_0 \leq Z^{1/2}/111E_0$ ). Due to the relatively low value of  $Z$  (12) and the relativistic electron energies we are dealing with, the above assumptions are justified. In addition, photon-electron showers in the magnesium layer have been neglected because its thickness is much smaller than one radiation length ( $> 10$  cm for Mg) (Ref. 8).

The quantity of interest at this stage of the calculation is the number of photons  $N(k)dk$  in the energy interval  $k$  and  $k + dk$ , given by

$$N(k)dk = \frac{d\sigma}{dk} nt dk \quad (1)$$

where

$\frac{d\sigma}{dk}$  = bremsstrahlung cross-section, differential with respect to photon energy  $k$ , in units of square centimeters per atom per incident electron per MeV

$n$  = number of atoms/cm<sup>3</sup> of material

$t$  = thickness of material, cm

Since  $d\sigma/dk$  is a function of the electron energy, the cross-section will, therefore, depend on the depth in the magnesium wall, at which the bremsstrahlung is produced. Since the electron enters with kinetic energy  $E_0$  and leaves with a mean energy  $E'_0$ , an approximate  $d\sigma/dk$  can be evaluated corresponding to electron kinetic energy

$$\frac{E_0 + E'_0}{2}$$

<sup>1</sup>Private communication between N. Divine and J. Barengoltz.



which is applied only for that part of the magnesium in which the electron kinetic energy has not degraded below 0.1 MeV. This adjustment will be unnecessary for  $E_0 > 30$  MeV since  $E'_0 \approx E_0$ , while for  $E_0 < 30$  MeV, it provides a correction.

One remark about the attenuation of the photons in the magnesium material is in order. The energy absorption coefficients  $\mu$  (Ref. 9) are high *only* for low energy photons. For photon energies  $k > 0.1$  MeV and assuming even a thickness of 500 mils (magnesium) the transmission factor  $e^{-\mu t}$  is never smaller than 0.99. For  $k < 0.03$  MeV, however,  $\mu$  increases rapidly and total attenuation is approached (e.g., for  $k = 0.01$  MeV,  $e^{-\mu t} \approx e^{-40}$ ). Therefore, the lower end of the photon spectrum can be neglected in evaluating the effect of the photons on the silicon dioxide layer although the energy absorption coefficients are high for low photon energies. This point is in fact utilized by the establishment of a lower limit on the numerical integration over  $k$ .

**d. Step 4.** The bremsstrahlung dose on the silicon layer can be calculated after the photon distribution  $N(k)$  is obtained in Step 3. Since photons are absorbed in both the Al and Si layers, it is obvious that the energy absorbed in the Si layer due to  $N$  photons of energy  $k$  MeV is given by

$$E_{\text{brem}}(k) = k N(k) \exp(-\mu_{\text{Al}} x_1) (1 - \exp[-\mu_{\text{Si}} x_2]) \\ \approx k N(k) (1 - \exp[-\mu_{\text{Si}} x_2]). \quad (2)$$

The factor  $\exp(-\mu_{\text{Al}} x_1)$  is taken as unity because  $x_1$  is sufficiently small (approximately 6000 Å) to make this a good approximation for all but the very low energy photons, which have already been neglected in Step 3.

The approximate form of Eq. (2) is multiplied by a suitable factor to convert to dose and the integration over  $k$  is carried out by the trapezoidal approximation.

**e. Step 5.** As a result of Steps 2 and 4, the total dose absorbed by the silicon due to an incident unit electron fluence of energy  $E_0$  on the magnesium can be obtained by an addition of the direct electron and the bremsstrahlung contributions. In this paper, these contributions are not combined, so that their relative effects can be discussed. At this point, instead, an electron model for Jupiter's trapped radiation belts is introduced. The design

levels for electron fluence due to Jupiter are based on the spectrum<sup>2</sup>

$$\frac{d\phi}{dE} = 1.5 \times 10^9 E e^{-E/20.1}, \text{ electrons/cm}^2\text{-s-MeV} \quad (3)$$

A proper treatment of this isotropic flux in terms of the parallel beam assumption would result in a reduction by possibly as much as a factor of  $2\pi$ . The neglect of this factor adds another measure of conservatism to the estimate. The total dose received by a MOS device in the design level electron environment can now be calculated.

**f. Numerical results.** The results of calculations are presented for magnesium thicknesses of  $t = 50, 100, 200$ , and 500 mils, an aluminum thickness  $x_1 = 6000$  Å, and silicon thickness  $x_2 = 2000$  Å. Initial electron energies of 1, 3, 10, 30, and 100 MeV are used.

**g. Electron collision dose.** Table 1 summarizes the values of initial electron energy  $E_0$ , the reduced energy

<sup>2</sup>Divine, T. N., *The Planet Jupiter (1970)*, NASA Space Vehicle Design Criteria (Environment) Monograph (to be published).

**Table 1. Values of  $E_0$ ,  $E'_0(t, T)$ , and  $E_{\text{coll}}$**

$E_0$ , MeV	$E'_0(t, T)$ , MeV	$E_{\text{coll}}$ , MeV $\times 10^{-5}$
1	0.531 ( $t = 50$ mils, $T = 0.65$ )	4.94
	0 ( $t = 100$ mils, $T = 0$ )	0
	0 ( $t = 200$ mils, $T = 0$ )	0
	0 ( $t = 500$ mils, $T = 0$ )	0
3	2.41 ( $t = 50$ mils, $T = 1$ )	7.18
	2.04 ( $t = 100$ mils, $T = 0.96$ )	6.84
	1.12 ( $t = 200$ mils, $T = 0.57$ )	4.03
	0 ( $t = 500$ mils, $T = 0$ )	0
10	9.40 ( $t = 50$ mils, $T = 1$ )	7.88
	8.82 ( $t = 100$ mils, $T = 1$ )	7.82
	7.68 ( $t = 200$ mils, $T = 1$ )	7.77
	4.45 ( $t = 500$ mils, $T = 1$ )	7.48
30	29.14 ( $t = 50$ mils, $T = 1$ )	8.43
	28.28 ( $t = 100$ mils, $T = 1$ )	8.41
	23.18 ( $t = 200$ mils, $T = 1$ )	8.33
	22.05 ( $t = 500$ mils, $T = 1$ )	8.31
100	98.22 ( $t = 50$ mils, $T = 1$ )	8.93
	96.44 ( $t = 100$ mils, $T = 1$ )	8.92
	94.66 ( $t = 200$ mils, $T = 1$ )	8.91
	83.56 ( $t = 500$ mils, $T = 1$ )	8.87

$E'_0$  of the electrons after passing through a magnesium layer of thickness  $t$ , the corresponding transmission factor  $T$ , and  $E_{coll}$ , the energy per electron deposited in the silicon (2000 Å) by its collisions.

The energy deposited in a slab of silicon, as given in Table 1, is presented graphically in Fig. 1 as a function of magnesium thickness  $t$  and incident electron energy  $E_0$ . Also note that the intercepts on the vertical axis in Fig. 1 suggest that for zero shielding, electron-induced ionization is fairly unchanged over a wide incident energy range. This conclusion is borne out by experimental data on several ionization sensitive MOSFETs, where the induced degradation for a given fluence of electrons was insensitive to their energies in the range 4 to 20 MeV (Ref. 1).

The results of the direct dose calculation are presented in Fig. 2 where, for simplicity, only two curves are drawn corresponding to  $t = 50$  mils (A) and  $t = 500$  mils (B). The curves represent a folding of the electron spectrum of Eq. (3) and the dose-thickness dependence of Fig. 1, for a given thickness. Integration of these curves yields a dose of  $1.8 \times 10^4$  rad for curve A and  $1.7 \times 10^4$  rad for curve B.

**h. Bremsstrahlung dose.** Table 2 summarizes the values of initial electron energy  $E_0$ , the thickness  $t$  of the magnesium layer, and the resultant dose per unit incident electron fluence. For a given  $E_0$ , entries for  $t$  are presented only where the dose is non-negligible.

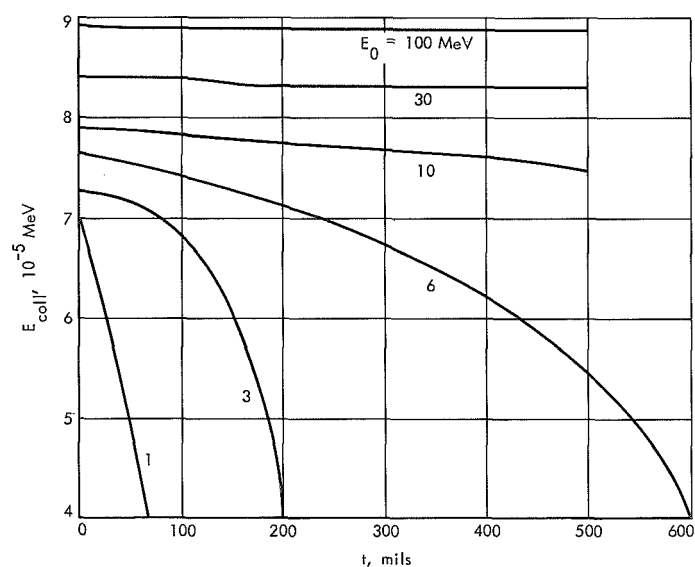


Fig. 1. Energy deposited by electrons in Si after penetrating Mg thickness  $t$

Table 2. Values of initial electron energy, thickness of layer, and dose

$E_0$ , MeV	$t$ , mils	Dose, rad/unit fluence
1	50	$1.2 \times 10^{-11}$
3	50	$1.3 \times 10^{-11}$
	100	$2.6 \times 10^{-11}$
	200	$7.4 \times 10^{-12}$
10	50	$2.7 \times 10^{-11}$
	100	$5.4 \times 10^{-11}$
	200	$7.3 \times 10^{-11}$
	500	$1.6 \times 10^{-10}$
30	50	$6.9 \times 10^{-11}$
	100	$1.4 \times 10^{-10}$
	200	$2.4 \times 10^{-10}$
	500	$6.0 \times 10^{-10}$
100	50	$2.5 \times 10^{-10}$
	100	$5.0 \times 10^{-10}$
	200	$9.7 \times 10^{-10}$
	500	$2.4 \times 10^{-9}$

The results of Table 2 show that the bremsstrahlung dose increases with increasing electron kinetic energy  $E_0$ . They also show that the dose increases linearly with the thickness  $t$  of the magnesium shield, as expected

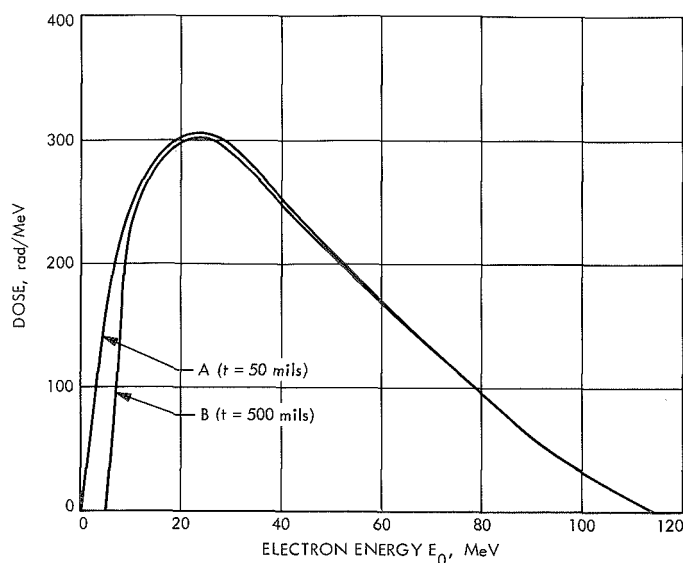


Fig. 2. Differential dose for direct electron collisions, using the spectrum of Eq. (3): curve A is for a 50-mil Mg shield; curve B is for a 500-mil Mg shield

from Eq. (1), until the shield becomes thick enough to significantly attenuate the photons created within it.

As shown in Fig. 3, the differential bremsstrahlung dose has a rather complex form, which is a consequence of folding the electron spectrum (Eq. 3) and the dose-thickness dependence presented in Table 2. For clarity, only the  $t = 50$  mils (A) and  $t = 500$  mils (B) curves are included. Integration of these curves yields 50 rad and 500 rad, respectively.

### 3. Conclusions

For currently envisioned outer planet missions, Jupiter's electron-induced ionization dose in MOS structures be-

hind a magnesium wall at least 50 mils thick does not exceed  $2 \times 10^4$  rad. This dose is near the threshold for degradation of the most ionization-sensitive MOS devices.

The bremsstrahlung dose is a small part of the total for magnesium walls from 50 to 500 mils thick. Even with a proper treatment of the bremsstrahlung produced in the aluminum layer of a MOS device, estimates show that it is less than 3% of the total dose.

In laboratory testing of MOS devices, where the spacecraft wall shield is absent, 3-MeV electrons are sufficient to simulate the effect of higher energy electrons. In such tests, the weak energy dependence of electron-induced ionization for energies in the range of 3 to 100 MeV may be accurately estimated.

### References

1. Palmer, E. E., Miles, J. K., and Lambert, D. L., *Effects of High-Energy Electrons on Candidate Jupiter Probe Electronic Components*, FZM-4910, General Dynamics, 1967.
2. Andre, B., Buxo, J., Esteve, D., and Martinot, H., "Effects of Ionizing Radiation on MOS Devices," *Solid State Electronics*, Vol. 12, pp. 123-131, 1969.
3. Nerman, P. A., Jr., "Low Flux Level Neutron Effects on MOSFET's," *Proc. IEEE Conf. NSRE 16*, 1970.
4. Berger, M. J., and Seltzer, S. M., "Additional Stopping Power and Range Tables for Protons, Mesons, and Electrons," NASA SP-3536, National Aeronautics and Space Administration, Washington, D.C., 1966.
5. Perkins, J. F., "Monte Carlo Calculation of Transport of Fast Electrons," *Phys. Rev.*, Vol. 126, No. 5, p. 1781, 1962.
6. Schiff, L. I., "Energy-Angle Distribution of Thin Target Bremsstrahlung," *Phys. Rev.*, Vol. 83, No. 2, p. 252, 1951.
7. Koch, H. W., and Motz, J. W., "Bremsstrahlung Cross-Section Formulas and Related Data," *Rev. Mod. Phys.*, Vol. 31, No. 4, p. 920, 1959.
8. *The Encyclopedia of Physics*, Reinhold Publishing Corp., p. 85, 1966.
9. Hubbell, L. H., "Photon Cross-Sections, Attenuation Coefficients, and Energy Absorption Coefficients From 10 keV to 100 GeV," NSRDS-NBS 29, National Bureau of Standards, 1969.

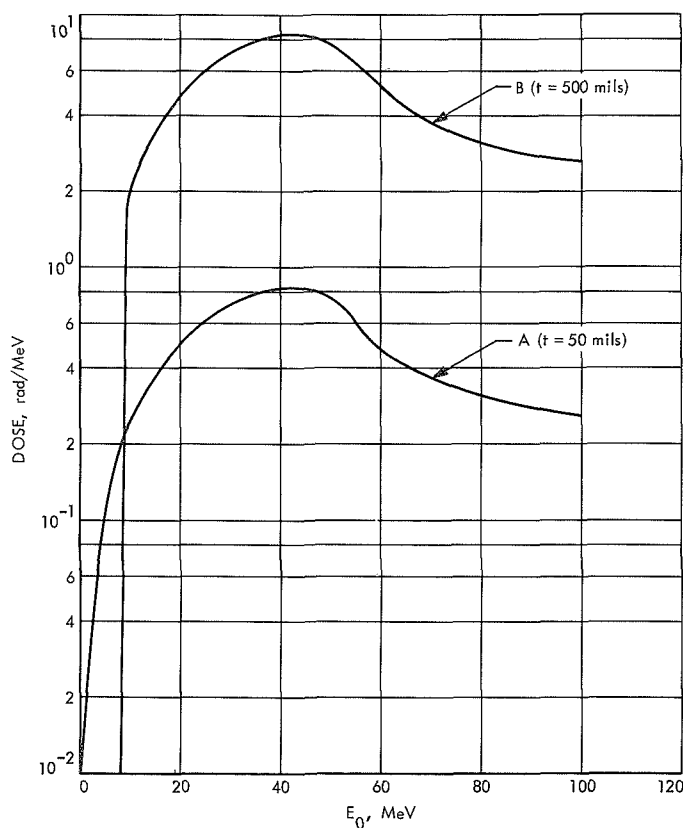


Fig. 3. Differential bremsstrahlung dose, using the spectrum of Eq. (3): curve A is for a 50-mil Mg shield; curve B is for a 500-mil Mg shield

## XVI. Space Simulation

### ENVIRONMENTAL SCIENCES DIVISION

#### A. Holographic Study of Operating Compact-Arc Lamp, C. G. Miller and C. L. Youngberg

##### 1. Introduction

Solar heatloads on spacecraft are simulated by using the radiation from xenon compact-arc lamps and causing the radiation to impinge in a nearly parallel beam on the structure under study. In order to get enough intensity in a nearly parallel beam whose divergence approaches that of sunlight at earth's orbit (32 sec of arc), the xenon compact-arc lamp must be operated at the material and structural physical limits for highest luminous output and brightness. In order to design to the material limits, we are interested in efficient heat transfer by fluids. The fluids involved are the pressurized xenon gas which transfers heat convectively, and the water flow which cools the interior of the cathode and anode of the lamp. Heat fluxes at the anode of an operating lamp are about  $2 \times 10^7$  Btu hr<sup>-1</sup>F<sup>-1</sup>ft<sup>-2</sup> and to design for such high rates, we need to know the mode of action of gaseous convective cooling, and of liquid heat transfer inside the water-cooled elements of the lamp.

Double exposure holographic techniques are used to determine the conditions in the gas stream between the cathode and the anode of an operating lamp.

##### 2. Arc Characteristics of Concern

The molecular density in the arc discharge column is the quantity of great importance in the understanding and utilization of compact-arc lamps. While the gas pres-

sure in the lamp is 12 atm when the arc is burning, having started from 4 atm cold, the arc core itself is at a temperature of 6000°K, and thus the particle density is only  $1.7 \times 10^{19}$  molecules cm<sup>-3</sup> at 12 atm, virtually the same as for a gas at 1 atm and room temperature.

The appearance of an operating arc (Fig. 1) shows the well-developed center core of highest radiance, and the surrounding arc sheath, which is also luminous, and whose outer edges are sharply defined in the pressure regime above atmospheric.

It is reasonable to consider the luminous core as the locus of high temperature, and thus low density partly ionized gas. Because of the great mobility of electrons compared to that of positive ions, ion pairs formed in the center core should lose their electrons faster than their positive ions, and thus the core has a net positive makeup, while the surrounding arc sheath should have a net negative charge.

A separate experiment, done for another purpose and previously reported (SPS 37-63, Vol. III, pp. 161-166), used a lamp with magnets external to the body as shown in Fig. 2. The magnets were powered by ac and caused the arc flame to move back and forth in the  $\mathbf{j} \times \mathbf{B}$  direction.

High-speed motion pictures of the back-and-forth motion of the arc brush and core caused by the arc currents interacting with the time-varying magnetic field of the



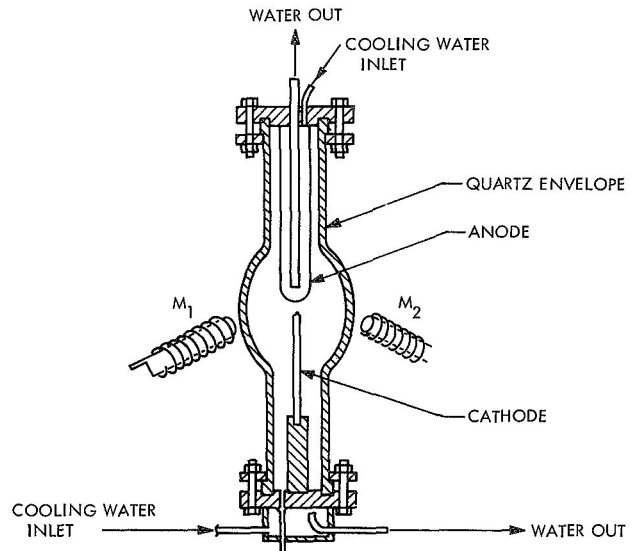
**Fig. 1. Representative operating arc**

magnets (shown in Fig. 2) showed that the brush and core were deflected in opposite directions upon application of varying magnetic fields, the core responding as though there were a resistive viscous force. This is consonant with the view that the cathode core is a predominantly positive ion stream, traveling from cathode to anode against the electric field by momentum transfer from the electron stream, and that the brush is predominantly negatively charged. An interferometric method to determine density profiles across the discharge gap would be valuable to support the physical picture developed here.

Another important aspect of the arc discharge concerns the hot gas flow about the anode. For highest heat transfer rates, it is desirable to develop turbulent flow and an interferometric method of examination should also be useful here.

### 3. Holographic Interferometry to Measure Aerodynamic Quantities

A means to examine the problem mentioned in the preceding subsection is at hand in the method of flash



**Fig. 2. Arrangement of magnets to deflect arc**

holographic interferometry. In this method (Ref. 1), the interferometer uses a common optical path in which beam separation is achieved by two exposures separated in time. This method allows a great relaxation in the usual interferometry requirements for accurate alignments and precision optical elements, and gives a complete three-dimensional record of the interference phenomenon, permitting post-exposure focusing and examination from various directions.

Such a system was used to examine an operating arc in this laboratory. The experimental details have been described elsewhere (Ref. 2). We will not repeat the experimental details here but will discuss the limitations on the use of information derived from flash holograms and the results of the experiments done in regard to the physical picture developed in *Subsection 2*.

Reference 3 discusses some sensitivity factors in plasma diagnostics, which are summarized below.

- (1) Optical quality of the lamp and windows may be low and still result in interferograms. Thus, actual lamp bodies may be employed.
- (2) Relative motion of parts of the objects of interest can completely obscure the results. Thus, the small relative motion caused by turbulence of the cooling water inside the cathode and anode of the lamp completely obscure the interferogram. It was

necessary to use a special lamp termed Variarc in which the electrodes are mounted very securely to minimize this difficulty.

- (3) The holographic method is insensitive to intrinsic plasma radiation. Since the plasma radiation is incoherent, it does not contribute to the formation of the hologram, but only leads to some illumination of the hologram, and degradation of contrast. In our case, the use of narrow band interference filters centered on the laser wavelength retained the contrast by suppressing most of the plasma illumination from the photographic film.
- (4) The holographic interferometry method has the same sensitivity as normal interference measurements. It is difficult to detect path differences smaller than  $\lambda/10$  or greater than, say,  $100 \lambda$ .
- (5) In the general case, far from absorption lines, the refraction of the plasma is given by

$$n - 1 = \left( A + \frac{B}{\lambda^2} \right) N_a - \left( \frac{e^2}{2 m \epsilon_0} \frac{1}{\omega_0^2 - \omega^2} \right) N_e$$

where  $N_a$  denotes the density of atoms,  $N_e$  denotes the density of electrons, and  $n$  is the index of refraction.

Using Cauchy's formula values of  $A$  and  $B$  for noble gases, we find that the gas refraction is comparable in magnitude to the refraction of an electron gas at a particle density

$$N_a/N_e \sim 10$$

Now changes in the optical path length of the order of  $\lambda/10$  (in a layer thickness of 1 cm and with  $\lambda = 0.5 \mu$ ) correspond to electron densities of  $5 \times 10^{16} \text{ cm}^{-3}$ , and atom densities of  $5 \times 10^{17} \text{ cm}^{-3}$ .

We should thus be able to distinguish density variations of 3% (from the  $1.7 \times 10^{19} \text{ atoms cm}^{-3}$  found in Subsection 2) at the  $\lambda/10$  criterion, or density variations of 30% at a  $1\lambda$  criterion.

To show that this should not be obscured by a decrease in refraction due to electrons, we note that from Saha's equation:

$$\frac{px^2}{1-x^2} = 3.16 (10^{-7}) T^2 \sqrt{T} \exp \left( \frac{-eV_i}{kT} \right)$$

which gives the degree of ionization  $x$  prevailing in a gas in thermal equilibrium at  $T^\circ\text{K}$ , and where  $p$  is the pressure in atmospheres, and  $V_i$  is the ionization potential of the gas. The electron density is  $(x)(1.7 \times 10^{19})$  or  $1.7 \times 10^{16} \text{ electrons cm}^{-3}$ , or only one third the number needed for the  $\lambda/10$  criterion.

#### 4. Interpretations of Holograms

A double flash hologram made of an operating Variarc (Fig. 3) shows no evidence of gas density variation in the vicinity of the center core, and no evidence of a highly localized region of very low density gas. Furthermore, an examination of the anode area appears to show no turbulence developed at the arc stream-anode interface. Such turbulence would be desirable to increase heat transfer rates.

Figure 4 shows a double exposure of an operating arc with an interval of about 2 min between exposures. The

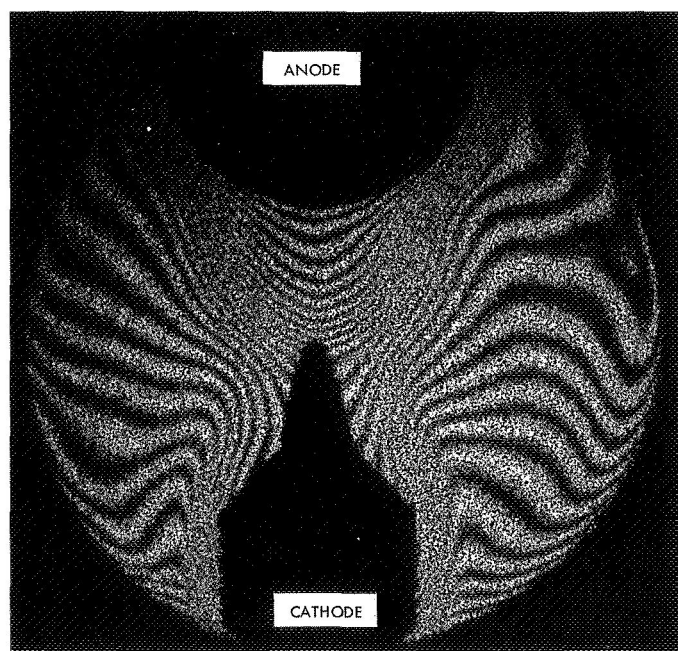


Fig. 3. Hologram of operating arc



purpose was to investigate the random variations in density within a continually operating lamp. The fringe pattern is nondescript, implying no long-term stability in the gas flow pattern.



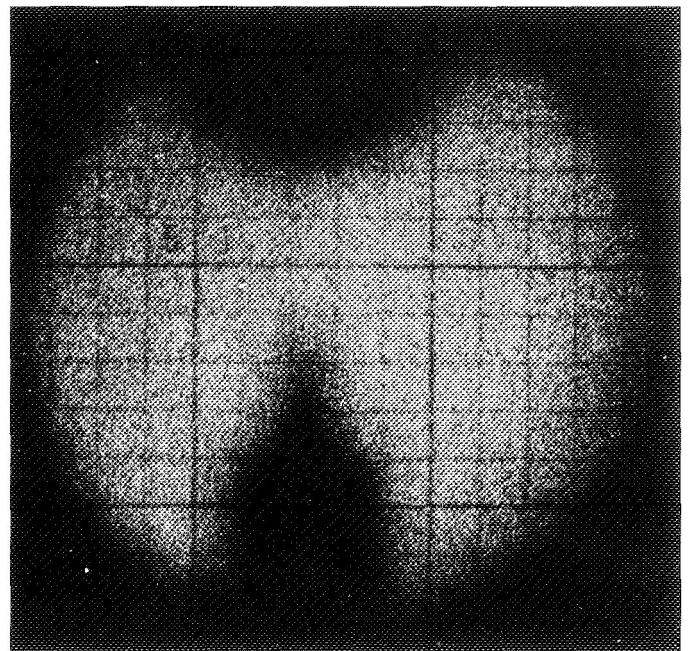
**Fig. 4. Double flash hologram with long interval between exposures**



**Fig. 5. Asymmetric pattern of arc with imposed magnetic field**

Figure 5 is a hologram reconstruction of an interferogram in which a magnetic field was introduced to deflect the arc. The interference fringes show a nonsymmetrical pattern in agreement with the visual observation that the magnetic field did deflect the arc.

Finally, Fig. 6 shows a hologram reconstruction of an interferogram of an operating lamp with a transparent grid placed on the holocamera diffuser, considerably behind the plane of the arc discharge. In the photograph, the camera is focused on this grid, and is looking through the arc. The electrodes are out of focus. The intent of this hologram was to determine if there was significant refraction by the arc. The lack of distortion of the grid lines indicate that there was not.



**Fig. 6. Hologram focused on grid pattern behind lamp showing no significant refraction by arc stream**

#### References

1. Heflinger, L. O., Wuerker, R. F., and Brooks, R. E., "Holographic Interferometry," *J. Appl. Phys.*, Vol. 37, pp. 642-649, 1966.
2. Miller, C. G., and Stephens, J. B., "Flash Holography Applied to Fluid Heat-Transfer Phenomena," in *Holographic Instrumentation Applications*, NASA SP-248, National Aeronautics and Space Administration, Washington, 1970.
3. Zaidel, A. N., Ostrovskaya, G. V., and Ostrovskii, Yu. I., "Plasma Diagnostics by Holography (Review)," *Sov. Phys.-Tech. Phys.*, Vol. 13, pp. 1153-1163, 1969.

## XVII. Solid Propellant Engineering

### PROPULSION DIVISION

#### **A. Nondestructive Testing of 1-W, 1-A Electro-Explosive Devices, V. J. Menichelli**

##### **1. Introduction**

A program is underway to study the response of various electro-explosive devices to the "transient pulse" and "thermal follow display" nondestructive tests. A description of the circuitry and the theory associated with these tests were reported in SPS 37-62, Vol. III, pp. 228-233. This article describes the isolation and dissection of a squib which displayed an abnormal response to the two nondestructive tests.

##### **2. Test Results**

During the tests, the dual bridgewire squib identified as #42901 displayed highly irregular and unstable heating curves for both bridgewire circuits. Due to the instability of response at the time of testing, no statistical data were taken. Figure 1 shows heating curves obtained by the transient pulse test. Included is a heating curve obtained on a squib from the same lot (#42906) which gave what is considered a normal response. Figure 2

shows heating curves obtained by the thermal follow display test and also a typically normal response expected from a healthy squib.

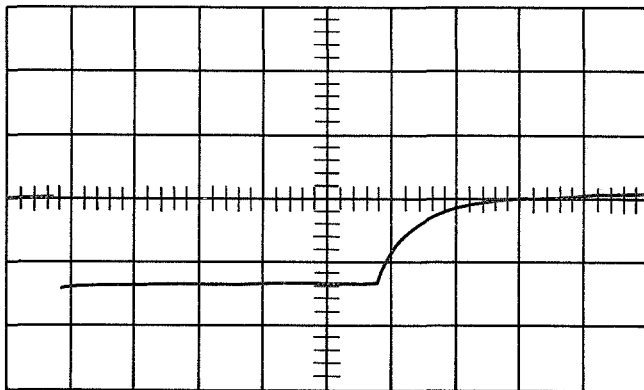
Examination of the transient pulse heating curves shows some sharp rapid changes early in the heating and later some slower more gradual changes. The same conditions are reflected in the thermal follow display curves. The early distortions in the heating curves are believed due to ohmic heating resulting from poor welds or bridgewire defects. The later more gradual changes in the heating curve are attributed to poor contact between the bridgewire and explosive material. A poor bridgewire weld results in high contact resistance, which is an important factor in generating the heating curve. A lack of intimate contact between the bridgewire and explosive material can cause non-uniform heat transfer resulting from wire expansion as the temperature rises.

##### **3. Investigation**

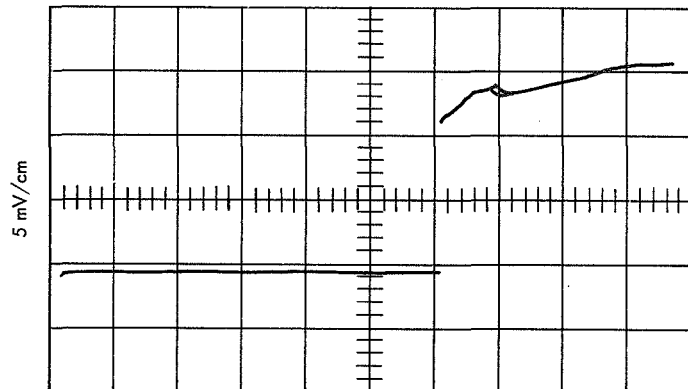
The squib was dissected to attempt to correlate the abnormal response which occurred during testing with



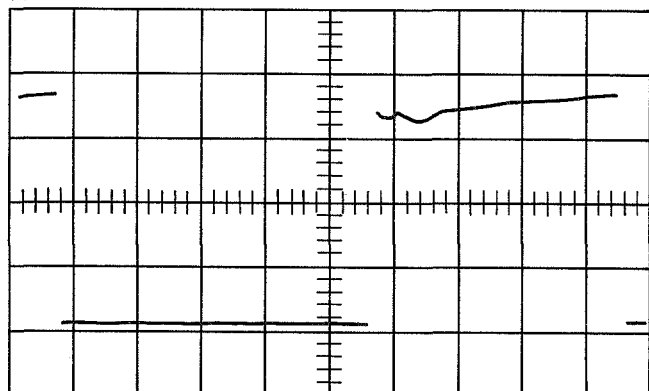
(a) #42906AB



(b) #42901AB

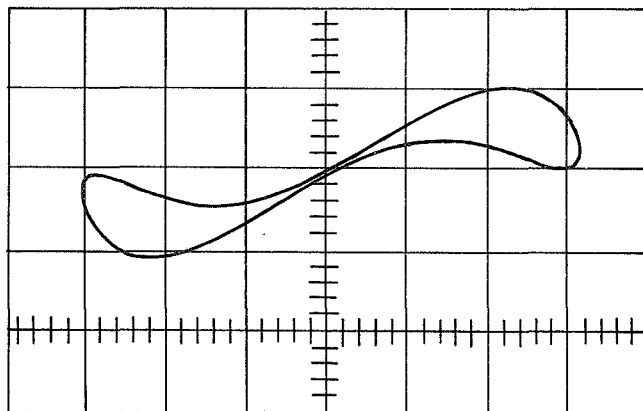


(c) #42901CD

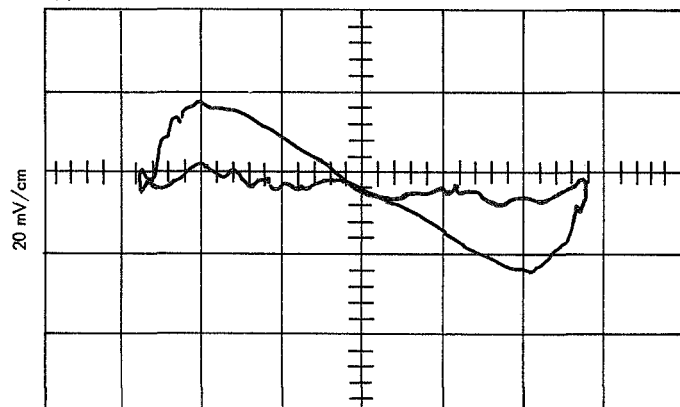


**Fig. 1. Heating curves generated by the transient pulse technique**

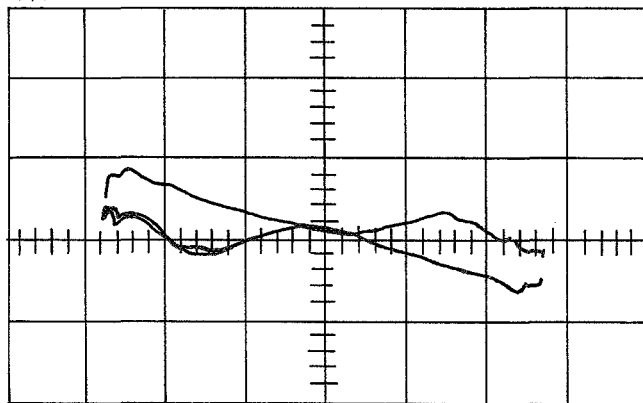
(a) TYPICAL NORMAL RESPONSE



(b) #42901AB



(c) #42901CD



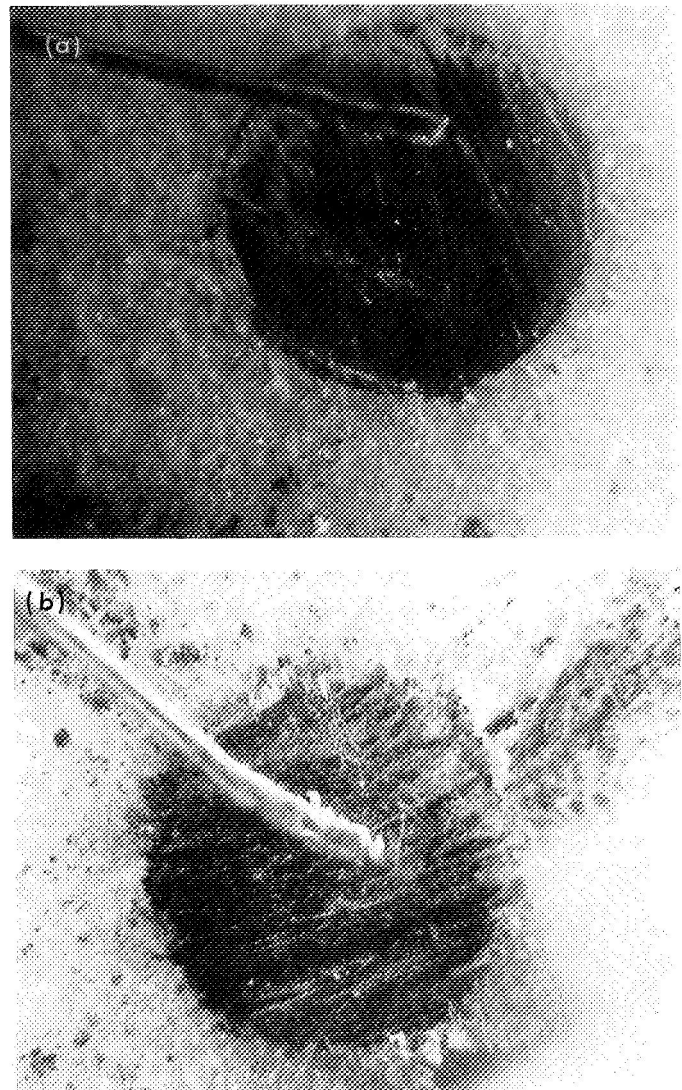
**Fig. 2. Heating curves generated by the thermal follow display technique**

some internal defect. Dissection began by carefully machining away the end seal from the squib and removing the base charge. Special care was taken not to disturb the matchhead material loaded on the bridgewire. It was observed that the density of the matchhead material was lower than good ordnance practice would dictate. The squib was then subjected to the transient pulse test, with the exception that the driving voltage was increased so as to heat the wire above the normal operating test temperature but not high enough to cause initiation. The pulsing at this level would stress the bridgewire beyond what it sees in a normal test. Also, the pulsing would continue for a much longer period of time. The purpose was to accelerate any defects to the point of failure. One would not normally use the transient pulse test in this manner when making a nondestructive measurement. However, for this application the equipment became an appropriate tool to study suspected faults. After approximately 15 min of pulsing, bridgewire AB opened without causing ignition of the matchhead. The same procedure was used on bridgewire CD, and after about 10 min, it also opened without causing the charge to initiate.

The next step in the investigation was the removal of the matchhead material from the bridgewire. This was carefully done by soaking out the material with acetone and cleaning with a camel hair brush. The material washed out quite easily, indicative of the low compaction density. The bridgewires and posts were examined microscopically. Examination of the AB bridgewire circuit showed that one end of the wire had been barely attached to the post (Fig. 3a). The weld was obviously very poor and the pulsing action of the transient pulse test caused it to fail. The other end of the circuit (Fig. 3b) showed the wire still welded to the post, but the weld is considered substandard. Bridgewire circuit CD showed a different failure mode. The bridgewire appeared to be welded securely at both ends (Figs. 4a and 4b). However, there appeared to be a necking down of the bridgewire at the inside weld which can be attributable to a wire defect or possibly corrosion. Figure 4c appears to show the open circuit resulting from a mechanical break. The break took place at the point where maximum stress was taking place due to the pulsing action.

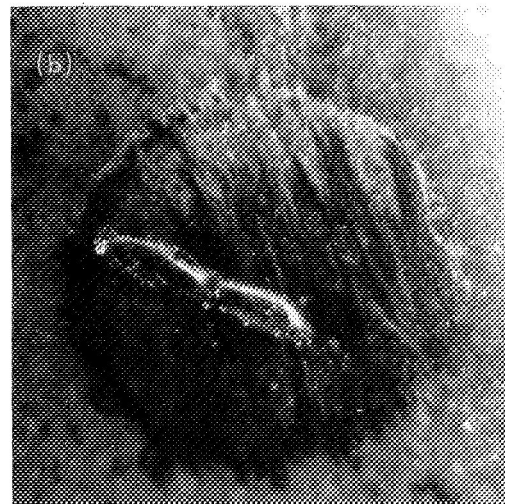
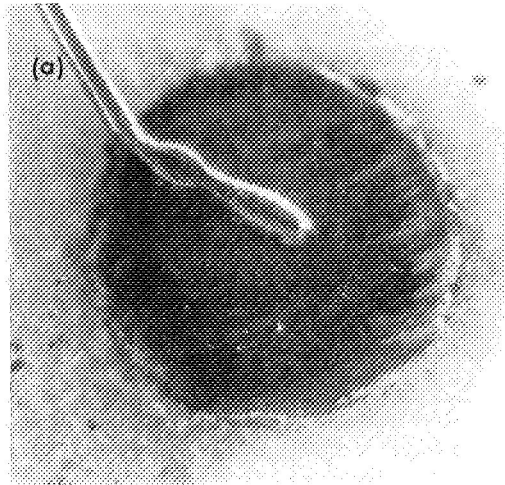
#### 4. Conclusions

The results of this investigation, although not conclusive, certainly demonstrate that the observed abnor-



**Fig. 3. Photo micrographs of AB bridgewire and posts for squib #42901**

mal heating curve effect can be related to a cause. The transient pulse and thermal follow display tests did indeed cull out an undesirable squib. It can be argued that this squib failed because of the unusual conditions under which it was tested. However, this and similar methods are valid in determining shelf life or ruggedness of a specific design. It is probable that the same or similar conditions could be encountered by temperature cycling from environments on earth or in space. Visual inspection of the welds showed that they were far from being adequate and could have failed from other causes such as vibration and shock. One would not want to use such a squib in a high reliability application.



**Fig. 4. Photo micrographs of CD bridgewire and posts for squib #42901**

## B. Low-Pressure $L^*$ -Combustion Limits, L. D. Strand

### 1. Introduction

As part of a development effort on long-burning, high-performance motors (SPS 37-63, Vol. III, pp. 188-196), several propellant variables have been investigated in an attempt to develop propellants with lower burning rates. These variables have included propellant binder, oxidizer coarseness, and the addition of a coolant additive. Since the motors being considered have rather severe conditions at propellant ignition with regard to propellant combustion stability, i.e., low  $L^*$  (motor free volume to nozzle throat area ratio) and chamber pressure, tests were performed to determine how these propellant variables affected the low-pressure/ $L^*$  extinction limit. The tests were performed in the  $7.62 \times 10^{-2}$  m (3 in.) ID test motor in the same manner as described in SPS 37-41, Vol. IV, pp. 85-91, and in Ref. 1.

### 2. Test Results

Three propellants were examined, each one containing 16% Al and ammonium perchlorate as the oxidizer: (1) Saturethane, with a saturated polybutadiene binder and the standard bi-modal oxidizer distribution (70% coarse—170- $\mu$ m mass median diam, 30% fine—17- $\mu$ m mass median diam); (2) JPL 540J, with the standard JPL 540 polyether-polyurethane binder, but a tri-modal oxidizer distribution (50%—400  $\mu$ m, 34%—200  $\mu$ m, 16%—50  $\mu$ m), and (3) EB-81, similar to JPL 540J, but containing 2½% of a burning rate suppressant, oxamide (one-half substituted for binder, one-half for oxidizer).

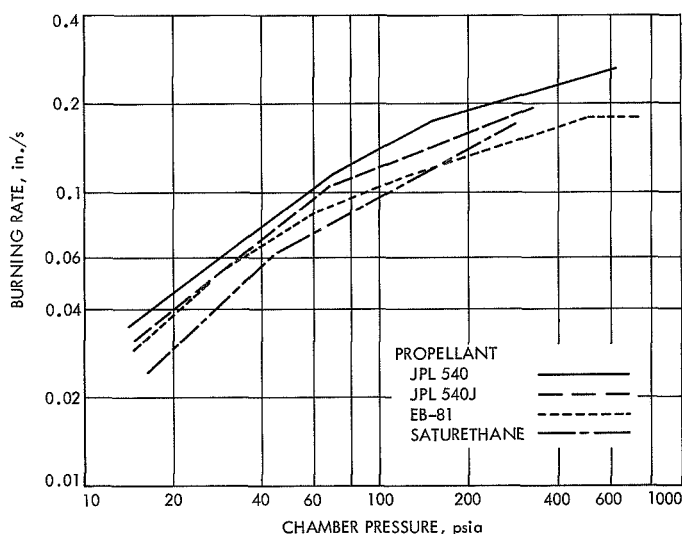


Fig. 1. Strand bomb burning rate vs pressure data

Strand bomb burning rate versus pressure data for the three propellant formulations are compared with the control JPL 540 propellant data (Fig. 1). In the pressure region of  $6.9 \times 10^5$ – $10.3 \times 10^5$  N/m<sup>2</sup> (100–150 psia), the propellants, in order of descending burning rates, are JPL 540, JPL 540J, EB-81, and Saturethane. The curves for the latter two propellants cross over at approximately  $10.3 \times 10^5$  N/m<sup>2</sup> (150 psia).

Figures 2–4 show the measured  $L^*$  versus mean chamber pressure at combustion extinction data for Saturethane, JPL 540J, and EB-81, respectively. The Saturethane data fell about the curve for JPL 540 (Fig. 2). Figure 3 shows that the tri-modal oxidizer in the JPL 540J propellant lowered the extinction pressure boundary approximately  $4.1 \times 10^4$  N/m<sup>2</sup> (6 psi) below that of JPL 540. The addition of the oxamide to the tri-modal oxidizer propellant had a rather strong effect, raising the extinction pressure boundary approximately  $6.9 \times 10^4$  N/m<sup>2</sup> (10 psi) so that it exceeded that of JPL 540 as well (Fig. 4). As can be seen from Fig. 4, the slopes of the curves for the different propellants are almost identical. This is in agreement with the conclusion of Ref. 1 that the slope of the  $L^*$ – $P_{c_{ex}}$  relationship is affected

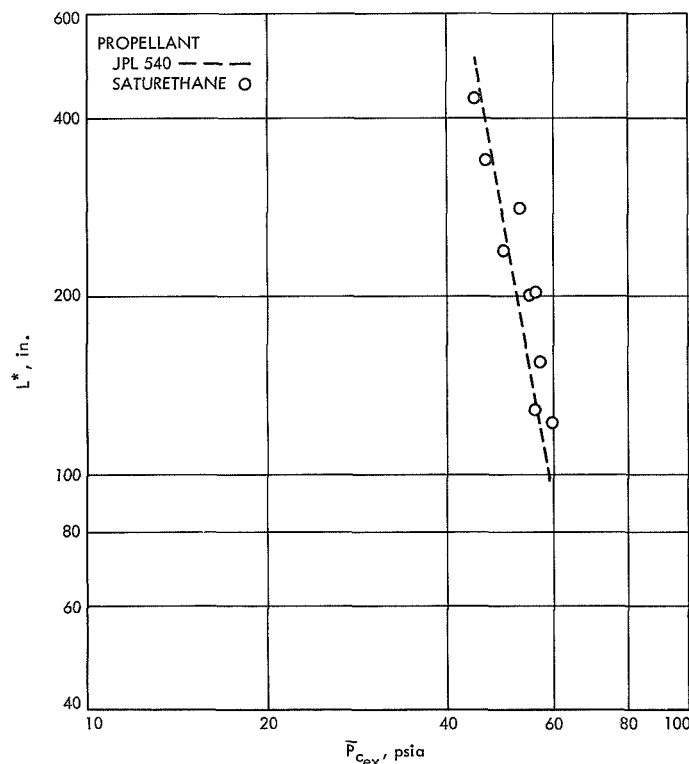


Fig. 2.  $L^*$  vs mean chamber pressure at combustion extinction  $\bar{P}_{c_{ex}}$  (saturethane propellant)

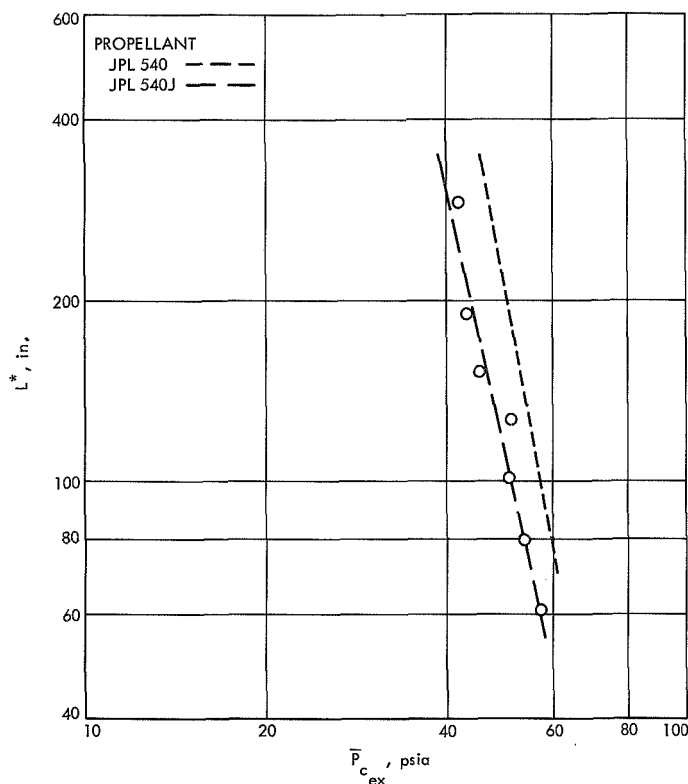


Fig. 3.  $L^*$  vs mean chamber pressure at combustion extinction  $\bar{P}_{c_{ex}}$  (JPL 540J propellant)

principally by the presence of aluminum, the concentration of aluminum, and the particle size of the aluminum in the propellant.

### 3. Discussion

In Ref. 1, a theoretical analysis, based on a highly simplified combustion model, was presented which predicted a relationship between the ease of extinguishment and propellant burning rate, namely, that the greater the low-pressure burning rate of a propellant, the lower were the motor  $L^*$ s and pressures required before combustion instability and extinction would occur. Listing the propellants tested in the order of ascending  $L^*$ -extinction pressure limit yields the following: JPL 540J, JPL 540 and Saturethane about equal, and EB-81. Referring to Fig. 1, it is obvious that the simple burning rate/stability relationship does not hold rigorously. Other factors must be considered, such as the nature of the binder decomposition, transparency of the propellant to thermal radiation, mechanisms of burning rate catalysts, and depressants, etc.

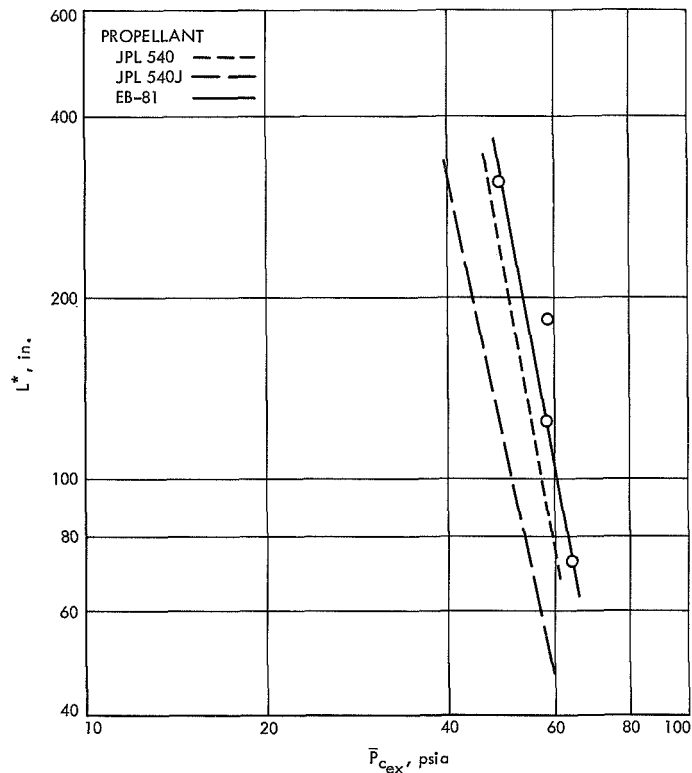


Fig. 4.  $L^*$  vs mean chamber pressure at combustion extinction  $\bar{P}_{c_{ex}}$  (EB-81 propellant)

Looking at the propellants individually: Differences in binder degradation mechanisms result in Saturethane and JPL 540 propellants having essentially identical  $L^*$ -extinction boundaries, even though Saturethane has a significantly lower burning rate. The greater transparency to thermal radiation of the slower burning, coarser oxidizer JPL 540J propellant may be the reason for its greater resistance to extinguishment, as postulated in Ref. 2. Finally, the addition of coolants, such as oxamide, which supposedly decompose endothermically on or near the propellant surface, significantly raises the  $L^*$ -extinction pressure limit.

### References

1. Strand, L. D., *Summary of a Study of the Low-Pressure Combustion of Solid Propellants*, Technical Report 32-1242. Jet Propulsion Laboratory, Pasadena, Calif., April 15, 1968.
2. Micheli, P. L., *A Stop-Start Study of Solid Propellants*, Final Report, Contract NAS 1-6600, NASA CR-66487. Aerojet-General Corporation, Propellant Research and Development Division, Sacramento, Calif.

## C. Diffraction of a High-Order Gaussian Beam by an Aperture, A. C. R. Livanos<sup>1</sup>

### 1. Introduction

In a number of important laser applications, e.g., laser explosive initiation (Refs. 1 and 2), holography, etc., it is desirable to obtain uniform laser intensity distributions. Because of the inhomogeneities of the laser rod, imperfections of alignment, and multiple reflections in the optical delivery system, the output beam usually contains higher spacial frequencies. By positioning an aperture of the proper diameter, one gets rid of the intensity wings. However, by using this technique, interference rings appear in the intensity distribution.

In this article, the diffraction of a laser beam of the most general gaussian mode by an aperture is described. The beam considered is the TEM<sub>pln</sub> mode<sup>2</sup> of a laser, which, after leaving the optical resonator, is focused by a lens and then diffracted by an aperture located at the focal point.

Huygen's principle is used to calculate the near- and far-field radiation patterns. The assumptions used to

simplify the diffraction integral permit a calculation for field distributions over the entire space to the right of the aperture, in contrast with the Fresnel and/or the Fraunhofer approximations.

In the work reported here, the ultimate intensity of the beam was determined as a function of the aperture size and the oscillating mode index. The results show that for all practical purposes the use of the lowest possible mode is desirable. In this case, the output looks like a simple gaussian when the aperture is large and like the Airy pattern when the aperture is small. For intermediate sizes of aperture, the intensity distribution is a mixture of the two limits.

### 2. Derivation of Intensity Distribution at Aperture

The resonator-lens-aperture configuration considered is sketched in Fig. 1. Given a resonator with mirrors of curvature  $R_1$ ,  $R_2$  and length  $L$ , one can find the unique gaussian which will just fit between these mirrors, assuming that the mirror aperture is larger than the spot sizes. The distribution at an arbitrary point  $P(r, \theta, z)$  inside the resonator is of the form

$$U_{pl}(r, \theta, z) = \frac{2}{(1 + \delta_{0\ell})^{1/2}} \left[ \frac{p!}{\pi(\ell + p)!} \right]^{1/2} \frac{1}{w(z)} \left[ \frac{(2r)^{1/2}}{w(z)} \right]^\ell L_p^\ell \left[ \frac{2r^2}{w^2(z)} \right] \left( \frac{\cos \ell \theta}{\sin \ell \theta} \right) \exp \left[ -j \frac{\kappa}{2} \frac{r^2}{\tilde{q}(z)} \right] \times \exp(-j\kappa z) \exp[j(2p + \ell + 1)\Psi(z)] \quad (1)$$

where

$$j = (-1)^{1/2}$$

$\kappa$  = wave vector

$\lambda$  = wavelength

$p, \ell$  = integer indices

$L_p^\ell$  = associated Laguerre polynomial

$$\delta_{0\ell} = \begin{cases} 1 & \text{for } \ell = 0 \\ 0 & \text{for } \ell \neq 0 \end{cases}$$

<sup>1</sup>The author would like to thank Dr. N. George of Caltech for his guidance and help throughout the project, and Dr. F. E. C. Culick, also of Caltech, for his assistance in expressing the manuscript in its final form.

<sup>2</sup>TEM = transfer electromagnetic mode.

$w(z)$  = spot size at arbitrary plane  $z$ ;

$$w(z) = w_0 \left[ 1 + \left( \frac{\lambda z}{\pi w_0^2} \right)^2 \right]^{1/2}$$

$w_0$  = spot size of gaussian wave at plane  $z = 0$ ; also referred to as waist

$\tilde{q}(z)$  = complex radius of curvature of gaussian wave at arbitrary plane  $z$

$$\Psi(z) = \tan^{-1} \left( \frac{\lambda z}{\pi w_0^2} \right)$$

$$\frac{1}{\tilde{q}(z)} = \frac{1}{R(z)} - \frac{j\lambda}{\pi w^2(z)}$$

$$R(z) = z + \left( \frac{\pi w_0^2}{\lambda} \right)^2 \left( \frac{1}{z} \right)$$

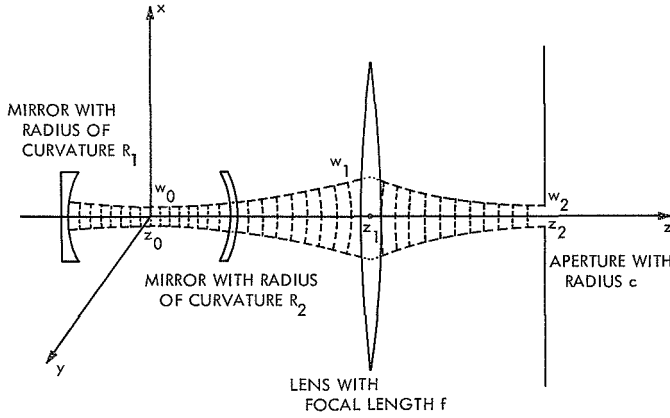


Fig. 1. Resonator-lens-aperture configuration

These symbols are the same as the ones used by Siegman (Ref. 3).

Now, if one of the mirrors is partially transparent, one can assume that the transmitted beam just before the lens will be of the same form as Eq. (1) but with reduced amplitude, and  $z$  replaced by  $z_1$ .

Immediately after the lens, the intensity distribution  $U'_{pl}$  is

$$U'_{pl}(r_1, \theta_1, z_1) = t(r_1, \theta_1) U_{pl}(r_1, \theta_1, z_1) \quad (2)$$

where for an ideal lens, having focal length  $f$ , the transmittance function is

$$t(r_1, \theta_1) = B \exp\left(-j \frac{\kappa r_1^2}{2f}\right) \quad (3)$$

The finite dimensions of the lens may be accounted for if one assumes the lens to have a gaussian transmittance function of the form

$$B \exp\left(-j \frac{\kappa r_1^2}{2\rho}\right) \quad (4)$$

where

$$\frac{1}{\rho} = \frac{1}{f} + \frac{4\lambda}{i\pi D^2}$$

and where  $D$  is the lens diameter.

In the calculations that follow, the first assumption will be used. However, the results obtained by using the second kind of transmittance function are equivalent, since a gaussian when Fourier-transformed will again yield a gaussian. Thus,  $U'_{pl}(r_1, z_1)$  is the same as  $U_{pl}(r_1, z_1)$  but with  $\tilde{q}(z_1)$  being replaced by  $\tilde{q}^1(z)$  where

$$\frac{1}{\tilde{q}^1(z_1)} = \frac{1}{\tilde{q}(z_1)} + \frac{1}{f} \quad (5)$$

At point  $z_2 = z_1 + f$ , it can be shown, for example, by use of the Collins chart, that the wave will be a plane wave with waist  $w_2$  ( $w_i$  = spot size of plane  $z = z_i$ ) given by

$$w_2 = \frac{\left[1 + \left(\frac{\lambda z_1}{\pi w_0^2}\right)^2\right]^{1/2}}{\left[1 + \left(\frac{w_0^4 \pi^2 + \lambda^2 z_1^2}{\lambda \pi w_0^2 S_1}\right)^2\right]^{1/2}} \quad (6)$$

where

$$S_1 = \frac{f \left[ z_1 + \left( \frac{\pi w_0^2}{\lambda} \right) \frac{1}{z_1} \right]}{f + z_1 + \left( \frac{\pi w_0^2}{\lambda} \right)^2 \frac{1}{z_1}} \quad (7)$$

Thus, the spot size at the focal point can be expressed in terms of  $z_1$ ,  $f$ ,  $w_0$ , which are all known parameters. The distribution at the aperture will then be of the form

$$U_{pl}(r_2, \theta_2) = \frac{A}{(1 + \delta_{0l})^{1/2}} \left[ \frac{p!}{\pi(\ell + p)!} \right]^{1/2} \left[ \frac{(2)^{1/2}}{w_2} \right]^\ell r_0' L_p' \left( \frac{2r_2^2}{w_2^2} \right) \exp(i\ell\theta_2) \exp\left(-\frac{r_2^2}{w_2^2}\right) \quad (8)$$

### 3. Derivation of Intensity Distribution at Observation Plane

Computation of the intensity at the screen is based on direct application of Huygen's principle and inclusion of a pupil function for the aperture.

$$U(P) = \frac{1}{j\lambda} \int_s U(P_2) \frac{\exp(-j\kappa R)}{\bar{R}} \cos(\bar{n}, \bar{R}) dS \quad (9)$$

Assuming that

$$r^2 + z^2 \ll r_2^2 \quad (\text{Fig. 2})$$

and expanding  $R$ , one obtains

$$U(P) = \frac{jz}{\lambda} \exp[-j\kappa(r^2 + z^2)^{1/2}] \int_s \int \frac{U(P_2)}{R^2} \exp\left\{-j\kappa\left[\frac{r_2^2}{2(r^2 + z^2)^{1/2}} - \frac{rr_2 \cos(\theta - \theta_2)}{(r^2 + z^2)^{1/2}}\right]\right\} r_2 dr_2 d\theta_2 \quad (10)$$

where  $z_2$  has been set equal to zero. In the exponent, only the second term of the binomial expansion was kept. The next order term is much less than 1 rad, for all cases in which  $r_2^2$  is much less than  $r^2$  or  $z^2$  or their sum. Substituting Eq. (8) into Eq. (10), and by approximating  $R^2$  by  $(r^2 + z^2)$ , permitted under the above assumptions, one obtains

$$U_{pl} = \frac{jz \exp[-j\kappa(r^2 + z^2)^{1/2}]}{\lambda(r^2 + z^2)} \frac{A}{(1 + \delta_{0l})^{1/2}} \left[ \frac{p!}{\pi(\ell + p)!} \right]^{1/2} \left[ \frac{(2)^{1/2}}{w_2} \right]^{\ell} \\ \times \int_0^{2\pi} \int_0^{\infty} r_2^{\ell} L_p^{\ell} \left( \frac{2r_2^2}{w_2^2} \right) \exp\left(-\frac{r_2^2}{w_2^2}\right) \exp(i\ell\theta_2) \exp\left\{-j\kappa \frac{r_2^2}{2(r^2 + z^2)^{1/2}}\right\} \exp\left\{\frac{j\kappa rr_2 \cos(\theta - \theta_2)}{(r^2 + z^2)^{1/2}}\right\} r_2 dr_2 d\theta_2 \quad (11)$$

Performing the integration over  $\theta_2$ , and including the pupil function which will just change the limits of integration over  $r_2$ , one obtains

$$U_{pl}(r, z) = \frac{jz \exp[-j\kappa(r^2 + z^2)^{1/2}]}{\lambda(r^2 + z^2)} \frac{A}{(1 + \delta_{0l})^{1/2}} \frac{p!}{[\pi(\ell + p)!]^{1/2}} \left[ \frac{(2)^{1/2}}{w_2^2} \right]^{\ell} \\ \times \int_0^c r_2^{1+\ell} \mathcal{J}_{\ell} \left[ \frac{\kappa rr_2}{(r^2 + z^2)^{1/2}} \right] L_p^{\ell} \left( \frac{2r_2^2}{w_2^2} \right) \exp\left\{-r_2^2 \left[ \frac{1}{w_2^2} + \frac{j\kappa}{2(r^2 + z^2)^{1/2}} \right]\right\} dr_2 \quad (12)$$

where  $\mathcal{J}_{\ell}$  is the Bessel function of integer order of the first kind.

Let

$$G = \int_0^c r_2^{1+\ell} \mathcal{J}_{\ell} \left[ \frac{\kappa rr_2}{(r^2 + z^2)^{1/2}} \right] L_p^{\ell} \left( \frac{2r_2^2}{w_2^2} \right) \exp\left\{-r_2^2 \left[ \frac{1}{w_2^2} + \frac{j\kappa}{2(r^2 + z^2)^{1/2}} \right]\right\} dr_2 \quad (13)$$

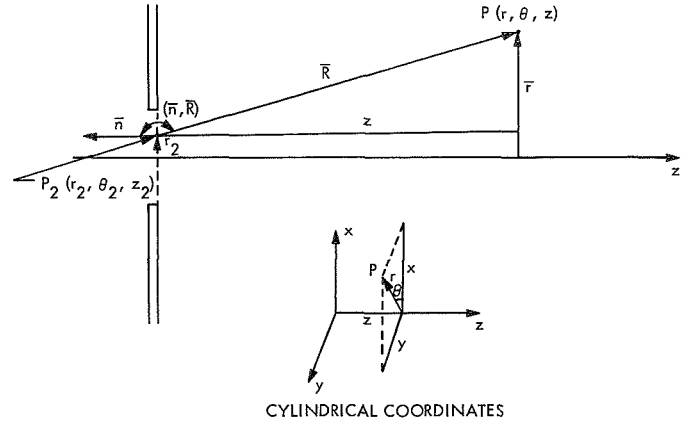


Fig. 2. Geometry of relationship between aperture and observation plane



and

$$\frac{r_2}{c} = x, \quad \frac{\kappa r c}{\lambda(r^2 + z^2)} = a, \quad \left[ \frac{c^2}{w_2^2} + \frac{j\pi c^2}{\lambda(r^2 + z^2)^{1/2}} \right] = \alpha, \quad \frac{2c^2}{w_2^2} = \beta \quad (14)$$

then

$$G = c^{2+i} \int_0^1 L_p^i(\beta x^2) \mathcal{J}_i(ax) \exp(-\alpha x^2) x^i x dx \quad (15)$$

but

$$L_p^i(z) = \sum_{k=0}^{k=p} (-1)^k \binom{p+\ell}{p-k} \frac{z^k}{k!}$$

where  $k$  is an integer index.

Hence,

$$G = c^{2+i} \sum_{k=0}^{k=p} (-1)^k \binom{p+\ell}{p-k} \frac{\beta^k}{k!} \int_0^1 x^{2k+i+1} \mathcal{J}_i(ax) \exp(-\alpha x^2) dx \quad (16)$$

The integral may be carried out if the Bessel function is written as a power series in  $ax$ . One finds eventually that

$$G = \frac{d^i}{2\alpha} c^{2+i} (p+\ell)! \sum_{k=0}^{k=p} \frac{\xi^k}{(\ell+k)!(p-k)!} \sum_{m=0}^{m=\infty} \frac{g^m \gamma(\ell+m+k+1, \alpha)}{m!(m+\ell)!} \quad (17)$$

where  $m$  is an integer index,  $\gamma(n, z)$  is the incomplete gamma function of order  $n$  and argument  $z$ ,

$$d = \frac{a}{2\alpha}, \quad \xi = \frac{-\beta}{\alpha}, \quad g = \frac{-a^2}{4\alpha}$$

Substitution of Eq. (17) into Eq. (12) will yield the distribution function  $U_{pl}$ . However, since the intensity  $I_{pl}$ , defined as  $|U_{pl}|^2$ , is of interest to us by being directly measurable, the latter will be presented.

$$I_{pl} = M \frac{1}{4\pi} \frac{c^4 z^2}{\lambda^2(r^2 + z^2)^2} \frac{p!(\ell+p)!}{(1+\delta_{0\ell})} h^i \left| \frac{1}{\alpha^{i+1}} \sum_{k=0}^{k=p} \frac{\xi^k}{(\ell+k)!(p-k)!} \sum_{m=0}^{m=\infty} \frac{g^m \gamma(\ell+m+k+1, \alpha)}{m!(m+\ell)!} \right|^2 \quad (18)$$

where

$$h = \frac{c^2 a^2}{2w_2^2}, \quad M = \text{normalization constant}$$

A number of important cases can be derived directly from Eq. (18):

(1) When  $p = 0$ , then Eq. (18) reduces to

$$I_{0l} = M \frac{1}{4\pi} \frac{c^4 z^2}{\lambda^2 (r^2 + z^2)^2} \frac{\beta^l}{l! (1 + \delta_{0l})} \left| \frac{e^{-\alpha}}{\alpha} \sum_{m=1}^{m=\infty} \frac{g_{l+m}(a)}{d^m} \right|^2 \quad (19)$$

(2) When  $p = l = 0$ , then Eq. (19) reduces to

$$I_{00} = M \frac{1}{8\pi} \frac{c^4 z^2}{\lambda^2 (r^2 + z^2)^2} \left| \frac{e^{-\alpha}}{\alpha} \sum_{m=1}^{m=\infty} \frac{g_m(a)}{d^m} \right|^2 \quad (20)$$

#### 4. Conclusions

From Eq. (18), it can be seen that the intensity is a function of the  $2(l+2)$  power of the radius of the aperture. Thus, for practical applications, where the efficiency  $c^4$  the system is important,  $l$  must be small. When  $p$  increases, then the number of rings or lobes increases as it is expected, except for the case of a very small aperture. But this is also undesirable, since even for  $l = 0$

the intensity is again a function of the fourth power of the radius of the aperture. The conclusion is, therefore, that the TEM<sub>00</sub> mode is the most desirable with which to work. The intensity distribution in this case, given by Eq. (20), which for a small aperture looks like the Airy pattern and for a large one looks like a simple gaussian, is the same result as the one obtained by Kauffman (Ref. 4) and Buck (Ref. 5) with a direct numerical calculation of the diffraction integral.

A computer program in Fortran IV has been prepared to calculate the intensity distribution according to the formulation presented here. Optional output in the form of graphic display is also available.

A number of characteristic plots illustrating the above-mentioned conclusions are included (Figs. 3-5).

#### References

1. Cox, C. H., *Uniform Laser Beam Profile Study*, under Contract No. NAS 7-744. Consumer and Technical Products Division, Owens-Illinois, Inc., Mar. 31, 1970.
2. Menichelli, V. J., Yang, L. C., *Sensitivity of Explosives to Laser Energy*, Technical Report 32-1474. Jet Propulsion Laboratory, Pasadena, Calif., Apr. 30, 1970.

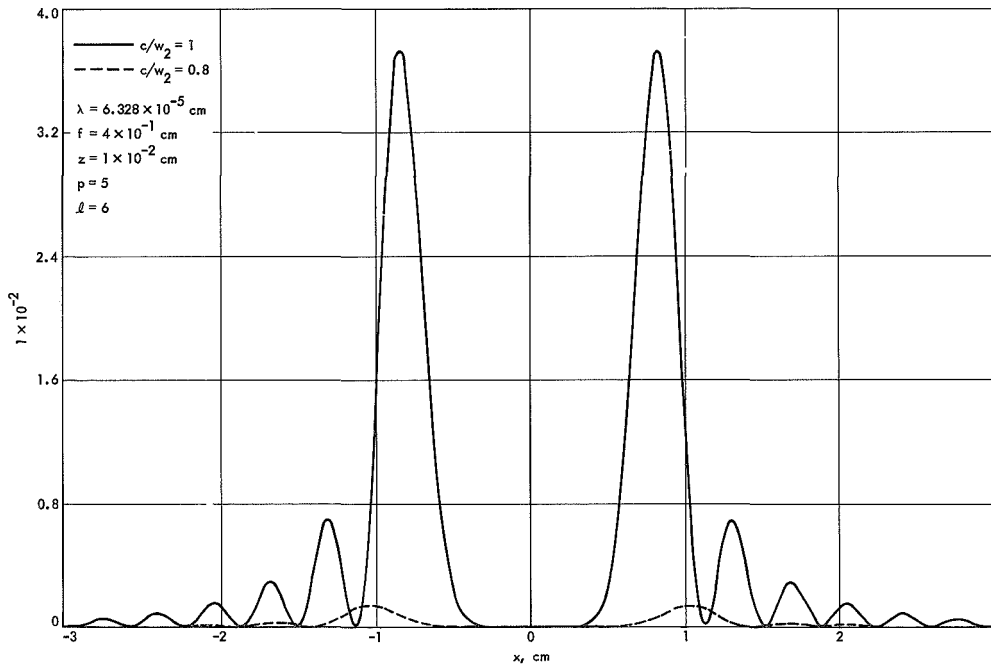
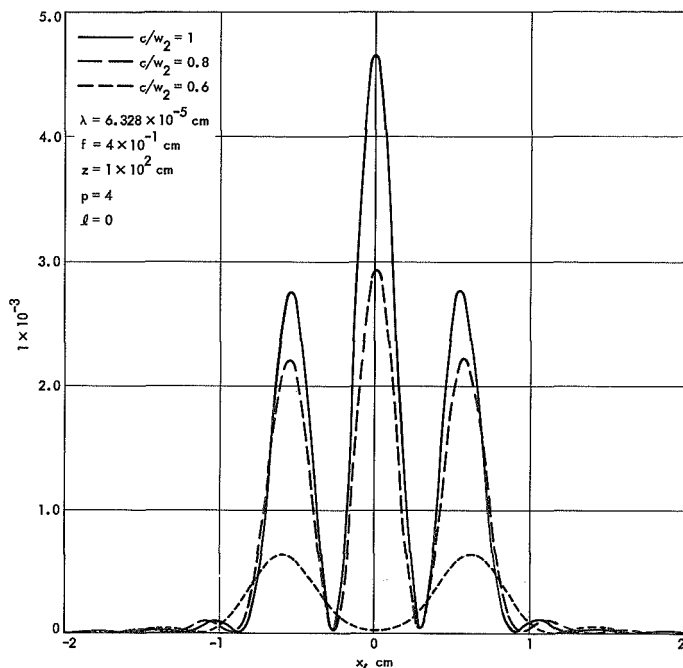
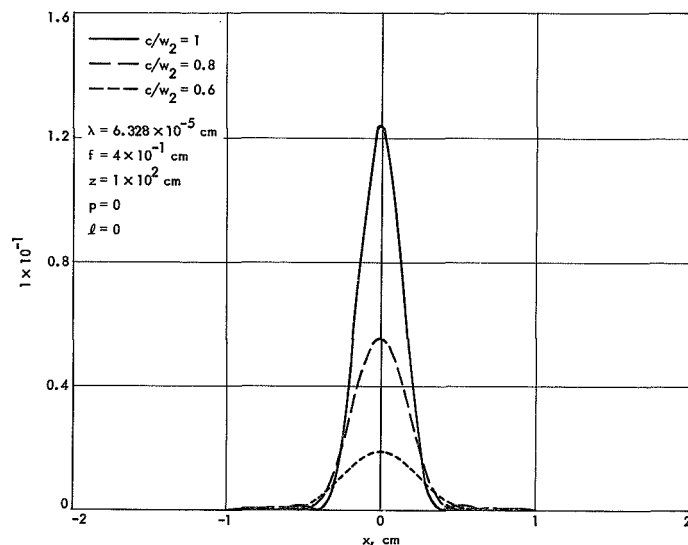


Fig. 3. Relative intensity  $I_{56}$  vs distance  $x$  from  $z$ -axis (intensity drops very fast because of high index  $l$ )



**Fig. 4. Relative intensity  $I_{40}$  vs distance  $x$  from  $z$ -axis (intensity distribution exhibits number of undesirable lobes)**



**Fig. 5. Relative intensity  $I_{00}$  vs distance  $x$  from  $z$ -axis (intensity distribution exhibits very low intensity wings in comparison with maximum, which clearly illustrates desirability of this mode)**

3. Siegman, A. E., *An Introduction to Lasers and Masers*, Prelim. Ed., pp. 8-15 to 8-53. McGraw-Hill Book Co., Inc., New York, 1968.

4. Kauffman, J. F., "The Calculated Radiation Patterns of a Truncated Gaussian Aperture Distribution," *IEEE Trans. Antennas and Propagation (Commun.)*, Vol. AP-13, pp. 473-474, May 1965.

5. Buck, A. L., "Radiation Pattern of a Truncated Gaussian Aperture Distribution," *Proc. IEEE (Proc. Letters)*, pp. 448-450, Mar. 1967.

## XVIII. Polymer Research

### PROPULSION DIVISION

#### A. Viscoelastic Behavior of Elastomers Undergoing Crosslinking Reactions,

*J. Moacanin and J. J. Aklonis<sup>1</sup>*

Previously (SPS 37-54, -57, -58, -59, -61, and -63, Vol. III) a framework was developed for the prediction of the physical behavior of polymers which are reacting while being subjected to stress relaxation, creep, or dynamic relaxation experiments. These predictions, however, were made for systems where scission was the only chemical reaction taking place. Here we extend the previous results to the prediction for systems where creep or stress relaxation experiments are conducted while crosslinking reactions are occurring. Since this situation is considerably more complicated than is the case of the scission reaction, we will deal only with a single instantaneous change in crosslink density  $\nu_1 \rightarrow \nu_2$ , where  $\nu_2 > \nu_1$ . Calculations will be made to allow prediction of behavior for this single transition system; equations for cases where the

crosslinking reaction is continuous will be derived but will not be solved.

Consider a creep experiment on a sample with crosslink density  $\nu_1$  subjected to a stress  $\sigma$ . At any time, the length of this sample may be given as

$$L(t) = [1 + \epsilon(t)] L_0 = [1 + D_{\nu_1}(t)\sigma] L_0 \quad (1)$$

At  $t^*$ , the change  $\nu_1$  to  $\nu_2$  is effected at virtually constant length and the behavior of the system at  $t > t^*$  will depend on  $D_{\nu_2}(t)$  and no longer on  $D_{\nu_1}(t)$ . The plausibility of this basic assumption has been argued previously (SPS 37-59, Vol. III).

Therefore, at times greater than  $t^*$  all of the relaxation mechanisms are governed by the character of a  $\nu_2$  network. But the chains are of two types. The fraction  $\nu_1/\nu_2$  of all the chains is supporting stress at  $t^*$ , whereas the fraction  $(\nu_2 - \nu_1)/\nu_2$  is not. The latter fraction represents the new chains which are always relaxed when formed (Ref. 1). Proceeding to introduce  $\bar{m}$  by equating

<sup>1</sup>Department of Chemistry, University of Southern California, Los Angeles, California.

sample length before and after the crosslink density change at  $t^*$ , we get

$$L(t^*) = [1 + D_{\nu_1}(t^*)\sigma]L_0 = \left[1 + \left(\frac{\nu_2}{\nu_1}\right)D_{\nu_2}(t^* + \bar{m})\sigma\right]L_0 \quad (2)$$

Here the factor  $\nu_2/\nu_1$  accounts for the fact that only a fraction of the chains of character  $\nu_2$  are supporting stress at  $t^*$ . Rearrangement of Eq. (2) yields

$$D_{\nu_1}(t^*) = \frac{\nu_2}{\nu_1} D_{\nu_2}(t^* + \bar{m}) \quad (3)$$

which defines the time shift  $\bar{m}$  in terms of  $\nu_1$ ,  $\nu_2$  and the corresponding compliance functions.

As the sample creep continues at  $t > t^*$ , the stress  $\sigma$  gradually shifts from the original  $\nu_1/\nu_2$  fraction of chains to the newly added chains. Therefore the distribution of the total stress  $\sigma$  between the two fractions of the network must be described.

According to the Boltzmann Superposition Principle (Ref. 1), the time-dependent strain  $\epsilon(t)$ , which results from any time-dependent stress  $\sigma(t)$  applied to a system whose compliance  $D(t)$  is known, may be calculated according to the equation

$$\epsilon(t) = \int_0^t \dot{\sigma}(\tau)D(t-\tau) d\tau \quad (4)$$

For times greater than  $t^*$ , the strain  $\epsilon_1(t)$  on the  $\nu_1/\nu_2$  chains now at crosslink density  $\nu_2$  may be given as

$$\epsilon_1(t) = \sigma \frac{\nu_2}{\nu_1} D_{\nu_2}(t + \bar{m}) + \int_{t^*}^t \dot{\sigma}_1(\tau) \frac{\nu_2}{\nu_1} D_{\nu_2}(t-\tau) d\tau \quad (5)$$

The first term on the right-hand side describes the response of the  $\nu_1/\nu_2$  chains subjected to the total stress  $\sigma$ , and the second term accounts for the stress transfer between the two networks.

Similarly for the  $(\nu_2 - \nu_1)/\nu_2$  newly formed chains, the strain  $\epsilon_2(t)$  for times greater than  $t^*$  would be given as

$$\epsilon_2(t) = \int_{t^*}^t \dot{\sigma}_2(\tau) \frac{\nu_2}{\nu_2 - \nu_1} D_{\nu_2}(t-\tau) d\tau \quad (6)$$

For a creep experiment, however, the total stress is constant; hence,

$$\text{or} \quad \left. \begin{aligned} \sigma &= \sigma_1 + \sigma_2 \\ \dot{\sigma}_1 &= -\dot{\sigma}_2 \end{aligned} \right\} \quad (7)$$

and Eq. (6) becomes

$$\epsilon_2(t) = - \int_{t^*}^t \dot{\sigma}_1(\tau) \frac{\nu_2}{\nu_2 - \nu_1} D_{\nu_2}(t-\tau) d\tau \quad (8)$$

Additional constraints are imposed by the fact that both networks are part of the same macroscopic system; accordingly, for times greater than  $t^*$ , the definitions depicted in Fig. 1 show that

$$D_1(t) = \frac{\Delta L^* + \Delta L(t)}{L_0} \quad (9)$$

For the  $(\nu_2 - \nu_1)/\nu_2$  new chains which are relaxed at  $t^*$ , however, the unstretched length which should appear in the denominator is  $L_0 + \Delta L^*$ . Accordingly, we define

$$D_2(t) = \frac{\Delta L(t)}{L_0 + \Delta L^*} \quad (10)$$

Elimination of  $\Delta L(t)$  from Eqs. (9) and (10) yields the relationship

$$L_0 \epsilon_1(t) - \Delta L^* = \epsilon_2(t)(L_0 + \Delta L^*) \quad (11)$$

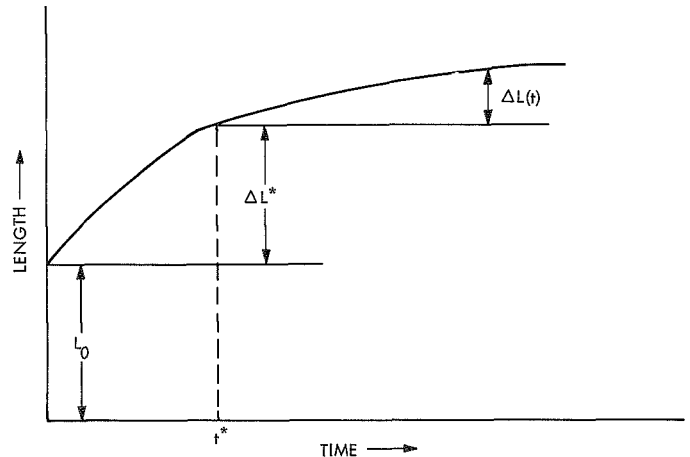


Fig. 1. Creep of an elastomer with a single crosslinking step at  $t^*$

Substitutions of Eqs. (5) and (8) into Eq. (11) yield

$$\int_{t^*}^t \dot{\sigma}(\tau) D_{\nu_2}(t - \tau) d\tau = \frac{\Delta L^* - L_0 \sigma_T \frac{\nu_2}{\nu_1} D_{\nu_2}(t + \bar{m})}{L_0 \frac{\nu_2}{\nu_1} + L_0 \frac{\nu_2}{\nu_2 - \nu_1} + \Delta L^* \frac{\nu_2}{\nu_2 - \nu_1}} \quad (12)$$

In principle Eq. (12) may be solved for  $\dot{\sigma}(t > t^*)$ ; this function may then be used for  $\dot{\sigma}_1(\tau)$  in Eq. (5) to calcu-

late  $\epsilon_1(t)$ , which can then be related to the length of the total sample as a function of time. To make the algebra tractable for computer calculations, we have generated from Eq. (12) recursion relationships in terms of finite differences. Parametric calculations are in progress.

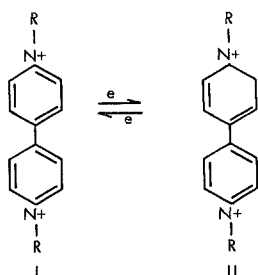
## Reference

1. Tobolsky, A. V., "Properties and Structure of Polymers," Chap. 3. John Wiley & Sons, Inc., New York, 1960.

## B. Energy Transfer in Bipyridilium (Paraquat)

Salts, A. Rembaum, V. Hadek, and S. P. S. Yen

The high electron affinity of 4,4'-bipyridilium (paraquat) salts, well-known herbicides (Ref. 1), has been demonstrated in 1933 by Michaelis and Hill (Ref. 2) who have shown that bipyridilium bases can be reduced to form colored species. The colorless bivalent cation accepts an electron to form II, a violet univalent cation, and this process may be represented by the following equilibrium:



where R is generally an aryl or alkyl group or a hydrogen atom.

II has several resonating structures and this fact is responsible for the absorption maximum in the long wavelength range of the spectrum (between 5700 to 6000 Å, depending on the structure of R).

The ease of reduction of paraquat is attributed to its high electron affinity which is also responsible for the

formation of a variety of charge transfer complexes between paraquats and phenols, quinones, hydroquinones, etc. (Refs. 3-5), the latter acting as electron donors and the paraquats as electron acceptors.

We report a charge transfer interaction between paraquat and 8,8,8'-tetracyanoquinodimethane radical anion (TCNQ $^{\cdot-}$ ). The purpose of our investigation was to determine whether an electron transfer occurs from TCNQ $^{\cdot-}$  to structure I and to compare the electron transport properties of aliphatic with aromatic tetracyanoquinodimethane salts. We have found that:

- (1) The paraquat salt I in which R = H when reacted with 2 molecules of LiTCNQ $^{\cdot-}$  exhibits a surprisingly low electrical resistivity (about 1 to 10  $\Omega$ -cm) and that on addition of neutral TCNQ the electrical resistivity remains unaltered. Furthermore, the same salt with R = CH $_3$  formed under identical conditions is characterized by a resistivity of the order of 10 $^6$   $\Omega$ -cm. However, the addition of neutral TCNQ to the latter reduces the electrical resistivity by about five orders of magnitude.
- (2) The visible spectra of the aliphatic diammonium and bipyridilium TCNQ salts exhibited the characteristic features previously reported for a number of other TCNQ compounds (Refs. 6 and 7).
- (3) The electronic transport properties of aliphatic diammonium TCNQ salts do not differ significantly from the corresponding bipyridilium salts,

Table 1. Aliphatic diammonium TCNQ salts

Diammonium salt	Ratio used in synthesis					
	1 mole diammonium salt:2 moles $\text{Li}^+\text{TCNQ}^{\cdot-}$			1 mole diammonium salt:2 moles $\text{Li}^+\text{TCNQ}^{\cdot-}$ : 2 moles $\text{TCNQ}^0$		
	Resistivity $\rho$ , $\Omega\text{-cm}$	Activation energy $\epsilon$ , eV	Seebeck coefficient $\alpha$ , $\mu\text{V}/^\circ\text{C}$	Resistivity $\rho$ , $\Omega\text{-cm}$	Activation energy $\epsilon$ , eV	Seebeck coefficient $\alpha$ , $\mu\text{V}/^\circ\text{C}$
$(\text{CH}_3)_2\text{N}(\text{CH}_2)_5\text{N}(\text{CH}_3)_2, 2\text{HBr}$	$4.0 \times 10^6$	0.29	+410	$1.2 \times 10^2$	0.22	-150
$(\text{CH}_3)_2\text{N}(\text{CH}_2)_6\text{N}(\text{CH}_3)_2, 2\text{HI}$	$5.0 \times 10^{10}$	0.73	—	$1.0 \times 10^2$	0.21	-143
$(\text{CH}_3)_2\text{N}(\text{CH}_2)_8\text{N}(\text{CH}_3)_2, 2\text{HBr}$	$3.7 \times 10^7$	0.41	+720	7.0	0.090	-10

Table 2. Bipyridilium TCNQ salts

Bipyridilium salt reference No.	Ratio used in synthesis					
	1 mole BP salt:2 moles $\text{Li}^+\text{TCNQ}^{\cdot-}$			1 mole BP salt:2 moles $\text{Li}^+\text{TCNQ}^{\cdot-}$ :2 moles $\text{TCNQ}^0$		
	Resistivity $\rho$ , $\Omega\text{-cm}$	Activation energy $\epsilon$ , eV	Seebeck coefficient $\alpha$ , $\mu\text{V}/^\circ\text{C}$	Resistivity $\rho$ , $\Omega\text{-cm}$	Activation energy $\epsilon$ , eV	Seebeck coefficient $\alpha$ , $\mu\text{V}/^\circ\text{C}$
4,4'-BP,2HCl	1.5	0.090	-66	1.2	0.098	-65
2,2'-BP,2HCl	1.4	0.060	-56	1.0	0.058	-52
4,4'-BP,2HI	12.0	0.110	-65	9.0	0.103	-66
4,4'-BP,2CH <sub>3</sub> I	$1.1 \times 10^6$	0.65	+150	22.0	0.130	-37

with the exception of the case where electron transfer occurs (when  $\text{R} = \text{H}$ ).

The electronic transport properties of aliphatic diammonium compounds and bipyridilium TCNQ salts are shown in Tables 1 and 2, respectively. The investigated samples were prepared by reacting one mole of the diammonium salt in the form of dichloride or diiodide with two moles of  $\text{LiTCNQ}^{\cdot-}$ . Another series of samples was prepared by reacting the above diammonium salts with two moles of  $\text{LiTCNQ}^{\cdot-}$  and two moles of neutral TCNQ (referred to as  $\text{TCNQ}^0$ ).

Table 1 confirms the previous results (Ref. 7), namely, the electrical resistivity of diammonium TCNQ salts in which both positive nitrogens are associated with a  $\text{TCNQ}^{\cdot-}$  is five to six orders of magnitude higher than the resistivity of the same compounds containing, in addition to the radical ion ( $\text{TCNQ}^{\cdot-}$ ), a neutral TCNQ molecule.

Table 2 illustrates the same facts for N,N' dimethyl-4,4'-bipyridilium TCNQ compound (4,4'-BP,2CH<sub>3</sub>I). How-

ever, the N,N'-dihydro-4,4' and 2,2'-bipyridilium TCNQ salts (4,4'-BP,2HCl, and 2,2'-BP,2HCl) constitute exceptions of what appeared to be a general rule. This anomalous behavior was elucidated by examination of the absorption spectra in the visible range in conjunction with the electrical properties.

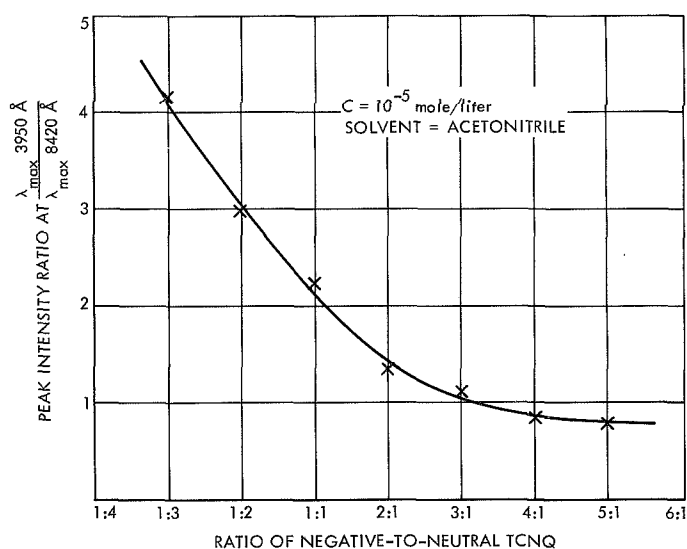
A calibration curve (Fig. 1) was determined using synthetic mixtures of  $\text{LiTCNQ}^{\cdot-}$  and  $\text{TCNQ}^0$ . The plot of the intensity ratio of  $\lambda_{\text{max}}$  at 3950 Å to 8420 Å versus the ratio of  $\text{TCNQ}^{\cdot-}:\text{TCNQ}^0$  served to estimate the composition of TCNQ salts and compare the latter with the electrical resistivity (Table 3).

On the basis of the above results it is possible to conclude that electron transfer occurs from  $\text{TCNQ}^{\cdot-}$  in structure III to yield structure IV, while in the case where a  $\text{CH}_3$  group is bonded to the positive nitrogen no electron transfer is observed.

The electron transfer results in the oxidation of  $\text{TCNQ}^{\cdot-}$  to  $\text{TCNQ}^0$  and is therefore responsible for the low

**Table 3. Correlation of resistivity with the ratio TCNQ<sup>-</sup>:TCNQ<sup>0</sup>**

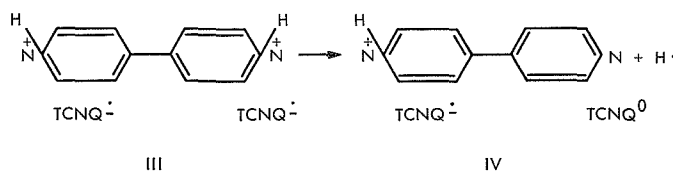
Bipyridilium salt reference No.	Intensity ratio at $\frac{\lambda_{\max} 3950 \text{ \AA}}{\lambda_{\max} 8420 \text{ \AA}}$	Number of moles in BP salt		Resistivity $\rho$ , $\Omega\text{-cm}$
		TCNQ <sup>-</sup>	TCNQ <sup>0</sup>	
4,4'-BP,2HCl	2.25	1	1	1.5
2,2'-BP,2HCl	2.25	1	1	1.4
4,4'-BP,2HI	2.25	1	1	$1.2 \times 10$
4,4'-BP,2CH <sub>3</sub> I		2	0	$1.1 \times 10^6$
4,4'-BP,2CH <sub>3</sub> I	1.2	2	1	$2.2 \times 10$



**Fig. 1. Calibration curve to determine negative-to-neutral TCNQ ratio according to peak-height ratio of spectral bands 3950 Å and 8420 Å**

resistivity of salts in which the positive nitrogen is bonded to a hydrogen. Thus, the reaction of structure I with LiTCNQ leads to a salt containing both the radical anion as well as a neutral TCNQ molecule; therefore its high electronic conductivity is not surprising. The formation of structure II in the electron transfer process does not appear likely on the basis of the observed absorption

spectra (Ref. 2). The following mechanism is therefore proposed:



It is assumed that the hydrogen atom reacts with the solvent or dimerizes and the absence of an oxidation-reduction reaction in methyl-substituted species may be due to the inductive effect of the methyl group. Alternatively the N—CH<sub>3</sub> bond being stronger than the N—H bond, the splitting of the hydrogen atom in structure III could be explained on the basis of bond strength. This assumption appears more plausible since the oxidation reduction potentials of 4,4'-BP,CH<sub>3</sub>I and 4,4'-BP,HCl determined polarographically were found to be practically identical, indicating the same electron affinity.

The increase in conductivity of salts containing TCNQ<sup>0</sup> is consistent with the theoretical derivations of LeBlanc (Ref. 8) who accounted for the high conductivity of the complex salts by hole-electron pairs in the Heitler London ground state configuration without placing two electrons on the same TCNQ site and therefore avoiding formation of a dianion. In this interpretation the presence of a neutral TCNQ molecule is necessary for efficient electron transport.

## References

1. Boon, W. R., *Endeavour*, Vol. XXVI, p. 27, 1967.
2. Michaelis, L., and Hill, E. S., *J. Gen. Phys.*, Vol. 16, p. 859, 1933.
3. MacFarlane, A. J., and Williams, R. T. P., *J. Chem. Soc., Part A*, p. 1517, 1969.
4. White, B. G., *Trans. Faraday Soc.*, Vol. 65, p. 2000, 1969.
5. Ledwith, A., and Woods, H. J., *J. Chem. Soc.*, p. 1422, 1970.
6. Melby, L. R., et al., *J. Amer. Chem. Soc.*, Vol. 84, p. 3374, 1962.
7. Rembaum, A., et al., *J. Phys. Chem.*, Vol. 73, p. 513, 1969.
8. LeBlanc, O. H., *J. Chem. Phys.*, Vol. 42, p. 4307, 1965.



## C. Electrical Properties of TCNQ Salts of Ionene Polymers and Their Model Compounds,

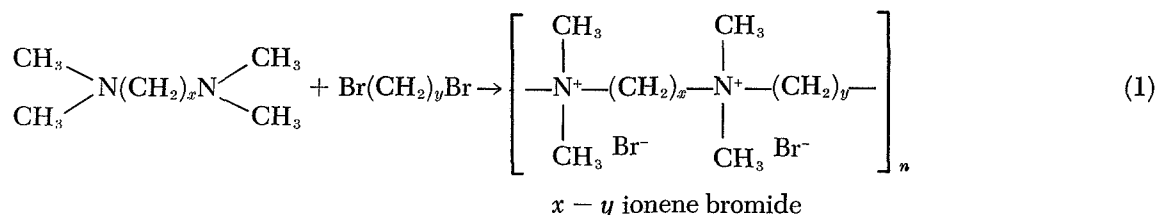
V. Hadek, H. Noguchi, and A. Rembaum<sup>1</sup>

### 1. Introduction

The discovery of highly conductive 7,7',8,8'-tetracyanoquinodimethane (TCNQ) paramagnetic salts constitutes an important advance in the study of electronic conductivity of purely organic materials. This high conductivity is exhibited by monosalts or polysalts, provided the TCNQ radical anion (TCNQ<sup>•-</sup>) is associated with a neutral TCNQ molecule (TCNQ<sup>0</sup>) (Refs. 1-3). Although the electronic conductivity of this type of material has

already been well established (Ref. 4), no data are available as yet on the electrical properties of TCNQ polymers as a function of polymer structure.

The present investigation was undertaken in order to gain insight into the mechanism of electronic conductivity of polymeric TCNQ salts, formed from polyelectrolytes, the structure of which could be systematically altered. Ionene polymers (Ref. 5) of well-defined structure seemed to be ideally suited for this purpose. The latter was therefore synthesized by means of the following reaction:



where  $x$  and  $y$  are integers, and  $n$  is the number of unit segments. The kinetics of formation and some solution properties of these polymers were recently described (Ref. 6).

Electrically conducting polysalts were prepared by the reaction of  $x-y$  ionene bromides with LiTCNQ in presence or in absence of TCNQ<sup>0</sup>. It was found that the specific resistivity and the activation energy for conduction of a 6-6 ( $x = 6$  and  $y = 6$ ) ionene polysalt was considerably higher than a number of similar polysalts with different  $x$  and  $y$  values. In order to elucidate this anomaly, a homologous series of model TCNQ compounds representing unit segments of the polymers was synthesized from polymethylene *bis*-(trimethylammonium halides) and their electrical properties were compared with those of TCNQ polysalts. The large variations in electrical properties were found to depend mainly on the geometrical configuration and crystal packing and

not on the distances between positively charged nitrogen atoms in the polyelectrolytes.

### 2. Results

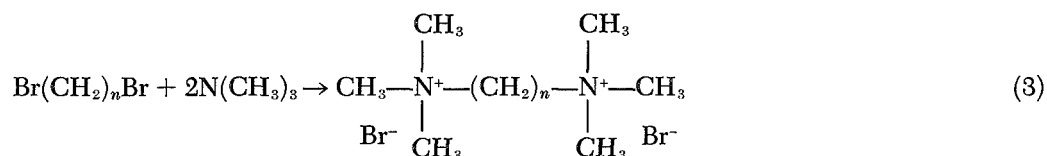
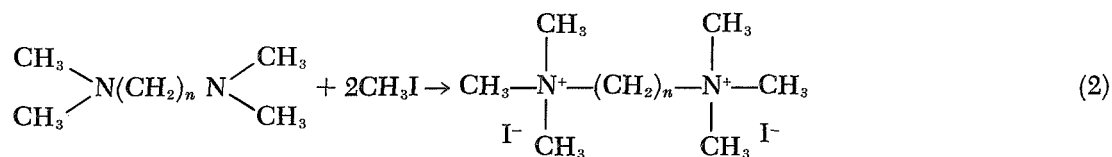
**a. Ionene polymers.** The first series of polymeric TCNQ salts was prepared by reacting ionene bromides (intrinsic viscosity  $[\eta] = 0.154$  to  $0.791$  dl/g measured in 0.4 mole KBr aq. solution) with LiTCNQ, using a molar ratio of unit segment to LiTCNQ of 1:2. The intrinsic viscosity of the polymers, the yields, and chemical analyses are recorded in Table 1.

The second series of polysalts was prepared by reacting ionene bromides with a mixture of LiTCNQ and TCNQ<sup>0</sup> using a molar ratio of 1:2:2.

The electrical properties of the polysalts, i.e., resistivity ( $\rho$ ), activation energy ( $\epsilon$ ), and Seebeck coefficient ( $\alpha$ ) in absence and in presence of TCNQ<sup>0</sup>, are recorded in Table 2. The dramatic changes in resistivity of salts in which  $x = 4$  or  $6$  and  $y = 3, 5, 6, 8, 10$ , and  $16$  are shown in Fig. 1.

<sup>1</sup>Acknowledgement made to the National Academy of Sciences, National Research Council, Washington, D. C., for the Postdoctoral Resident Associateships supporting this research, and to Dr. R. Bau, University of Southern California, for assistance in the x-ray investigations.

*b. Model compounds (DiTCNQ salts).* The model compounds, namely, polymethylene *bis*-(trimethylammonium halides), were prepared by means of reaction (2) or (3):



The details of synthesis and chemical analysis are shown in Table 3.

Table 1. Synthesis of TCNQ polysalts (in absence of TCNQ<sup>0</sup>)

x-y	[η], dl/g	Temperature, °C	Yield, %	Elemental composition						
				Calculated, <sup>a</sup> %			Found, %			
				C	H	N	C	H	N	Br
3-4	0.154	60-65	85.1	70.69	5.76	23.55	67.01	6.36	19.95	2.13
6-3	0.246	22-30	81.9							
6-3	0.246	60-65	84.7							
6-5	0.482	22-30	82.2							
6-5	0.482	60-65	91.9							
6-6	0.299	22-30	—							
6-6	0.299	60-65	96.1							
6-8	0.577	22-30	91.0							
6-8	0.577	60-65	97.5							
6-8	0.577	60-65	85.9	72.80	6.98	20.21	68.62	7.68	15.73	3.71
6-10	0.791	22-30	94.4							
6-10	0.791	60-65	90.3	73.30	7.27	19.43	70.23	7.90	15.70	3.23
6-16	0.157	22-30	51.6							
6-16	0.157	60-65	63.0	74.59	8.01	17.40	73.17	8.32	14.71	0.13

<sup>a</sup>Calculated for the idealized structure:

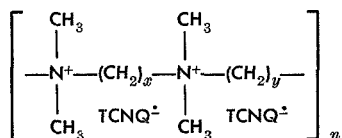


Table 2. Electrical properties of the TCNQ polysalts

Numbers of CH <sub>2</sub> group (x-y)	One-unit segment to 2TCNQ <sup>-</sup>			One-unit segment to 2TCNQ <sup>-</sup> 2TCNQ <sup>0</sup> with negative and neutral TCNQ		
	Resistivity $\rho$ , $\Omega$ -cm	Activation energy $\epsilon$ , eV	Seebeck coefficient $\alpha$ , mV/°C	Resistivity $\rho$ , $\Omega$ -cm	Activation energy $\epsilon$ , eV	Seebeck coefficient $\alpha$ , mV/°C
3-4	$2.8 \times 10^4$	0.22	+0.12	$1.5 \times 10^8$	0.20	+0.09
6-3	$9.5 \times 10^5$	0.30	+0.31	$3.4 \times 10^8$	0.24	+0.06
6-5	$1.5 \times 10^8$	0.56	Positive <sup>a</sup>	$7.7 \times 10^1$	0.140	-0.027
6-6	$3.2 \times 10^8$	0.58	Positive <sup>a</sup>	$5.2 \times 10^7$	0.46	+0.93
6-8	$7.2 \times 10^7$	0.52	+0.72	$1.4 \times 10^7$	0.45	+0.66
6-10	$3.0 \times 10^6$	0.39	+0.39	$7.3 \times 10^4$	0.29	+0.27
6-16	$9.6 \times 10^6$	0.48	+0.04	$4.7 \times 10^2$	0.23	-0.008

<sup>a</sup>The Seebeck coefficient was positive but its exact value could not be determined because of polarization effects.

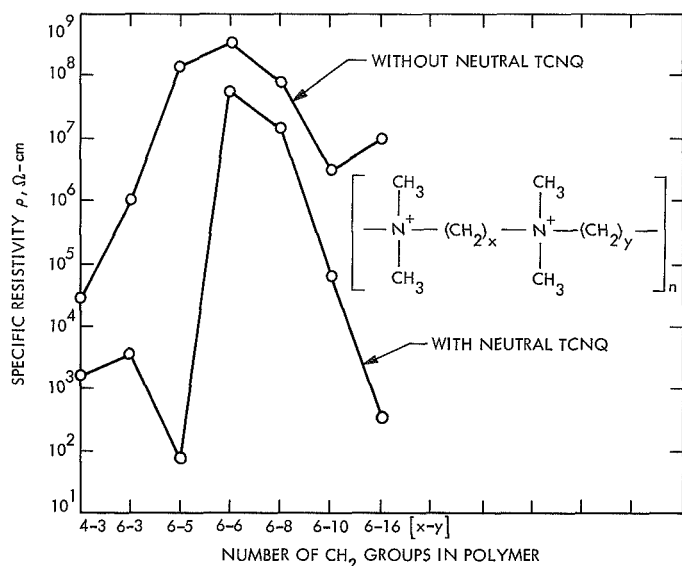


Fig. 1. Specific resistivities of TCNQ salts derived from the model compounds as a function of the number of CH<sub>2</sub> groups

The model compounds were subsequently reacted with LiTCNQ and the obtained disalts had the composition  $M^{2+}(TCNQ^-)_2$ , where  $M^{2+}$  represents the polymethylene bis-(trimethylammonium) dication, which is generally in excellent agreement with the chemical analysis (Table 4). Similarly, when the model compounds were reacted with LiTCNQ and TCNQ<sup>0</sup> using the three reagents in a molar ratio of 1:2:2, another series of disalts was obtained. The electrical properties of both types of model salts are recorded in Table 5 and the resistivity versus number of CH<sub>2</sub> groups is shown in Fig. 2. The resistivity values of

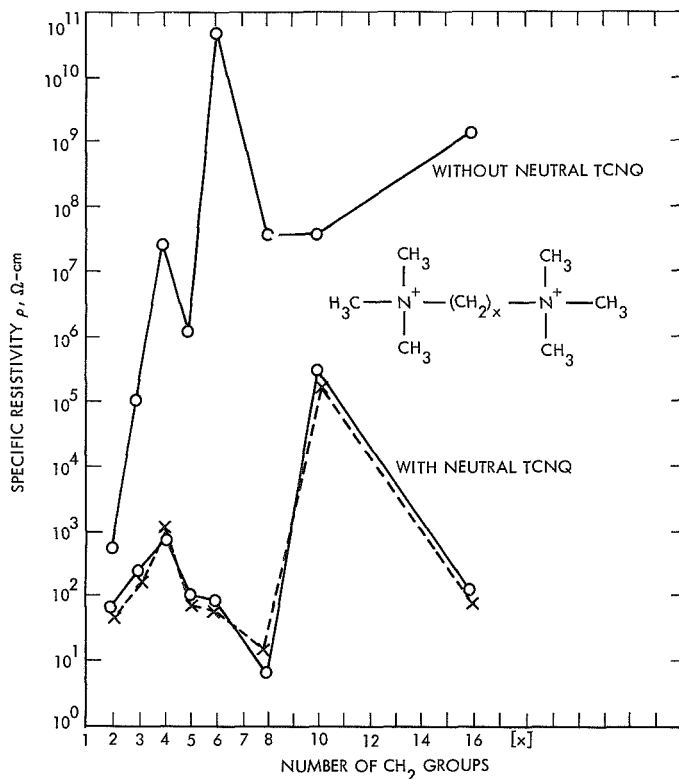


Fig. 2. Specific resistivities of TCNQ salts derived from ionene polymers as a function of the numbers of CH<sub>2</sub> groups

the same batch of materials could be reproduced with an accuracy of  $\pm 10\%$ .

The broken line (Fig. 2) represents the resistivity of salts prepared at 65°C and allowed to crystallize for

Table 3. Synthesis of polymethylene bis-(trimethylammonium halides)

n	Reaction	Concentration		Reaction time, h	Yield, %		Recrystallization solvent	Elemental composition								
								Calculated, %				Found, %				
								Type of halogen X	C	H	X <sup>-</sup>	N	C	H	Halide	N
2	2	Diamine, mol/ml	CH <sub>3</sub> I, mol/ml	120	94.5	4.9	Methanol	I	24.02	5.54	63.44	7.00	24.16	5.47	63.27	6.97
3	2	0.100/100	0.220/100	120	0	99.6	95% ethanol	I	26.10	5.84	61.29	6.76	26.23	5.87	61.36	6.72
4	2	0.100/100	0.220/100	120	71.4	28.6	95% ethanol	I	28.05	6.12	59.28	6.54	27.94	6.16	59.20	6.41
5	3	N(CH <sub>3</sub> ) <sub>3</sub> <sup>a</sup> , mol	Dibromide/CH <sub>3</sub> OH, mol/ml	260	0	99.2	93% ethanol	Br	37.95	8.11	45.90	8.05	38.03	8.14	45.90	7.99
8	3	0.252	0.100/150	360	0	99.9	Abs ethanol	Br	43.09	8.78	40.95	7.18	43.11	8.69	40.95	7.14
10	3	0.244	0.100/200	260	0	113.3	Abs ethanol	Br	45.95	9.16	38.21	6.70	45.86	9.14	38.08	6.68
16	3	0.249	0.100/200 + 30Bz <sup>c</sup>	480	0	92.6	n-butanol	Br	52.59	10.03	31.81	5.58	52.44	10.08	31.85	5.50
		0.0450	0.0150/40 + 40Bz													

<sup>a</sup>I: product filtered off from the mixture at the end of the reaction.  
<sup>b</sup>II: product isolated from the mother liquor.  
<sup>c</sup>Bz: benzene.

<sup>a</sup>I: product filtered off from the mixture at the end of the reaction.

<sup>b</sup>II: product isolated from the mother liquor.

<sup>c</sup>Bz: benzene.

Table 4. Synthesis of the complexes with TCNQ anions from polymethylene bis-(trimethylammonium halides)<sup>a</sup>

Methonium		Compound, mmol	LiTCNQ, mmol	CH <sub>3</sub> OH, <sup>b</sup> ml	Tempera- ture, °C	Yield, %	Elemental composition						
n	Halo- gen						Calculated, %			Found, %			
							C	H	N	C	H	N	Halogen
2	I	1.00	2.00	50	RT	97.4							
		1.50	3.00	50	55–60	93.3							
		1.50	3.00	50	60–65	90.6	69.29	5.45	25.25	67.16	5.48	24.94	
3	I	1.50	3.00	50	40–50	76.2	69.70	5.67	24.63	69.32	5.71	24.89	
		1.50	3.00	50	40–50	84.3							
		1.50	3.00	100	60–65	79.9							
4	I	0.75	1.50	50	60–65	80.5							
		1.50	3.00	50	40–50	77.2	70.08	5.88	24.04	70.31	5.83	24.08	
5	Br	0.75	1.50	50	60–65	71.9							
		1.50	3.00	50	40–50	74.0	70.45	6.08	23.47	70.36	6.04	23.68	
6	I	0.75	1.50	50	60–65	78.2				70.60	6.38	22.70	Trace (I)
		1.50	3.00	50	40–50	83.3	70.80	6.27	22.93	70.94	6.20	22.88	
8	Br	0.75	1.50	50	60–65	84.7							
		1.50	3.00	50	40–50	91.5	71.45	6.63	21.93	71.64	6.55	21.92	
10	Br	0.75	1.50	50	60–65	82.7				72.02	7.10	20.68	Trace (Br)
		1.50	3.00	50	40–50	93.4	72.04	6.95	21.00	72.04	6.95	21.03	
16	Br	0.50	1.00	25	RT	91.3							
		0.50	1.00	25	40–50	88.0	73.57	7.78	18.65	73.43	7.80	18.58	
		0.75	1.50	50	60–65	79.3							

<sup>a</sup>The reactions were carried out at room temperature (RT: ~22°C) for 1.5 h. The reactions at elevated temperatures were carried out for 20 min and left at room temperature for 20 h before filtration.

<sup>b</sup>Amount of methanol used to dissolve each reactant.

Table 5. Electrical properties of the TCNQ complexes derived from polymethylene bis-(trimethylammonium halides)

Number of CH <sub>2</sub> Group (n)	With negative TCNQ			With negative and neutral TCNQ		
	$\rho$ , $\Omega$ -cm	$\epsilon$ , eV	$\alpha$ , mV/°C	$\rho$ , $\Omega$ -cm	$\epsilon$ , eV	$\alpha$ , mV/°C
2	$1.9 \times 10^3$	0.12	+0.12	$9.0 \times 10^1$	0.10	+0.043
3	$1.3 \times 10^5$	0.24	+0.59	$4.2 \times 10^2$	0.16	-0.060
4	$3.1 \times 10^7$	0.41	+0.52	$6.5 \times 10^2$	0.26	+0.132
5	$4.0 \times 10^6$	0.29	+0.41	$1.2 \times 10^2$	0.22	-0.150
6	$5.0 \times 10^{10}$	0.73	Positive <sup>a</sup>	$1.0 \times 10^2$	0.21	-0.143
8	$3.7 \times 10^7$	0.41	+0.72	$7.0 \times 10^0$	0.090	-0.010
10	$3.6 \times 10^7$	0.44	+0.69	$4.1 \times 10^5$	0.28	+0.64
16	$1.4 \times 10^9$	0.46	+0.75	$1.5 \times 10^2$	0.16	0

<sup>a</sup>The Seebeck coefficient was positive but its exact value could not be determined because of high impedance.

15 hours while the parallel solid line represents the resistivity of salts prepared at room temperature and allowed to crystallize for 3 days. In both cases the same relation between resistivity and the number of CH<sub>2</sub> groups was found.

The model compounds exhibited large variation in electrical properties. For the first series with TCNQ anions, only the specific resistivity varied in a broad range, reaching the maximum of  $5 \times 10^{10} \Omega\text{-cm}$  for the salt with 6 methylene groups. The positive Seebeck coefficient exhibited by the polymeric salts indicates hole-type conductivity probably due to the electrically active lattice defects.

For the second series of complexes with neutral TCNQ, the magnitude of resistivity substantially decreased, especially where negative Seebeck coefficient was observed, indicating appearance of electron conductivity. The same relation between the negative Seebeck coefficient, electron conductivity, low resistivity, and low activation energy of conductivity was observed for polymeric complexes (Table 2 and Fig. 1).

**c. Spectral properties.** The spectra of all salts containing TCNQ<sup>-</sup> only exhibited two typical major bands at 4200 and 8420 Å. The intensity ratio of the two bands ( $\sim 0.57$ ) was independent of the number of CH<sub>2</sub> groups. The salts containing TCNQ<sup>0</sup> exhibited a  $\lambda_{\text{max}}$  at 3950 Å due to the neutral TCNQ molecule. The intensity ratio of the peak values at  $\lambda_{\text{max}}$  3950 Å to that at  $\lambda_{\text{max}}$  8420 Å varied from 0.9 to 1.6, the highest value representing the most conducting salt (with 8 CH<sub>2</sub> groups).

**d. X-ray analysis.** The materials investigated by single-crystal x-ray diffraction methods were TCNQ salts of model compounds of the following composition:

$[(\text{CH}_3)_3\text{N}^+(\text{CH}_2)_5\text{N}^+(\text{CH}_3)_3]$	$2(\text{TCNQ}^-)(\text{TCNQ}^0)$	I
$[(\text{CH}_3)_3\text{N}^+(\text{CH}_2)_5\text{N}^+(\text{CH}_3)_3]$	$2(\text{TCNQ}^-)$	II
$[(\text{CH}_3)_3\text{N}^+(\text{CH}_2)_6\text{N}^+(\text{CH}_3)_3]$	$2(\text{TCNQ}^-)(\text{TCNQ}^0)$	III
$[(\text{CH}_3)_3\text{N}^+(\text{CH}_2)_6\text{N}^+(\text{CH}_3)_3]$	$2(\text{TCNQ}^-)$	IV

Other crystals, listed in Table 5, were twinned and were too small to yield satisfactory diffraction patterns. I and IV yielded the most distinct and intense patterns. These two salts were found to have triclinic unit cells (Table 6). The dimensions of the unit cell of IV were substantially smaller than those of I, evidently due to the effect of two extra TCNQ molecules in the unit cell.

**Table 6. Crystal structure**

I. Model compound with 6 CH <sub>2</sub> groups and negative TCNQ.		
Unit cell: Triclinic		
$a = 11.500 \text{ Å}$	$\alpha = 93^\circ 33'$	
$b = 11.690 \text{ Å}$	$\beta = 125^\circ 17'$	
$c = 7.675 \text{ Å}$	$\gamma = 95^\circ 40'$	
$V = 829.53 \text{ Å}^3$		
II. Model compound with 5 CH <sub>2</sub> groups and negative and neutral TCNQ.		
Unit cell: Triclinic		
$a = 14.151 \text{ Å}$	$\alpha = 98^\circ 29'$	
$b = 7.829 \text{ Å}$	$\beta = 108^\circ 07'$	
$c = 13.813 \text{ Å}$	$\gamma = 72^\circ 28'$	
$V = 1384.39 \text{ Å}^3$		

Two other salts, II and III, gave less satisfactory diffraction patterns, which indicated nevertheless one-dimensional structures.

The unit cell parameters of I are recorded in Table 6. The (h o l) precession photograph showed a strong series of spots in the  $(\bar{1} 0 4)$  direction (i.e., spots  $\bar{1} 0 4$ ,  $\bar{2} 0 8$ ,  $\bar{3} 0 12$  were very intense). This could be interpreted as stacking of electron density regions roughly perpendicular to the axis and separated by a repeat distance of 3.21 Å.

Salt II also showed evidence of 3.21 Å stacking as in salt I, but unfortunately crystal disorder effects prevented determination of unit cell dimensions.

As in the previous case, salt III showed a one-dimensional stacking effect (this time with a repeat distance of 3.11 Å), but further analysis was hampered by crystal twinning.

Salt IV gave very good diffraction patterns, but no evidence of any stacking phenomena was observed (i.e., no intense series of spots could be found in the diffraction pattern). The unit cell parameters are shown in Table 6.

### 3. Discussion

The values of  $\rho$ ,  $\epsilon$ , and  $\alpha$  have been found to be reproducible within 30%. Further confidence in the accuracy of the results was gained by the use of disalts or polysalts synthesized under different experimental conditions, i.e., varying the temperature or solvent system. Essentially the same results were obtained indicating that the trends depicted in Figs. 1 and 2 are correct. The high values of

specific resistivities of 6-6 ionene polymers were independent of molecular weight (samples of approximate weight average molecular weight of 10,000, 25,000, and 80,000 were tested). In addition, the properties of  $x$ - $y$  ionenes were identical to those of  $y$ - $x$  ionenes.

In general, the examination of Figs. 1 and 2 and Tables 1, 2, 3, and 4 lead to the following conclusions:

- (1) The highest specific resistivity and activation energy were observed for salts derived from model compounds or polymers containing 6 CH<sub>2</sub> groups.
- (2) The salts with both negative and neutral TCNQ derived from the model compounds with 16 CH<sub>2</sub> groups and 6-16 ionene bromide exhibited very low specific resistivities, in spite of the fact that they contain the longest methylene chain between positive nitrogen atoms.
- (3) The salts from the model compounds and the polymers with high specific resistivities invariably exhibited hole-type conductivity or positive Seebeck coefficients, while those with low specific resistivities were characterized by electron-type conductivity or negative Seebeck coefficients.
- (4) The salts with 16 CH<sub>2</sub> groups exhibited low Seebeck coefficients. The lowest concentrations of the charge in these salts gave rise to a mixed-type conductivity with the same concentration of holes and electrons.
- (5) The characteristic variations of specific resistivity as a function of the number of CH<sub>2</sub> groups can be attributed to the conformations of polymethylene chains in the solid state, the high conductivity being due to electron delocalization in one-dimensional TCNQ columns.

## D. Evaluation of Magnetic Recording Tapes: A Method for the Quantitative Determination of Stick-Slip, R. H. Silver, S. H. Kalfayan, and J. K. Hoffman<sup>1</sup>

### 1. Introduction

The behavior known as "stick-slip," encountered with magnetic recording tapes, was discussed previously (SPS 37-63, -64, and -65, Vol. III). The quantitative deter-

The resistivity value of the polysalt containing 6 CH<sub>2</sub> groups between positive nitrogens and only TCNQ<sup>-</sup> is also highest in the analogous model salts (Fig. 2). The fact that TCNQ salt of the model compound (salt IV, see *Subsection d*) exhibits a resistivity value of the order of 10<sup>2</sup> Ω-cm seems to be out of line. It can be understood on the basis of the x-ray investigations. The geometry of the salts formed is the dominant factor in the electrical values of the materials under investigation and it is only when one-dimensional column stacking of the TCNQ moiety occurs that a very high electronic conductivity is observed. It is therefore not surprising to note that a decrease of concentration of positive charges in the polymeric chain or in the model compounds does not necessarily increase the values of  $\rho$  or  $\epsilon$  (Figs. 1 and 2).

The preliminary x-ray data support the anomalously high resistivity of salt IV as compared with salt II. Salts I and II have even lower resistivities and lower activation energy of conductivity. This is consistent with the x-ray results, for salt IV is the only one of the four compounds studied which shows no evidence of one-dimensional stacking.

### References

1. Melby, L. R., et al., *J. Am. Chem. Soc.*, Vol. 84, p. 3374, 1962.
2. Rembaum, A., et al., *J. Phys. Chem.*, Vol. 73, p. 513, 1969.
3. Rembaum, A., *J. Polym. Sci.*, Part C, No. 29, p. 157, 1970.
4. Kepler, R. G., *Phonons and Phonons Interactions*, p. 579. Edited by T. A. Bak. W. A. Benjamin, Inc., New York, 1964.
5. Rembaum, A., Baumgartner, W., and Eisenberg, A., *J. Polym. Sci.*, Part B, No. 6, p. 169, 1968.
6. Rembaum, A., Rile, H., and Somoano, R., *J. Poly Sci.*, Part B, p. 457, 1970.

mination of this phenomenon, that is, the measuring of the degree or extent of stick-slip, under different conditions would be most desirable. This article discusses a method developed to achieve this end *without using a tape transport*.

### 2. Discussion

Stick-slip will cause a variation in the tape-to-head speed. Sticking will decrease and slipping will increase the speed of the tape passing over the magnetic heads.

<sup>1</sup>Spacecraft Data Systems Section, Astrionics Division.

If the instantaneous speed of the tape could be determined, then this would serve as a measure of the degree of stick and slip.

Instead of attempting the very difficult direct measurement of the instantaneous speed of the tape, we have measured a factor which is a manifestation of the change in speed. This factor is the apparent period of a cycle in the signal recorded on the tape. The explanation is as follows.

A sinusoidal electrical signal is recorded<sup>2</sup> on a portion of magnetic tape to be evaluated. The tape is then mounted, along with the heads in the Instron test apparatus, in the same manner as described previously (SPS 37-63, Vol. III, pp. 209-214), and as shown in Fig. 1. Electrical connections are then made to the pickup coil of the magnetic head, which is connected to an amplifier.<sup>3</sup> The signal therefrom is fed into a digital period counter and a paper tape printout.<sup>4</sup> The cycle period, in seconds, of the recorded signal can be measured accurately by oscilloscopic examination of the tape during initial recording.

If the speed of the tape passing over the heads remains constant, then the period of the recorded electrical signal will also be constant. If, however, the speed of the tape should vary because of sticking and slipping, the apparent period of the recorded signal will also vary. The extent of the variation of the period from its original value (the standard deviation  $\sigma$ ) will be a measure of the degree of stick-slip.

### 3. Experimental Procedures and Results

A length of new CEC-W4 magnetic recording tape was prerecorded with a  $9000 \pm 30$ -Hz signal at 15-in./s speed, and at saturation voltage. The tape was then placed in the Instron test apparatus at ambient temperature and humidity, and passed over 2 brass heads for a total of 40 passes at a crosshead speed of 0.166 in./s. During each pass both the force experienced by the load cell and the period of every tenth cycle of the electrical signal were recorded. There were approximately 75 cycle periods recorded for each pass, a pass being the downward journey of the 110-g weight pulling the tape over the magnetic heads.

<sup>2</sup>Model 651B test oscillator (Hewlett-Packard)/Ampex Model FR-100 magnetic tape test recorder.

<sup>3</sup>Astrodata Model 120 nanovolt amplifier.

<sup>4</sup>Model 5532A electronic counter (Hewlett-Packard)/Model H43562Z digital recorder (Hewlett-Packard).

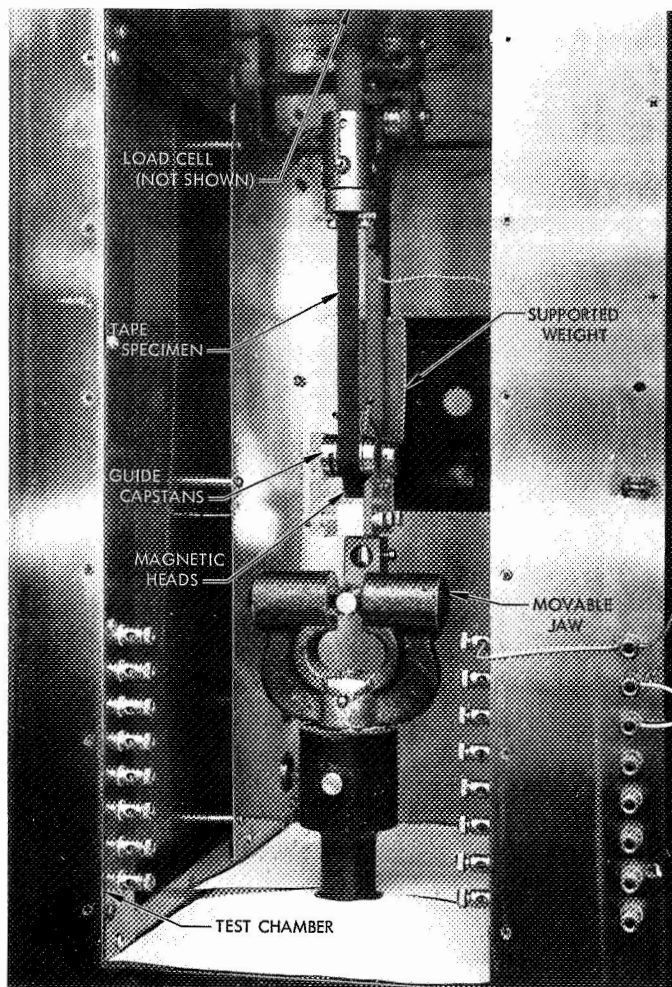


Fig. 1. Test apparatus

Frequency  $f$  is inversely proportional to period  $\pi$ :

$$f = \frac{1}{\pi} \quad (1)$$

The following relationships are assumed to hold true and should yield the standard deviation for speed:

$$\frac{f}{\sigma_f} = \frac{S_{CH}}{\sigma_s}$$

or because of Eq. (1):

$$\frac{\sigma_\pi}{\pi} = \frac{S_{CH}}{\sigma_s}$$



or

$$\sigma_s = \frac{S_{CH} \times \pi}{\sigma_\pi} \quad (2)$$

where

$\sigma_f$  = standard deviation of frequency values

$S_{CH}$  = crosshead speed

$\sigma_\pi$  = standard deviation of period values

$\sigma_s$  = standard deviation of speed values

The crosshead speed  $S_{CH}$  was held constant at 0.166 in./s and the average period was found to be 0.0100 seconds, which agreed well with the calculated value of 0.01002 seconds, obtained from the known frequency and speed:

$$\begin{aligned} f &= \frac{\text{Recorded frequency (Hz)} \times \text{Crosshead speed (in./s)}}{\text{Speed of tape at recording (in./s)}} \\ &= \frac{9 \times 10^3 \text{ Hz} \times 0.166 \text{ in./s}}{15 \text{ in./s}} = 99.6 \text{ Hz} \end{aligned}$$

Since

$$\begin{aligned} \pi &= \frac{1}{f} \\ \pi &= \frac{1}{99.6} = 0.01002 \text{ seconds} \end{aligned}$$

The  $\sigma_\pi$  was computed from the test data, and, therefore,  $\sigma_s$  could be calculated from Eq. (2).

The plot of  $\log \sigma_s$  versus the number of passes (Fig. 2) indicates how severely the standard deviation changes; therefore the stick-slip increases by the repeated passage of this tape over the magnetic heads. It must be kept in mind, however, that the worst conditions were selected

for this experiment, in order to point out the measurability of stick-slip from apparent changes in the cycle period. *The conditions used were different from those customarily associated with spacecraft tape recorder applications.* Experiments are in progress to determine  $\sigma_s$  for tapes under normal use conditions.

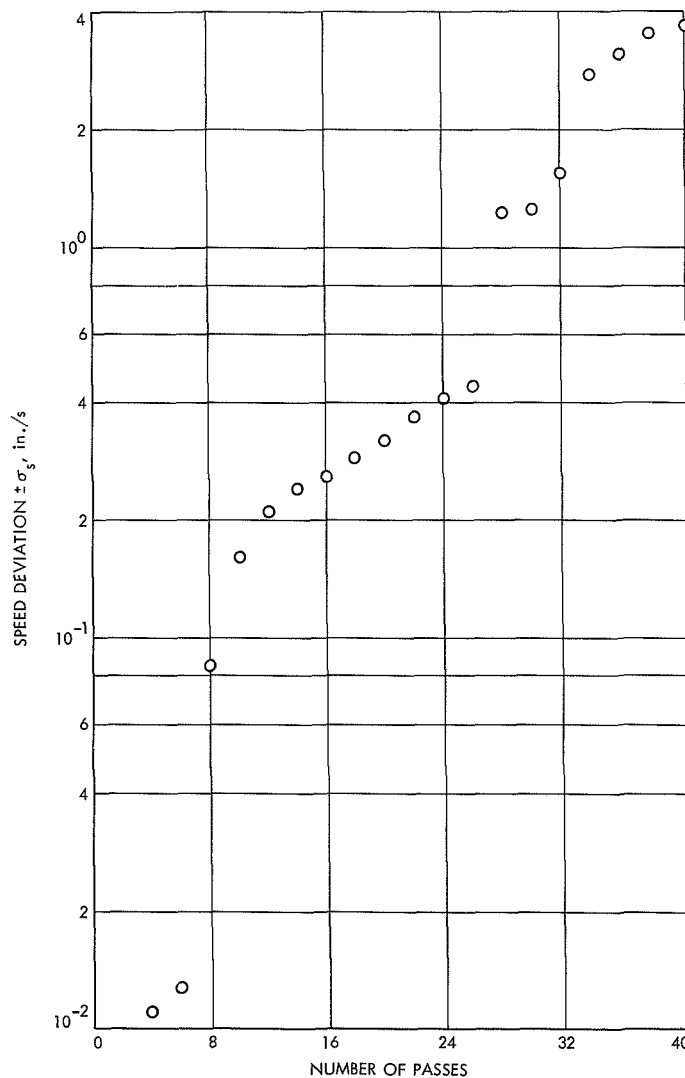


Fig. 2. Speed deviation vs number of passes

## XIX. Research and Advanced Concepts

### PROPULSION DIVISION

#### A. Ion Thruster Electron Baffle Sizing, E. V. Pawlik

##### 1. Introduction

A 20-cm-diameter electron-bombardment ion thruster of the type described in a JPL Technical Memorandum<sup>1</sup> is currently in use in the solar-electric propulsion system studies under way at JPL (Ref. 1). This thruster employs a two-grid ion accelerating geometry because of the proven reliability of this grid type and consequently does not readily obtain the low level of arc chamber losses that have been obtained by other researchers using a coated single grid (Refs. 2 and 3). In a continuing product improvement effort some optimization of the present design was attempted in order to reduce arc chamber losses. The effect of one component, the electron baffle, was studied and modifications in its geometry were found to present some performance gains.

##### 2. Apparatus

A sketch of the ion thruster used in the program is shown in Fig. 1. The nominal operating range for this thruster is 1000 to 2000 W (0.5 to 1.0-A ion beam current at a net accelerating voltage of 2000 V) of throttleable output beam power at a constant specific impulse near

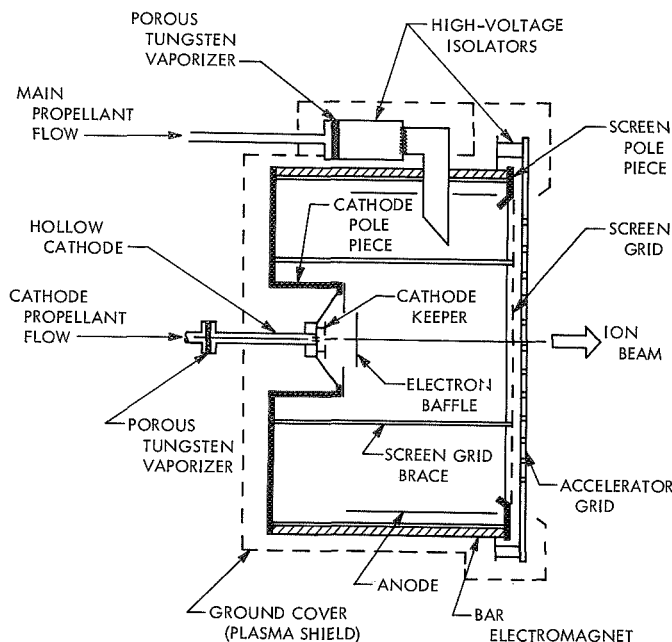
4000 s. Details of thruster geometry and performance can be found in Footnote 1. The thruster is operated in a  $3 \times 7$ -ft vacuum tank while maintaining pressures on the order of  $2 \times 10^{-6}$  to  $8 \times 10^{-6}$  torr.

The hollow cathode and the mounting arrangement within the pole piece are shown in Fig. 2. The hollow cathode is essentially a plasma device which is supplied gaseous mercury propellant from a vaporizer. It consists of a hollow tube with a 0.100-cm-diameter orifice at the end. The discharge is initiated and maintained by a keeper electrode.

An electron baffle is placed between the cathode and anode to provide a distributed source of electrons to the arc chamber. This electron baffle approach is the same as was first used in Ref. 4 and is necessary in order to minimize arc chamber power losses. The baffle also serves to distribute the mercury propellant flow from the cathode into the arc chamber, directing it radially outward from the cathode pole piece region. This radial flow direction increases the residence time of the mercury atoms within the arc chamber, consequently increasing their probability of ionization.

Four electron baffle sizes were investigated. As the different diameters were tested, the distance between the pole piece and the baffle was varied so as to maintain a constant open area of 7.2 cm<sup>2</sup>.

<sup>1</sup>Pawlik, E. V., *Performance of a 20-cm-Diameter Electron-Bombardment Hollow-Cathode Ion Thruster*, Technical Memorandum 33-468, Jet Propulsion Laboratory, Pasadena, Calif. (to be published).

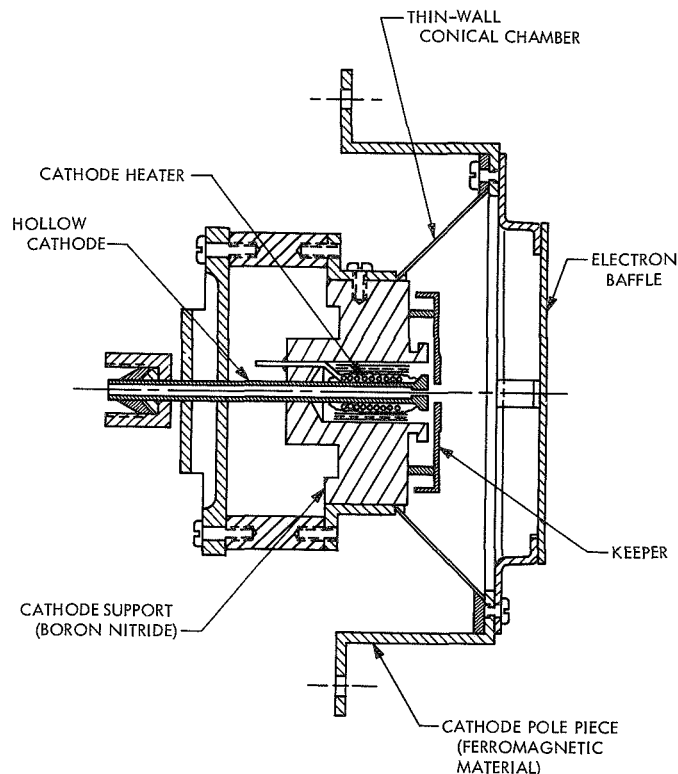


**Fig. 1. Sketch of 20-cm-diameter electron-bombardment ion thruster**

The cathode mounting was designed to minimize heat conduction to the thruster. This was accomplished by using a cone of thin stainless steel to support the boron nitride in which the cathode was held.

### 3. Results and Discussion

Arc chamber losses were measured as the thruster was operated at both the high and low levels of the nominal output power range. These losses, for a propellant utilization near 90% and an arc voltage of 35 V, are plotted in Fig. 3a. For the largest baffle tested the arc was prone to extinguishing at low output power levels. Increasing the current drawn to the keeper electrode from the nominal 0.5 to 1.0 A reduced this tendency somewhat. As the baffle diameter was reduced, the arc chamber losses were reduced, minimizing near a diameter of 4.5 cm for low-power output operation. At high thruster output power, operation with smaller baffles produced lowest arc chamber losses as the diameter was reduced but this was accompanied by extremely noisy arc chamber operation for the two smallest baffle sizes. The cause of this noise is not understood at present. The cathode flow necessary to maintain the arc chamber conditions is shown in Fig. 3b and was found to decrease with baffle diameter. It is believed that the arc voltage could be maintained with less flow as the baffle was reduced because the open area was removed further from the magnetic field at the end



**Fig. 2. Cathode pole piece details**

of the pole piece, which served as a magnetic restriction to the electron flow in a manner similar to that reported in Ref. 5. The lower cathode flow results in increased main vaporizer flow, creating a more favorable flow distribution and reduced arc chamber losses.

A baffle near 4.7 cm in diameter appears to be an optimum size inasmuch as arc chamber noise problems are avoided. The arc chamber losses are nearly constant with output power for this baffle size and therefore the penalties in throttling thruster output power over the 2:1 range of interest are minimized. Arc chamber losses for this geometry are essentially the same as reported for the thrusters using the single-coated grid.

The keeper voltage as a function of arc current is plotted in Fig. 4 for each of the four baffle sizes investigated. The voltage decreased with increasing current as would be expected for increasing plasma density. The voltage level was lower than that observed for similar hollow-cathode thrusters (Ref. 6). Because of this lower keeper voltage, it is expected that the plasma potential and subsequent cathode orifice wear should be considerably reduced at these levels. Future studies will include measurements of the rate of cathode erosion.

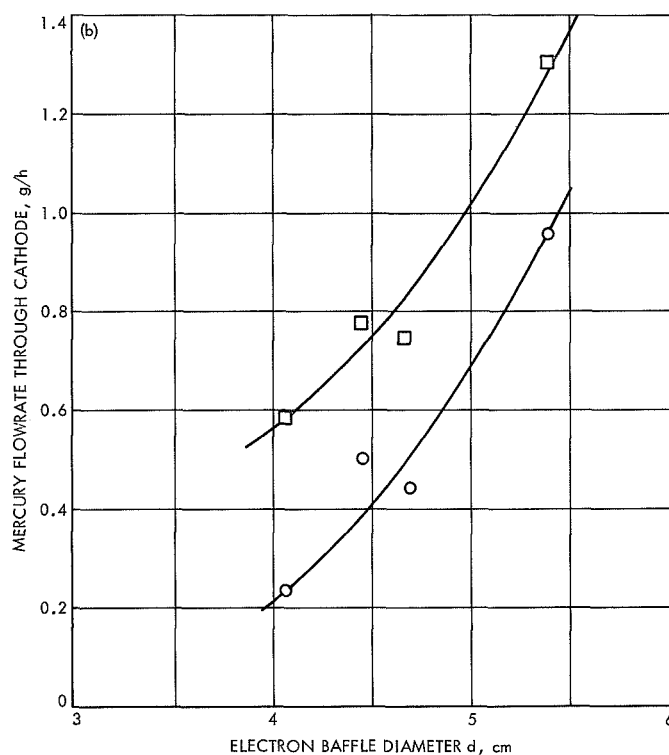
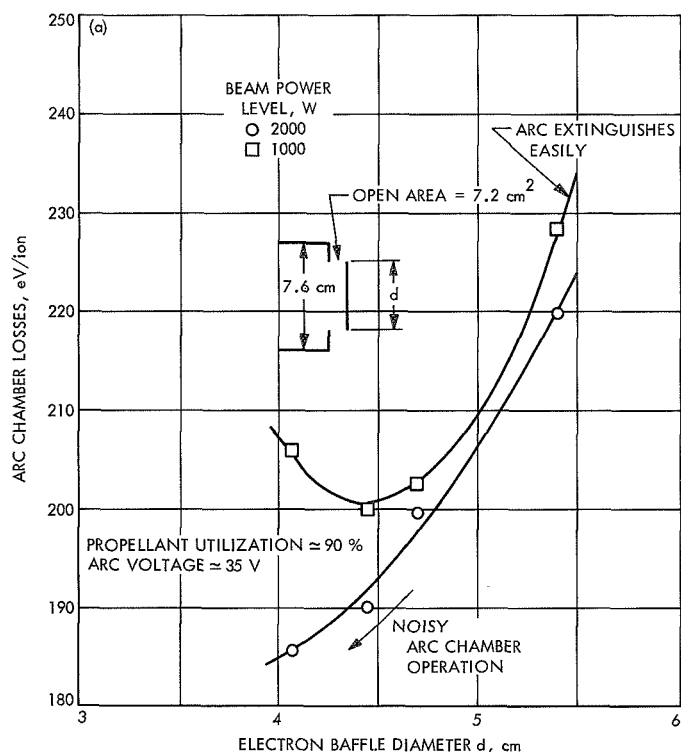


Fig. 3. Effects of electron baffle diameter on (a) arc chamber losses, (b) cathode flow

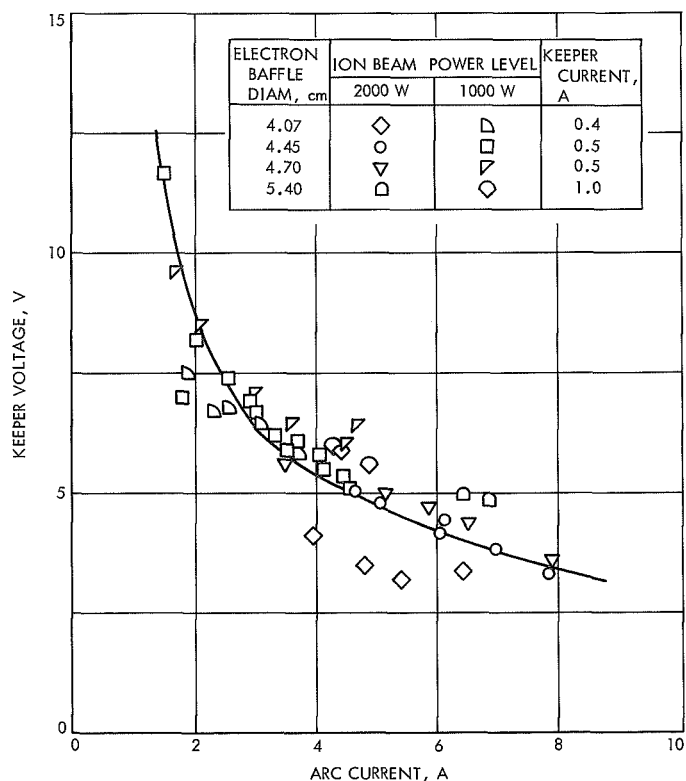


Fig. 4. Effect of arc current on keeper voltage

## References

1. Masek, T. D., "Solar Electric Propulsion System Technology," AIAA Paper 70-1153, presented at the 8th Electric Propulsion Conference, Stanford, Calif., Aug. 1970.
2. King, H. J., and Poeschel, R. L., *Low Specific Impulse Ion Engine*, Final Report, NASA CR-72677. Hughes Research Laboratories, Malibu, Calif., Feb. 1970.
3. Bechtel, R. T., "Performance and Control of a 30-Centimeter Diameter Low Impulse Kaufman Thruster," AIAA Paper 69-238, presented at the 7th Electric Propulsion Conference, Williamsburg, Va., Mar. 1969.
4. Eckhardt, W. O., et al., *4,000-hour Life Test of a Liquid-Mercury Cathode in a 20-cm LeRC Ion Thruster*, special report prepared for NASA (Contract NAS 3-6262). Hughes Research Laboratories, Oct. 1966.
5. Poeschel, R. L., and Knauer, W., "A Variable Magnetic Baffle for Hollow Cathode Thrusters," AIAA Paper 70-175, presented at the 8th Aerospace Sciences Meeting, New York, N.Y., Jan. 1970.
6. Bechtel, R. T., "Component Testing of a 30-Centimeter Diameter Electron Bombardment Thruster," AIAA Paper 70-1100, presented at the 8th Electric Propulsion Conference, Stanford, Calif., Aug. 1970.

## B. An Ion Thruster Utilizing a Combination Keeper Electrode and Electron Baffle,

E. V. Pawlik

### 1. Introduction

Intense interest has recently been focused on using solar-electric-powered spacecraft for deep space missions. Experimental systems work is therefore being conducted (Ref. 1)<sup>1</sup> at JPL using an ion thruster with a hollow-cathode electron source and employing mercury as the propellant. Some thruster improvements, such as a reduction of both cathode operating temperature and arc chamber losses, have been described in SPS 37-64, Vol. III, pp. 146-149, and in the preceding article, *Section A*. A portion of the present systems work has been directed toward experimentally examining the effects on thruster

operation of eliminating the chamber in which the cathode is customarily mounted (Refs. 2 and 3). This is of interest since the ions generated within this region are trapped within it and contribute mainly to thruster losses (Ref. 4).

### 2. Test Setup

A sketch of the 20-cm-diameter ion thruster used in the program is shown in Fig. 1. The nominal operating level for this thruster is 1000 to 2000 W (0.5 to 1.0-A ion beam current at 2000 V) of throttleable output beam power at a constant specific impulse near 4000 s. Details of thruster performance can be found in Footnote 1. The thruster is operated in a  $3 \times 7$ -ft vacuum tank while maintaining pressures on the order of  $2 \times 10^{-6}$  to  $8 \times 10^{-6}$  torr.

Two ways of mounting the hollow cathode within the thruster are shown in Fig. 2. Case I contains the conventional cathode ion chamber and Case II employs a combination keeper electrode and electron baffle. The keeper is necessary to initiate and maintain the cathode discharge. The electron baffle serves to distribute electrons into the arc chamber plasma in a manner so as to minimize arc

<sup>1</sup>Also Pawlik, E. V., *Performance of a 20-cm-Diameter Electron-Bombardment Hollow-Cathode Ion Thruster*, Technical Memorandum 33-468, Jet Propulsion Laboratory, Pasadena, Calif. (to be published).

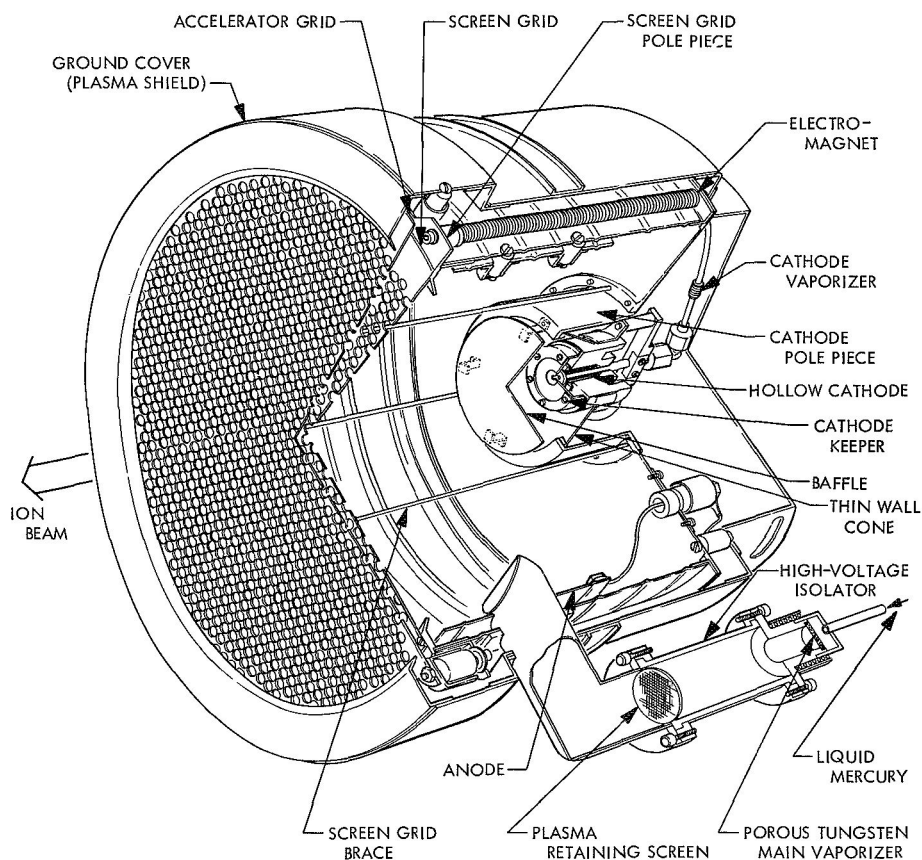
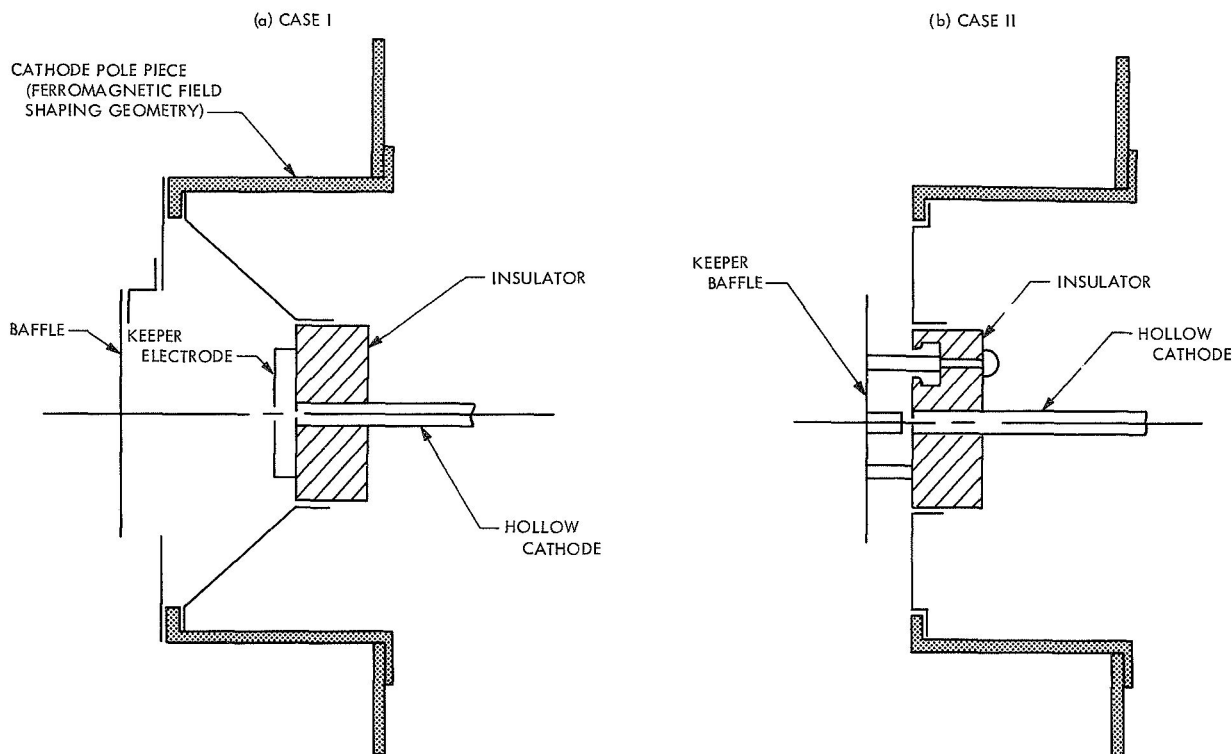


Fig. 1. 20-cm diameter electron-bombardment ion thruster

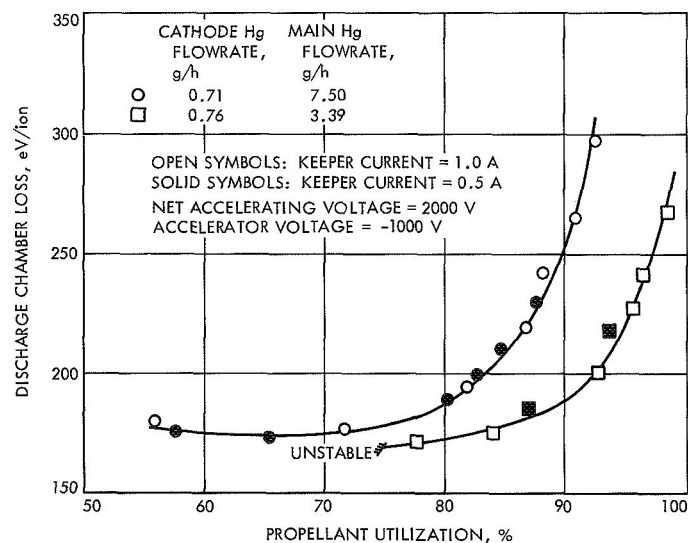


**Fig. 2. Cathode mounting details: (a) Case I, cathode mounted within pole piece; (b) Case II, cathode mounted at end of pole piece and using a combination keeper electrode and electron baffle**

chamber losses. These two items have been combined into a single unit. The cathode orifice (1.0-mm diam) and electron baffle (4.7-cm diam) used were the optimum sizes as determined in SPS 37-64, Vol. III, and the preceding article, *Section A*.

### 3. Test Results

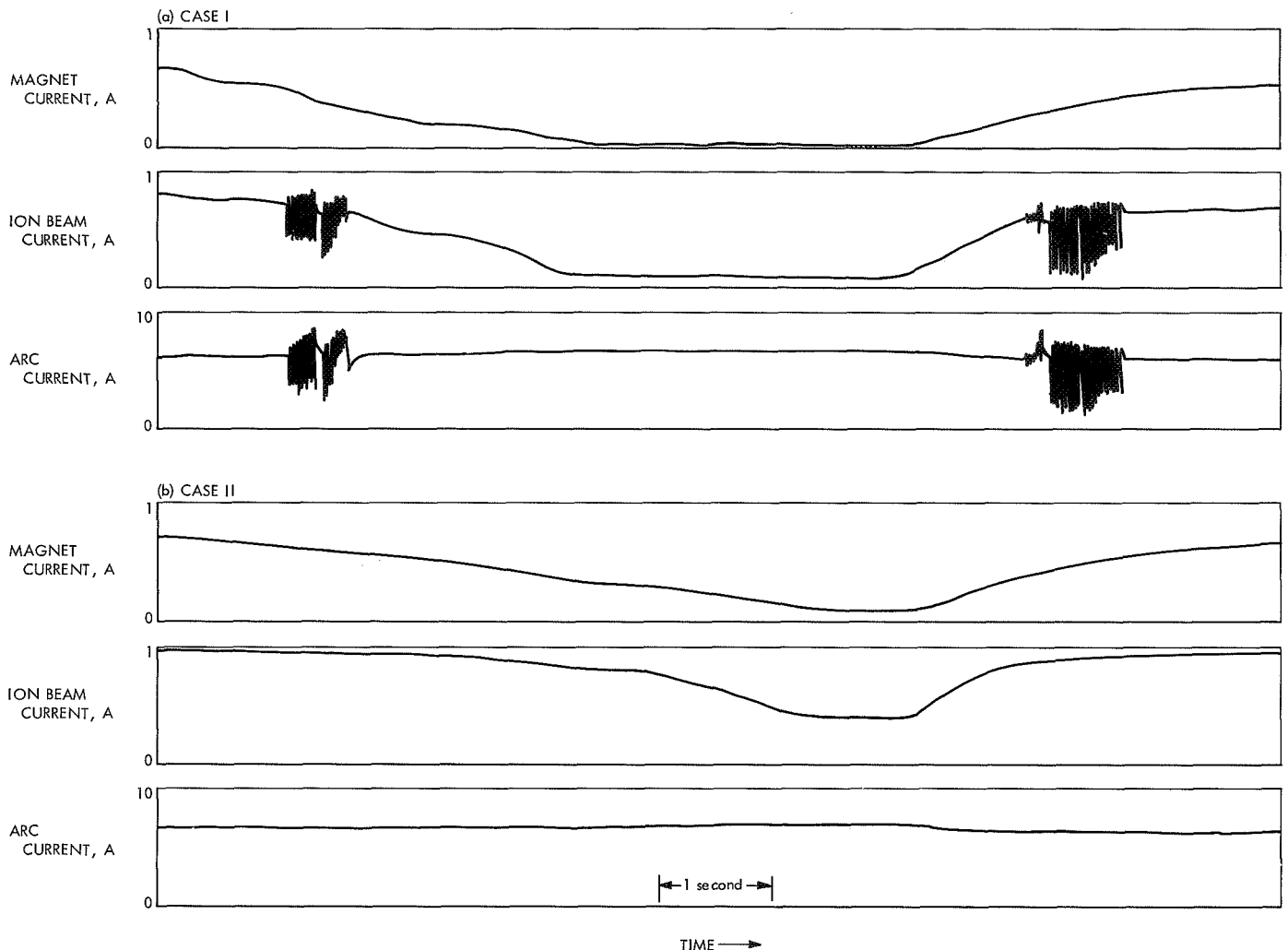
The thruster was operated at both the high and low regions of the normal operating range. The discharge chamber loss (the most significant thruster loss) is shown plotted against propellant utilization in Fig. 3. Two values of current to the keeper electrode were examined. Thruster operation was well behaved and insensitive to keeper current levels over the majority of the thruster operating range that was investigated. Some high amplitude oscillations of the arc current were found at low propellant utilizations when operating at low beam power. This instability is well outside of the nominal 90% propellant utilization operating region. The arc chamber losses were essentially the same as those obtained for this configuration shown as Case I in Fig. 2 (reference preceding article, *Section A*). This suggests that significant losses still exist within the cathode-to-keeper region or that there are additional arc chamber losses by electrons being attracted to the ion chamber side of the baffle. The fact that these



**Fig. 3. Thruster arc chamber performance**

losses do not change with keeper current levels indicates that electrons withdrawn from the arc chamber are not, however, a major loss mechanism.

An advantage that was found in using the combined keeper-baffle geometry was the elimination of arc current



**Fig. 4. Thruster response to variations in current to the electromagnets: (a) Case I, cathode mounted within pole piece; (b) Case II, keeper-baffle combination**

oscillations that have been observed at low values of the thruster magnetic field. Typical thruster operation for variations in the current for the electromagnets is shown in Fig. 4 for both cases. The absence of the oscillations for the combination keeper-baffle geometry (Case II) should result in more stable thruster operation under conditions where a reduction in magnetic field is desirable, such as during automatic operation with a power conditioning unit (Ref. 1). This unit provides the controls and proper voltages to a thruster operating from a solar array. Reduction of the magnetic field provides a convenient way to restore thruster operation after a momentary shutdown to protect the power conditioner from arcing between the grids of the ion-accelerating geometry.

Operation with the combination baffle appeared to be more insensitive to the level of current drawn to the keeper. No regions of thruster operation were observed

where it was necessary to draw high levels of keeper current in order to maintain the arc chamber discharge.

#### References

1. Pawlik, E. V., Costogue, E. N., and Schaefer, W. C., "Operation of a Lightweight Power Conditioner with a Hollow Cathode Ion Thruster," AIAA Paper 70-648, presented at the 6th Propulsion Joint Specialist Conference, San Diego, Calif., June 1970.
2. Bechtel, R. T., "Performance and Control of a 30-Centimeter Diameter Low Impulse Kaufman Thruster," AIAA Paper 69-238, presented at the 7th Electric Propulsion Conference, Williamsburg, Va., Mar. 1969.
3. King, H. J., and Poeschel, R. L., *Low Specific Impulse Ion Engine*, Final Report, NASA CR-72677. Hughes Research Laboratories, Malibu, Calif., 1970.
4. *Research on Hollow Cathodes in Mercury Ion Thrusters*, Final Report. Colorado State University, Fort Collins, Colo., Sept. 1970.

## C. Thrust Subsystem Design for Nuclear Electric Spacecraft, T. D. Masek

### 1. Introduction

Nuclear-powered, electrically propelled, unmanned spacecraft are of interest for the 1980s (Ref. 1). The purpose of this study is to provide a method for estimating the mass and dimensions of the thrust subsystem for such a nuclear electric vehicle. A Jupiter orbiter was chosen as a reference mission. A 13,600-kg spacecraft, launched to a 700-nmi orbit by a *Titan* booster, spirals to escape velocity. A similar spiral maneuver is performed to obtain the Jupiter orbit. Mission duration is approximately 2 yr with 12,000 h of full-power operation at 240 kW into the thruster array. The general guidelines used in determining thrust subsystem size are as follows:

Total conditioned power to thrusters	240 kW
True specific impulse	5000 s
Thruster redundancy	20%
Attitude control	Electric propulsion system
Maximum envelope diameter	3.05 m
Thrust duration	12,000 h
Technology	Estimated for 1980

A design based on these parameters is shown in Fig. 1. The major features are illustrated in the drawing. The principle source of design information was Ref. 2.

### 2. Thruster

A typical thruster is shown schematically in the system drawing. This is a mercury bombardment-type thruster and is similar to that described in Refs. 3 and 4. Mercury was chosen over other propellants because of the present relatively well-developed technology of mercury systems.

Calculations, based on the present SE-20C thruster design (SPS 37-51, Vol. III, pp. 124-128), result in the following equation for thruster mass:

$$m_t = 1.85 + 57D_b^2 \quad (1)$$

where

$m_t$  = individual thruster mass, kg

$D_b$  = ion beam diameter, m

If the maximum current density of the SE-20C is assumed ( $32 \text{ A/m}^2$ ), the beam diameter in meters is given by

$$D_b = \frac{6.30 \times 10^2}{I_{sp}} (\eta_m \eta_t p_t)^{1/2} \quad (2)$$

where

$p_t$  = individual thruster input power, kW

$I_{sp}$  = true specific impulse, s

$\eta_m$  = propellant utilization efficiency

$\eta_t = \eta_m \eta_p$  = total thruster efficiency

$\eta_p$  = power efficiency

Combining Eqs. (1) and (2), we find

$$m_t = 1.85 + 2.26 \times 10^7 \frac{\eta_m \eta_t p_t}{I_{sp}^2} \quad (3)$$

The choice of the number of thrusters for a constant power system is mainly a trade-off of reliability (spares) against complexity (switching), and logic and total thruster mass. A convenient number of thrusters for this power level is 37. This provides reasonable packaging and six spares. Considering Ref. 5, this number of spares seems appropriate although reliability calculations were not made here. Considering switching and power conditioning requirements (and the associated complexity), this number of spares would provide a spare thruster for groups of four or five operating thrusters. Switching, logic, and spare power conditioning (PC) units could also be grouped in this way to reduce the extremely large number of possible thruster-PC combinations. Such a grouping scheme would make each group about the size and complexity of the JPL SEPST<sup>1</sup> III system (Ref. 6).

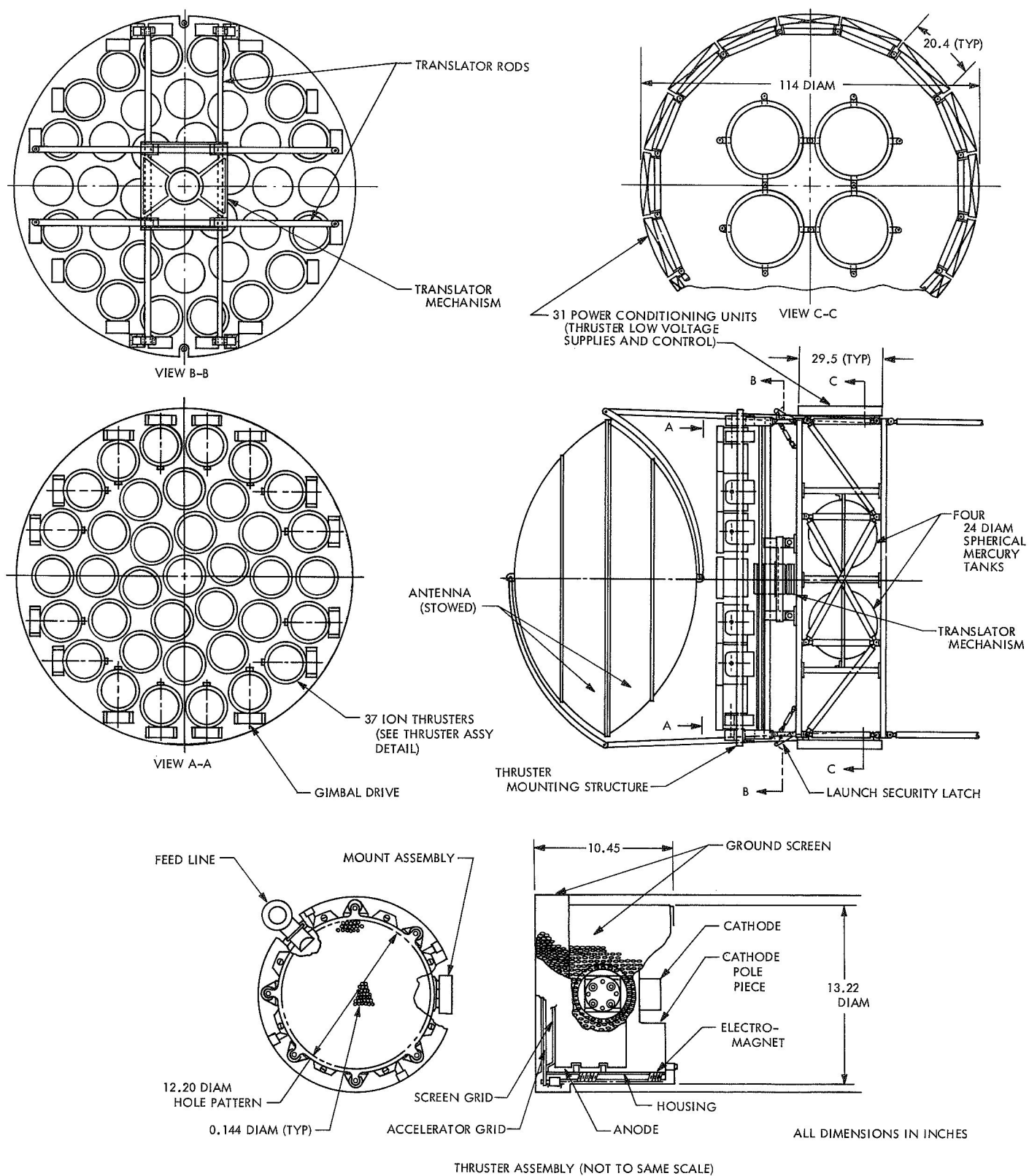
With 31 equal size operating thrusters, the input power per thruster is 7.75 kW. Using typical values of efficiency of  $\eta_m = 0.90$  and  $\eta_p = 0.92$  ( $\eta_t = 0.83$  projected from Ref. 3 at a specific impulse of 5000 s), the thruster beam diameter is about 0.31 m and the mass per thruster is approximately 7.1 kg. These values were used in the present design.

The total thruster mass and the dependence of mass on the number of thrusters (for a given power) is of interest. In general, several thruster sizes could be used in a system. In such a case, the total thruster mass is

$$M_t = \sum_i m_{t,i} n_{t,i} \quad (4)$$

<sup>1</sup>SEPST = Solar Electric Propulsion System Technology.





**Fig. 1. 240-kWe nuclear-powered electric propulsion system**

where  $m_{t,i}$  is the mass of the  $i$ th thruster and  $n_{t,i}$  is the number of  $i$ th thrusters. The total number of thrusters is given by

$$N_t = n_t + n_s$$

where

$n_t$  = number of initially operating thrusters

$n_s$  = number of initial spare thrusters

Similarly, the total thruster power is

$$P_t = \sum_i p_{t,i} n_{t,i} \quad (5)$$

where  $p_{t,i}$  is the power level of the  $i$ th thruster. Using Eqs. (3) and (5), the total mass is given by

$$M_t = 1.85(n_t + n_s) + 2.26 \times 10^7 \frac{\eta_m \eta_t}{I_{sp}^2} \left( P_t + \sum_i p_{t,i} n_{s,i} \right) \quad (6)$$

If all thrusters are the same size, Eq. (6) becomes

$$M_t = 1.85(n_t + n_s) + 2.26 \times 10^7 \frac{\eta_m \eta_t P_t}{I_{sp}^2} \left( 1 + \frac{n_s}{n_t} \right)$$

or

$$M_t = \left( 1.85n_t + 2.26 \times 10^7 \frac{\eta_m \eta_t P_t}{I_{sp}^2} \right) \left( 1 + \frac{n_s}{n_t} \right) \quad (7)$$

For a fixed percentage redundancy, i.e., a constant value of  $1 + (n_s/n_t)$ , the trade-off of system weight versus thruster number is 1.85 kg per operating thruster. In the present case (with  $\eta_m = 0.9$  and  $\eta_t = 0.83$ ), the total mass is about 266 kg. A change in  $n_t$  of  $\frac{1}{3}$  ( $\Delta n_t = 10$ ) results in a change in  $M_t$  of about 8%. Thus, the total thruster mass is not too sensitive to the choice of  $n_t$ .

As a first choice, equal size thrusters were chosen. This simplifies thruster and PC development. However, if the system detail design showed advantages in packaging or reliability with multiple sizes, the mass trade-off is small. In addition, the grouping scheme mentioned previously would minimize the effect of the reduced interchangeability with multiple sizes.

The beam voltage is determined by specific impulse and propellant utilization and is given approximately by

$$V_b = 10^{-4} \frac{I_{sp}^2}{\eta_m} \quad (8)$$

For  $I_{sp} = 5000$  s and  $\eta_m = 0.9$ ,  $V_b = 3090$  V, and  $\eta_p = 0.92$ . Hence, for the 7.75-kW thruster, the beam current is about 2.3 A. This corresponds to a propellant flowrate of about  $5.33 \times 10^{-4}$  g/s.

### 3. Power Conditioning

Typical PC units are shown in Fig. 1. These were sized assuming 90% thruster power efficiency and that the units shown provide only the low power supplies and thruster controls. The beam power is assumed to be provided by a PC unit near the reactor. The area of the PC units shown assumes a panel temperature of 25°C and a conversion efficiency of about 90% (Ref. 5).

As an estimate of the PC mass, the specific mass is taken as 3.0 kg/kW of input (0.775 kW at 250 Vdc) power (Ref. 5). Thus, the mass per PC unit is about 2.3 kg.

### 4. Thrust Vector Control

A typical thrust vector control (TVC) design is illustrated in Fig. 1. This design provides three-axis attitude control (two-axis translation, one-axis gimbal). Calculations were not made to determine the needed translation distances or gimbaling angles. Such calculations depend upon the spacecraft design. Thus, the weight and size of the mechanical portion of the TVC depend on the complete design.

As a first approximation, the TVC system shown in the system drawing was chosen. This design was based mainly on the 3.05-m envelope constraint. Assuming a translator actuator size, the translation distance is fixed. In this case, the translation distance along the translation axis is about 0.76 m.

An alternate approach to attitude control may be through the use of differential thruster throttling combined with gimbaling. The large number of operating thrusters could possibly provide adequate control. However, multiple-axis coupling might require a substantial amount of logic for choosing the correct thrusters for throttling. Study would be needed to assess feasibility.

## 5. Switching Matrix

Switches are estimated to weigh about 2.7 kg per PC unit assuming six to eight-position switches are used. Such switches are consistent with the grouping scheme discussed previously. Thus, with 37 PC units, the switch weight would be about 220 lb. The switching logic is estimated to be about 13.6 kg. Each switch can be expected to be 3 or 4 in. in diameter and about 6 in. long.

## 6. Propellant Storage and Distribution

An estimate for the size of this element can be obtained by assuming the mass to be a fraction of the propellant mass. The total propellant flowrate during operation, defined in unconventional but more useful dimensions of kg/h, is

$$\begin{aligned}\dot{M}_p &= \frac{0.75 \times 10^{-2} I_b}{\eta_m} \\ &= \frac{0.75 \times 10^{-2} \eta_p P_t}{\eta_m V_b}\end{aligned}$$

or

$$\dot{M}_p = \frac{7.5 \times 10^4 \eta_c P_t}{I_{sp}^2} \quad (9)$$

where

$I_b$  = total beam current, A

$V_b$  = net beam voltage, kV

The total propellant used is

$$M_p = 7.5 \times 10^4 \int_{t_i}^{t_f} \frac{\eta_t P_t}{I_{sp}^2} dt \quad (10)$$

where  $t_i$  and  $t_f$  are propulsion start and finish times, respectively. If we now relate  $M_p$  to the propellant storage and distribution (PSD) mass  $M_{ps}$  by

$$M_{ps} = \gamma M_p$$

we find

$$M_{ps} = 7.5 \times 10^4 \gamma \int_{t_i}^{t_f} \frac{\eta_t}{I_{sp}^2} P_t dt \quad (11)$$

The factor  $\gamma$  (approximately 0.03) includes tankage, valves, feed lines, and miscellaneous fittings. Assuming that  $\eta_t$ ,  $P_t$ , and  $I_{sp}$  do not vary with time,

$$M_p = 7.5 \times 10^4 \frac{n_t P_t}{I_{sp}^2} \Delta t \quad (12)$$

For a 12,000-h mission,  $M_p = 7200$  kg. With a 10% contingency for operation off the desired utilization efficiency, the total propellant mass would be approximately 7920 kg. Using this, the PSD mass is about 240 kg.

The element masses are summarized in Table 1. Within the accuracy of the present estimates, a specific mass of 4 kg/kW appears to be reasonable.

Table 1. Element mass summary

Element	Unit mass, kg	Total, kg
Thruster (37) (including neutralizer)	7.1	263
Power conditioning (31)	2.3	71
Thrust vector control:		
Gimbals (16)	1.4	22
Translator	23.0	23
Structure	180.0	180
Electronics	23.0	23
Switching:		
Switches (31)	2.7	84
Electronics	13.6	13.6
Propellant storage and distribution	240.0	240
Miscellaneous (wiring, launch caging, adapters, etc.)	50.0	50
		<u>969.6</u>

## References

1. Mondt, J. F., and Prickett, W. Z., "External-Fuel and Flashlight Thermionic Reactor Potential for Electric Propulsion," AIAA Paper 70-644, presented at the 6th Propulsion Joint Specialist Conference, San Diego, Calif., June 1970.
2. Masek, T. D., "Sizing a Solar Electric Thrust Subsystem," Technical Report 32-1504, Jet Propulsion Laboratory, Pasadena, Calif., Nov. 15, 1970.
3. Masek, T. D., and Pawlik, E. V., "Thrust System Technology for Solar Electric Propulsion," AIAA Paper 68-541, presented at the 4th Propulsion Joint Specialist Conference, Cleveland, Ohio, June 1968.
4. Masek, T. D., and Womack, J. R., "Experimental Studies with a Mercury Bombardment Ion Engine System," AIAA Paper 67-698, presented at the AIAA Electric Propulsion and Plasma Dynamics Conference, Colorado Springs, Colo., Sept. 1967. Also available as Technical Report 32-1280, Jet Propulsion Laboratory, Pasadena, Calif., July 15, 1968.
5. *Solar Powered Electric Propulsion Spacecraft Study*, Final Report SSD-50094-R, Hughes Aircraft Company, Culver City, Calif., Dec. 1965.
6. Masek, T. D., and Macie, T. W., "Solar Electric Propulsion System Technology," AIAA Paper 70-1153, presented at the 8th Electric Propulsion Conference, Stanford, Calif., Aug. 1970.

## D. Ion Thruster Connectors, E. T. Hopper

### 1. Introduction

Solar electric propulsion system tests are now being conducted (Refs. 1 and 2) that involve integration of a power conditioner with an ion thruster. A small but important component in the system that has been of concern in flight tests (Ref. 3) is the single cable connector used to transfer power to the thruster. In the use of commercially available connectors, it was found that arcing occurred between disconnect pins and leads due to proximity of high-potential and low-potential lines. This resulted in the breakdown of components and of the systems life tests.

Arcing also occurred when high-pressure areas developed in the connector due to poor pump-out conditions or out-gassing of components under normal operational testing.

Failures experienced while using a conventional disconnect prompted development of a design using an approach similar to that used in Ref. 4. Successful operation has been accomplished with no breakdowns over more than 200 h of operation.

### 2. Apparatus

The connector was used during the normal integration tests of a 20-cm thruster (Fig. 1) with a power conditioner system. Table 1 shows the voltages and grouping in-

volved. The connector selected (MS 3102A-28-21P) was a standard production type, modified as follows:

- (1) All aluminum parts were chemically stripped of zinc chromate, cadmium, and copper.
- (2) The plastic insulator parts were removed and replaced with boron nitride (Fig. 2).

Table 1. Voltage and groups

Voltage	Group
High voltage (+ 2000 V)	Magnet Anode Cathode vaporizer Cathode vaporizer thermocouple Cathode keeper Cathode heater Common
Low voltage (ground potential voltage)	Main vaporizer Main vaporizer thermocouple Neutralizer vaporizer Neutralizer cathode Neutralizer keeper Neutralizer common Neutralizer vaporizer thermocouple
Accelerator high voltage (- 1000 V)	Accelerator grid

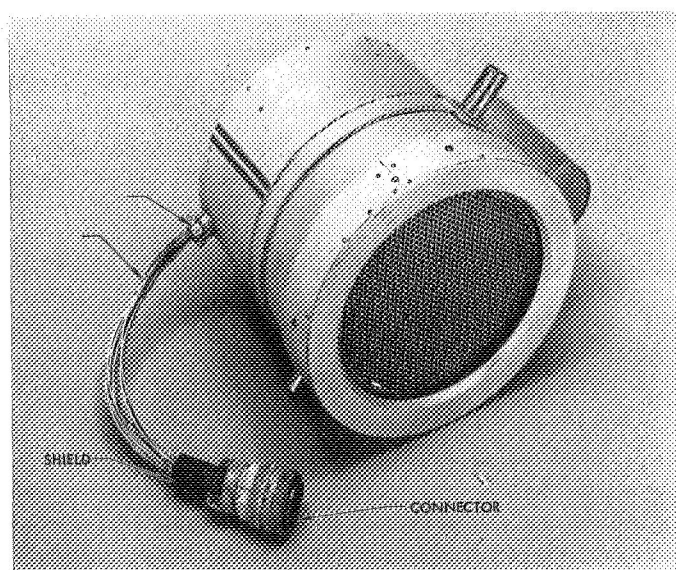


Fig. 1. 20-cm-diam ion thruster and cable connector

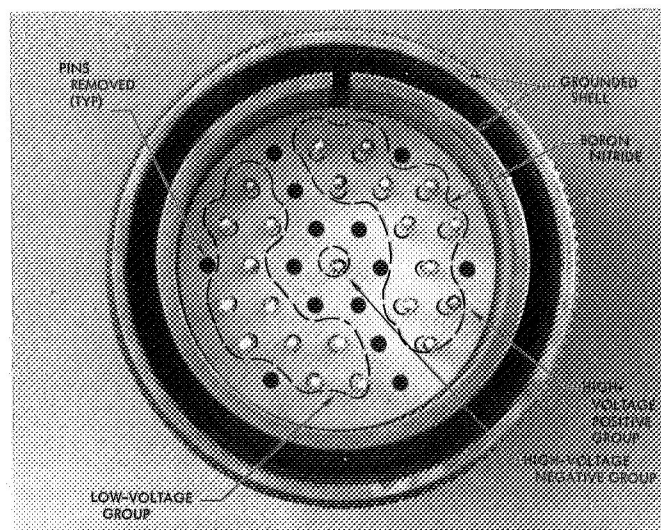


Fig. 2. Internal view of connector, showing various voltage groups

- (3) Connector pins (Fig. 2) were removed in a pattern so as to leave the remaining pins in three separated groups, each group having a different potential. The removal of pins allowed a free pump-out path in all areas of the connector to avoid local high pressures. The removal pattern also allowed much longer creepage paths between different potential groups.

In addition, special precautions were instituted to minimize interwire breakdowns in the cable. Power and thermocouple leads both have Kapton Conformal-wrapped insulation. All leads were also sheathed with Teflon sleeving at the thruster feedthrough clamp, while, at the connector, the conventional plastic insulator sleeve was used as a shadow shield against back sputtered vacuum tank material and left unclamped to facilitate faster pump-out.

Connector pins were silver-soldered into the pins and all flux removed. All sharp edges and points were then smoothed to large radii.

### 3. Test Results

The connector was bench checked to an applied voltage of 5 kV between voltage groups at ambient temperature and atmospheric pressure. Less than  $0.1 \mu\text{A}$  leakage was noted in this test. The connector was then installed in the thruster system with the connector body electrically grounded. The portion of the thruster system located within the vacuum tank (i.e., feedthroughs, thruster connector, thruster, insulators, wiring) was high-voltage checked at atmospheric conditions prior to installation within the vacuum chamber. At 3.5 kV, less than  $0.5 \mu\text{A}$  leakage existed within this system.

The thruster was operated in 10- to 20-h intervals as part of a test program involving internal thruster geometry changes. At the end of each test interval the con-

necter was examined and the system was retested as part of the normal thruster checkout procedure. No appreciable deposits of metal were noted at the rear of the connector when the protective shadow shield was removed. Thruster checkout including cable and connector was performed after each test operation of the thruster system and the cable and connectors were examined for breakdowns. No damage to the insulation was observed. The creepage paths between voltage groups were free from contamination and arc trails. High-voltage checking of the system showed no deviation from the original test setup.

### 4. Conclusions

The connector tests indicate its applicability for use in a systems life test of 1000 to 10,000 h.

The same configuration could possibly be considered for flight applications if  $\text{Al}_2\text{O}_3$  were used in place of boron nitride as the insulator material and stainless steel in place of the aluminum shell. Such a part is commercially available.

### References

1. Masek, T. D., and Macie, T., *Solar Electric Propulsion System Technology*, AIAA Paper 70-1158, presented at the 8th Electric Propulsion Conference, Stanford, Calif., Aug. 1970.
2. Pawlik, E. G., Costogue, E. N., and Schaefer, W. C., *Operation of a Lightweight Power Conditioner With a Hollow Cathode Ion Thruster*, AIAA Paper 70-648, presented at the AIAA 6th Propulsion Joint Specialist Conference, June 1970.
3. Rulis, R. J., *Design Considerations and Requirements for Integrating an Electric Propulsion System Into the SERT II and Future Spacecraft*, AIAA Paper presented at the 8th Electric Propulsion Conference, Stanford, Calif., Aug. 1970.
4. Paul, F. W., and Burrow Bridge, D., "The Prevention of Electrical Breakdown in Spacecraft," NASA SP-208, Goddard Space Flight Center, Greenbelt, Md., 1969.

## XX. Liquid Propulsion

### PROPULSION DIVISION

#### A. High-Thrust Throttleable Monopropellant Hydrazine Reactors, T. W. Price

##### 1. Introduction

Studies of rendezvous and landing missions have shown advantages for throttleable monopropellant hydrazine rocket engines to provide the terminal velocity correction. These thrusters would be used in much the same manner as the bipropellant vernier engines on JPL's *Surveyor* spacecraft. A requirement for a rocket with maximum thrust of 500 to 1000 lbf appears in several conceptual missions for unmanned spacecraft, the most immediate of which is the *Viking* Mars lander.

There are three primary factors which led to the choice of hydrazine, and all are related to the extra-terrestrial life experiments planned for the landing missions. These are: (1) the stagnation temperature of the hydrazine decomposition products is much lower than that from a bipropellant engine (2000°F compared to 5000°F for the  $N_2H_4-N_2O_4$  system) and this reduces the likelihood of destroying soil organisms in the vicinity of the landing site, (2) heat sterilization of the complete propulsion system, including the fuel, before launch is

feasible (SPS 37-47, Vol. III, pp. 153-162), and (3) the exhaust gases contain only small amounts of the two species<sup>1</sup>, carbon and water, of most interest to the life scientists. The inherent, superior reliability of monopropellant propulsion systems, as compared to bipropellants, further enhances their desirability in spite of the lower specific impulse of the monopropellant.

Among the problem areas associated with large monopropellant reactors of this type, some would be expected, at least *a priori*, to be similar to those experienced with the *Mariner*-size engines. Chief among these would be catalyst retention, structural support of the catalyst and adequate distribution of the fuel over the catalyst bed. Lander mission requirements, such as sterilization and dynamic throttling, may pose additional problems. Although none of these problems appears insurmountable, only a limited amount of experience with either high thrust or throttled hydrazine reactors has been accumulated to date.

---

<sup>1</sup>Both are present as impurities (the carbon as aniline) and could be almost completely removed, if necessary.

Therefore, an applied research program to investigate the broad technology area of high-thrust throttleable monopropellant hydrazine engines is in progress. This program was divided into two major parts, both under JPL technical direction: a contracted effort with TRW Systems Group and an in-house effort at JPL. The contracted program has been completed. The results are the subject of a formal report (Ref. 1) prepared by TRW and will not be discussed here. The initial phase of the in-house program, two series of subscale tests, has been completed and is the subject of this article. For these tests, surplus *Mariner* Mars 1969 reactors, which provide 50 lbf of thrust in vacuum, were used. The subscale tests were directed at a fuller understanding of the effects of dynamic throttling and sterilization. Each series is discussed separately.

## 2. Dynamic Throttling Tests

**a. Objectives.** The primary objectives of this group of tests were to assess the problems involved in controlling a subscale throttle valve/reactor chamber by means of a signal pre-recorded on magnetic tape (in preparation for test with larger chambers) and to determine if there were any fundamental problems involved in dynamically throttling a hydrazine decomposition chamber. In addition, the subscale hardware used was of interest in itself for potential use on one version of a contemplated Jupiter orbiter spacecraft. Therefore, secondary objectives were the evaluation of the dynamic characteristics of the specific throttle valve/reactor combination and of the performance of the catalytic reactor at various throttle levels.

**b. Apparatus and procedures.** For the tests to be described, thrust chambers from the *Mariner* Mars 1969 program were mated with a throttle valve from the *Surveyor* program (Ref. 2). Surplus components were used throughout because of cost and availability considerations. The basic *Mariner* Mars 1969 engine configuration is shown in Fig. 1. An engine/valve combination, as set up for testing, is shown in Fig. 2.

Two comments are pertinent to the setup of Fig. 2. First, the hardware was originally assembled and supplied to the Langley Research Center for a series of soil contamination tests. The piping appears somewhat cramped because it was arranged so that the engine/valve assembly could be test-fired at JPL, packaged, and shipped with a minimum of disassembly. Second, the *Surveyor* propulsion system utilized both a fuel and an oxidizer; consequently, the throttle valve consists of two

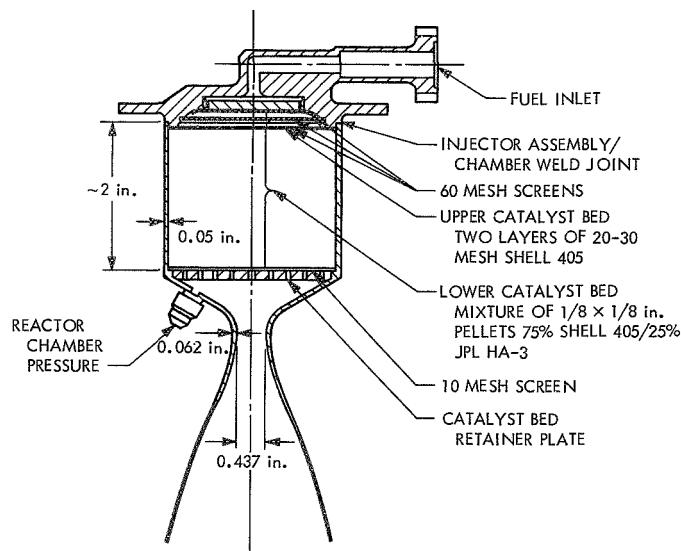


Fig. 1. *Mariner* Mars 1969 engine, internal configuration

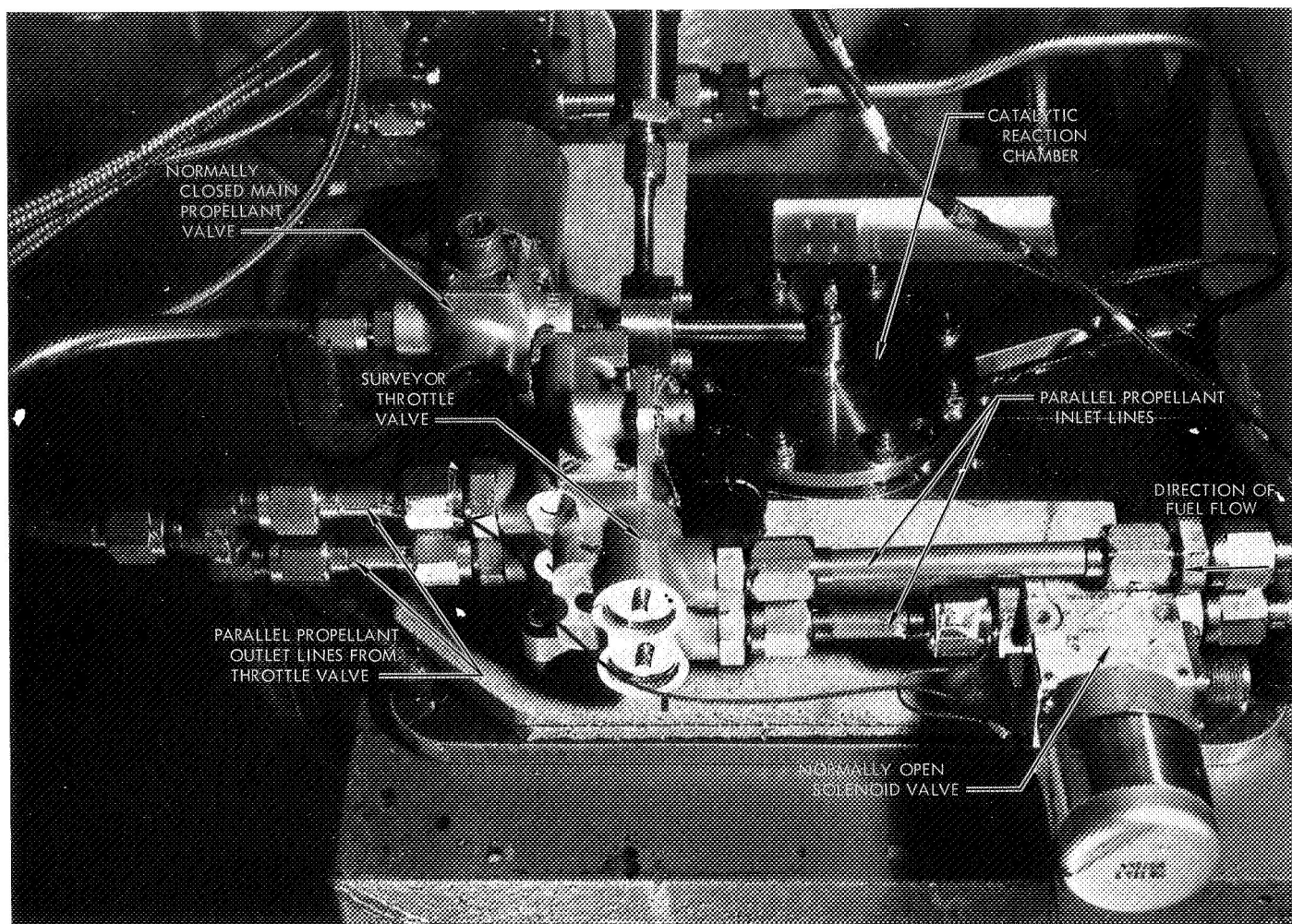
variable area fluid systems, with both the fuel and oxidizer pintles being positioned by a single actuator. Neither of the two valve circuits by itself can accommodate the design flowrate (0.22 lbm/s) of the monopropellant reactor without an excessive pressure drop, but both sides in parallel are adequate. Therefore, the single hydrazine supply line is divided just upstream of the *Surveyor* valve (Fig. 2), hydrazine is routed to both parts of the valve, and the two lines are rejoined downstream of the valve. In addition, a normally open solenoid valve was placed in one of the two parallel lines so that flow through one side of the throttle valve could be remotely closed off. Thus, the range of flow rates possible with a single supply tank pressure was increased.

Two different reactors, of basically the same design, were used. For tests 4-6<sup>2</sup> and 23-26 engine SN C-101 was used. Engine SN B was employed for all other tests in this series. The significant dimensions and catalyst bed configuration were the same for both engines and are as shown in Fig. 1. The major difference between the two is that the injector of SN C-101 bolts to the chamber instead of being welded.

During the initial tests (numbers 4-6) the throttle valve position was controlled manually by means of a

<sup>2</sup>All tests are numbered sequentially regardless of the hardware being used. Thus, the first tests of these series do not begin with 1, and test numbers for the same hardware do not necessarily follow sequentially.





**Fig. 2. Surveyor throttle valve for *Mariner Mars 1969* engine ready for test**



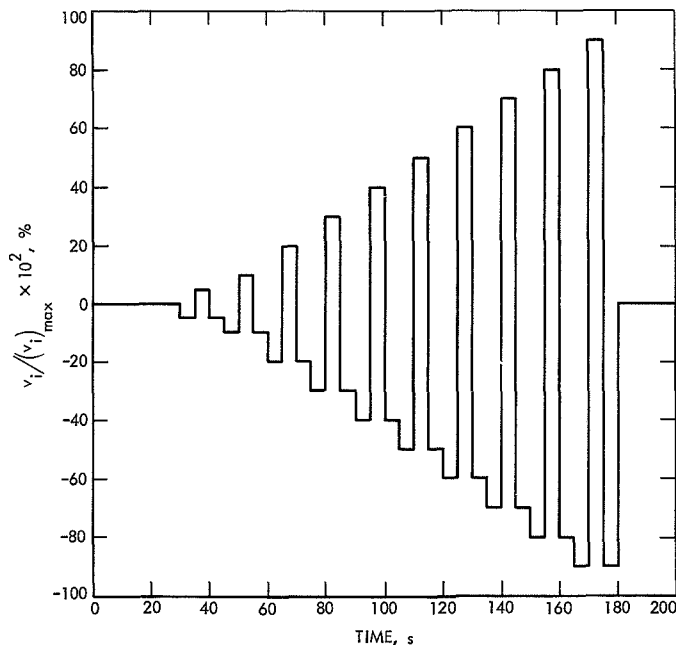


Fig. 3. Square wave duty cycle for throttle valve tests

potentiometer. For the remaining tests (numbers 23–26 and 66–68) throttle valve control was from magnetic tape. The two duty cycles used for remote control of the throttle valve are shown in Figs. 3 and 4. The signals are expressed as the ratio of valve input voltage to the value of input voltage corresponding to the maximum travel in either direction about the null point. The throttle valve itself is designed so that at zero input voltage (i.e., the null point) the pintles are approximately halfway between the travel limits. Square and sinusoidal (two amplitudes and five frequencies about four different mean levels) signals were employed even though the idealized duty cycles of Figs. 3 and 4 would most likely not be employed during a mission. Rather, the response to such signals defines the system dynamic characteristics and provides data directly applicable to the design of the necessary control circuitry.

In addition to throttle valve position, the hydrazine supply tank pressure is another variable by which the absolute engine output level can be varied. The tank pressure was chosen so that with the valve in the null position the reactor pressure would be slightly under the nominal *Mariner* Mars 1969 operating pressure of 200 psia. Since one-half the valve stroke is above the null point, the fuel flow rate could be increased above that corresponding to the null position, and the engines operated at pressures above, as well as below, the design value. The *Mariner* Mars 1969 thrust chamber can be

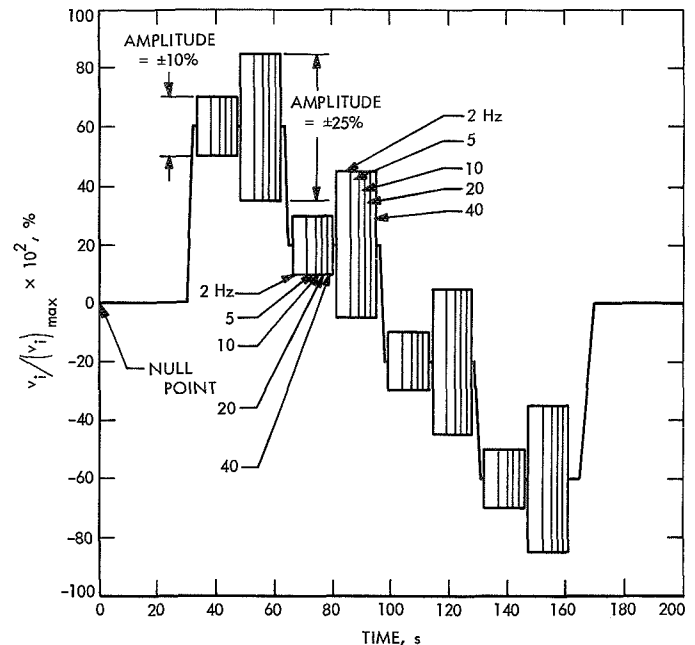


Fig. 4. Sinusoidal duty cycle

operated at 1.5 times the design pressures, so that such throttling was still well within the engine's capabilities.

**c. Results and Discussion.** The primary objectives of this test series were attained. Dynamic control of a throttle valve from a pre-recorded signal was demonstrated, and no problems in the rapid throttling of a hydrazine decomposition chamber were encountered. The response of the engine and valve to the step and sinusoidal signals was measured, but since the valve has no positive position indicator, the characteristics of the chamber cannot be separated from those of the valve, and the data are applicable only to the specific valve/chamber combination tested.

The planned test sequence for the square-wave input of Fig. 3 was to repeat this duty cycle three times without an engine shutdown between repetitions. This was chosen as grossly representative of the duty cycle for the contemplated Jupiter orbiter spacecraft. Because of a minor facility problem a shutdown was required just before completion of the first cycle. (Thus, test 66 was slightly shorter than either tests 67 or 68.) After correction of the problem, tests 67 and 68 were completed as originally planned.

The results of the step input tests (66, 67, and 68) are presented in Figs. 5 through 8. Figures 5 and 6 are tracings from an oscillogram and show behavior during

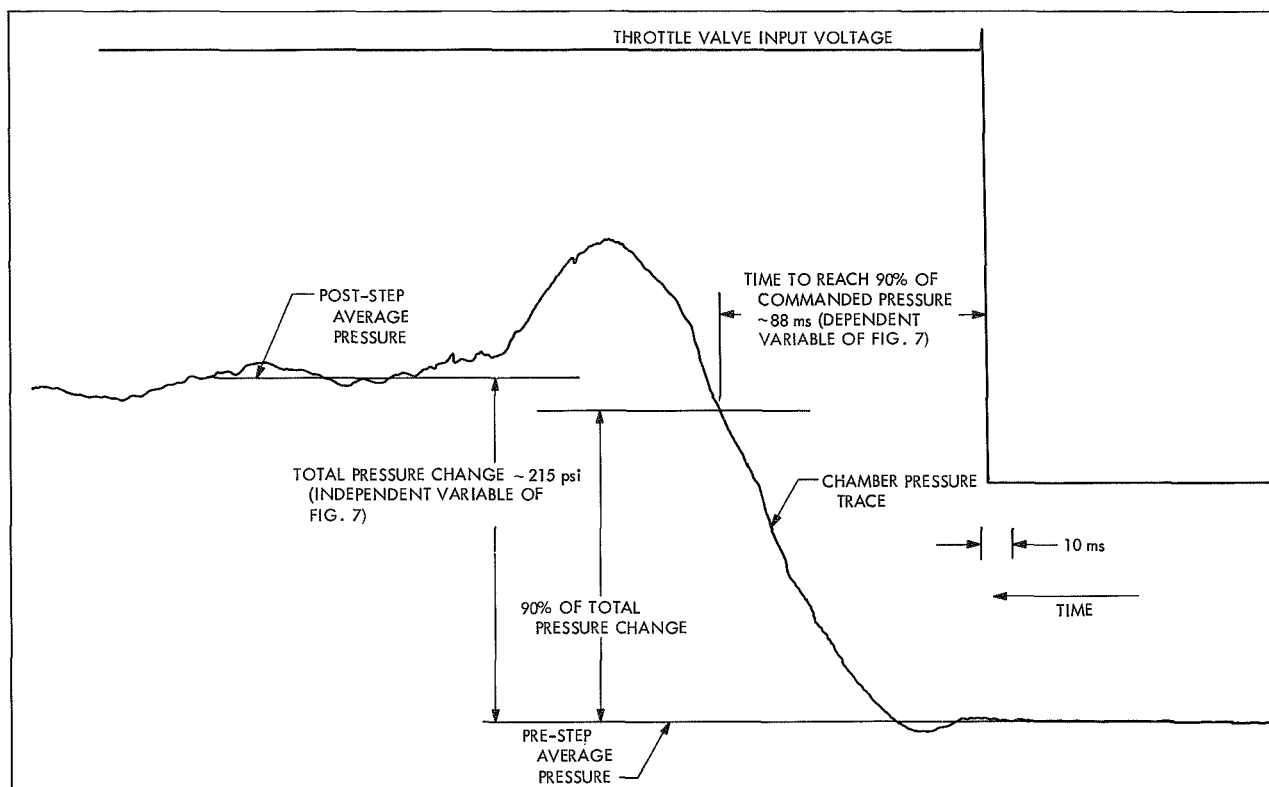


Fig. 5. Typical "up-throttle" step

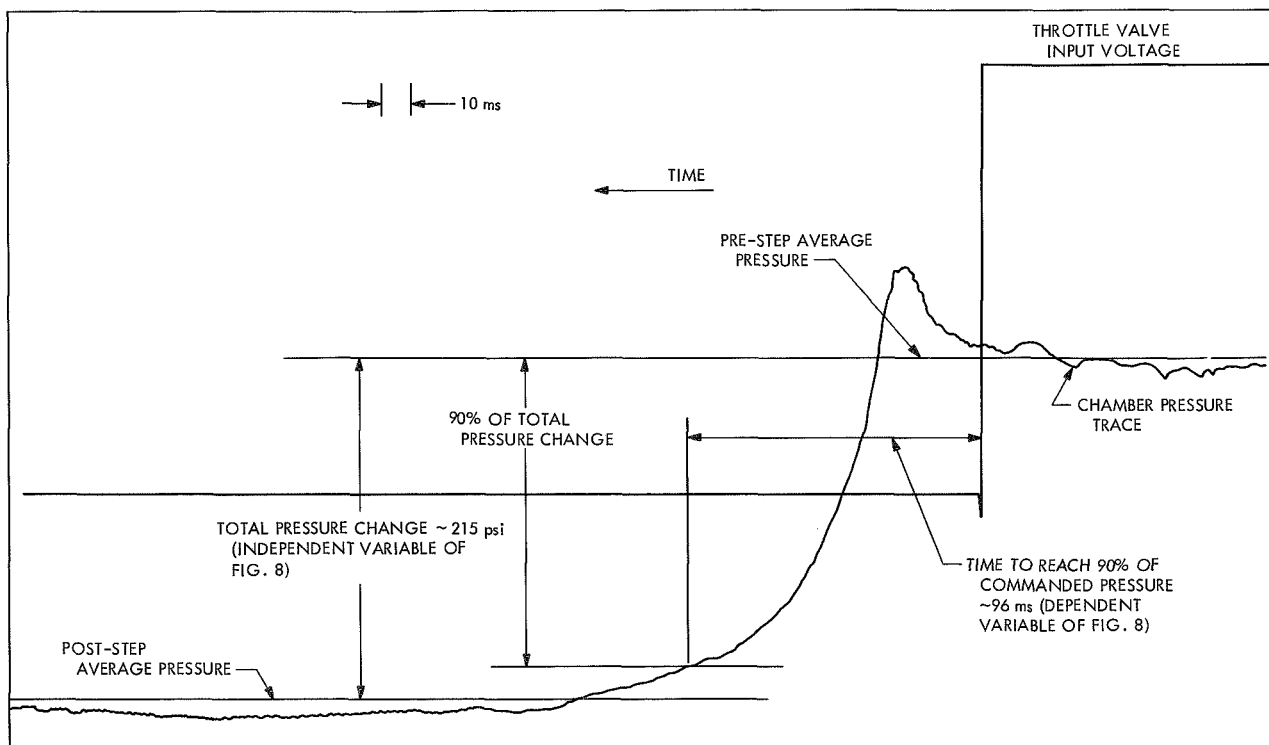


Fig. 6. Typical "down-throttle" step

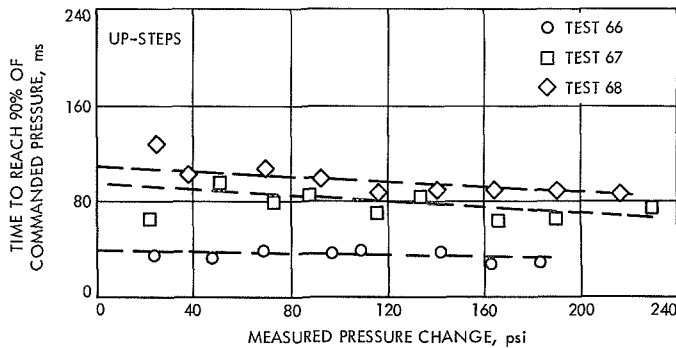


Fig. 7. Response to pressure increase

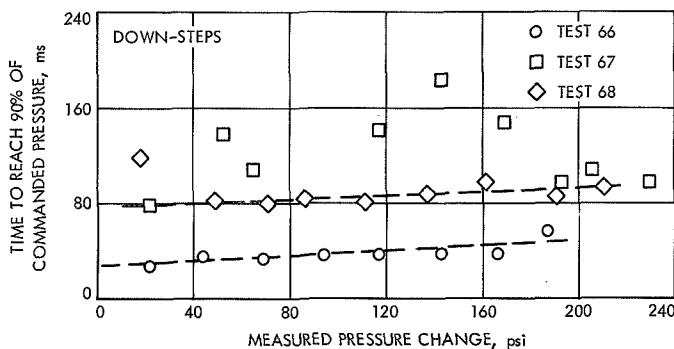


Fig. 8. Response to pressure decrease

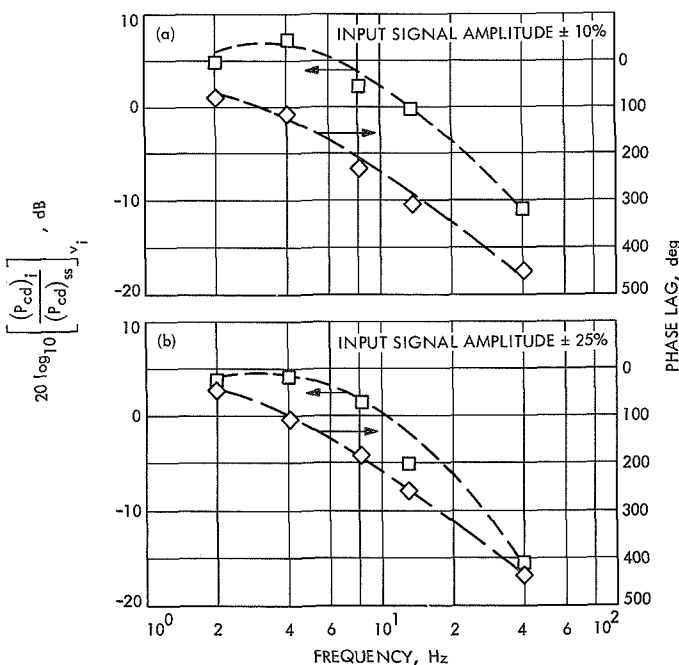


Fig. 9. Amplitude ratio and phase lag versus frequency input signal: (a) amplitude  $\pm 10\%$ , (b) amplitude  $\pm 25\%$

a typical "up-step" and "down-step," respectively. Also defined on Figs. 5 and 6 are the total chamber pressure change resulting from the step input, and the time required to reach 90% of this value. These two variables are correlated for all three tests in Figs. 7 and 8.

Figures 7 and 8 are plots of the engine/valve response time versus the pressure change. As used here the response time is defined as the time increment from a change in the valve command signal until the chamber pressure reaches 90% of the final steady-state value (Figs. 5 and 6). The pressure change is simply the difference between the initial and final steady-state reactor chamber pressures. The straight lines of Figs. 7 and 8 were fit to the data by a least-squares method. No attempt was made to fit a curve to the data from the down-throttle step of Test 67 (Fig. 8), because of its erratic nature. The response times are nearly independent of the size of the pressure steps within each test, but there was a substantial increase in the level of response times between tests 66 and 67. Neither the cause of this change nor the source of the erratic data from the down throttle steps of test 67 have yet been determined.

By way of comparison, the response requirement for the contracted effort with TRW was 75 ms. This requirement is moderately difficult, but within the current state-of-the-art, and is representative of the *Viking* requirement. The *Mariner* engine/*Surveyor* valve easily met the 75 ms during test 66, marginally during test 67, and not at all during test 68. Such shifts in response would be unacceptable in a flight application. Because of the lack of a positive valve position indication these response changes cannot be ascribed with certainty to either the chamber or valve. The same effect will be searched for during the full-scale testing to be conducted, and, if observed, the source should be determined since the larger throttle valves are equipped with positive position transducers.

Typical engine/valve response to the sinusoidal input (Fig. 4) are shown in Fig. 9 as standard Bode plots. Figure 9a illustrates the response to an input signal amplitude of  $\pm 10\%$ , while Fig. 9b shows the response for  $\pm 25\%$ . Because of the quantity of data gathered not all was reduced and plotted. The phase lag of Fig. 9 is defined as the time between identical positions (e.g., maximum points) on the recorded throttle valve command signal and the reactor chamber pressure oscillo-

gram traces and is expressed in degrees. The amplitude ratio is defined as

$$20 \log_{10} \left[ \frac{(P_{cd})_i}{(P_{cd})_{ss}} \right]_{v_i}$$

in decibels, where  $(P_{cd})_i$  and  $v_i$  are point values of reactor chamber pressure and throttle valve input voltage, respectively, and are evaluated at any arbitrary point during the sinusoidal period.  $(P_{cd})_{ss}$  is the reactor chamber pressure that would result from the application of  $v_i$  to the valve under steady-state conditions. The steady-state relationship between  $P_{cd}$  and  $v$  was determined from tests 66 to 68, described above.

Note that the data resulting from the  $\pm 10\%$  amplitude input signal show a greater phase lag and larger amplitude ratio than that resulting from the  $\pm 25\%$  input signal. This is as one might expect. The larger amplitude signal requires a greater mechanical movement of the valve in the same time period, and hence inertia effects are greater. Similarly, greater changes in the propellant flow rate are required, so that flow inertia effects are also greater. The valve/chamber/propellant feed system exhibits less damping when subjected to the lower amplitude input signal.

Again, for the purposes of comparison, the sinusoidal response requirement used in the TRW contract was, that in response to a sinusoidal throttle valve input signal of 5 Hz and amplitude of  $\pm 10\%$ , the thrust should have a phase lag no greater than 45 deg and the amplitude ratio should be greater than or equal to 0.7 (footnote 3); i.e., less negative than -3.1 dB, or even positive. From Fig. 9 it is clear that the *Mariner engine/Surveyor* valve combination could not meet the moderate phase lag limits, but could satisfy those for the amplitude ratio. However, included in the specifications for any flight system would probably be an *upper* limit on the amplitude ratio. A value of +0.2 dB would be typical. Such limitation of amplification might also prevent the present engine from meeting the flight criteria at 5 Hz, as may be seen from Fig. 9.

<sup>3</sup>The contract requirement is in terms of a thrust ratio,  $F_i/F_{ss}$ . Since thrust is directly proportional to reactor chamber pressure,  $(F_i/F_{ss} \cong P_i/P_{ss})$  the 0.7 figure can be converted to decibels by the relationship above. The equivalent value is -3.1 dB.

Included within all the duty cycles, both manually and remotely controlled, were many periods of relatively short, but steady-state operation. From this steady-state operation it is possible to determine the reactor performance at various throttle conditions. This type of data is needed for control of a landing maneuver. Given the spacecraft altitude, velocity, and acceleration, a throttle valve position can then be chosen to yield a thrust that will, in turn, result in a "soft" landing. The steady-state reactor performance, as measured by the characteristic velocity (one of the classical indices of rocket performance), is summarized in Fig. 10. Data from all the tests in this series are included there.

Also included on Fig. 10 are the performance predictions derived from Refs. 3, 4, and 5. Given the reactor geometry, fuel flow rate, chamber pressure, and the catalyst particle size and distribution, these references predict the composition of the gases leaving the catalyst bed. For monopropellant hydrazine the theoretical characteristic velocity is a unique function of the gas composition. Thus, for a given reactor,  $c^*$  as a function of mass flow rate can be predicted. Experimental  $c^*$ s are generally a large fraction of the theoretical values since catalytic reactors are not nearly as susceptible to mixing and atomization losses as are bipropellant injectors. However, while the predictions from Refs. 3, 4, and 5 are in reasonable agreement among themselves<sup>4</sup>, none predict the observed performance degradation at the lower flow rates evident in Fig. 10.

Since all three referenced correlations are based on data taken at bed loadings as low as 0.1 (lbm/(in.<sup>2</sup>-s)) it is concluded that the present performance degradation, as well as the erratic nature of the data at such low flow rates, result from a source other than excessive ammonia dissociation. Although it is known that propellant atomization is less important for hydrazine catalytic reactors than for a bipropellant engine, it is hypothesized that this may be the source. This same problem was not observed by TRW (Ref. 1) during the very limited amount of testing they completed with a 600-lbf-thrust

<sup>4</sup>Note that Ref. 5, strictly speaking, is not applicable to the Shell 405 catalyst used in these tests. Comparison between the correlation of Ref. 5 and the present data was attempted, however, because of the lack of agreement between experiment and the theories of Refs. 3 and 4 at low flow rates. As expected, Ref. 5 predicts a higher  $c^*$ , but again the dependence on  $\dot{m}$  is slight and of the same order of magnitude as in the other two works.

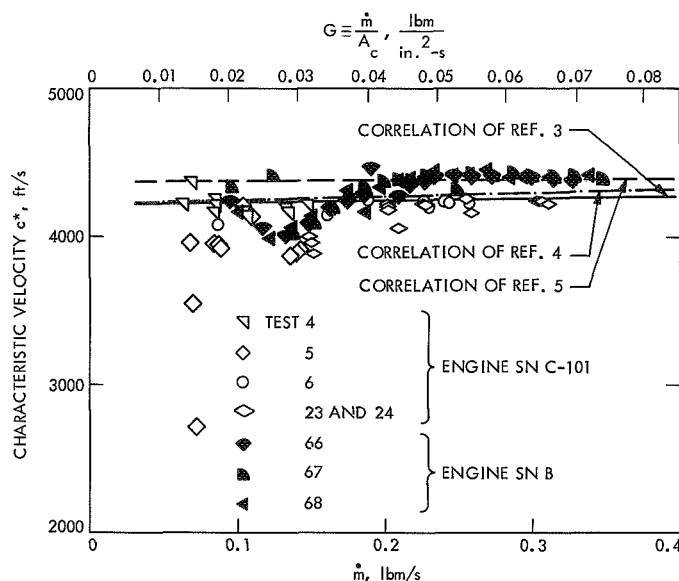


Fig. 10. Steady-state performance of Mariner Mars 1969 engine as a function of flow rate

engine. This phenomenon will be investigated further in future work.

### 3. Heat Sterilization Tests

**a. Objective.** In order to significantly reduce the possibility of introducing terrestrial organisms to another planet, all spacecraft which might impact or are intended to soft-land must be sterilized. Heat sterilization is the method most often used. The vehicle must be subjected to six 64-h periods at 275°F in a nitrogen atmosphere. Since a hydrazine decomposition chamber typically operates at temperatures of about 2000°F, there is no reason to expect that the sterilization requirements outlined above should significantly change an engine's operating characteristics. A short series of reactor tests was conducted to verify this.

**b. Apparatus and procedures.** The following steps were followed in the sterilization investigation. Two surplus Mariner Mars 1969 flight engines, numbers SN 005 and SN 008, were used for these tests. SN 005 was sterilized, while SN 008 served as a control. Since the engines had slightly different test histories, they were first repacked with fresh catalyst. The repacking was accomplished by machining open the engines at the injector-chamber weld joint (Fig. 1), filling both the upper and lower beds with unfired catalyst (the same size and types were used as shown in Fig. 1), and then rewelding. Each engine was then test-fired continuously for 45 s at fixed

thrust before sterilization, so as to obtain baseline performance measurements. The tests were made as identical as possible by conducting both on the same day, using hydrazine from the same lot, using the same instrumentation, etc. Following the baseline tests, all propellant and instrumentation openings of engine SN 008 were capped, the nozzle exit was covered with tape, and the engine was stored in its wooden transportation container. Engine SN 005 was sterilized in the materials research process laboratory at JPL. The temperature-time schedule noted above was followed. After the sterilization, which required 19 days to complete, all openings of engine SN 005 were closed in the same manner as for SN 008, and the engine returned to the test cell. The fixed-thrust performance tests were then repeated, again under identical conditions.

**c. Results and discussion.** Three criteria were used to judge the effect of heat sterilization on reactor operation. These were: (1) the characteristic velocity,  $c^*$ , (2) roughness, as measured by the normalized chamber pressure fluctuations,  $\Delta P_{cd}/2P_{cd}$ , and (3) the start transient characteristics, such as ignition delay (the time between the electrical signal to open the propellant valve and the first indication of chamber pressure), pressure rise time (the time between the first indication of chamber pressure and the attainment of 90% of the final steady-state pressure), and the peak overpressure (the maximum pressure attained during the start transient). The results of the four tests are summarized in Table 1.

Only the start transient characteristics exhibited a noticeable change. The ignition delay and pressure rise times both increased noticeably. The overpressure in itself is not considered particularly significant, since it is a direct result of the increased ignition delay and pressure rise-time. The fuel flow rate is fixed by the pressure drop across the injector, and since this pressure differential is large until chamber pressure is built up, the fuel flow rate is also large during the ignition delay/pressure rise-time periods. The net result is that the reactor fills with a gas mixture which is largely vaporized, unreacted hydrazine. Once ignition is achieved, this hydrazine reacts very rapidly, and hence the overpressure. Similar sterilization tests in which no overpressure was noted were subsequently conducted at TRW and are reported in SPS 37-47, Vol. III. However, the test setup used there included a throttle valve incorporating a cavitating venturi which prevented the fuel flow rate from going above the final steady-state value and thus reduced the quantity of fuel accumulated in the chamber.

Table 1. Summary of results of heat sterilization experiments

Test No.	Engine No.	$P_c$ (avg), psia	$\dot{m}$ , lbm/s	$c^*$ , ft/s	$\frac{\Delta P_{cd}}{2P_{cd}}$	Ignition delay, ms	Peak pressure, psia	Pressure rise time, ms
21	SN 005	206.8	0.234	4360	2.4	98	245	67
22	SN 008	187.6	0.214	4320	2.0	100	206	50
37	SN 008	189.8	0.216	4310	1.9	129	204	60
38	SN 005 <sup>a</sup>	206.9	0.234	4370	2.0	159	454	85

<sup>a</sup>After sterilization.

The details of the start transient of a catalytic monopropellant hydrazine engine are not well understood at the present; in fact, this process is the subject of a JPL-managed NASA Advanced Technology Contract (Ref. 6). Work done under that contract shows that the catalyst activity may be reduced when it is exposed at elevated temperatures to an inert gas, such as nitrogen. However, the activity reduction is reversible and has no discernible effect on steady-state operation. For a planetary lander application such as *Viking*, where the engine control is through an active guidance system, this initial pressure overshoot should cause no great problem, since the additional impulse can be compensated for later in the mission. However, for a passive, preprogrammed engine firing duration this could be a more significant problem. Such passive control has been used on all the JPL *Mariner* missions completed to date, and is currently planned for the *Mariner* Mars 1973 spacecraft.

Based on present results and those of other workers, it is reasonable to conclude that heat sterilization has a significant effect on only subsequent reactor start transient characteristics.

#### Definition of terms

- $A_c$  cross-sectional area of the cylindrical chamber, in.<sup>2</sup>
- $A_t$  area of nozzle throat, in.<sup>2</sup>
- $c^*$  characteristic velocity  $\equiv \frac{P_{cd} A_t g_0}{\dot{m}}$ , ft/s
- $g_0$  units conversion factor; numerically equal to  $32.174 \frac{\text{lbm-ft}}{\text{lbf sec}^2}$
- $F$  thrust, lbf

$G$  catalyst bed loading  $G \equiv \frac{\dot{m}}{A_c}$ , lbm/in.<sup>2</sup>-sec

$\dot{m}$  propellant mass flow rate, lbm/sec

$P_{cd}$  reactor chamber pressure measured downstream of the catalyst bed (Fig. 1), lbf/in.<sup>2</sup>

$\Delta P_{cd}$  average amplitude of fluctuations in chamber pressure about their mean value, lbf/in.<sup>2</sup>

$v_i$  valve input emf, V

$(v_i)_{\max}$  value of  $v_i$  corresponding to maximum valve travel, in either direction, about the null position, V

#### References

1. Kenney, R. J., and Reeves, D. F., "Throttleable Thruster System," Final Report, JPL Contract 952344, TRW Report No. 70-4726.3-28, April 28, 1970.
2. *Surveyor Project Final Report, Part I. Project Description and Performance, Vol. II*, Technical Report 32-1265, Jet Propulsion Laboratory, Pasadena, Calif., July 1, 1969.
3. "Development of Design and Scaling Criteria for Monopropellant Hydrazine Reactors Employing Shell 405 Spontaneous Catalyst," Final Report, NASA Contract NAS 7-372, Rocket Research Corp. Report RRC-66-R-76-Vol. II, Jan. 18, 1967.
4. Kesten, A. S., "Analytical Study of Catalytic Reactors for Hydrazine Decomposition," Second Annual Progress Report, NASA Contract NAS 7-458, United Aircraft Research Laboratories Report No. UARL G910461-24, May 1968.
5. Grant, A. F., Jr., "Basic Factors Involved in the Design and Operation of Catalytic Monopropellant-Hydrazine Reaction Chambers," Report 20-77, Jet Propulsion Laboratory, Pasadena, Calif., Dec. 31, 1954.
6. Sangiovanni, J. J., and Kesten, A. S., "Study of Hydrazine Reactor Vacuum Start Characteristics," First Annual Progress Report for NASA Contract NAS 7-696, United Aircraft Research Laboratories Report No. UARL H910758, December 1969.

## **B. Injector Hydrodynamics Effects on Baffled-Engine Stability—A Correlation of Required Baffle Geometry With Injected Mass Flux,** **R. M. Clayton**

### **1. Introduction**

It is well known that baffles can inhibit the development of sustained high-amplitude transverse wave systems in liquid rocket combustion chambers. Further, research conducted by JPL and others indicates that for conventional high-performance combustors, baffles are found to be most effective when placed in the early reaction region near the injector. Thus it can be argued that a primary variable, controlling baffle effectiveness, is the spatial relationship of the baffle geometry with the distribution of unreacted propellants. However, there are virtually no design criteria to reliably prescribe the necessary baffle/combustion-distribution relationship to provide a high margin of stability against resonance-initiating disturbances imposed either artificially or spontaneously.

The effective axial length of the early reaction region presently remains difficult to define for any given combustor because of the complexities of the overall combustion process; but it is clear that liquid phase atomization and propellant distribution largely determine the eventual combustion distribution, and that these properties are injector performance variables. Moreover, any substantial response of injector performance to perturbed combustion chamber conditions can, in general, only further complicate the dynamic behavior of the overall combustor. This responsiveness is believed to exist for most dynamic stability experiments with the conventional flight-type injectors for which other system requirements usually dictate minimized injector pressure drops. Thus, correlation of stability testing results, in terms of injector design variables, historically has been nearly impossible, and trial and error methods for injector "fixes" have prevailed.

In contrast, the JPL resonant combustion research has emphasized injection schemes producing steady reproducible hydrodynamic properties by virtue of the high flow-impedance and developed flow characteristics of long-tube (length = 100  $D$ ) orifices. Hence, the hydrodynamic boundary conditions governing spray formation have been controlled so that meaningful correlations should be possible between injector performance variables as "causes" and observed combustor behavior as "effects." Invariably, combustors using these JPL injectors

have exhibited a low dynamic-stability margin without baffles (i.e., stable to ordinary low levels of combustion noise, but unstable to discrete disturbances) producing a fully developed spinning detonation-like wave described in Ref. 1. But just as invariably, they have been reliably stabilized with baffles without modifications to the injection schemes. This stabilization procedure is contrary to nearly all contemporary flight engine development histories.

The observations stated above have led to a JPL conclusion that the essential qualitative requirement for maximizing the stability control effectiveness of baffles is as follows: provide sufficient hydrodynamic control of the injected flow so that the spatial distribution of mass and energy sources in the early reaction region is reproducibly compact and steadily maintained, relative to the baffle geometry, under perturbed combustion conditions.

For the conclusion to be very useful, however, it is presumed to be necessary to quantize two important criteria: (1) the critical spatial relationship between baffle geometry (length and spacing) and a fixed combustion distribution, and (2) the lower limit of control over the injection hydrodynamics required to maintain that critical spatial relationship in the presence of high amplitude combustion disturbances. The latter design limit information is important for minimizing possible design complications associated with long orifices while at the same time providing the lower limit of compromise to hydrodynamic control. Consequently, an investigation seeking these criteria was initiated with the intent of using combustion hardware from previous studies and a new "variable-control" injector. However, funding limitations have forced a termination of the investigation (first reported in SPS 37-62, Vol. III, pp. 272-279), and further work is expected to be delayed for an indefinite time.

This article presents the results of a final series of experiments before work was terminated that concern the first criterion itemized above and compares them with results from previous experiments. It will be shown that a correlation between baffle geometry and at least one injector performance variable is demonstrated and that this correlation can be stated in terms of rudimentary design criteria. The correlation holds, hence the design criteria is valid even though the relationship does not yet include terms that represent the combustion distribution specifically.

## 2. Description of Experiments

The dynamic stability of the baffled long-tube research injector, RMIR-5, was determined in order to supplement existing data for two other JPL injectors having similar stiff hydraulic configurations: injector RC-1 (SPS 37-62, Vol. III, pp. 272-279); and injector RMIR-7 (Ref. 2). Other design characteristics common to the respective combustors are uniform average mass flux distributions (non-humped distributions), unlike doublet injection elements, and cylindrical chambers nominally 16 in. (41 cm) long with an area contraction ratio of 2.0. Nominal values of design and operational differences for each combustor are listed in Table 1. Specific injection scheme variables are listed in Table 2.

Both 3 and 5 blade baffle arrays were used with the 11-in. (28 cm) diam RMIR-5 combustor. Blade lengths ranging from 1.5 to 7.8 in. (3.81 to 19.81 cm) were investigated for the three-blade array, and lengths of 2.0, 2.9, and 4.0 in. (5.08, 7.37, and 10.16 cm) were used for

the five-blade array. These configurations are shown in Fig. 1, where they can be compared with the RC-1 and RMIR-7 arrays used previously. For all of the experiments, the blades were welded to the injector face and were hand fitted so that no significant end gap (not more than several thousands of an inch) existed at the chamber wall. The somewhat nonradial placement of the blades was dictated by available space within the orifice arrangements of the respective injectors. The generally nonsymmetrical circumferential spacing stemmed partially from the same consideration, but also from intuitive (and unproven) reasoning that gas motion coupling between baffle cavities would be reduced by the non-symmetry.

The stability of the RMIR-5 engine was evaluated using an explosive-bomb technique. The effectiveness of each length baffle was tested by imposing bomb discharges of various sizes [2, 6.9, 13.5 gr (0.13, 0.45, 0.87 mkg) of PETN in the form of electrically initiated blasting caps] from various chamber locations. In practice, as

Table 1. Combustor operating conditions (nominal values for design conditions)

	RMIR-5	RC-1	RMIR-7
Propellants	SFNA + Corporal fuel <sup>d</sup>	N <sub>2</sub> O <sub>4</sub> + 50/50 (UDMH/N <sub>2</sub> H <sub>4</sub> )	N <sub>2</sub> O <sub>4</sub> + N <sub>2</sub> H <sub>4</sub>
Overall mixture ratio $r$	2.80	2.00	1.20
Main mixture ratio $r_m$	2.80	2.11	1.20
Boundary mixture ratio $r_b$	None	1.27	None
Total propellant flowrate $\dot{m}_t$ , lbm/s ( $4.54 \times 10^{-1}$ kg/s)	95.8	79.4	83.5
Mass fraction of boundary flow, $\dot{m}_b/\dot{m}_t$	—	0.10	—
Average chamber mass flux $\bar{G}_{ch}$ , lbm/s-in. <sup>2</sup> ( $7.04 \times 10^2$ kg/s-m <sup>2</sup> )	1.01	0.31	0.88
Chamber pressure, psia ( $6.89 \times 10^3$ N/m <sup>2</sup> )	300	100	300
Chamber diameter, in. (2.54 cm)	11.04	18.04	11.04
Combustion efficiency <sup>a</sup> , %	97	97	96
Combustion roughness, psi rms ( $6.89 \times 10^3$ N/m <sup>2</sup> )	2.5	<1.0	15.0
Acoustic velocity <sup>b</sup> , ft/s ( $3.05 \times 10^{-1}$ m/s)	3500	3960	4030
Measured frequency of spinning tangential wave (without baffles), Hz	2120	1445	2420
Measured maximum amplitude of sustained spinning wave (without baffles) <sup>c</sup> , psi p/p ( $6.89 \times 10^3$ N/m <sup>2</sup> )	~2490	~750	~1000

<sup>a</sup>Based on ratio of measured to theoretical equilibrium  $c^*$ .

<sup>b</sup>Based on measured  $c^*$ .

<sup>c</sup>From Kistler measurements near injector "corner" of chamber.

<sup>d</sup>SFNA (stabilized fuming nitric acid) consists of a mixture of the following compounds with percentage by weight as noted: 81.3-84.5 % HNO<sub>3</sub>; 14.0 ± 1.0 % NO<sub>2</sub>; 2.5 ± 0.5 % H<sub>2</sub>O; 0.6 ± 0.1 % HF.

Corporal fuel consists of a mixture of the following compounds with percentage by weight as noted: 46.5 ± 0.2 % furfuryl alcohol (C<sub>4</sub>H<sub>3</sub>OCH<sub>2</sub>OH); 7.0 ± 0.2 % N<sub>2</sub>H<sub>4</sub>; 1.5 % max H<sub>2</sub>O; 0.7 % max impurities; remainder, aniline (C<sub>6</sub>H<sub>5</sub>NH<sub>2</sub>).



Table 2. Injection conditions (nominal values)

Condition	RMIR-5	RC-1 <sup>c</sup>		RMIR-7
		Main flow	Boundary flow	
Number of elements	52	84	24	47
Oxidizer orifice diameter $D_{ox}$ , in. (2.54 cm)	0.173	0.142	0.073	0.173
Fuel orifice diameter $D_f$ , in. (2.54 cm)	0.0986	0.101	0.073	0.173
Oxidizer injection velocity $V_{ox}$ , ft/s ( $3.05 \times 10^{-1}$ m/s)	86.5	58.0	68.0	64.2
Fuel injection velocity $V_f$ , ft/s ( $3.05 \times 10^{-1}$ m/s)	137.6	86.0	92.0	73.5
Element flow rate ( $\dot{m}_{el}$ ), lbm/s ( $4.54 \times 10^{-1}$ kg/s)	1.84	0.85	0.33	1.81
Impingement distance from face, in. (2.54 cm)	0.75	0.625	0.125	0.75 or 0.875
Included impingement angle, deg. ( $1.75 \times 10^{-2}$ rad)	44	60	60	45
Normalized peak mass flux sampled for element ( $G_{max}$ ) <sub>SN</sub> <sup>a</sup> , (lbm/s-in <sup>2</sup> )/ (lbm/s), [( $1.55 \times 10^3$ kg/s-m <sup>2</sup> )/(kg/s)]	0.232	0.104	Not determined	0.342
True angular location of peak mass flux ( $\sigma$ ) <sup>a</sup> , deg ( $1.75 \times 10^{-2}$ rad)	4.78	0.0	Not determined	0.0
Mixing factor ( $\eta_m$ ) <sup>a</sup> , %	85.0	88.7	80.0	73.7
Mass median drop diameter for fuel ( $\bar{D}_f$ ) <sup>b</sup> , $\mu$ m	466	392	330	712

<sup>a</sup>From nonreactive flow measurements described in Ref. 4.

<sup>b</sup>Based on correlations from Ref. 6:  $\bar{D}_f = 10^5 \left( \frac{D_f^{0.27}}{V_f^{0.74}} \frac{D_{ox}^{0.023}}{V_{ox}^{0.33}} \right) (K_p) (K_\alpha)$

where  $K_p$  = correction for physical properties  
 $K_\alpha$  = correction for impingement angle

<sup>c</sup>Mass flux correlations shown in Fig. 6 based on main flow elements.

many as six bombs of a selected size were discharged sequentially, nominally 100 ms apart, during each engine firing. Bomb locations, order of firing, and construction details were essentially the same as for the RC-1 experiments (SPS 37-62, Vol. III) and will not be repeated here. The RMIR-7 experiments did not use the bombs, but this combustor (using  $N_2O_4$ - $N_2H_4$ ) characteristically exhibits extremely noisy combustion and is spontaneously unstable to these abnormally high combustion disturbances when not equipped with baffles (Ref. 2). Thus, it is assumed that these spontaneous disturbances are equivalent to artificially imposed disturbances for the purpose of testing the combustor's dynamic stability with various baffle configurations.

Provisions for Photocon or Kistler measurements were available for all firings. Pertinent location information is specified where necessary as the results are presented. The physical and performance details of these measurement systems are described in Ref. 1.

The RMIR-5 injector was operated near its design mixture ratio of 2.80 (Table 1) for SFNA + *Corporal* fuel

for most of the runs ( $2.69 < r < 2.82$ ); however, inadvertent off-mixture ratios as low as 2.39 were observed for some firings. In view of the lack of sensitivity of baffle effectiveness to operating mixture ratio exhibited by RC-1 (SPS 37-62, Vol. III), these off-mixture ratio firings are not considered to influence the present results. The RMIR-7 data to be shown were also for near design conditions.

As with mixture ratio, total flow rate for the RMIR-5 experiments was held nominally constant for most firings at between 92.0 and 96.7 lbm/s (41.7 and 43.9 kg/s) producing a mean chamber pressure range of 292 to 307 psig (2.01 to 2.12 MN/m<sup>2</sup>) at the injector face. Again, several firings inadvertently extended beyond this nominal range but are not considered sufficiently off-nominal to modify the results. Under these near-design flow conditions,  $c^*$  performance (corrected for stagnation pressure, but uncorrected for heat losses) was consistently 97% of theoretical equilibrium flow. Steady state combustion noise was generally less than 3 psi (20.7 kN/m<sup>2</sup>) RMS with no tendency for popping. Analogous performance characteristics for the RC-1 and RMIR-7 combustors are summarized in Table 1 for comparison.

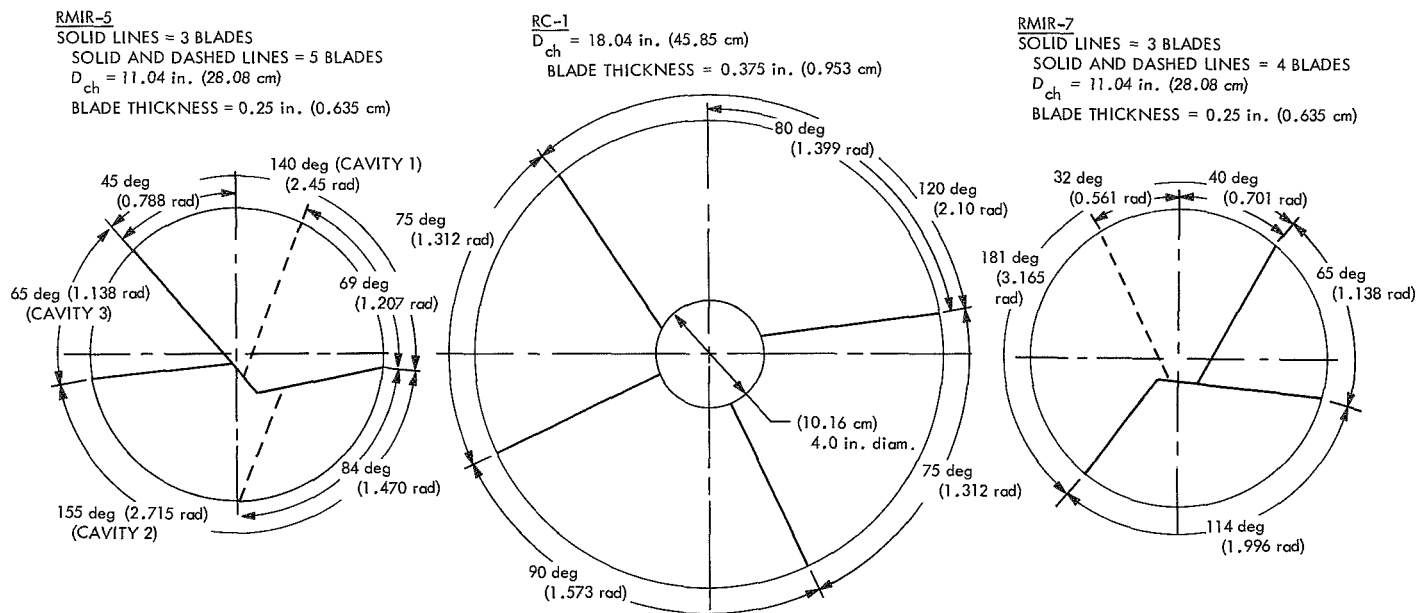


Fig. 1. Baffle array schematics for injectors RMIR-5, RC-1, and RMIR-7

### 3. Results for RMIR-5 Combustor

The overall results showing the dynamic stability characteristics of the RMIR-5 engine are summarized in Table 3. The engine was classified stable if for the particular baffle configuration all bomb disturbances damped with no discernible tendency for resonance at a sustained amplitude. Only the 5-blade array with lengths  $\geq 2.9 \text{ in. (7.37 cm)}$  were found to provide a high stability margin. For these stable configurations bomb pulse amplitudes between 165 and 1240 psi (1.14 and 8.55  $\text{MN/m}^2$ ) above mean chamber pressure, depending upon bomb and pressure measurement location, were damped after  $\sim 16 \text{ ms}$  with a spread of 12 to 22 ms. The higher pulse amplitudes and longer damping times were associated with bombs located within the baffle cavities. Note that the rather large incremental length variations for the 5-blade array do not resolve the minimum length required by less than 0.9 in. (2.3 cm); therefore, the minimum length probably falls between 2.0 in. (5.08 cm) which was unstable and the 2.9 in. (7.37 cm) length which was stable.

For the 3-blade array, lengths in the range of 2.9 to 3.4 in. (7.37 to 8.64 cm) provided marginal stability in that the 2-gr (0.13 mkg) bomb damped (average damping time 12 ms) but the 6.9 and 13.5 gr (0.45 and 0.87 mkg) bombs did not in all cases. Thus, a lower level of stability was provided by the most effective 3-blade configuration compared to the most effective 5-blade array. Pulse amplitudes ranged 75 to 475, 100 to 720, and

160 to 610 psi (0.52 to 3.28, 0.69 to 4.96, and 1.10 to 4.21  $\text{MN/m}^2$ ), respectively, for the 2, 6.9, and 13.5-gr (0.13, 0.45, and 0.87 mkg) bomb pulses that did damp for this 0.5 in. (1.27 cm) range of baffle length. Bombs located within the baffle cavities were generally the most apt to initiate resonance when stability was marginal.

Lengths of the 3-blade array shorter than 2.9 in. (7.37 cm) did not provide stability to all 2-gr (0.13 mkg) bombs [amplitudes 135 to 390 psi (0.93–2.69  $\text{MN/m}^2$ )]. Lengths greater than 3.4 in. (8.64 cm) were not all tested with the smallest bomb, but the results (Table 3) suggest that the greater lengths produced only marginal stability because they did not dampen all of the 13.5-gr (0.87 mkg) bombs [pulse amplitudes ranging from 65 to 735 psi (0.45 to 5.07  $\text{MN/m}^2$ )].

The amplitudes quoted above are for pulses that *did* damp. Those pulses that *did not* damp were generally of substantially greater amplitude. For instance, one 13.5 gr (0.87 mkg) bomb pulse that failed to damp for a 3.4-in. (8.64 cm) baffle length was detected at 1690 psi (11.65  $\text{MN/m}^2$ ) even though the bomb was located 10.8 in. (27.4 cm) downstream of the injector. All pressures quoted above were measured by three Photocon transducers located circumferentially at a chamber station 4.0 in. (10.2 cm) from the injector face.

The resonance mode observed for the 5-blade array with the 2.0-in. (5.08 cm) length was a first tangential

spinning wave with a maximum amplitude of about 400 psi (2.76 MN/m<sup>2</sup>) peak-to-peak and a frequency of 1735 Hz. This mode is termed a chamber mode herein because the baffle cavities as well as the chamber volume downstream of the cavities exhibited the same frequency. This exhibition of chamber modes rather than distinct baffle cavity modes was also noted for the RC-1 combustor with its 4-blade baffles (SPS 37-62, Vol. III). For the RMIR-5 3-blade array both chamber and cavity modes were encountered depending on baffle length.

For baffle lengths  $\leq 4.5$  in. (11.43 cm), chamber modes were always observed for the unstable firings with the 3-blade array. The observed modes varied from a spinning first tangential motion at a baffle length of 1.5 in. (3.81 cm), to a standing first transverse for 2.4 in. (6.10 cm), to a standing second transverse for 2.9 in. (7.37 cm), to a first radial mode for 3.4 in. (8.64 cm), and back to a second standing transverse motion at 4.0 and 4.5 in. (10.20 and 11.43 cm) baffle lengths. Frequencies as measured by Photocons at the 4.0 in. (10.20 cm) station are summarized in Table 2. The frequency of the first tangential modes (spinning or standing) decreased with increasing baffle length, whereas no such variation was noted for the other chamber modes.

For lengths  $\geq 5.0$  in. (12.70 cm) only cavity modes were present. In this case, cavity 1 (Fig. 1) generally exhibited a sharp-fronted wave traveling back and forth in the transverse direction of the cavity with a nearly constant frequency of from 2255 to 2315 Hz independent of baffle lengths between 5.0 and 7.8 in. (12.70 and 19.81 cm). Intermittently, cavity 2 would often simultaneously exhibit a similar wave motion, but with a consistently lower frequency in the range of 2000 Hz. Under these conditions cavity 3 appeared very noisy and occasionally exhibited a frequency of  $\sim 4740$  Hz. Phase measurements between the oscillations in the individual cavities always indicated that the cavities were uncoupled gas-dynamically. From similar measurements, no definite relationship between chamber and cavity oscillations were apparent. Finally, the chamber oscillations did not correspond with a natural chamber mode, even though the cavity frequencies were near the first tangential chamber frequency. Rather, the downstream oscillations appeared generally noisy, occasionally exhibiting all the cavity mode frequencies simultaneously; i.e., a chamber mode did not appear to be coupled to the cavity modes.

In summary, the effectiveness of baffles for stabilizing the RMIR-5 combustor can be stated as follows. The

5-blade array with length  $\geq 2.9$  in. (7.37 cm) provides a high margin of dynamic stability, prohibiting the development and sustenance of either cavity or chamber modes of resonance. On the other hand, the 3-blade array provides substantially less stability margin, allowing cavity modes to be sustained when the baffles were  $\geq 5$  in. (12.70 cm) long and failing to prevent sustained chamber modes when  $\leq 4.5$  in. (11.43 cm) long.

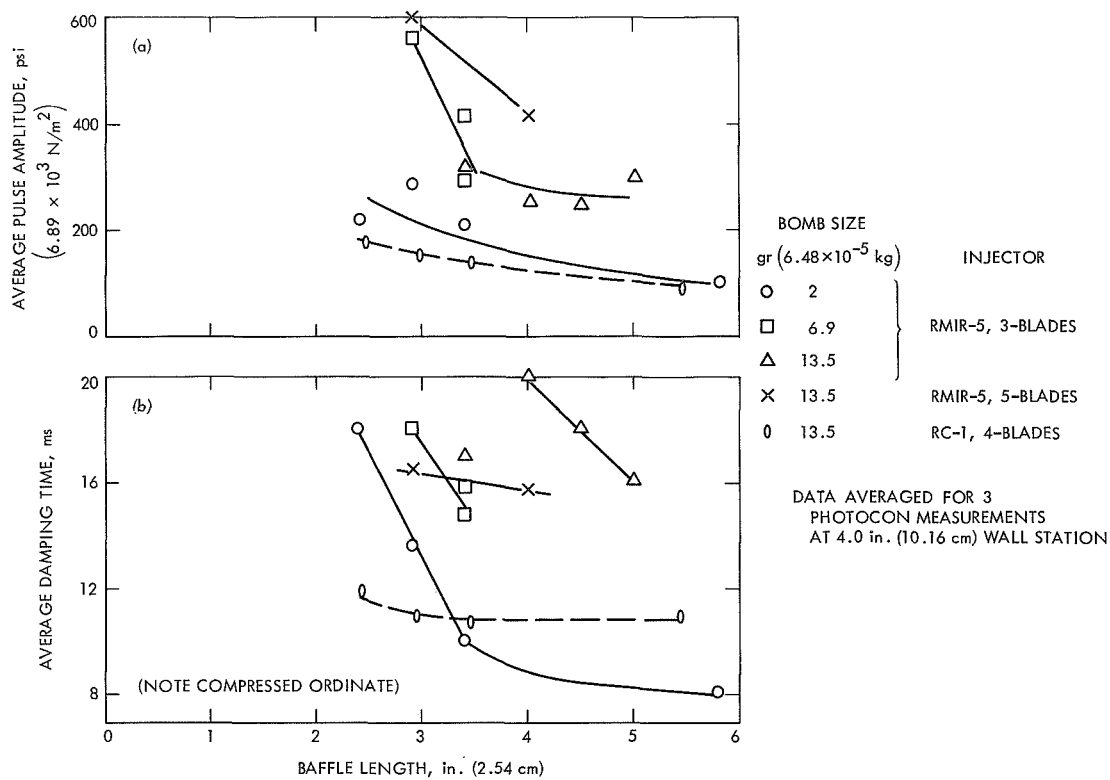
#### 4. Comparisons With Previous Results

*a. Bomb pulse amplitude and damping time.* For the adequate or nearly adequate baffle configurations, it was possible to measure damping times<sup>1</sup> and initial pulse amplitudes. Both were found to vary (for a fixed baffle condition) with bomb size and position, and measurement location; however, average values for 3 peripheral locations were taken to be a measure of the damping behavior of the engine. Figure 2 shows typical results for the RMIR-5 and RC-1 engines. (Analogous data for the RMIR-7 engine is not available.) If it is recalled that the 3-blade array for RMIR-5 did not provide a high margin of stability, it can be seen by comparing the damping times of the 3-blade array with those of the 5-blade and RC-1 arrays that damping time per se does not provide a positive measure of stability margin when the injected flow response is minimal. For this situation, damping time appears to be highly correlated with the time constant of the chamber-nozzle combination.

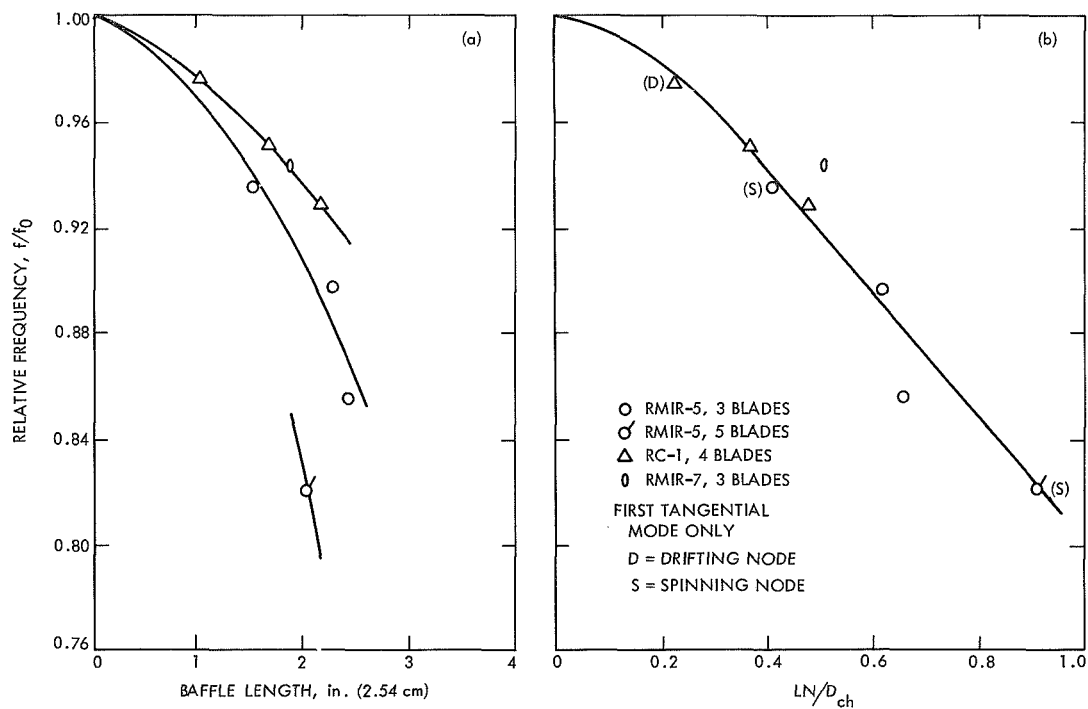
*b. Resonance frequency and amplitude reduction.* Even though inadequate baffling fails to prohibit the development of the first transverse chamber mode, the frequency as well as the amplitude of the wave is attenuated compared to the unbaffled situation.

*Frequency reduction.* The reduction of frequency of the first transverse chamber mode when chamber and baffle cavity oscillations are gas-dynamically coupled is expected both from analysis and past experimental observations. However, published theoretical analysis of the phenomena are limited to two-dimensional geometries without combustion sources or sinks (Ref. 3), while at the same time published data from engine firings are fragmentary. Figure 3 shows results for the three JPL

<sup>1</sup>Damping time is frequently used as a measure of stability margin. As used here, this term is defined as the total time (to the nearest millisecond) for the chamber pressure from Photocon measurements at the 4.0 in. (10.20 cm) chamber station to return to its pre-disturbance appearance on a high time-resolution oscillograph record.



**Fig. 2. Average damping time and pulse amplitude versus baffle length for injectors RMIR-5 and RC-1**



**Fig. 3. Relative frequency versus baffle length and the parameter  $LN/D_{ch}$  for injectors RMIR-5, RC-1, and RMIR-7**

combustors, where relative frequency  $f/f_0$  is shown versus baffle length  $L$  (Fig. 3a), and the parameter  $LN/D_{ch}$  (Fig. 3b), where  $N$  is the number of baffle blades and  $D_{ch}$  is the chamber diameter. As indicated in Ref. 3, the relative frequency is not a function of baffle length alone. These data suggest that  $f/f_0$  is nearly linearly related to  $LN/D_{ch}$ . Chamber length is not a significant variable in this data since all the engines had chamber lengths from injector nozzle entrance between 15.86 and 16.43 in. (40.30 and 41.70 cm). Interestingly, the reduction in frequency does not appear to be influenced by whether the mode is spinning or standing.

**Amplitude reduction.** The limited analytical results from Ref. 3, based on a model of pressure coupling between the chamber and the baffle cavities indicate that pressure amplitude in the cavities tends to increase relative to the amplitude for the chamber space, thus suggesting a *destabilizing* effect of baffles. No such effect has been observed in the present experimental results, nor is it generally observed by others for liquid rocket engines. Resonance amplitude variations as a function of chamber wall axial station for several lengths of baffles for the RC-1 engine are shown in Fig. 4. These results typify analogous trends for the RMIR-5 and 7 engines. The pressures were measured with Kistler transducers located near the pressure antinode of the first tangential waves; therefore, they are maximum amplitudes for each baffle condition. The amplitude distribution for the unbaffled chamber is also shown. The latter distribution was obtained for the fully developed spinning detonation-like wave, characteristic of the unbaffled engines. In addition to preventing this severely destructive mode from developing, the baffles clearly reduced the ampli-

tude of the standing mode as their length was increased toward the 2.4 in.-length (6.10 cm) required to stabilize the RC-1 engine. Note that this length corresponds to the axial station where the slope of the unbaffled pressure distribution curve nears a minimum value.

**c. Correlation of baffle performance results.** Initial comparisons of the results, particularly the apparent amplitude/critical baffle length relationship noted in Fig. 4, indicated that variations in energy and mass source distributions for each combustor must be accounted for if a correlation of stabilizing performance with baffle geometry and axial combustion distribution indeed exists. However, prior to using extensive computer calculations, based on assumed initial spray properties, to evaluate the evaporatively controlled combustion rates along the length of the combustors, it was decided to attempt a correlation of the data on the basis of existing cold flow measurements of the mass distribution produced by the specific elements used in each injector. That is, if the axial mass concentration of unreacted propellants is a first-order variable controlling baffle stabilizing effectiveness, it is reasoned that at least a partial measure of such a variable is contained in a comparison of local values of injected mass flux with the average chamber mass flux.

For this purpose, a variable termed  $\mathcal{G}_{rel}$  was selected that is defined as the value of the element peak axial mass flux  $(G_{max})_{L'}$  on a plane perpendicular to the chamber axis a distance  $L'$  from the impingement point divided by the average chamber mass flux,  $\bar{G}_{ch}$ ; where  $L'$  is the axial distance of the downstream edge of the baffle from the impingement point. Thus,

$$\mathcal{G}_{rel} \equiv \frac{(G_{max})_{L'}}{\bar{G}_{ch}}$$

The evaluation of  $(G_{max})_{L'}$  is based on cold flow (non-reactive fluids) measurements of the mass distributions of the particular elements, for which the experimental methods are described in Ref. 4. The essential features of the evaluation are shown in Fig. 5. Pertinent values of the sampled spray properties are listed in Table 2.

As noted previously, the RMIR-5 experiments showed a dependence of stabilizing performance on baffle spacing as well as length; therefore, any correlation of baffle effectiveness with combustion distribution must also account for this second spatial property of the baffle

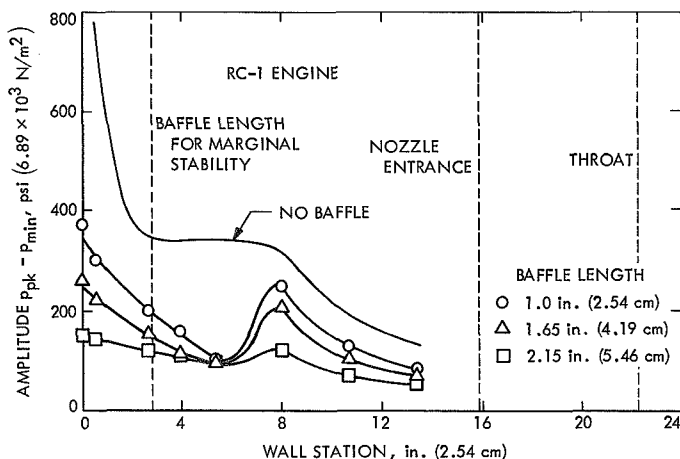
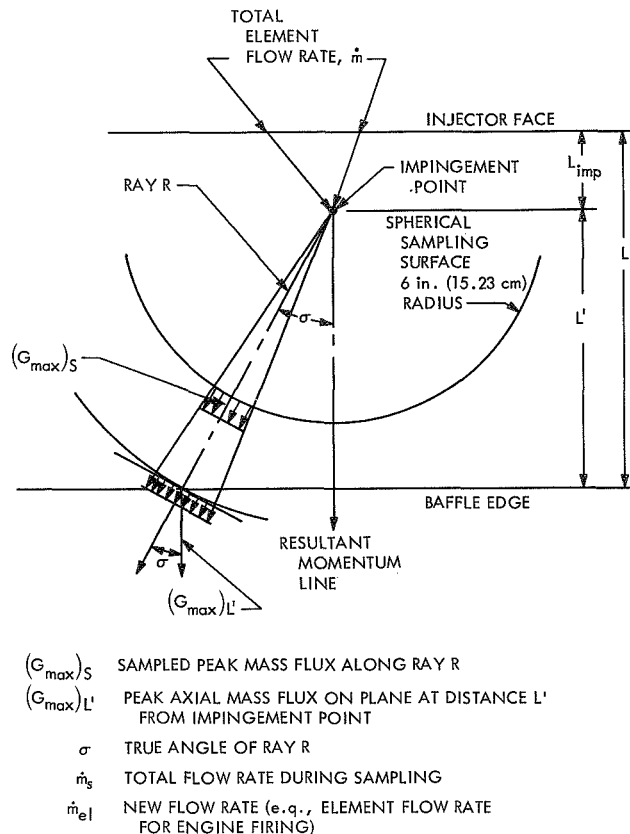


Fig. 4. Resonant pressure distribution for various baffle lengths, RC-1 engine



SEE TABULATED VALUES IN TABLE 2

Fig. 5. Spray mass flux relationships

array. Based primarily on heuristic arguments, different arrangements of baffle cavities can be shown to be more or less compatible with particular acoustic modes of gas motion in the unbaffled portion of the chamber. For example, Ref. 5 states that an odd number of radial blades is most effective against the tangential modes and that the order of the mode divided by the number of blades should not be an integer. However, clear evidence of a relationship of this kind has not been observed in the present results.

For the RMIR-5 engine with the 3-blade array both the first and second tangential modes were observed for particular ranges of baffle length (Table 3). For the RC-1 engine with 4 blades, only the first tangential mode was observed and the velocity nodes were not aligned with a blade but, rather, bisected opposing cavities. And for RMIR-7, the engine was equally well stabilized against the first tangential with either 3- or 4-blade arrays of adequate length. Therefore, while some relationship of baffle cavity dimension to chamber dimension is evidently important to stabilizing effectiveness, it is not

$$\mathcal{G}_{rel} \equiv \frac{(G_{\max})_{L'}}{\bar{G}_{ch}}$$

WHERE FROM GEOMETRY AND CONSERVATION OF SPRAY PARTICLE MASS:

$$(G_{\max})_{L'} = (G_{\max})_S \frac{\dot{m}_{el}}{\dot{m}_S} \left( \frac{6}{L'} \right)^2 (\cos \sigma)^2$$

$$\text{OR IF } \frac{(G_{\max})_S}{\dot{m}_S} \equiv (G_{\max})_{SN}$$

THEN

$$(G_{\max})_{L'} = \frac{36 \dot{m}_{el} (G_{\max})_{SN} (\cos \sigma)^2}{(L')^2}$$

clear from these data that it is strictly an acoustic mode relationship. However, for the lack of a more definitive spatial correlating parameter, one of this nature will be used.

The apparent interaction of the relative local mass flux, baffle length, and blade spacing on baffle performance for the three combustors is shown in Fig. 6. A data point for each combustor/baffle configuration is plotted as the ratio of average cavity-peripheral-width  $\bar{w}$  to the first tangential wave length  $\lambda_0$  versus  $1/\mathcal{G}_{rel}$ . The width  $\bar{w}$  was taken as the chamber circumference divided by the number of radial baffles less the blade thickness, and  $\lambda_0$  was based on the measured first tangential frequency for each unbaffled engine. The data are plotted as  $1/\mathcal{G}_{rel}$  so that the direction of increasing  $L'$  is to the right.

Each combustor/baffle configuration was classified as stable, marginally stable, or unstable according to its observed response to combustion chamber disturbances,

**Table 3. RMIR-5 dynamic stability results**

Blade array	3			5
Disturbance <sup>a</sup>	2.0	6.9	13.5	13.5
Blade length <sup>b</sup> , in. (2.54 cm)				
1.5	U 1T-1982 (Drifting)			
2.0				U 1TS-1735
2.4	U 1T-1812			
2.9	S	U 2T-3380		S
3.4 <sup>c</sup>	S	U 1R-4160 ----- S	U 1R-4160	
4.0			U 2T-3380	S
4.5			U 2T-3370	
5.0			U C1-2255	
5.8 <sup>c</sup>	U C1-2270		U C1-2280 ----- U C1-2270 ----- U C1-2255	
7.8 <sup>c</sup>			U C1-2260 ----- U C1-2315	

<sup>a</sup>Numerals indicate bomb size.

<sup>b</sup>U, unstable; S, stable; mode designations with frequency, 1TS-first spinning tangential, 1T-first standing tangential, 2T-second standing tangential, 1R-first radial, C1-transverse mode in cavity 1.

<sup>c</sup>More than one classification per baffle length denotes replicate firings.

whether artificially or spontaneously imposed. A configuration was classified as stable if all disturbances damped with no discernible tendency for resonance at a sustained amplitude. Marginal stability was assigned to the engine if any discernible resonance occurred intermittently,

either during a single run or a succession of runs. This classification was also assigned to engines that were sensitive to some disturbances, but insensitive to others. The unstable classification was assigned if the engine consistently exhibited continuous resonance from the time of its initiation, by whatever means, through the remainder of the firing.

As can be seen from Fig. 6, baffle configurations provided dynamic stability when blade length was sufficient to reduce  $\mathcal{G}_{rel}$  to approximately 3; if, simultaneously, the number of blades was sufficient to keep  $\bar{w}/\lambda_0$  below approximately 0.57. It is noted that the latter restriction is based only on the 3-blade RMIR-5 engine which exhibited higher-order chamber and cavity modes with baffle lengths beyond those required to prevent the first tangential mode. This may reflect a relatively low combustion rate for the SFNA + *Corporal* fuel propellants allowing high concentrations of unreacted propellants within the cavities. It is further noted that one stable RMIR-7 configuration (with  $N_2H_4$ ) occurred to the left of the stability limit line which suggests that its combustion rate may be relatively more rapid than implied by the simple relative mass flux parameter.

Since these results were obtained for relatively large engines with a fixed contraction ratio, and since the injectors produced a relatively large thrust per element [135 to 425 lbf (600 to 1890 N)], it is of interest to compare analogous baffle performance data reported in Ref. 6 for the Rocketdyne *Apollo* lunar module (LM) ascent engine. This engine offers a contrast to the JPL research engines in that chamber diameter [7.79 in. (19.78 cm)], contraction ratio (2.9), orifice size scale ( $<20$  lbf (89 N)/element) and orifice length ( $<4D$ ) are significantly different. However, the impingement angle and diameter ratio of the LM unlike doublet core elements are sufficiently similar to the RC-1 element design to allow an assumption of the same values for  $(C_{max})_{SN}$  and the angle  $\sigma$  (see Table 2 and Fig. 5) for the two elements. With this assumption and  $\dot{m}_{e1}$  for the LM elements,<sup>2</sup> it is then possible to compute  $\mathcal{G}_{rel}$  for various baffle lengths. The results for the LM engine are shown in Fig. 7, where it can be seen that they are consistent with the limit lines noted for the JPL engines. This agreement is possibly fortuitous, but on the other hand, it lends credence to the applicability of the work reported here to stable-engine design.

<sup>2</sup>Dimensional and operational data for the LM injector was furnished by Carl L. Oberg via Rocketdyne letter 70RC10366, dated September 24, 1970.





## 5. Conclusions

The baffle length required for dynamic stability is primarily controlled by the concentration of unreacted propellants in the early reaction region of the chamber.

Even though combustion rates for specific propellants and atomization conditions affect this concentration, a rudimentary design criterion is to limit the value of  $\mathcal{G}_{rel}$  to 3, while at the same time limiting  $\bar{w}/\lambda_0$  to less than about 0.57 by using at least three baffle blades.

In order to assure reproducible stability, sufficient hydraulic control of the injection scheme must be exercised to maintain the critical spatial relationship between a fixed baffle geometry and the unreacted propellant distribution under perturbed chamber conditions. The lower limit of control sufficiency remains to be determined.

## References

1. Clayton, R. M., Rogero, R. S., and Sotter, J. G., *An Experimental Description of Destructive Liquid Rocket Resonant Combustion*, Technical Report 32-1293, Jet Propulsion Laboratory, Pasadena, Calif. (Reprinted from *AIAA Journal*, Vol. 6, No. 7, pp. 1252-1259, July 1968).
2. Rupe, J. H., *An Experimental Correlation of the Nonreactive Properties of Injection Schemes and Combustion Effects in a Liquid-Propellant Rocket Engine: Part V. On the Influence of Vanes on Combustion and Combustion Stability*, Technical Report 32-255, Jet Propulsion Laboratory, Pasadena, Calif., Sept. 15, 1967.
3. Oberg, C. L., Wong, T. L., and Schmeltzer, R. A., *Analysis of the Acoustic Behavior of Baffled Combustion Chambers*, Report R-8076 (NASA CR-72625), North American Rockwell Corporation, Rocketdyne Division, Canoga Park, Calif., January 1970.
4. Rupe, J. H., *An Experimental Correlation of the Non-reactive Properties of Injection Schemes and Combustion Effects in a Liquid Propellant Rocket Engine: Part I. The Application of Nonreactive-Spray Properties to Rocket-Motor Injector Design*, Technical Report 32-255, Jet Propulsion Laboratory, Pasadena, Calif., July 15, 1965.
5. *Design Guide for Stable  $H_2/O_2$  Combustors: Volume 1. Design Application*, Report 20672-P2D, Aerojet Liquid Rocket Company, Sacramento, Calif., May 1970.
6. *Lunar Module Ascent Engine Acoustic Cavity Study*, Report R-7935, North American Rockwell Corporation, Rocketdyne Div., Canoga Park, Calif., Aug. 1, 1969.
7. Knight, R. M., Nurick, W. H., *Correlation of Spray Dropsizes Distribution and Injector Variables*, Interim Report, R-7995, North American Rockwell Corporation, Rocketdyne Div., Canoga Park, Calif., Sept. 30, 1969.

# XXI. Tracking and Orbit Determination

## MISSION ANALYSIS DIVISION

### A. Spectral Factorization in Discrete Systems,

*T. Nishimura*

#### 1. Introduction

The spectral factorization is the basic problem arising in the Wiener-Shannon filtering theory. Usually the characteristic equation is solved and the roots having negative real parts are picked up in order to formulate a realizable filter in the continuous system. If the system is discrete, the characteristic roots lying inside the unit circle around the origin on the complex plane are selected for this purpose.

In contrast to these static filters, dynamic filters such as Kalman filters have been introduced in recent years which can be extensively applied to estimation problems in real time.

The sequential (Kalman) filter is asymptotically stable if the system is uniformly and completely observable and controllable (the definitions of these terms are contained in Ref. 1). Particularly when the system is time-invariant, the error covariance  $P(t)$  converges to a constant matrix which is identical to the one obtained by the spectral factorization of the Wiener-Shannon theory.

According to Potter and Anderson (Refs. 2 and 3), this spectral factorization problem can be solved for continuous systems by finding the eigenvectors of a matrix which

is constructed by the coefficient matrices of the Riccati equation (covariance equation).

In this article, the counterparts of their theorems in discrete systems are derived and the technique is demonstrated by an example.

#### 2. Mathematical Results

Let the process equation of parameter  $x$  ( $n \times 1$ ) and equation of data  $y$  ( $m \times 1$ ) of a discrete system be

$$x(k+1) = \Phi(k)x(k) + G(k)w(k) \quad (1)$$

$$y(k) = H(k)x(k) + v(k) \quad (2)$$

where  $w$  and  $v$  are independent, white noises with zero mean. Hence the correlation of noise is

$$E[w(j)w'(k)] = Q(k)\delta_{jk} \quad (3)$$

$$E[v(j)v'(k)] = R(k)\delta_{jk} \quad (4)$$

$$E[w(j)v'(k)] = 0 \quad (5)$$

where  $\delta_{jk}$  is the Kronecher delta and  $E[\ ]$  indicates an ensemble average.

The error covariance  $P(k)$  of the optimal estimate  $x^*(k)$  is provided by the following Riccati equation (Ref. 1):

$$\begin{aligned} P(k+1) = & \Phi(k) \{P(k) - P(k)H'(k) \\ & \times [H(k)P(k)H'(k) + R(k)]^{-1} \\ & \times H(k)P(k)\} \Phi'(k) + G(k)Q(k)G'(k) \end{aligned} \quad (6)$$

with the initial condition

$$P(0) = E[x(0)x'(0)] \quad (7)$$

where the prime indicates the transpose.

In this article we are concerned with the solution of Eq. (6) when the system is time-invariant and the noises are stationary.

According to Kalman, Eq. (6) converges to a steady-state solution, provided the system is uniformly and completely observable and controllable. Assuming that this observability and controllability condition is satisfied, the solution of the following equation is sought in this article:

$$P = \Phi \{P - PH' [HPH' + R]^{-1} HP\} \Phi' + GQG' \quad (8)$$

As principal results of this article, three theorems are derived concerning the solution of Eq. (8). The proof of theorems may be referred to a JPL internal document.<sup>1</sup>

**THEOREM 1.** *Let the matrix  $A$  ( $2n \times 2n$ ) be described by*

$$\begin{aligned} A = & \left[ \begin{array}{c|c} \Phi + GQG'\Phi'^{-1}H'R^{-1}H & GQG'\Phi'^{-1} \\ \hline \Phi'^{-1}H'R^{-1}H & \Phi'^{-1} \end{array} \right] \\ = & \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \end{aligned} \quad (9)$$

Also, let  $T$  be a matrix such that  $T^{-1}AT$  reduces to a Jordan canonical form  $J$ . When  $T$  and  $J$  are similarly partitioned as  $A$ , the solution of Eq. (8) is given by

$$P = T_{11} T_{21}^{-1} \quad (10)$$

if  $T_{21}$  is non-singular.

<sup>1</sup>Nishimura, T., *Spectral Factorization in Discrete Systems*, May 12, 1970 (JPL internal document).

**THEOREM 2.** *For an eigenvalue  $\lambda_i$  of  $A$ , there exists an eigenvalue  $\lambda_j = 1/\lambda_i$ . Namely, one-half of eigenvalues of  $A$  lie outside the unit circle on the complex plane while the other half lie inside the unit circle.*

**THEOREM 3.** *Assuming all the eigenvalues of  $A$  are distinct, a real, symmetric, non-negative-definite solution  $P$  of Eq. (8) can be uniquely determined by arranging the sub-matrix*

$$\begin{bmatrix} T_{11} \\ T_{21} \end{bmatrix}$$

*such that it consists of eigenvectors of  $A$  corresponding to the eigenvalues which lie outside the unit circle.*

### 3. Example

The solution provided by Theorem 3 is applied to the tracking problem of a spacecraft moving along a straight line by means of the ranging system at the ground station. The probe undergoes a small acceleration which is white with variable magnitude. The steady-state solution is sought in order to get the limiting accuracy of such tracking system.

The range data are taken every 1000 seconds and the data noise is taken as 10, 5, and 1 m, which is also assumed white. Let  $x_1$  and  $x_2$  be position and speed, respectively, of the spacecraft. Then

$$x_1(k+1) = x_1(k) + x_2(k) \quad (11)$$

$$x_2(k+1) = x_2(k) + w(k) \quad (12)$$

$$y(k) = x_1(k) + v(k) \quad (13)$$

Hence the coefficient matrices are

$$\Phi = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \quad (14)$$

$$H = [1, 0] \quad (15)$$

$$R = \sigma_w^2 \quad (16)$$

$$Q = \begin{bmatrix} 0 & 0 \\ 0 & \sigma_w^2 \end{bmatrix} \quad (17)$$

The limiting accuracies of this ranging system are demonstrated in Figs. 1 and 2. Figure 1 shows the uncertainty of position of the spacecraft versus the magnitude of acceleration while Fig. 2 does that of speed for different values of data noise.

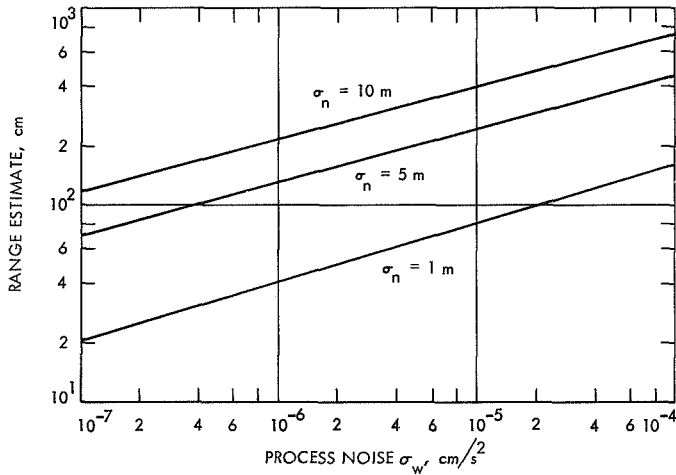


Fig. 1. Limiting accuracy of range estimate

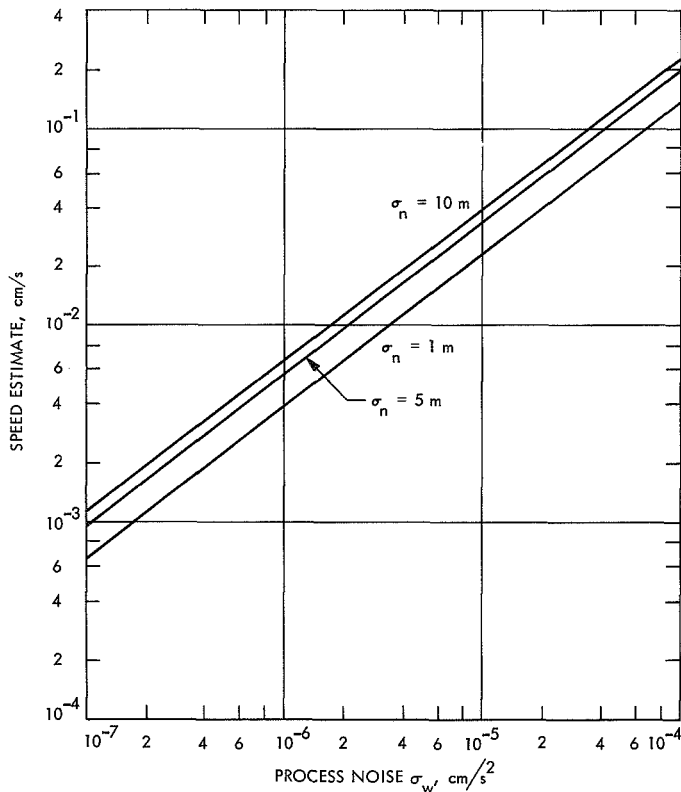


Fig. 2. Limiting accuracy of speed estimate

The range of process noise is such that  $10^{-7} \sim 10^{-6} \text{ cm/s}^2$  is approximately the magnitude of solar disturbances and leakage in gas jets and  $10^{-3} \sim 10^{-4}$  the magnitude of thrust fluctuations of solar electric spacecraft.

The discrete Riccati Eq. (8) for the system described by Eqs. (14-17) reduces to, for  $p_{11}$  (variance of estimate of  $x_1$ ),

$$p_{11}^4 = \sigma_w^2 (p_{11} + \sigma_n^2) (p_{11} + 2\sigma_n^2)^2 \quad (18)$$

This is difficult to solve analytically. However, when  $p_{11} \ll \sigma_n^2$ , the solution is derived by

$$\sqrt{p_{11}} \approx 2^{1/4} \sigma_w^{1/4} \sigma_n^{3/4} \quad (19)$$

Therefore the standard deviation of the estimate of position is proportional to  $\sigma_w^{1/4}$  of process noise and to  $\sigma_n^{3/4}$  of data noise.

This agrees with the analysis in the continuous case for the same ranging system (Ref. 4). Since the inclination of straight lines in Fig. 1 is  $1/4$ , it proves that  $\sqrt{p_{11}}$  is proportional to the fourth root of standard deviation of process noise.

Similarly an approximate solution for  $p_{22}$  becomes

$$\sqrt{p_{22}} \approx 2^{1/4} \sigma_w^{3/4} \sigma_n^{1/4}$$

Namely, the standard deviation of estimate of speed is proportional to  $\sigma_w^{3/4}$  of process noise and to  $\sigma_n^{1/4}$  of data noise.

Again Fig. 2 shows that  $\sqrt{p_{22}}$  is proportional to the three-fourth power of standard deviation of process noise.

## References

1. Kalman, R. W., *New Methods and Results in Linear Prediction and Estimation Theory*, Technical Report 61-1. Research Institute for Advanced Study, Baltimore, Md., 1961.
2. Potter, J. F., "Matrix Quadratic Solutions," *Soc. Ind. Appl. Math.*, Vol. 14, pp. 496-501, May 1966.
3. Anderson, B. D. O., "An Algebraic Solution to the Spectral Factorization Problems," *IEEE Trans. Automat. Contr.*, Vol. AC-12, No. 4, pp. 410-414, Aug. 1967.
4. Nishimura, T., "A New Approach to Estimation of Initial Conditions and Smoothing Problems," *IEEE Trans. Aerosp. Electron. Sys.*, Vol. AES-5, No. 5, pp. 828-836, Sept. 1969.

## Subject Index

Subject	Pages	Subject	Pages
<b>Antennas and Transmission Lines</b>		<b>Fluid Mechanics</b>	
spacecraft antenna tolerances . . . . .	43-49	response of supersonic laminar boundary	
development of a conical-gregorian		layer to a moving external pressure field . . . . .	13-16
high-gain antenna . . . . .	118-119	injector hydrodynamics effects on baffled-	
high-gain antenna for Thermoelectric		engine stability: correlation of required	
Outer-Planet Spacecraft (TOPS) . . . . .	157-159	baffle geometry with injected mass flux . . . . .	222-232
<b>Chemistry</b>		<b>Information Theory</b>	
multi-phase ammonia-water system . . . . .	4-9	digital clean-up loop transponder for	
viscoelastic behavior of elastomers under-		sequential ranging system . . . . .	18-27
going crosslinking reactions . . . . .	187-189	performance of a phase-locked loop	
energy transfer in bipyridilium		during loss of signal . . . . .	28-32
(paraquat) salts . . . . .	189-191	optimum configurations for phase-shift	
electrical properties of TCNQ salts of		keyed/phase modulated systems . . . . .	33-36
ionene polymers and their model		critical problem of combinatorial	
compounds . . . . .	192-198	geometry and coding theory . . . . .	36-42
<b>Computer Applications and Equipment</b>		stability of second-order tracking loops	
automated test techniques for control		with arbitrary time delay . . . . .	50-53
and guidance subsystems . . . . .	90-91	signal design for single-sideband	
use of <i>m</i> -ary linear feedback shift		phase modulation . . . . .	54-58
registers with binary devices . . . . .	161-163	steady-state performance of a data-	
<b>Control and Guidance</b>		transition type of first-order	
proposed laser obstacle detection sensor		digital phase-locked loop . . . . .	59-68
for a Mars rover . . . . .	80-89	optimum modulation index for a data-aided	
automated test techniques for control and		phase-coherent communication system . . . . .	69
guidance subsystems . . . . .	90-91	use of <i>m</i> -ary linear feedback shift	
attitude control of a flexible solar-electric		registers with binary devices . . . . .	161-163
spacecraft . . . . .	92-99	command prefix code for Thermoelectric	
finite element modeling for appendage		Outer-Planet Spacecraft (TOPS) . . . . .	164-166
interaction with spacecraft attitude		spectral factorization in discrete systems . . . . .	233-235
control . . . . .	100-105	<b>Lunar Interior</b>	
digital gyro system for Thermoelectric		lunar elevation correction for gravity	
Outer-Planet Spacecraft (TOPS)		measurements . . . . .	1-3
single-axis simulator . . . . .	105-107	<b>Masers and Lasers</b>	
stepper motor drive electronics for		proposed laser obstacle detection sensor	
solar-electric thrust vector		for a Mars rover . . . . .	80-89
control subsystem . . . . .	108-110	holographic study of operating xenon	
<b>Electronic Components and Circuits</b>		compact-arc lamp . . . . .	171-174
digital clean-up loop transponder for		diffraction of a high-order gaussian	
sequential ranging system . . . . .	18-27	laser beam by an aperture . . . . .	181-186
stepper motor drive electronics for		<b>Materials, Nonmetallic</b>	
solar-electric thrust vector		spacecraft adhesives for long life and	
control subsystem . . . . .	108-110	extreme environment . . . . .	111-114
radiation effects on electronic parts . . . . .	145-147	viscoelastic behavior of elastomers under-	
evaluation of recording tape and heads		going crosslinking reactions . . . . .	187-189
for spacecraft . . . . .	160	energy transfer in bipyridilium	
effect of Jupiter electron dose on		(paraquat) salts . . . . .	189-191
metal oxide semiconductors . . . . .	166-170	electrical properties of TCNQ salts of	
evaluation of magnetic recording tapes . . . . .	198-200	ionene polymers and their model	
ion thruster connectors . . . . .	211-212	compounds . . . . .	192-198

Subject	Pages
<b>Mathematical Sciences</b>	
unitary similarity transformation of a hermitian matrix to a real symmetric tridiagonal matrix .....	10-12
solution of partial differential systems .....	17
general solution to problem of performance of a phase-locked loop during loss of signal .....	28-32
optimum configurations for phase-shift keyed/phase modulated systems .....	33-36
critical problem of combinatorial geometry and coding theory .....	36-42
expressions for spacecraft antenna tolerances .....	43-49
analysis of stability of second-order tracking loops with arbitrary time delay .....	50-53
signal design for single-sideband phase modulation .....	54-58
analysis of steady-state performance of a data-transition type of first-order digital phase-locked loop .....	59-68
linear equations of motion and stability analysis relative to attitude control of a flexible solar-electric spacecraft .....	92-99
finite element modeling for appendage interaction with spacecraft attitude control .....	100-105
derivation of equations governing heated shallow shells of revolution .....	141-144
derivation of intensity of a high-order gaussian beam as a function of aperture size and oscillating mode index .....	181-186
spectral factorization in discrete systems.....	233-235
<b>Mechanics</b>	
lunar elevation correction for gravity measurements .....	1-3
linear equations of motion and stability analysis relative to attitude control of a flexible solar-electric spacecraft .....	92-99
finite element modeling for appendage interaction with spacecraft attitude control .....	100-105
<b>Mechanisms</b>	
mechanical devices for Thermoelectric Outer-Planet Spacecraft (TOPS) .....	148-156
<b>Optics</b>	
holographic study of operating xenon compact-arc lamp .....	171-174
diffraction of a high-order gaussian beam by an aperture .....	181-186
<b>Photography</b>	
holographic study of operating xenon compact-arc lamp .....	171-174

Subject	Pages
<b>Planetary Atmospheres</b>	
multi-phase ammonia-water system for modeling Jupiter and Saturn atmospheres .....	4-9
<b>Planetary Exploration, Advanced</b>	
radioisotope thermoelectric generators for outer-planet missions .....	70-75
proposed laser obstacle detection sensor for a Mars rover .....	80-89
effect of Jupiter electron dose on metal oxide semiconductors .....	166-170
thrust subsystem design for nuclear-electric Jupiter orbiter .....	207-210
<b>Power Sources</b>	
radioisotope thermoelectric generators for outer-planet missions .....	70-75
optimization and reliability calculations for multi-thermionic-converter systems .....	76-80
cracking of filter layers in a high- performance solar cell filter .....	115-117
<b>Propulsion, Electric</b>	
attitude control of a flexible solar-electric spacecraft .....	92-99
stepper motor drive electronics for solar-electric thrust vector control subsystem .....	108-110
ion thruster electron baffle sizing .....	201-203
ion thruster using combination keeper electrode and electron baffle .....	204-206
thrust subsystem design for nuclear- electric spacecraft .....	207-210
ion thruster connectors .....	211-212
<b>Propulsion, Liquid</b>	
creep failure of randomly excited pressure vessels .....	120-128
high-thrust throttleable monopropellant hydrazine reactors .....	213-221
injector hydrodynamics effects on baffled- engine stability: correlation of required baffle geometry with injected mass flux.....	222-232
<b>Propulsion, Solid</b>	
low pressure/low motor free volume to nozzle throat area ratio combustion limits .....	179-180
viscoelastic behavior of elastomers undergoing crosslinking reactions.....	187-189
<b>Pyrotechnics</b>	
nondestructive testing of pyrotechnic devices .....	175-178
<b>Quality Assurance and Reliability</b>	
optimization and reliability calculations for multi-thermionic-converter systems .....	76-80
reliability of randomly excited structures .....	128-136
radiation effects on electronic parts .....	145-147

<b>Subject</b>	<b>Pages</b>
<b>Solid-State Physics</b>	
creep failure of randomly excited structures . . . . .	120-128
reliability of randomly excited structures . . . . .	128-136
boundary layer equations of a heated constrained spherical shell . . . . .	136-140
derivation of equations governing heated shallow shells of revolution . . . . .	141-144
viscoelastic behavior of elastomers undergoing crosslinking reactions . . . . .	187-189
<b>Structural Engineering</b>	
creep failure of randomly excited structures . . . . .	120-128
reliability of randomly excited structures . . . . .	128-136
magnetometer boom for Thermoelectric Outer-Planet Spacecraft (TOPS) . . . . .	148-156
<b>Telemetry and Command</b>	
optimum configurations for phase-shift keyed/phase modulated systems . . . . .	33-36
steady-state performance of a data-transition type of first-order digital phase-locked loop . . . . .	59-68
optimum modulation index for a data-aided phase-coherent communication system . . . . .	69
command prefix code for Thermoelectric Outer-Planet Spacecraft (TOPS) . . . . .	164-166
<b>Temperature Control</b>	
scientific instrument heat control pump for Thermoelectric Outer-Planet Spacecraft (TOPS) . . . . .	148-156
<b>Test Facilities and Equipment</b>	
holographic study of solar simulation arc lamps . . . . .	171-174
<b>Thermodynamics</b>	
boundary layer equations of a heated constrained spherical shell . . . . .	136-140
derivation of equations governing heated shallow shells of revolution . . . . .	141-144

<b>Subject</b>	<b>Pages</b>
<b>Thermoelectric Outer-Planet Spacecraft (TOPS)</b>	
automated test techniques for TOPS control and guidance subsystems . . . . .	90-91
digital gyro system for TOPS single-axis simulator . . . . .	105-107
radiation effects on electronic parts . . . . .	145-147
magnetometer boom . . . . .	148-156
mechanical devices . . . . .	148-156
scientific instrument heat control pump . . . . .	148-156
high-gain antenna . . . . .	157-159
command prefix code . . . . .	164-166
<b>Tracking</b>	
digital clean-up loop transponder for sequential ranging system . . . . .	18-27
performance of a phase-locked loop during loss of signal . . . . .	28-32
optimum configurations for phase-shift keyed/phase modulated systems . . . . .	33-36
stability of second-order tracking loops with arbitrary time delay . . . . .	50-53
steady-state performance of a data-transition type of first-order digital phase-locked loop . . . . .	59-68
spectral factorization in discrete systems . . . . .	233-235
<b>Wave Propagation</b>	
digital clean-up loop transponder for sequential ranging system . . . . .	18-27
performance of a phase-locked loop during loss of signal . . . . .	28-32
optimum configurations for phase-shift keyed/phase modulated systems . . . . .	33-36
spacecraft antenna tolerances . . . . .	43-49
signal design for single-sideband phase modulation . . . . .	54-58

1. Report No. 37-66, Vol. III	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle  SPACE PROGRAMS SUMMARY, VOL. III SUPPORTING RESEARCH AND ADVANCED DEVELOPMENT		5. Report Date December 31, 1970	
		6. Performing Organization Code	
7. Author(s) JPL Staff		8. Performing Organization Report No.	
9. Performing Organization Name and Address JET PROPULSION LABORATORY California Institute of Technology 4800 Oak Grove Drive Pasadena, California 91103		10. Work Unit No.	
		11. Contract or Grant No. NAS 7-100	
		13. Type of Report and Period Covered Space Programs Summary Oct. 1 to Nov. 30, 1970	
12. Sponsoring Agency Name and Address NATIONAL AERONAUTICS AND SPACE ADMINISTRATION Washington, D.C. 20546		14. Sponsoring Agency Code	
15. Supplementary Notes			
16. Abstract  <p>The Space Programs Summary is a multivolume, bimonthly publication that presents a review of technical information resulting from current engineering and scientific work performed, or managed, by the Jet Propulsion Laboratory for the National Aeronautics and Space Administration. The Space Programs Summary is currently composed of four volumes:</p> <div style="margin-left: 100px;"> Vol. I. Flight Projects (Unclassified)  Vol. II. The Deep Space Network (Unclassified)  Vol. III. Supporting Research and Advanced Development (Unclassified)  Vol. IV. Flight Projects and Supporting Research and Advanced Development (Contents Confidential) </div>			
17. Key Words (Selected by Author(s))  Not applicable for this type of report		18. Distribution Statement  Unclassified -- Unlimited	
19. Security Classif. (of this report)  Unclassified	20. Security Classif. (of this page)  Unclassified	21. No. of Pages  238	22. Price



## HOW TO FILL OUT THE TECHNICAL REPORT STANDARD TITLE PAGE

Make items 1, 4, 5, 9, 12, and 13 agree with the corresponding information on the report cover. Use all capital letters for title (item 4). Leave items 2, 6, and 14 blank. Complete the remaining items as follows:

3. Recipient's Catalog No. Reserved for use by report recipients.
7. Author(s). Include corresponding information from the report cover. In addition, list the affiliation of an author if it differs from that of the performing organization.
8. Performing Organization Report No. Insert if performing organization wishes to assign this number.
10. Work Unit No. Use the agency-wide code (for example, 923-50-10-06-72), which uniquely identifies the work unit under which the work was authorized. Non-NASA performing organizations will leave this blank.
11. Insert the number of the contract or grant under which the report was prepared.
15. Supplementary Notes. Enter information not included elsewhere but useful, such as: Prepared in cooperation with... Translation of (or by)... Presented at conference of... To be published in...
16. Abstract. Include a brief (not to exceed 200 words) factual summary of the most significant information contained in the report. If possible, the abstract of a classified report should be unclassified. If the report contains a significant bibliography or literature survey, mention it here.
17. Key Words. Insert terms or short phrases selected by the author that identify the principal subjects covered in the report, and that are sufficiently specific and precise to be used for cataloging.
18. Distribution Statement. Enter one of the authorized statements used to denote releasability to the public or a limitation on dissemination for reasons other than security of defense information. Authorized statements are "Unclassified-Unlimited," "U. S. Government and Contractors only," "U. S. Government Agencies only," and "NASA and NASA Contractors only."
19. Security Classification (of report). NOTE: Reports carrying a security classification will require additional markings giving security and downgrading information as specified by the Security Requirements Checklist and the DoD Industrial Security Manual (DoD 5220.22-M).
20. Security Classification (of this page). NOTE: Because this page may be used in preparing announcements, bibliographies, and data banks, it should be unclassified if possible. If a classification is required, indicate separately the classification of the title and the abstract by following these items with either "(U)" for unclassified, or "(C)" or "(S)" as applicable for classified items.
21. No. of Pages. Insert the number of pages.
22. Price. Insert the price set by the Clearinghouse for Federal Scientific and Technical Information or the Government Printing Office, if known.